Structural and Functional Investigations to Address Challenging Drug Targets

Dissertation

Zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von
Patrick David Fischer

Saarbrücken, 2021

Tag des Kolloquiums:	23.03.2022
Dekan:	Prof. Dr. J. Walter
Berichterstatter:	Prof. Dr. C. Ducho
	Prof. Dr. A. K. Kiemer
Vorsitzender:	Prof. Dr. C. Jacob
Akad. Mitarbeiter:	Dr. J. Zapp

Die vorliegende Arbeit wurde an der Universität des Saarlandes an der Naturwissenschaftlich-Technischen Fakultät in der Fachrichtung Pharmazie im Zeitraum von Juli 2016 bis November 2021 angefertigt.

Erstgutachter: -

Zweitgutacther: -

"Papa, you Arbeit?"

- Liam Nico Fischer

Abstract

The rational development of drugs against diseases such as infections, immunological disorders, cancer, inflammation etc. requires a detailed understanding of the proteins involved in these diseases. This PhD project is subdivided into six projects, each of which focuses on a different protein involved in diseases which has been difficult to interfere with due to a variety of challenging aspects.

Project 1 is focused on the bacterial membrane-bound enzyme MraY. Studies on the inhibition of different MraY isoforms in divergent purification systems, resistance development and protein-protein interactions of the enzyme were investigated. Project 2 revolved around NFAT. Novel NMR experiments were developed, and protein dynamics influenced by phosphorylation were analyzed. In addition, protein engineering using IDRs was investigated. Project 3 furnished new avenues to target the master regulator for eukaryotic protein translation eIF4E for cancer treatment using a novel small-molecule inhibitor and solving the structure of phosphorylated eIF4E. Project 4 aimed to elucidate the specificity in substrate recognition by AHL-synthases of acyl carrier proteins loaded with different cargoes. In project 5, the oligomeric nature of GILZ was shown and experiments were conducted to solve the interface with HDAC7. Project 6 leveraged the interaction between Keap1 and Nrf2 to develop an open-source software in silico screening tool.

Zusammenfassung

Die rationale Entwicklung von Arzneistoffen gegen Krankheiten wie Infektionen, Immunerkrankungen, Krebs, Entzündungen etc. erfordert ein detailliertes Verständnis der Proteine die an diesen Erkranknungen beteiligt sind. Diese Arbeit ist unterteilt in sechs Projekte, von denen jedes auf ein anderes in Krankheiten involviertes Protein fokussiert ist, welches aufgrund verschiedener herausfordernder Aspekte schwer zu untersuchen ist.

Projekt 1 behandelt das bakterielle, membranständige Enzym MraY. Studien der Inhibition verschiedener MraY-Isoformen in divergenten Präparationen, die Möglichkeit der Resistenzentwicklung und Protein-Protein-Interaktionen des Enzyms wurden untersucht. Projekt 2 thematisiert NFAT. Neuartige NMR-Experimente wurden entwickelt und Beeinflussing der Proteindynamik durch Phosphorylierung wurde analysiert. Zusätzlich wurde Protein-Enginerring mittels IDRs untersucht. Projekt 3 lieferte neue Wege, das Regulatorprotein eukaryotischer Translation eIF4E mittels kleiner Moleküle zu inhibieren und die Struktur von phosphoryliertem eIf4E wurde gelöst. Projekt 4 zielte darauf ab, die Spezifität der Substraterkennung von ACPs zu Cosubstraten durch AHL-Synthasen zu verdeutlichen. Projekt 5 zeigte den oligmeren Status von GILZ, und Experimente wurden konstruiert, um die Bindeschnittstelle mit HDAC7 zu lösen. Projekt 6 nutzte die Interaktion zwischen Keap1 und dem Transkriptionsfaktor Nrf2, um ein Open-Source in silico Screening Tool zu entwickeln.

Danksagung

Ich danke meinem Doktorvater Prof. Dr. Christian Ducho für die spannende Themenstellung der Doktorarbeit, stete Diskussionsbereitschaft, Unterstützung wissenschaftlicher und persönlicher Natur und das Ermöglichen, durch den Besuch verschiedener Konferenzen meine Expertise zu erweitern und wissenschaftliche Ergebnisse zu präsentieren.

Prof. Dr. Alexandra K. Kiemer danke ich für die Übernahme des Zweitgutachtens und die Betreuung als wissenschaftliche Begleiterin. Weiterhin danke ich Prof. Dr. Haribabu Arthanari und Prof. Dr. Gerhard Wagner für die Unterstützung meiner binationalen Promotion in Boston.

Ich danke allen ehemaligen und aktuellen Mitarbeiten der Arbeitskreise Ducho, Wagner und Arthanari für die direkte Zusammenarbeit in den vorgestellten Projekten. Besonderer Dank geht an Dr. Giuliana Niro, Dr. Melissa Wojtyniak, Theresa Manz, Katrin Voos, Stefan Koppermann, Stefanie Weck, Jannine Ludwig und Reem Fathalla für die enge Zusammenarbeit im Arbeitskreis Ducho in Saarbrücken. Gleichermassen danke ich Dr. Sandeep Chhabra, Dr. Andras Boeszoermenyi, Dr. Jon M. Dempersmier, Dr. Krishna M. Padmanabha Das, Huel Cox III, Siddanth Ramesh and Sandra Goll für die Unterstützung in Boston.

Einen persönlichen Dank möchte ich an Stefanie Weck, Dr. Giuliana Niro und Dr. Melissa Wojtyniak richten, deren Unterstützung weit über wissenschaftliche Diskussion hinaus ging. Ich danke euch für eure Freundschaft und stete Hilfe in jeglichen Fragestellungen. Des Weiteren möchte ich einen persönlichen Dank an zwei meiner ältesten und besten Freunde, Nathalie Kirsch und Tina Sander richten, die mich seit nunmehr über einem Jahrzehnt in jeglichen Lebenssituationen unterstützen.

Ich möchte weiterhin Dank ausrichten an die Arbeitskreise von Steven van Lanen und Georgios Skiniotis für die Kollaboration in MraY-Projekten, Daniella Goldfarb und Martin Blackledge für die Kollaboration in NFAT-Projekten, Sirano Dhe-Paganon für alle

kristallographischen Projekte, Rajesh Nagarajan für die Kollaboration in ACP-Projekten und Alexandra K. Kiemer für die Kollaboration in GILZ-Projekten.

Ein weiterer Dank geht an den Fonds der chemischen Industrie und die Chleck Foundation für die finanzielle Unterstützung in Form von Promotionsstipendien.

Den grössten Dank möchte ich meiner Familie aussprechen. Meiner Mutter Christina Fischer, meinem verstorbenen Vater Klaus Fischer und meiner Schwester Melanie Fischer für finanzielle und moralische Unterstützung während meiner gesamten wissenschaftlichen Laufbahn. Meiner Frau Caroline Fischer, meinem Sohn Liam Nico Fischer sowie meiner ungeborenen Tochter gilt mein allergrösster Dank. Ohne euch wäre nichts davon möglich gewesen, und ich bin überglücklich euch immer an meiner Seite zu wissen. Danke für eure Geduld und eure Liebe.

Abbreviations and symbols

°C	degrees Celsius
3D	three dimensional
4EBP	4E binding protein
4E-SE	4E sensitive element
Å	Ångström
aa	amino acid
aaRS	aminoacyl-tRNA synthetase
ACP	acyl carrier protein
AHL	acyl homoserine lactones
AKT/Akt	protein kinase B
AML	acute myeloid leukemia
AP1	activating protein 1
ARCA	anti-reverse cap analogue
ATP	adenosine triphosphate
BLI	bio-layer interferometry
BME	β-mercaptoethanol
Bn	benzyl
bp	base pairs
BTB	*bric-a-brac*, tramtrack and broad complex
ca.	*circa* (approximately)
CamKII	Ca^{2+}/calmodulin-dependent protein kinase II
CBP	cyclic adenosine monophosphate response element binding protein binding protein
CCND1	cyclin D1
ccpNMR	collaborative computing project for nuclear magnetic resonance
CD	cluster of differentiation or circular dichroism
CDK/Cdk	cyclin dependent kinase

CFU	colony-forming units
Chip	C-terminus of Hsc70 interacting protein
CK	casein kinase
CL	cardiolipin
cm	centimeter(s)
CMC	critical micelle concentration
Cn	calcineurin
COX-2	cyclooxygenase 2
CPMG	Carr-Purcell-Meiboom-Gill
CPU	central processing unit
CRISPR	clusters of regularly interspaced short palindromic repeats
CRM1	chromosomal maintenance 1
CSA	chemical shift anisotropy (NMR)
CSP	chemical shift perturbation (NMR)
CST	cross-saturation transfer (NMR)
CTD	C-terminal domain
CUL3	cullin-3
Da	Dalton
DD	dipole-dipole
DDM	dodecyl-β-D-maltopyranoside
DEAE	diethylaminoethyl
DEER	double electron-electron resonance
DHPC	1,2-diheptanoyl-*sn*-glycero-3-phosphocholine
DIP	degradable intake protein
DM	decyl-β-D-maltopyranoside
DMPC	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine
DMPG	1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol)
DMSO	dimethyl sulfoxide

DNA	desoxyribonucleic acid
DOPE	1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine
DPC	dodecylphosphocholine
DTT	dithiothreitol
DXO	decapping and exoribonuclease protein
DYRK	dual specificity tyrosine phosphorylation regulated kinase
E. coli	*Escherichia coli*
e.g.	*exempli gratia* (for example)
EDTA	ethylenediaminetetraacetic acid
eIF4A	eukaryotic translation initiation factor 4A
eIF4B	eukaryotic translation initiation factor 4B
eIF4E	eukaryotic translation initiation factor 4E
eiF4F	eukaryotic initiation factor 4F
eIF4G	eukaryotic translation initiation factor 4G
eIF4H	eukaryotic translation initiation factor 4H
EM	electron microscopy
EpoR	erythropoietin receptor
EPR	electron paramagnetic resonance
EPS	exopolysaccharide
ERK	extracellular-signal-regulated kinase
et al.	*et alia* (and others)
etc.	*et cetera* (and other similar things)
eV	electron Volt
EVH1	enabled/vasodilator-stimulated phosphoprotein homology-1
Fas	Fas cell surface death receptor
FasL	Fas ligand
FDA	Food and Drug Administration
Flox	flanked by locus of crossover in P1

FoxO3	forkhead box O3
FP	fluorescence polarization
Fwd	forward
g	gram(s)
GB1	globular protein B1
GC	glucocorticoid
GDP	guanosine diphosphate
GFP	green fluorescent protein
GILZ	glucocorticoid induced leucine zipper
GMP	guanosine monophosphate
gp130	glycoprotein 130 receptor
GPCR	G-protein coupled receptor
GR	glucocorticoid receptor
GRE	glucocorticoid response element
GSK-3	glycogen synthase kinase 3
GST	glutathione *S*-transferase
GTP	guanosine triphosphate
h	hour(s) or Planck's constant
HDAC	histone deacetylase
HEAT3	Huntington, elongation factor 3, PP2A, TOR 3
HEK293	human embryonic kidney cell line 293
HEPES	*N*-2-hydroxyethylpiperazine-*N*-ethanesulfonic acid
het	heteronuclear (NMR)
HIF-1	hypoxia-inducable factor 1
hmsIST	Harvard Medical School implementation of the iterative soft thresholding approach
HRP	horseradish peroxidase
HS	hot spot

HSP/Hsp	heat shock protein
HSQC	heteronuclear single quantum coherence (NMR)
Hz	Hertz
i.e.	*id est* (that is)
IC_{50}	inhibitory concentration (50 %)
ID	identification
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IgG	immunoglobulin G
IL	interleukin
INEPT	insensitive nuclei enhanced by polarization transfer
INF	interferon
IP	immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
IVA	*in vivo* assembly
IVR	intervening region
JAK	just another kinase or Janus kinase
JNK	c-Jun N-terminal kinase
K	Kelvin
K	kilo
kb	kilo bases
kcal	kilo calories
KD	dissociation constant
kDa	kilo Dalton
Keap1	Kelch-like ECH-associated protein 1
k_{ex}	exchange rate
kJ	kilo Joules

L	liter(s)
LepR	leptin receptor
log	logarithm (base 10)
Luc	luciferase
LZ	leucine zipper
M	molecular weight or molar
M. xanthus	*Myxococcus xanthus*
m7G(m)G(m)G	N^7-methyl-3'*O*-methyl-guanosine-5'-triphosphate-5'-(2'*O*-methyl)-guanosine-phosphate-guanosine
m7GA(m)G	N^7-methyl-guanosine-5'-triphosphate-5'-(2'*O*-methyl)-adenosine-phosphate-guanosine
m7GG	N^7-methyl-guanosine-5'-triphosphate-5'-guanosine
m7GG(m)G	N^7-methyl-guanosine-5'-triphosphate-5'-(2'*O*-methyl)-guanosine-phosphate-guanosine
MAPK	mitogen activated protein kinase
MAPPIT	mammalian protein-protein interaction trap
MB2	muraymycin B2
MB6	muraymycin B6
MBP	maltose binding protein
MD2	muraymycin D2
MDA5	melanoma differentiation-associated protein 5
mdeg	millidegree(s)
MDM2	mouse double minute 2 homolog
MEK	Mitogen-activated protein kinase kinase
MES	2-(*N*-morpholino)ethanesulfonic acid
mg	milligram(s)
MHz	megaHertz
MIC	minimal inhibitory concentration
min	minute(s)

MKK6	mitogen-activated protein kinase kinase 6
mL	milliliter(s)
mM	millimolar
Mnk	mitogen-activated protein kinase interacting protein kinases
mol	moles
$MraY_{AA}$	MraY from *Aquifex aeolicus*
$MraY_{CB}$	MraY from *Clostridium bolteae*
$MraY_{SA}$	MraY from *Staphylococcus aureus*
$MraY_{SP}$	MraY from *Streptomyces platensis*
mRNA	messenger ribonucleic acid
MSP	membrane scaffold protein
MST	microscale thermophoresis
mTOR	mammalian target of rapamycin
mTORC1	mammalian target of rapamycin complex 1
MTSSL/MTSL	2,5-dihydro-2,2,5,5-tetramethyl-3-[[(methylsulfonyl)thio]methyl]-1*H*-pyrrol-1-yloxy
MYC	myelocytomatosis
myoD	myoblast determination protein 1
NB	nanobody
NES	nuclear export signal
NFAT	nuclear factor of activated T-cells
NF-κB	nuclear factor kappa B
NLS	nuclear localization signal
nM	nanomolar
nm	nanometer(s)
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NPC	nuclear pore complexes

Nrf2	Nuclear factor erythroid 2-related factor 2
NTA	nitrilotriacetic acid
N-Ter	N-terminal
NUS	non uniform sampling (NMR)
OD_{600}	optical density at 600 nm
off res	off resonance
on res	on resonance
ORF	open reading frame
PA	phosphatidic acid
PABP	poly-(A) binding protein
PB-Cre4	probasin Cre recombinase 4
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PDB	protein database
PDCD4	programmed cell death protein 4
PDMS	phosphorylation dependent mobility shift
PE	phosphatidylethanolamine
PEG	polyethylene glycol
PER	proline glutamate rich region
PG	phosphatidylglycerol
PI(3)K	Phosphoinositide 3-kinase
PI-3	phosphoinositol 3
PKA	protein kinase A
PKB	protein kinase B
PKC	protein kinase C
Plk1	polo-like kinase 1
pM	picomolar
PMA	phorbol 12-myristate 13-acetate

PMLO	proteinaceous membrane-less organelle
POI	protein of interest
PP2A	protein phosphatase 2A
ppm	parts per million (NMR)
PRE	paramagnetic relaxation enhancement
PROTAC	proteolysis targeting chimeras
Pten	phosphatase and tensin homolog
PTM	posttranslational modification
R^1	longitudinal relaxation rate
R^2	transverse relaxation rate
RAF	rapidly accelerated fibrosarcoma or rat fibrosarcoma
RANBP2	ras-related nuclear protein binding protein 2
RAS	rat sarcoma
Rev	reverse
R^{ex}	chemical exchange relaxation (NMR)
RF	radio frequency (NMR)
RIG-1	retinoic acid-inducible gene I
RING	really interesting new gene
RIPA	radioimmunoprecipitation assay buffer
R_{min}	minimal radius
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
ROS	reactive oxidative species
s	second(s)
SAC	spindle assembly checkpoint
SAM	*S*-adenosylmethionine
SAR	structure activity relationship
SARM	selective androgen receptor modulators

SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sfp	surfactin phosphopantetheinyl transferase
shRNA	short hairpin ribonucleic acid
siRNA	small interfering ribonucleic acid
snRNA	small nuclear ribonucleic acid
SOC	super optimal broth
SP	serine/proline rich region
sp.	species
SPR	surface plasmon resonance
SRR	serine rich region
STAT	signal transducer and activator of transcription
STD	saturation transfer difference
SUMO1	small ubiquitin like modifier 1
T_1	spin-lattice relaxation time (NMR)
T_2	spin-spin relaxation time (NMR)
TAD	transactivation domain
TBST	tris-buffered saline with Tween-20
TCEP	tris(2-carboxyethyl)phosphine
TCR	t cell receptor
TEM	transmission electron microscope
TEV	tobacco etch virus
TGF-β	transforming growth factor-β
TM	transmembrane
Tm	melting temperature
TMG	trimethylguanosine
TMX	transmembrane helix X
TNF	tumor necrosis factor

TNfn	fibronectin type III-like repeat(s)
TOCSY	total correlated spectroscopy
TOR	target of rapamycin
tRNA	transfer ribonucleic acid
TROSY	transverse relaxation-optimized spectroscopy (NMR)
TSC	TGF-β stimulated clone box
TXO	triple resonance X1+X2 nucleus observe (NMR probe)
UDP	uridine diphosphate
US	United States
UTR	untranslated region
V. fischerii	*Vibrio fischeri*
V. harveyi	*Vibrio harvey*
VEGF	vascular endothelial growth factor
VHL	von Hippel-Lindau
vol	volume
wt	wild type
γ	gyromagnetic ratio (NMR)
δ	chemical shift (NMR)
θ	molar ellipticity (CD)
µcal	micro calorie(s)
µg	microgram(s)
µL	microliter(s)
µM	micromolar
ν	frequency
σ_{SH}	heteronuclear cross-relaxation rate
τ_m	correlation time

Table of Contents

PAPERS INCLUDED IN THIS THESIS

This thesis is divided into five publications and unpublished material discussed in the appendix. The publications will be referred to by their individual letters:

A. **^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins**
Sandeep Chhabra*, Patrick Fischer*, Koh Takeuchi, Abhinav Dubey, Joshua J. Ziarek, Andras Boeszoermenyi, Daniel Mathieu, Wolfgang Bermel, Norman E. Davey, Gerhard Wagner, and Haribabu Arthanari;
Proc. Natl. Acad. Sci. U.S.A. **2018**, 115, *8*, E1710-E1719.
** These authors contributed equally*

B. **Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**
Stefan Koppermann, Zheng Cui, Patrick D. Fischer, Xiachang Wang, Jannine Ludwig, Jon S. Thorson, Steven G. Van Lanen, and Christian Ducho;
ChemMedChem **2018**, 13, *8*, 779-784.

C. **An open-source drug discovery platform enables ultra-large virtual screens**
Christoph Gorgulla, Andras Boeszoermenyi, Zi-Fu Wang, Patrick D. Fischer, Paul Coote, Krishna M. Padmanabha Das, Yehor S. Malets, Dmytro S. Radchenko, Yurii S. Moroz, David A. Scott, Konstantin Fackeldey, Moritz Hoffmann, Iryna Iavniuk, Gerhard Wagner, Haribabu Arthanari;
Nature **2020**, 580, *7805*, 663-668.

D. **A biphenyl inhibitor of eIF4E targeting an internal binding site enables the design of cell-permeable PROTAC-degraders**
Patrick D. Fischer*, Evangelos Papadopoulos*, Jon M. Dempersmier, Zi-Fu Wang, Radosław P. Nowak, Katherine A. Donovan, Joann Kalabathula, Christoph Gorgulla, Pierre P. M. Junghanns, Eihab Kabha, Nikolaos Dimitrakakis, Constantine Mitsiades, Christian Ducho, Vladimir Gelev, Eric S. Fischer, Gerhard Wagner and Haribabu Arthanari; *Eur. J. Med. Chem.* **2021**, 219, *113435*.
** These authors contributed equally*

E. **VirtualFlow Ants – Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization**
Christoph Gorgulla, Suleyman Cinaroglu, Patrick D. Fischer, Konstantin Fackeldey, Gerhard Wagner and Haribabu Arthanari;
Int. J. Mol. Sci. **2021**, 22, 5807.

PAPERS NOT INCLUDED IN THIS THESIS

F. **A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening**
Christoph Gorgulla, Krishna M. Padmanabha Das, Kendra E. Leigh, Marco Cespugli, Patrick D. Fischer, Zi-Fu Wang, Guilhem Tesseyre, Shreya Pandit, Alec Shnapir, Anthony Calderaio, Minko Gechev, Alexander Rose, Noam Lewis, Colin Hutcheson, Erez Yaffe, Roni Luxenburg, Henry D. Herce, Vedat Durmaz, Thanos D. Halazonetis, Konstantin Fackeldey, J.J. Patten, Alexander Chuprina, Igor Dziuba, Alla Plekhova, Yurii Moroz, Dmytro Radchenko, Olga Tarkhanova, Irina Yavnyuk, Christian Gruber, Ryan Yust, Dave Payne, Anders M. Näär, Mark N. Namchuk, Robert A. Davey, Gerhard Wagner, Jamie Kinney, Haribabu Arthanari; *iScience* **2021**, 24, 2, 102021.

1 Introduction

1.1 Drug targets

Generally, the development of therapeutics against a disease is generated by the identification of the so-called "target". A target is a molecular or cellular structure involved in a pathological context that the potential drug is directed against. The vast majority of these targets are either human or pathogen-derived proteins, as was highlighted in a review article by SANTOS et al. published in 2017 on the analysis of 1,578 US-FDA-approved drugs.[1] Of 893 targets in total, 667 are human proteins involved in human diseases. Together with 189 pathogen proteins targeted for treatment of infectious diseases, proteins make up more than 95 % of all drug targets. Other biomolecules such as nucleic acids, heparin, lipids or peptides only account for a small percentage of all drug targets.[1]

To provide intelligent drug design and effectively improve binding affinities of ligands, structural characterization of said targets is of vital importance.[2–4] This information is traditionally provided by X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy or cryoelectron microscopy (cryo-EM). However, structural information is specifically hard to provide for some of the most interesting targets such as membrane proteins, intrinsically disordered proteins, proteins which undergo posttranslational modifications, proteins involved in protein-protein interactions or highly dynamic proteins.

1.1.1 Membrane proteins

Proteins that require the presence of a lipid bilayer membrane for stabilization of their tertiary structure are classified as membrane proteins. While being encoded by approximately 30 % of the genomes of most organisms[5,6], they make up at least half of all pharmaceutical drug targets (i.e. G-protein coupled receptors [GPCRs], Ion channels or nuclear receptors).[1] Nonetheless, less than 0.3 % of all unique entries in the protein database are occupied by membrane proteins.[7] There is therefore an evident discrepancy between the need for structural characterization and its provision. This is

largely due to the fact of the proteins' requirement for a bilayer-mimicking system for stabilization. There are several ways to stabilize membrane proteins, such as the use of detergent micelles, bicelles or detergent-free nanodiscs. Whereas detergents are usually able to preserve structural integrity, they tend to compromise enzymatic activity and/or ligand binding of membrane proteins. For example, when the β-barrel forming membrane protein OmpX is stabilized in dodecylphosphocholine (DPC) detergent micelles, 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC)/1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) bicelles or DMPC nanodiscs, the dynamic motions vary substantially between the different environments.[8] Sometimes, specific lipid interactions are required for membrane protein functions. Examples are the regulation of ion channels by cardiolipin[9], the role of phosphoinositides in cell migration[10] or the formation of signaling structures moderated by electrostatic factors[11].

In the past, membrane mimetic systems have been limited to vesicles or liposomes.[12–14] While providing a large lipid bilayer for membrane proteins, samples prepared from liposomes tend to be turbid, viscous and not stable for an extended period.[15] Solubilization of membrane proteins in detergent micelles has therefore become a popular technique that persists to this day.[16] More recently, lipid-detergent bicelles[17] and other extended bilayers such as amphipathic polymers (amphipols)[18] have been used to study protein structure and function. However, controlling sample size and homogeneity, as well as avoiding fusion still represent major difficulties with these systems.[19,20]

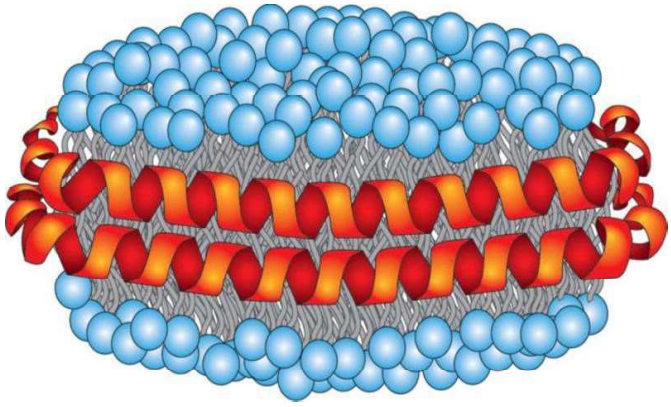


Fig. 1-1 *Schematic picture of an empty nanodisc: a bilayer of phospholipids (polar head groups in blue, lipophilic tails in grey) are encircled by helical membrane scaffold proteins (MSPs, in red). Taken from I. G. Denisov, S. G. Sligar,* Nanodiscs in Membrane Biochemistry and Biophysics. Chem. Rev. ***2017****,* 117*, 4669–4713.*

An even newer form of stabilized membrane protein, so-called nanodiscs, are comprised of phospholipids, enclosed by an aliphatic helical belt protein

called membrane scaffold protein (MSP, see **Fig. 1-1**). The initial MSP was based on the human ApoA1 protein, a major component of the high-density lipoprotein complex.[21] From detergent-solubilized mixtures of membrane proteins, phospholipids and MSPs, nanodiscs self-assemble upon removal of detergent, which can be realized either by adsorption on hydrophobic beads or dialysis.[22] If assembly is performed at the optimal target protein/lipid/MSP ratio, monodisperse and homogeneous lipoprotein samples can be obtained, as shown with negative stain electron microscopy by HAGN et al.[22]

1.1.2 Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) of proteins are defined as proteins or protein segments with a lack of stable secondary structure.[23] Due to their involvement in diverse cellular processes, such as signaling, translation or transcription, these regions mostly occur in human proteins associated with cancer (~79 %) and signaling (~66 %).[24] Mutations and changes in abundance of IDPs are linked to a variety of diseases, such as cancer[25–27], neurodegeneration[28,29], cardiovascular diseases[30,31] or diabetes[32]. The conformational heterogeneity and constitutional dynamics render the "classical" structure determination methods such as X-ray crystallography and cryo-EM ineffective.[33,34] Therefore, solution-NMR has persevered as the main tool to study structural and dynamic behavior of these proteins.[35–37]

An IDP is characterized by its biased amino acid distribution, low sequence complexity and low content of bulky, hydrophobic amino acids.[23] The resulting protein or region within a protein adopts an ensemble of conformations and displays interesting dynamics by fluctuation between those conformations. Apart from their important role in signaling, transcription and translation, IDPs partake in assembly of macromolecular machinery such as the ribosome, nuclear import, chaperone function, assembly and disassembly of microfilaments and microtubules, binding and transport of small molecules as well as serving as entropic flexible linkers between functional protein domains.[38–40]

Recently, IDRs have gained broader attention due to the ability of low-complexity domains to perform phase separation, thus creating proteinaceous membrane-less organelles (PMLOs).[41] These condensed lipid droplets represent an overcrowded milieu

within the cytoplasm or nucleoplasm with considerably higher concentrations of proteins.[42,43]

1.1.3 Posttranslational modifications

A posttranslational modification (PTM) describes the covalent, generally enzymatic modification of proteins after their biosynthesis. They can occur either on side chains of the amino acids or the C- and N-termini.[44] Modifications are used for a variety of purposes:

- facilitating membrane localization by increasing hydrophobicity through acylation with fatty acids[45]
- enhancing enzymatic activities through the addition of cofactors such as heme C or flavin moieties[46,47]
- numerous other effects through the addition of smaller functional groups such as acetyl, methyl or phosphate[48–50].

Due to the diverse possibilities of different posttranslational modifications, this introduction will focus on the specific alteration of hydroxyl side chains to phosphates (phosphorylation), as they play an important role for this work. This PTM also represents the most recurrently occurring protein modification according to the Swiss-Prot database, with a frequency of 57,191 entries compared to the next most common modification (acetylation) with a frequency of 6,656 entries.[51]

Phosphorylation of a specific amino acid can be used to alter the binding affinity to interaction partners[52], ion transport through membranes[53] or regulation of enzymes[54]. Multisite phosphorylation can have numerous reasons. The highly phosphorylated nuclear factor of activated T-cells (NFAT) serves as an example for the explanation of those possibilities, as is presents a major target of this work.

Increased sensitivity:

It is known that multisite phosphorylation can sometimes be used to increase a protein's sensitivity of biochemical processes to their signaling inputs.[55,56] SALAZAR and HOEFER[57] calculated the sensitivity of NFAT to calcineurin activity changes as a function of dephosphorylations required for an active NFAT conformation (see **Fig. 1-2**).

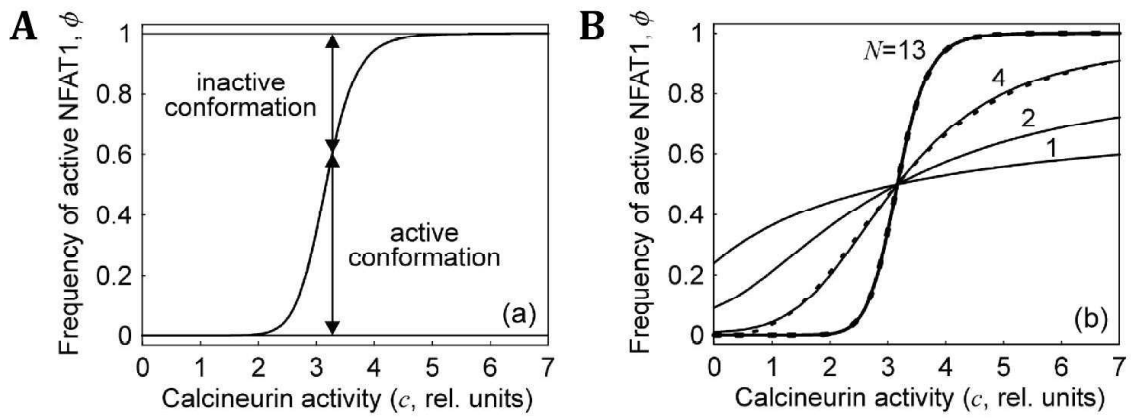


Fig. 1-2 A*. Frequency for the occurrence of active NFAT1 conformation for 13 phosphorylation sites.* ***B.*** *Stimulus-response curve with varying phosphorylation sites. More phosphorylation sites result in sharper thresholds. Taken from C. Salazar, T. Höfer,* Allosteric Regulation of the Transcription Factor NFAT1 by Multiple Phosphorylation Sites: A Mathematical Analysis. J. Mol. Biol. ***2003****,* 327*, 31–45.*

In a virtual simulation, the group simulated NFAT activity as a function of dephosphorylation events. Of reported 14 phosphoserines, 13 are dephosphorylated by the phosphatase calcineurin, thus enabling nuclear translocation of NFAT and activation of its downstream targets.[58] If dephosphorylations were set to the reported 13 phosphoserines, the dose-response curve was steep and cooperative, defining a threshold of calcineurin activity above which full NFAT activation was achieved and below which NFAT remained inactive (see **Fig. 1-2 B**). If dephosphorylations were set to 1-2 dephosphorylation sites instead, the dose-response curve was rather shallow. Therefore, to obtain a maximal response range, each phosphorylation site is required to make smaller energetic contribution for a large number of phosphorylation sites compared to a small number of sites.[57]

Electrostatic repulsion

Numerous phosphorylation sites result in a highly negatively charged protein. It is reasonable to assume that this can be used as a mechanism to electrostatically repel another negatively charged structure or increase the binding affinity to a highly positively charged structure, thus changing the conformation of the protein.[59] In NFAT, at least 14 negatively charged phosphoserines are located relatively close to each other.

It is a possibility that this arrangement induces an intramolecular repulsion, therefore changing the protein's conformation compared to an unphosphorylated state. On the other hand, an interaction of the regulatory domain with another structure, either a different protein or another domain of NFAT, can be imagined.

Time-dependent phosphorylation

The third possible purpose of multisite phosphorylation, especially by multiple kinases, is the use of time-dependent phosphorylation states in order to regulate activation. If the order of NFAT phosphorylation is regulated strictly, one can imagine that different stages of phosphorylation influence the level of calcineurin binding, conformational flexibility or the binding of NFAT kinases.[60] Time dependency could also be involved in the dephosphorylation process. For example, an early dephosphorylation of the serine-rich region (SRR) could promote further dephosphorylation needed for full activation.[58]

2 Literature Review

2.1 MraY

For decades, the bacterial cell wall biosynthetic pathway has been an important and traditional target for antibiotics (e.g. penicillins).[61] However, emerging antibiotic resistance calls for the development of novel antibacterial agents to fight infectious diseases.[62] Most of the conventional drugs that inhibit cell wall synthesis affect the extracellular steps of peptidoglycan formation.[63] These steps are catalyzed by soluble enzymes. Almost no drugs have been developed against integral membrane proteins involved in the intracellular synthesis of translocation precursors. MraY (Phospho-*N*-acetylmuramoyl-pentapeptide-transferase, Translocase I) is such an integral membrane enzyme that catalyzes the reaction of Park's nucleotide with the membrane-bound lipid carrier undecaprenyl phosphate to give lipid I (see **Fig. 2-1**).[64–67]

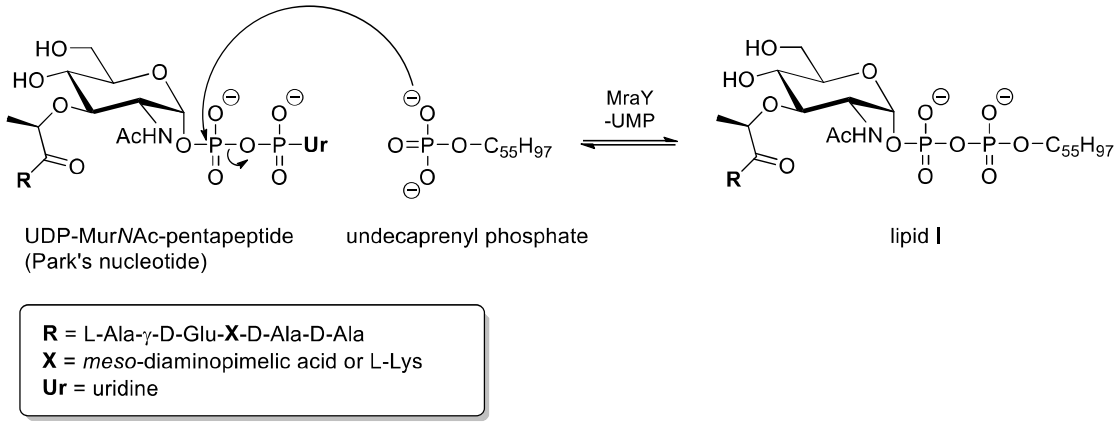


Fig. 2-1 *Reaction catalyzed by MraY (Translocase I).*

The reaction shown in **Fig. 2-1** is inhibited by nucleoside antibiotics, such as muraymycins[68], mureidomycins[69], caprazamycins[70] and others. These natural products all have a uridine-derived core structure, but are otherwise structurally rather diverse, including different peptide moieties. More information on the structures and

actions of nucleoside antibiotics targeting MraY can be found in reviews by WIEGMANN et al.[71], KIMURA et al.[72], WIKER et al.[73] or BUGG and KERR[74].

2.1.1 Structural information

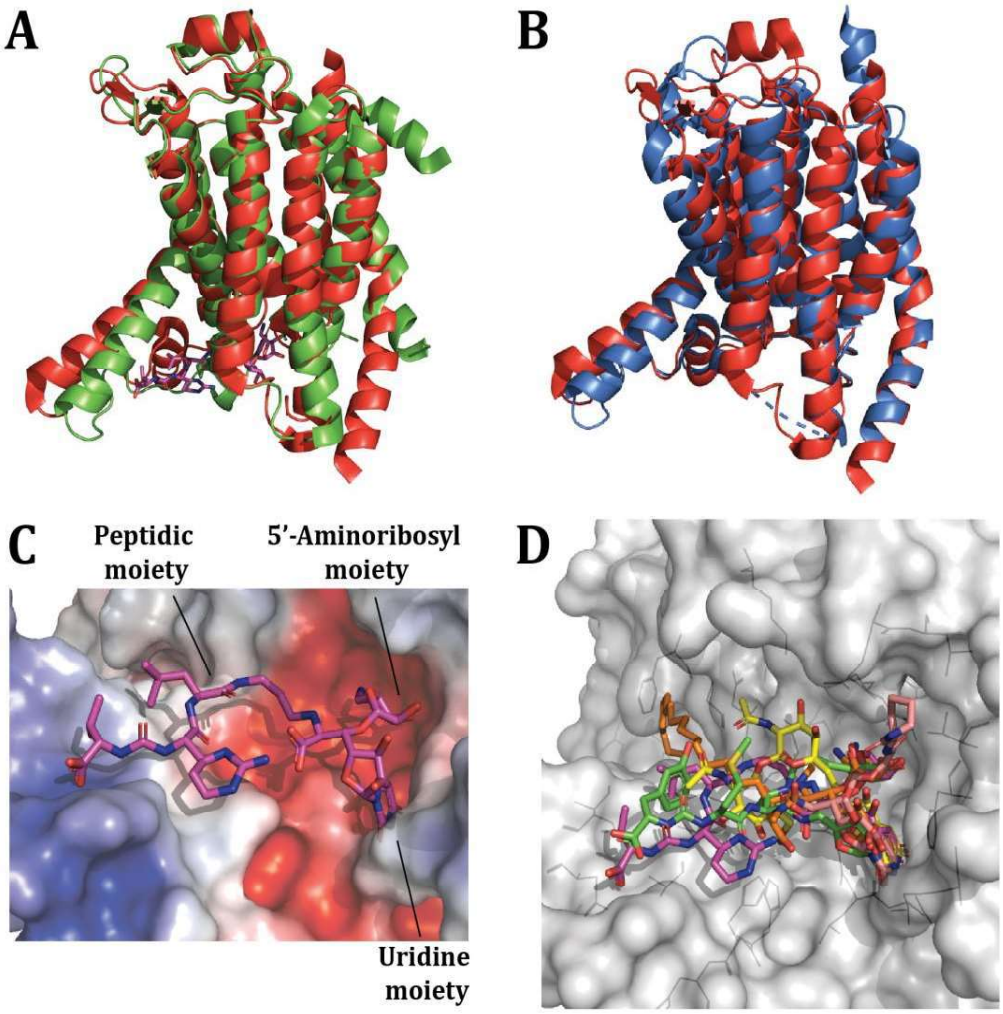


Fig. 2-2 *Overlay of apo-MraY$_{AA}$ (green, PDB code: 4J72) and MraY$_{AA}$ bound to muraymycin D2 (protein red, MD2 in magenta, PDB code: 5CKR) shows significant structural changes upon ligand binding.* ***B.*** *Overlay of MraY$_{AA}$ bound to muraymycin D2 (red) and MraY$_{CB}$ bound to tunicamycin (blue, PDB code: 5JNQ) shows a similar conformational change for the binding of different structures.* ***C.*** *Zoom in on the MraY – muraymycin D2 binding interface. Well-defined binding pockets are occupied by the uridine moiety and 5'-aminoribosyl moiety, while the peptidic moiety binds on the protein's surface.* ***D.*** *Overlay of all published MraY binders (MD2 [magenta], tunicamycin [yellow], carbacaprazamycin [orange, PDB code: 6OYH], capuramycin [salmon, PDB code: 6OYZ], 3'-hydroxymureidomycin A [green, PDB code: 6OZ6]).*

MraY has long been considered an interesting target for the development of new antibiotics[75], but the lack of structural information has hampered drug development efforts. Following the X-ray crystal structure of MraY (i.e., the ligand-free apoprotein, PDB ID: 4J72) from *Aquifex aeolicus* ($MraY_{AA}$)[76], an X-ray crystal structure of the protein in complex with its natural occurring inhibitor muraymycin D2 (MD2, PDB ID: 5CKR[77]), provided new insights into the ligand-target interaction. A complete picture of interaction "hot spots" between MraY and naturally occurring nucleoside antibiotics was provided by HAKULINEN et al.[78] and MASHALIDIS et al.[79] when they created more co-crystal structures of the enzyme. These works revealed tremendous conformational changes of MraY upon ligand binding; helices rotate away from the binding site to create room for the ligand.[77–79] It is worth noting that a conserved HHH motif at the C-terminus of the protein forms a helix upon ligand binding (see **Fig. 2-2 A**). The second and third histidine is conserved among most species of bacteria, while the first histidine is replaced by a glutamine in *Streptomyces sp*., the producers of MraY-inhibiting natural products such as muraymycins[68] or tunicamycin[80]. Implications of this observation will be discussed in detail later.

The second co-crystal structure in the protein database (i.e. the complex of MraY from *Clostridium bolteae* [$MraY_{CB}$] and tunicamycin, PDB ID: 5JNQ[78]) reveals a similar conformation of the enzyme when compared to the complex structure of *Aquifex aeolicus* MraY bound to muraymycin D2 (see **Fig. 2-2 A** and **B**). The binding mode for muraymycin D2 claims three major chemical motifs for efficient binding: The uridine moiety and the 5'-aminoribosyl moiety, which act as key interactions in well-defined binding pockets within the enzyme, and the peptidic moiety, which extends on the protein surface and contributes to binding and specificity (see **Fig. 2-2 C**). A closer look at an alignment of the two ligands muraymycin D2 and tunicamycin emphasizes the importance of the uridine moiety, as this specific motif aligns perfectly in both structures. More flexibility seems to be present in the 5'-aminoribosyl moiety, as it is not present in tunicamycin (see **Fig. 2-2 D**). More recently, MASHALIDIS et al. reported a method to obtain ternary crystal complexes of $MraY_{AA}$ with a nanobody labeled NB7 and the nucleoside antibiotics carbacaprazamycin, capuramycin and 3′-hydroxymureidomycin A.[79] The previously described structural rearrangement of the MraY-muraymycin D2 complex was also

found in the newly reported protein-ligand structures (PDB IDs: 6OYH, 6OYZ and 6OZ6). Variation between different nucleoside antibiotic classes can be found in the degree of bending in transmembrane helix 9b (TM9b) and conformation of loop E. These new findings revealed that the only defined common binding pocket for MraY-targeting nucleoside antibiotics is the cytoplasmic face of MraY that binds the uracil motif. Six "hot spots" (HS) adjacent to the uridine binding pocket were identified be MASHALIDIS et al.; HS1 consists of T75, N190, D193 and G264 and is involved in the binding of all five reported natural products, therefore representing the most druggable site on MraY. Noteworthy, the key interaction found between the 5'-aminoribosyl prong of muraymycin D2 and HS1 of MraY appears to be replaceable by a variety of different chemical moieties (see **Fig. 2-2 D**), as long as the geometry of the molecule allows for the "plug-like" shape that binds HS1 and the uridine binding site simultaneously.

2.1.2 Protein-protein interactions

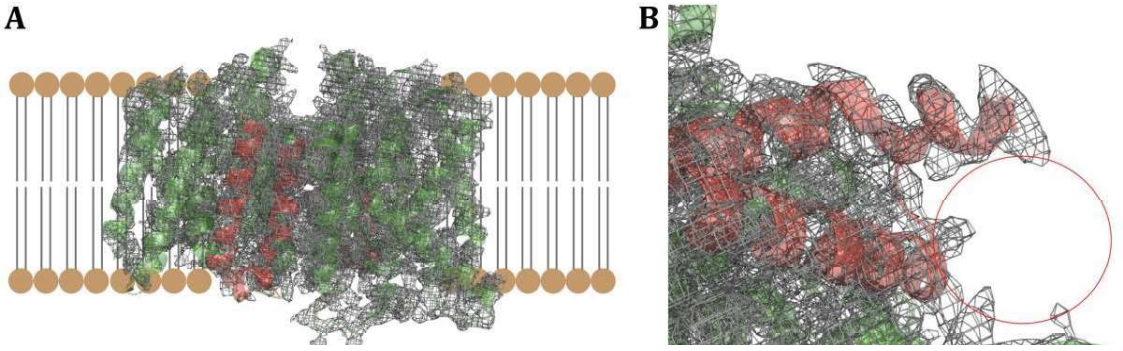


Fig. 2-3 A. *Electron density distribution obtained from apo-MraY$_{AA}$ crystal by B. C. Chung, J. Zhao, R. Gillespie, D. Y. Kwon, Z. Guan, J. Hong, P. Zhou, S.-Y. Lee,* Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. Biophys. J. ***2014***, 106, *14a. MraY across the membrane is shown in green, TM1 and TM2 are highlighted in red.* ***B.*** *Zoom in on the missing density between TM1 and TM2, highlighted with a red circle*.

It has already been reported that MraY displays several key interactions with other membrane embedded proteins such as the chaperone SlyD, which are disrupted by its interaction with the lysis protein E of bacteriophage ΦX174.[81,82] However, the binding interface on MraY has not been determined. Additionally, investigations of the bacterial morphogenetic proteins MreB and MreD showed the possibility of the formation of a multienzyme complex within bacterial cell walls including MraY, MurF, MurG, RodA,

RodZ, FtsZ and others.[83] Furthermore, the published crystal structures lack density in loop A, which connects TM1 and TM2 (see **Fig. 2-3**), suggesting the need for a stabilizing interaction partner such as MurF or MurG.[84] Also, this loop is directly adjacent to the nucleoside inhibitor binding site. It is therefore reasonable to assume that binding of another protein to this region can alter the interaction of MraY with natural products. MurF and MreB have already been characterized as scaffold proteins for other Mur ligases, including MurC, MurD, MurE and MurF.[85,86]

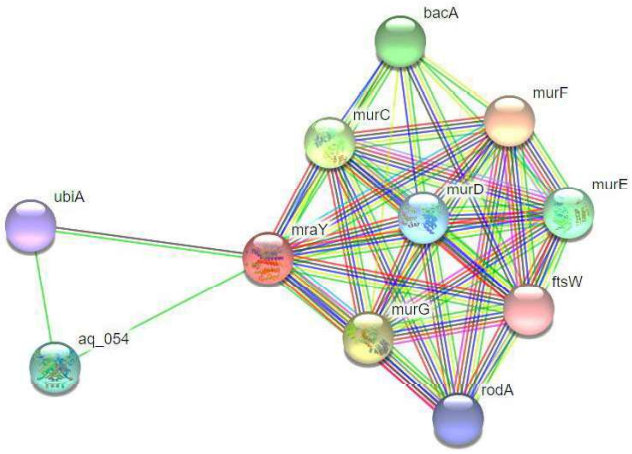


Fig. 2-4 *Interaction network of the mraY gene obtained from the open source software STRING.*

A similar interaction network is described by the bioinformatical software STRING. STRING scores potential protein-protein interactions based on genomic neighborhood, databases of known metabolic pathways, gene fusion, occurrence patterns of gene families and proteins whose genes are observed to be correlated in expression.[87] **Fig. 2-4** shows the interaction map of the *mraY* gene.

2.1.3 Antibiotic resistance mechanisms

Ever since sulfonamides were first introduced as drugs to treat infectious diseases caused by bacteria in 1936[88], microbes found ways to develop resistances against these treatments. Resistances towards these antibiotics developed by bacteria were reported as early as in the late 1930s.[89,90] Since antibiotics are developed to eliminate bacteria, selective pressure encourages the microorganisms to develop resistances.[91] This process occurs on a genetic level.[92] Through their ability to transfer genetic material directly between each other, resistances can easily spread within a population. The numerous different mechanisms which microbes employ to escape treatment will be discussed below.

2.1.3.1 Modifications of the drug

The biosynthesis of enzymes capable of altering antibiotic molecules by chemical modifications has proven to be an effective way to develop antibiotic resistance. Interestingly, this mechanism is mostly used against ribosome-targeting drugs.[93] The modifying enzymes display their actions through acetylation (aminoglycosides, chloramphenicol, streptogramins), phosphorylation (aminoglycosides, chloramphenicol) and adenylation (aminoglycosides, lincosamides).[94] The effect of the alteration is likely related to steric impediments in the original binding pockets.

Rather than being modified by the addition of chemical moieties, the alteration can also be caused through the enzymatic degradation of the drug. Most famously, β-lactamases are enzymes developed by bacteria to hydrolyze the β-lactam ring of penicillins. They have been discovered in the early 1940s, though they might have existed for even longer.[95,96] To overcome this mechanism, hydrolysis-resistant β-lactams such as methicillin, cephalosporins or carbapenems were developed. However, in the 1960s new β-lactamases emerged capable of hydrolyzing these new molecules.[97]

2.1.3.2 Decreased cell permeation and efflux

Decreased cellular uptake of antibiotics is a resistance mechanism which is particularly important for Gram-negative bacteria, since their cell walls include two plasma membranes, separated by a periplasmic space.[98,99] Especially hydrophilic substances are affected by changes in permeability of the outer membrane of Gram-negative bacteria. For this reason, antibiotics such as vancomycin for example are not effective against Gram-negative bacteria.[100] Porins needed for the uptake of substances can be altered by bacteria to achieve antibiotic resistance, e.g. by shifts in the expressed porin types, change in expression levels or impairment of porin function.[101]

Development of efflux systems capable of pumping toxins out of bacterial cells have been discovered as a resistance mechanism for tetracycline in *E. coli* in the early 1980s.[102] These complex machineries might be specific for substrates or recognize broader varieties of molecules, typically found in multi drug resistance strains.[103]

2.1.3.3 Changes of target sites

Tetracycline antibiotics exhibit their activity through binding of the 30S subunit of the bacterial ribosome, thus inhibiting protein translation.[104] Resistance mechanisms can be employed for example by safeguarding bacterial enzymes through expression of protecting proteins. A mechanism commonly used against tetracyclines is protection of the ribosome by expression of GTPases that act as elongation factors. Tet(M) and Tet(O) for example bind the ribosome and dislodge tetracycline, thus protecting the ribosomal machinery from inhibition.[105] Another example represents the quinolone resistance protein Qnr, which protects its target by competing with quinolones for the DNA binding site of the DNA gyrase and topoisomerase IV.[106,107]

One of the most common resistance mechanisms however is the modification of the target site itself. This modification can be introduced by enzymatic alterations of the binding site, replacement of the target or point mutations. For example, it has been suggested that macrolide resistance is facilitated by methylation of the ribosome in position A2058 of the domain V of the 23rRNA of the 50S ribosomal subunit.[108,109] Overproduction of the antibiotic target is another strategy employed to achieve trimethoprim-sulfamethoxazole resistance. In this case, genetic mutations lead to the massive overproduction of dihydrofolate reductase in response to trimethoprim treatment.[110] Also, a variety of different genetic changes have been reported as a response to single-point mutations in the *rpoB* gene after rifamycin treatment. The replacement of amino acids results in decreased affinity of the drugs, while not altering the catalytic activity of their target, the RNA polymerase in this case. Notably, *Amycolatopsis mediterranei* employs this strategy to generate autoresistance towards its synthetic natural product rifamycin B.[111]

As an essential membrane enzyme, MraY isoforms are found in all classes of bacteria. This raises the obvious questions of how a bacterial producer of natural products facilitates self-resistance mechanisms. Notably, the possibility of single-point mutations to render inhibitors ineffective has already been reported for MraY.[77] Of particular interest were the $MraY_{AA}$ mutants D193N and F262A, which retained enzymatic activity while reducing inhibitory activity of muraymycin D2 (MD2) at a concentration of 1 µM

by over 75 % (K_D of MD2 = 17.2 nM). Other mutants of $MraY_{AA}$ (N255A and Q305A) showed similar tendencies: Both mutants were enzymatically active, but protein inhibition of MD2 at 1 µM was reduced by ~ 50 %. Interestingly, the D193A mutant displayed no enzymatic activity, highlighting that persisting mutations cannot be random.[77] Recently, Cui et al. reported the role of covalently modified (i.e. phosphorylated and adenylated) precursors in self-resistance.[112] These modifications

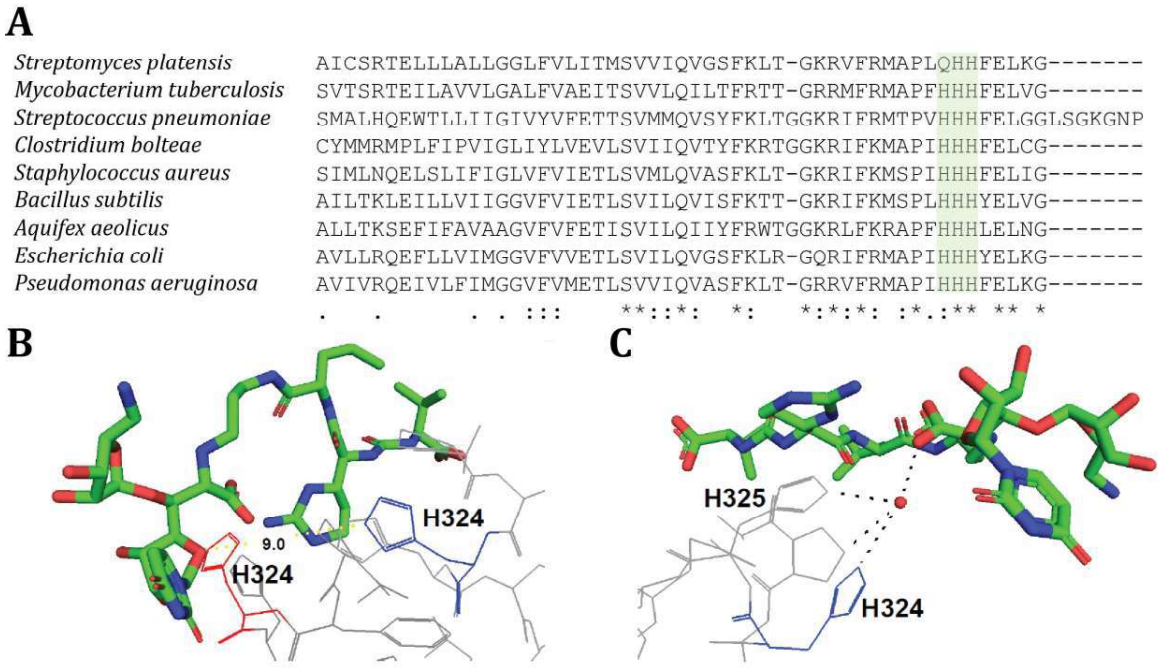


Fig. 2-5 ***A.*** *Alignment of the C-terminal region of Streptomyces platensis MraY with other MraY isoforms, generated using Clustal Omega (1.2.4.).* ***B.*** *Conformational change upon MD2 binding to $MraY_{AA}$ induces a repositioning of H324 over a distance of 9 Å. MD2 is shown in green, H324 from the apo-protein structure (B. C. Chung, J. Zhao, R. Gillespie, D. Y. Kwon, Z. Guan, J. Hong, P. Zhou, S.-Y. Lee,* Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. Biophys. J. ***2014***, 106, *14a.) in red and H324 from the MD2-bound structure, PDF code: 5CKR (B. C. Chung, E. H. Mashalidis, T. Tanino, M. Kim, A. Matsuda, J. Hong, S. Ichikawa, S.-Y. Lee,* Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. Nature ***2016***, 533, *557–560.) in blue.* ***C.*** *H324 and H325 participate in a water mediated hydrogen bond forming to MD2*.

lead to notable reductions in IC_{50} values of these precursors compared to their parent compounds determined in MraY activity assays. However, even though reduced, there is still inhibitory activity found in these precursors[112], suggesting additional mechanisms of self-resistance.

Due to high conservation of the catalytic center of MraY between *Streptomyces sp.* and other isoforms, protein mutations have not received a lot of attention in the past. However, emerging structural information of natural product binding modes reveal other residues involved in ligand binding.[77–79] Notably, *Aquifex aeolicus* MraY H324 is a very well conserved histidine among most MraY isoforms, except for MraY from *Streptomyces sp.*, where a glutamine residue is found instead (see **Fig. 2-5 A**).[113] Furthermore, alignments of *A. aeolicus* apo-MraY and *A. aeolicus* MraY bound to muraymycin D2 reveal the conformational flexibility and importance of this specific residue (see **Fig. 2-5 B** and **C**). This observation leads to the suggestion that a single point mutation from histidine to glutamine could compromise the conformational flexibility of MraY required for substrate specificity, whilst retaining enzymatic activity, a mechanism that has been described for other proteins in the past.[114–117]

2.2 NFAT

Coordinated assembling of transcription factors and DNA-binding proteins is crucial for achieving efficient regulation of gene expression in eukaryotes. The presence of diverse transcription factors in the emerging multiprotein DNA complexes, so called enhanceosomes, is necessary for time-related and tissue specific gene regulation. Therefore, numerous signaling pathways resulting in distinct activation of each transcription factor within enhanceosome-complexes are required and offer a logical explanation for specific cellular responses to various stimulations.[118]

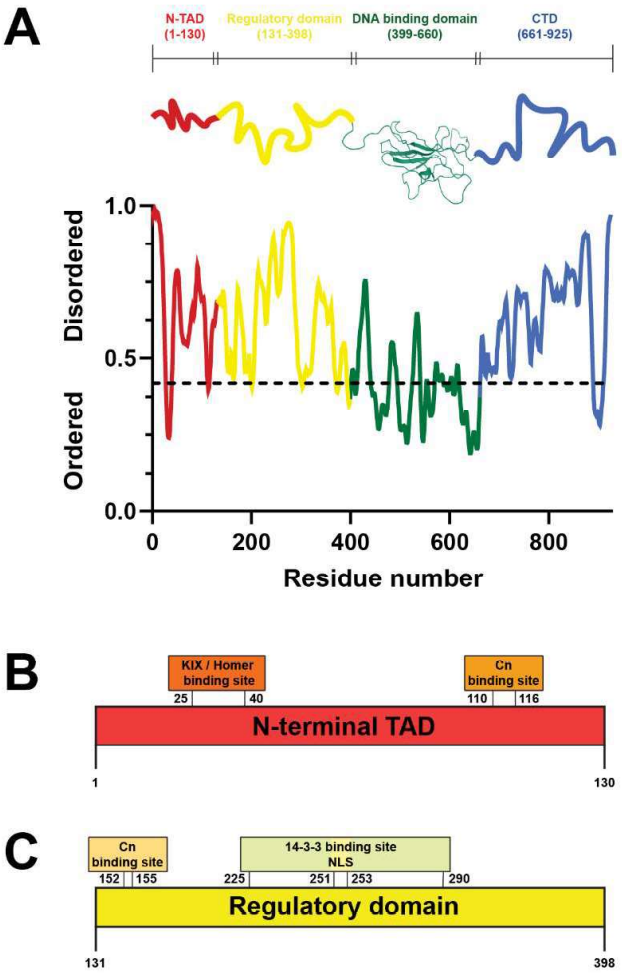


Fig. 2-6 A. *Domain architecture of NFAT: The DNA-binding domain (green) is the only structured domain of NFAT with available structural information. Structure taken from S. A. Wolfe, P. Zhou, V. Dötsch, L. Chen, A. You, S. N. Ho, G. R. Crabtree, G. Wagner, G. L. Verdine,* Unusual Rel-like architecture in the DNA-binding domain of the transcription factor NFATc. Nature ***1997***, 385, *172-176. The N- and C-terminal domains are predicted to be disordered by PrDOS (T. Ishida, K. Kinoshita,* PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. ***2007***, 35, *W460–W464).* ***B.*** *Sequence of NFAT's transactivation domain including interaction motifs.* ***C.*** *Sequence of NFAT's regulatory domain including interaction motifs and the position of the nuclear localization sequence (NLS).*

One of these transcription factor families involved in different immunological processes is the nuclear factor of activated T-cells (NFAT, see **Fig. 2-6**). They are present in T lymphocytes of vertebrates and control gene expression of several key genes involved in T cell activation, such as interleukin-2 (IL-2), interferon gamma (IFN-γ), tumor necrosis factor α (TNF-α) and many more.[119–122] Upon T cell receptor activation, calcium influx induces the serine/threonine phosphatase calcineurin (Cn) to dephosphorylate multiple serine residues in the ~ 200 amino acid long NFAT regulatory domain. Joint nuclear translocation of NFAT and Cn occurs once NFAT is dephosphorylated on all but one phosphoserine.[123] The mode of action of immunosuppressant drugs such as Cyclosporine and Tacrolimus is inhibiting Cn from dephosphorylating and activating NFAT after binding of the corresponding Calcineurin-inhibitor immunophilin complexes.[124]

So far, five different members of the NFAT family have been identified: NFATC1 (NFAT2), NFATC2 (NFAT1), NFATC3 (NFAT4), NFATC4 (NFAT3)[125] and NFAT5[126]. This work focuses only on NFAT1, which is the primary family member found in T lymphocytes and plays the most prominent role in T cell activation.[127] It has been shown that mice lacking NFAT1 and NFAT4 develop a hyperproliferative phenotype due to impaired FasL expression in thymocytes.[128] Moreover, NFAT-driven transcription of lineage-specific transcription factors is necessary for T-helper-cell differentiation[129] and for T cell activation[130].

The four conserved domains of NFAT1 (shown in **Fig. 2-6 A**), from here on referred to as NFAT, are described below.

1. **N-terminal transactivation domain (N-TAD, $NFAT_{1-130}$)**

The structure of this region is predicted to be intrinsically disordered by the PrDOS bioinformatics server.[131] It arbitrates recruitment of transcriptional co-activators such as CBP/p300 or Med15 to induce promoter-specific assembly of the transcriptional machinery.[132,133]

2. **Regulatory domain ($NFAT_{131-398}$)**

This domain is also predicted to be intrinsically disordered. It contains a nuclear localization signal (NLS), which gets exposed upon calcineurin-dependent dephosphorylation. The region is highly phosphorylated by multiple kinases in

serine/proline rich regions (SRR, SP1, SP2, and SP3). The two Cn binding sites are represented by the highly conserved motifs PxIxIT and LxVP, flanking the regulatory domain.[134,135]

3. **DNA-binding domain (NFAT$_{399-660}$)**

The DNA-binding domain is the only part of the protein which is predicted to be structured and for which structural information is available.[136–138]

4. **The C-terminal domain (NFAT$_{661-925}$)**

The C-terminal domain contains a nuclear export signal (NES) and is proposed to function as a TAD, but so far is poorly understood. It is also predicted to be disordered.

2.2.1 Phosphorylation of NFAT

NFAT is heavily phosphorylated in its regulatory domain in resting T cells to prevent it from entering the nucleus. Previous work identified at least five kinases that phosphorylate NFAT; casein kinase 1 (CK1), glycogen synthase kinase 3 (GSK-3), protein kinase A (PKA), dual specificity tyrosine phosphorylation regulated kinase (DYRK) and p38 mitogen activated protein kinase (p38 MAPK).[58,122,139–143] GSK-3 activity requires prior phosphorylation by a priming kinase such as PKA.[144] More kinases such as cyclin-dependent kinases (CDKs), casein kinase 2 (CK2), protein kinase C (PKC), other MAPKs (JNK, ERK) and Ca^{2+}/calmodulin-dependent protein kinase II (CamKII) are likely NFAT-targeting kinases as well, as suggested by phosphorylation prediction tools or reported phosphorylation events of other NFAT-family members.[145–150] A distinction is drawn between maintenance and export kinases; maintenance kinases assure phosphorylation status of NFAT in the cytosol, whereas export kinases phosphorylate nuclear NFAT and mediate its relocation to the cytoplasm.[122]

Phosphorylation takes place at a minimum of 14 conserved serine residues throughout the regulatory domain of NFAT.[58] Upon T cell receptor (TCR) activation, rising Ca^{2+} levels activate calmodulin, which in turn activates calcineurin to dephosphorylate all phosphoserine residues but one (in the NLS region).[58] The presence of several different export kinases suggests differential regulation of individual NFAT isoforms in a single cell type. The development of skeletal muscle cells for example shows selective nuclear localization of NFAT2, while NFAT1 and NFAT4 are located primarily in the

cytoplasm.[151] In T cells, T cell receptor stimulation triggers calcium influx which then causes NFAT nuclear import. However, the stimulation of the costimulatory receptor CD28 induces activation of the PI-3 kinase pathway, leading to activation of the kinase Akt/PKB which promotes inhibitory phosphorylation of GSK-3. As a result, NFAT nuclear export is inhibited by decreased GSK-3 activity.[152–154]

Once dephosphorylated, NFAT nuclear transport is accompanied by relocalization of calcineurin from the cytoplasm to the nucleus. It is required in stimulated cells to maintain the dephosphorylated status and nuclear localization of NFAT.[155] If Ca^{2+} entry is prevented, NFAT rephosphorylation is achieved quickly and phosphorylated NFAT rapidly leaves the nucleus ($t_{1/2} \sim 15$ min).[156–158]

2.2.2 Relevance of the intrinsically disordered N-terminal transactivation domain ($NFAT_{1\text{-}130}$)

The N-terminal portion of NFAT contains a highly conserved nine-amino-acid (9aa) transactivation domain (TAD) motif: $_{26}$DELDFSILF$_{34}$. The 9aaTAD is common to a large superfamily of eukaryotic transcription factors, including NFAT, NF-κB, p53 and others. These motifs are known to directly interact with coactivator and mediator proteins such as the KIX domains of Med15 or CBP/p300.[133,159,160]

Furthermore, the first calcineurin binding site is located within the N-terminal TAD of NFAT: $_{110}$SPRIEIT$_{116}$. It is located N-terminal relative to the highly phosphorylated regulatory domain and binds calcineurin with micromolar affinity.[134,161] Blocking the interaction of calcineurin by inhibition with a high-affinity VIVIT peptide decreased NFAT activity but did not prevent NFAT nuclear localization completely.[162] Thus, it was concluded that interaction with the PxIxIT site alone is not sufficient for efficient NFAT activation. A second binding site $_{369}$LLVP$_{372}$ was identified later.[135,163–166]

In addition to CBP/p300 and Med15, the EVH1 domain of Homer3 was previously reported to bind NFAT somewhere between amino acids 1-102.[167] The Homer family of cytoplasmic scaffolding proteins was found to negatively regulate T cell activation. HUANG et al. state that Homer3 binding is competitive with calcineurin binding.[167] However, more work needs to be done to elucidate the role of Homer proteins in the regulation of T cell responses and their function in anergy and antigen-response.

2.2.3 Relevance of the intrinsically disordered regulatory domain ($NFAT_{131-399}$)

The regulatory domain of NFAT harbors most of its phosphorylation and is mainly responsible for the subcellular distribution of the transcription factor.[58] The phosphorylation status of NFAT regulates DNA-binding affinity as well as exposure of the nuclear localization (NLS) and nuclear export (NES) signals.[58,168] As shown in **Fig. 2-6**, the NLS ($NFAT_{251-253}$) is located within the regulatory domain. In a phosphorylated state, this sequence is masked in a way that is not well understood yet. Dephosphorylation of all but one phosphoserine is required in order to cause full exposure of the NLS and keep NFAT in an active conformation. At the same time, the NES ($NFAT_{904-913}$), which is located within the C-terminal domain, is concealed in unphosphorylated NFAT and becomes exposed after phosphorylation.[58] How all of this occurs on a structural level is not well studied yet.

Furthermore, the regulatory domain harbors the second conserved calcineurin binding site $_{369}$LLVP$_{372}$. Two schools of thought exist on the mechanism of calcineurin binding to the second binding site. Early findings of RODRIGUEZ et al. located the LxVP binding site on the interface of the two calcineurin subunits A and B.[166] Later, GRIGORIU et al. showed that an LxVK sequence of the viral A238L protein, which is representative of NFAT's LxVP sequence, also binds at a different position than the PxIxIT motif.[164] On the other hand, GAL et al. presented data showing joint binding of both PxIxIT and LxVP motifs on the same region of calcineurin located at the calcineurin subunit A surface.[165] Therefore, many questions of the mechanism of NFAT dephosphorylation remain unanswered and require further investigation.

2.3 eIF4E

The eukaryotic translation initiation factor 4E (eIF4E) is a master regulator of cap-dependent protein synthesis. Binding of the mRNA cap to eIF4E is largely considered to be the rate-limiting step of translation initiation.[169,170] Although of vital importance for translation initiation, eIF4E is only one of three proteins that comprise the translation initiation complex eIF4F. eIF4F is, in addition to eIF4E, subdivided into eIF4A (a DEAD

box RNA helicase) and the large scaffold protein eIF4G.[171–174] Upon binding of m^7-GTP cap structures at the 5'-end of mRNA transcripts by eIF4E, the eIF4F complex assembles to recruit the 40S ribosomal subunit and unwind secondary structures of 5'-untranslated regions (5'-UTRs) of mRNAs.[175,176] Cap-dependent translation initiation is a process highly controlled by a variety of protein-protein interactions; eIF4A activating proteins eIF4B and eIF4H and eIF4A inhibiting proteins such as programmed cell death protein 4 (PDCD4)[177–181]; eIF4E inhibiting proteins 4EBP1-3, which themselves are regulated by kinases[182–186] and many eIF4G binding proteins (e.g. MAPK interacting protein kinases [Mnks] that phosphorylate eIF4E or poly(A)-binding protein PABP)[187–189].

Many mRNAs preferably selected by eIF4E are long, highly structured transcripts. These typically belong to cancer-related genes involved in cell growth, cell cycle progression or angiogenesis (c-MYC, VEGF, CCND1 etc.).[190]

2.3.1 Regulation of eIF4E

The eIF4F complex lies in the middle of major cell signaling regulating pathways such as the PI(3)K/AKT/mTOR pathway and the RAS/RAF/MEK/ERK/MNK MAPK pathway, both of which are highly involved in cancer development.[191,192] eIF4E is regulated by 4E binding proteins (4EBPs), which are substrates for mTORC1.[193,194] Once phosphorylated by mTORC1, 4EBPs are unable to bind eIF4E, thus enabling 4E to bind to eIF4G and form the translation initiation complex.[175,195] Mnk kinases (mitogen-activated protein kinases [MAPK] interacting protein kinases) are downstream targets in the ERK pathway.[196] Upon activation, Mnks bind to eIF4G and phosphorylate eIF4E on a single amino acid (S209).[189] The implications of this phosphorylation are complex and subject of heavy debate and will be discussed in more detail later.

Furthermore, eIF4E and the other two members of the eIF4F complex are regulated at the transcriptional level by one of the most prominent oncogenes: *c-myc*. The transcription factor MYC is frequently activated in most human cancers and was shown to increase the transcription rates of eIF4E, eIF4A and eIF4G.[197]

It has also been shown that eIF4E activation is regulated by ubiquitination through, for example, the carboxyl terminus of heat shock cognate protein 70-interacting protein (Chip), which targets proteins for proteasomal degradation. It has been suggested that

ubiquitination at Lys159 is reduced upon binding of 4EBPs or heat-shock protein 27 (Hsp27), adding another regulation mechanism to this system.[198,199]

2.3.2 Role of eIF4E beyond translation initiation

Notably, cellular localization of eIF4E is not restricted to ribosomes, but it is also found in the nucleus and in P bodies.[200,201] This suggests involvement of the protein in processes besides translation initiation. P bodies are aggregates within eukaryotic cells, consisting of enzymes that degrade mRNA.[202] eIF4E's function in P bodies is likely the protection of decay-dependent degradation of specific mRNAs.[203] Moreover, eIF4E is found in stress granules.[204] These are aggregations of proteins and RNA that appear when a cell experiences stress.[205] They contain stalled translation complexes that hold mRNAs until a cellular decision is made to continue translation, engage degradation or storage of the RNA.[206] In the nucleus, eIF4E's main function is the export of specific mRNAs.[201,207] These transcripts share a common 50-nucleotide sequence at the 3'-UTR known as the 4E-SE (4E sensitive element).[208] To facilitate this function, eIF4E binds to cofactors like chromosomal maintenance 1 (CRM1), which functions as a transport receptor.[209] Another control mechanism for nuclear mRNA export which is employed by eIF4E is its ability to significantly reduce RANBP2 (RAN binding protein 2) levels, which changes the composition of nuclear pore complexes (NPCs).[210]

Other, non-cap-binding related functions of eIF4E are its interaction with histone H4 through its N-terminal domain, which can be formed without the need of a cap.[211] Also, binding of $m^{2,2,7}G$ (trimethylguanosine, TMG) caps that are found at the 5'-end of small nuclear RNAs (snRNAs) was reported, indicating a novel physiological role for eIF4E that is not well understood yet.[207]

2.3.3 eIF4E phosphorylation

Generally, eIF4E phosphorylation is associated with cell growth and increased translation rates. Specifically, phosphorylation at Ser209 has been reported to carry tumorigenic properties.[212,213] FURIC et al. showed the importance of eIF4E phosphorylation in a mouse model with knock-in mice, in which serine 209 was mutated to alanine, thus creating an eIF4E-phosphodeficient organism.[214] Mice having a deletion

of the tumor suppressor *Pten* in the prostate epithelium (PtenFlox/Flox; PB-Cre4) would typically develop invasive carcinomas at 5-8 months of age. *Pten*-conditional knock-out mice crossed with eIF4E-S209A knock-in mice were shown to become resistant of induced prostate cancer, thus highlighting the important role of eIF4E phosphorylation.[214] In agreement with that, human prostate cancers display increasing amounts of eIF4E and phosphorylated eIF4E with development of those cancers.[214,215] More recently, eIF4E-S209A knock-in mice were reported to have become resistant towards polyoma middle T-driven mammary tumors. Additionally it was shown that eIF4E phosphorylation drives translation of Snail and Mmp3 mRNAs, which are required for metastasis.[216]

As mentioned before, eIF4E is phosphorylated at S209 by Mnk kinases. These are activated by p38 and ERK1/2 MAPKs. Not surprisingly, it has been shown that Mnk kinases are frequently upregulated in tumor cells.[217] Phosphorylation of eIF4E is achieved upon the formation of a ternary complex. Activated Mnk binds at the C-terminal region of eIF4G, thus providing proximity between Mnk and eIF4E, which is bound to the HEAT3-domain at the N-terminus of eIF4G.[189,218]

Although as of now no structure of phosphorylated eIF4E was solved, the structural implications of S209 phosphorylation have been addressed by molecular dynamics studies (see **Fig. 2-7**).[219] S209 is located in the loop between the β7 and β8 strands towards the C-terminus of the protein, near the cap binding site. It has been shown that phosphorylated eIF4E has reduced affinity to capped mRNA.[220] One possible explanation for this finding is the proximity of the phosphoserine to the triphosphate of the cap, resulting in repulsion of the negative charges. Another explanation, if it is assumed that eIF4E is phosphorylated before a cap structure is bound, is the existence of a salt-bridge between the phosphoserine and K159 (located in the loop segment between the β5 and β6 strands) on the other site of the interaction interface. Cap binding would interrupt this interaction, resulting in unfavorable binding conditions. This theory is supported by the findings of LAMA et al.[219] Furthermore, the β7β8 loop that harbors the phosphorylation site is highly flexible, as can be deducted from the fact that the loop is

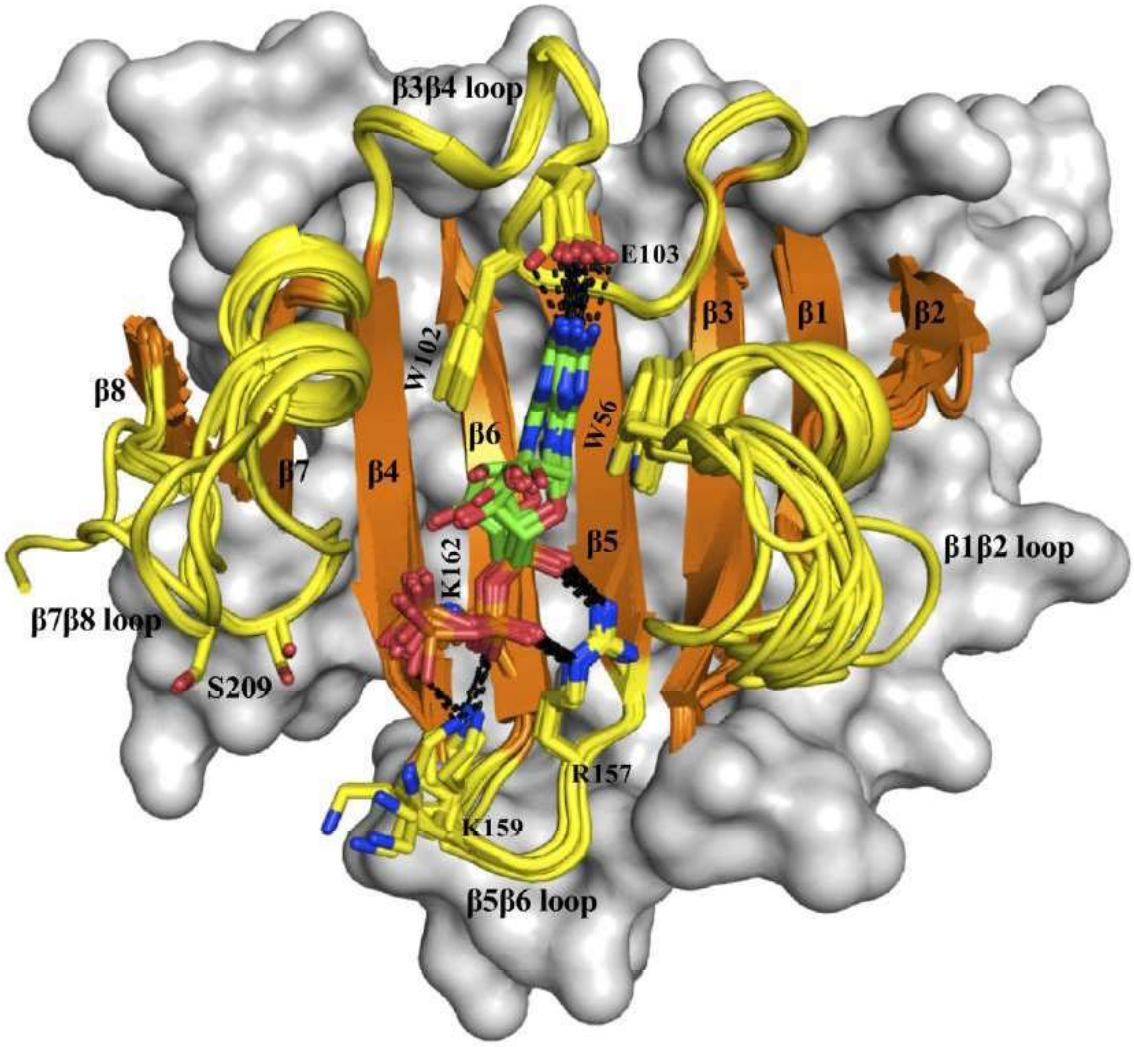


Fig. 2-7 *Overlay of sixteen eIF4E structures with m*^{7}*-GTP bound. Beta strands and loops are shown in orange and yellow, respectively, in cartoon representation. The rest of the protein is displayed in surface representation in grey. Figure taken from D. Lama, C. S. Verma,* Deciphering the mechanistic effects of eIF4E phosphorylation on mRNA-cap recognition**.** Protein Sci**.** ***2019****, 29, 1373-1386.*

not resolved in most crystal structures and large scale motions of the loop were reported in the only available solution NMR structure (PDB ID: 2GPQ).[221] However, these *in silico* findings have not sufficiently been validated experimentally and more work needs to be done to fully elucidate the implications of eIF4E phosphorylation.

2.3.4 Targeting eIF4E

Various approaches have been tried to target eIF4E as anticancer treatment. The first class of molecules targets the synthesis of eIF4E. In this case, eIF4E mRNA is directly targeted, either by siRNA/shRNA or antisense oligonucleotides.[222–224] This strategy has been proven effective for a variety of different cancers, such as ovarian cancer cells[225], colorectal cancers[226], non-microcytic lung carcinoma cells[227] and increased drug sensitivity in hepatocellular carcinoma cells[228]. However, RNA treatment still suffers from drawbacks such as disadvantageous pharmacokinetics[229], which is why additional therapeutic avenues are desirable.

The second type of drug targets the eIF4F complex. In particular, the protein-protein interaction interface between eIF4E and eIF4G presents an attractive target for specific inhibition efforts. 4EGI-1 is a small molecule that specifically inhibits this interaction by allosterically binding eIF4E, causing a conformational change which gives rise to the dissociation from eIF4G.[230,231] 4E1RCat is another small molecule inhibitor, which works similarly to 4EGI-1.[232] A competitive inhibitor of the eIF4E-cap interaction is represented by the antiviral drug Ribavirin as an m^7-GTP analogue. Ribavirin has been shown to have promising effects in AML patients and nasopharyngeal carcinoma cells.[233,234] Interestingly, one of the effects of the cardiac glycoside Ouabain that does not arise from its Na^+/K^+-ATPase inhibition, but rather its downregulation of HIF-1. Recently, it was shown that this phenomenon is due to a direct disruption of the eIF4E/eIF4G interaction by Ouabain.[235]

The last class of drugs that target eIF4E are those that affect its phosphorylation. Among these, Mnk inhibitors (i.e. CGP57380, Merestinib, and Cercosporamide) directly target the kinase responsible for eIF4E phosphorylation.[236–238] CGP57380, for example, has been shown to display activity in chronic myeloid leukemia cells.[239]

2.3.5 The PROTAC approach for eIF4E targeting

PROteolysis TArgeting Chimeras (PROTACs) have emerged over the last couple of years as promising potential therapeutics for a variety of different diseases.[240] PROTACs are bifunctional molecules that bind proteins of interest (POI) and connect them via linkers to a ligand for E3 ubiquitin ligases. A ternary complex between the PROTAC, the POI and the ubiquitin ligase is formed, resulting in ubiquitination of the POI, thus targeting it for proteasomal degradation (see **Fig. 2-8**).

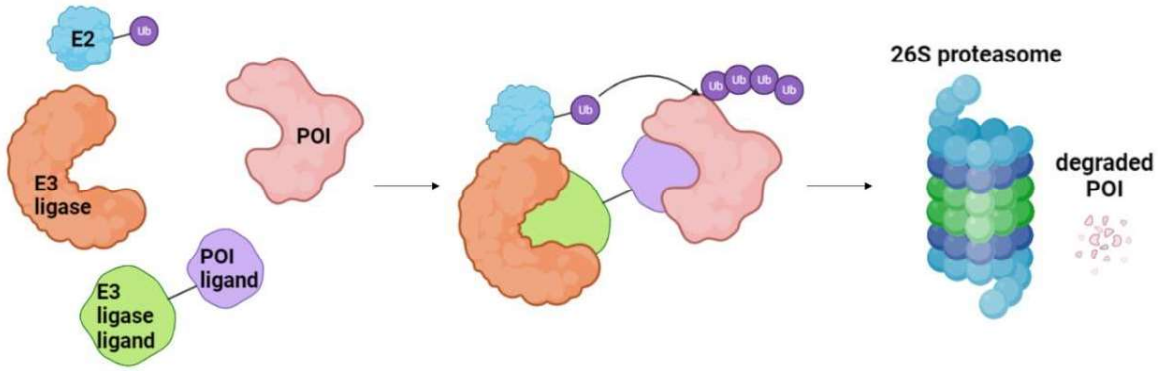


Fig. 2-8 *Schematic depiction of the PROTAC concept: Simultaneous binding of the POI ligand (violet) to the POI (salmon) and E3-ligase ligand (green) to an E3-ligase (orange) causes ubiquitination of the POI. Subsequent recruitment of the 26S proteasome leads to degradation of the POI.*

PROTACs display several advantages over traditional drug approaches: In contrast to antibody treatments for example, PROTACs can address intracellular targets.[241] They are delivered systemically, thus improving on alternative knockdown mechanisms such as siRNA.[242] Unlike regular small molecule drugs, PROTACs can target scaffolding proteins and eliminate pathogenic proteins.[243] All advantages are shared with the CRISPR technique for genome editing, but unlike this approach, PROTACs exhibit oral bioavailability.[244]

Ever since the first proof of concept of a small molecule PROTAC was realized by the CREWS group in 2008[245](SARM-nutlin PROTAC, see **Fig. 2-9**), a number of molecules with ligands for different E3-ligases have been developed against diseases such as cancer, viral infections or immunity-related diseases.[246] **Fig. 2-9** shows representative PROTAC structures featuring different E3-ligase ligands.[245,247–249]

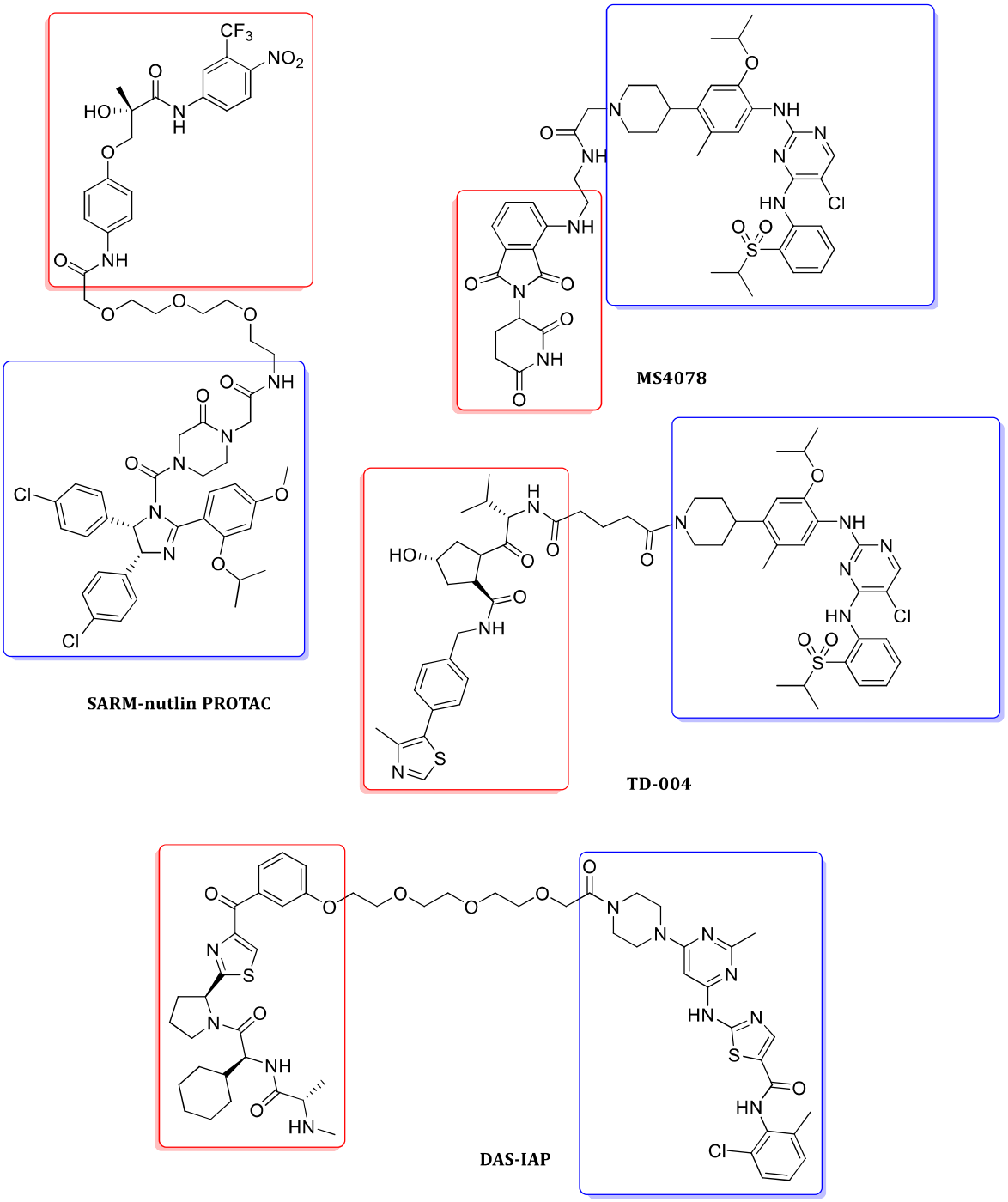
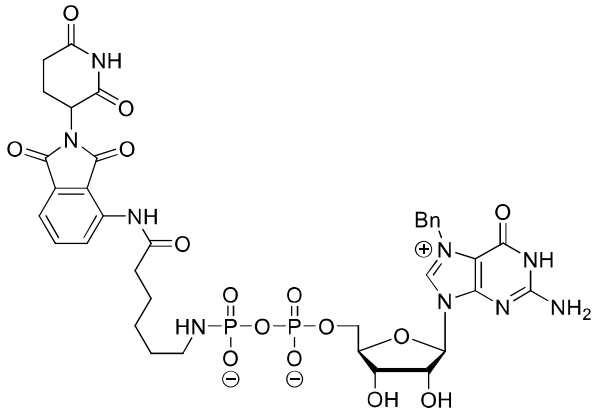
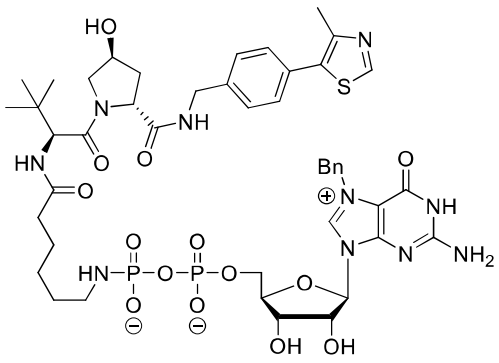


Fig. 2-9 *Representative PROTACs with different E3-ligase ligands (highlighted in red) connected via different linkers to POI ligands (highlighted in blue).*

eIF4E has been suggested as a target for the PROTAC approach by KAUR et al.[250], who attempted the creation of cap analogues linked to lenalidomide or von Hippel-Lindau (VHL) ligands.



eIF4E PROTAC 1



eIF4E PROTAC 2

Fig. 2-10 *eIF4E PROTACs developed by KAUR et al., T. Kaur, A. Menon, A. L. Garner, Synthesis of 7-benzylguanosine cap-analogue conjugates for eIF4E targeted degradation. Eur. J. Med. Chem. 2019, 166, 339–350. eIF4E-PROTAC 1 represents a cap-analogue linked to lenalidomide, eIF4E-PROTAC 2 represents a cap-analogue linked to a VHL ligand.*

In the assay used by the group, HEK293 lysates were incubated with either m^7-GMP or m^7-GDP agarose resin capable of binding and purifying eIF4E. From the library of linkers synthesized by them (representative structures shown in **Fig. 2-10**), it was shown that GDP conjugates were able to bind eIF4E, while GMP analogues failed to do so. Bn^7-GDP linked to lenalidomide (eIF4E-PROTAC 1) was shown to be an effective eIF4E binder *in vitro* at 50 µM, but all tested compounds failed to degrade eIF4E in cells.[250] The authors contribute this to insufficient cell permeability, a known problem for highly negatively charged cap analogues.[251] Therefore, there is a need for improvement to make effective eIF4E degraders.

2.4 Acyl carrier protein (ACP) and quorum sensing

Acyl carrier proteins (ACPs) are universally conserved proteins amongst different species and are involved in fatty acid synthesis.[252] In eukaryotes, such as mammals and yeast, ACP is represented as a domain of a fatty acid synthase protein complex, whereas prokaryotes such as bacteria have smaller, monomeric ACPs.[252] Furthermore, bacteria use ACPs as acyl carriers and donors for the synthesis of molecules such as endotoxins or acyl homoserine lactones (AHLs), which are used in quorum sensing mechanisms.[253,254] In general, quorum sensing is defined as the ability to adapt gene expression as a function of cell-population density.[255,256] As cell density increases, chemical signal molecules produced by bacteria increase in concentration until they reach a threshold after which they alter gene expression. This process can be used to regulate physiological activities such as biofilm formation, sporulation, virulence, production of antibiotics, motility and many more.[257–259] Gram-positive and Gram-negative bacteria use different types of molecules as signal molecules: the system employed by Gram-negative bacteria includes the aforementioned AHLs, while Gram-positive bacteria make use of processed oligopeptides. Since this work is focused on the interaction between ACP and AHL synthases, Gram-positive bacteria quorum sensing will not be discussed here.

2.4.1 LUXI/LUXR-type quorum sensing

Quorum sensing in Gram-negative bacteria (LUXI/LUXR-type quorum sensing) is named after the phenomenon found in the first microorganism it was discovered in: luminous bacterial species called *Vibrio fischeri* and *Vibrio harveyi.*[260] It was observed that light emission by these species was only detected at high cell-population densities, leading to the discovery of quorum sensing. To date, more than 25 different species of Gram-negative bacteria have been identified to employ quorum sensing mechanisms.[261] Except for *V. harveyi* and *M. xanthus*, all Gram-negative bacteria make use of the same signaling circuit that involves homologues of the two *V. fischeri* proteins called LuxI and LuxR.[261–263]

LuxI-type proteins, or AHL synthases, are enzymes that are involved in the biosynthesis of acylated homoserine lactone signaling molecules. LuxR-type proteins are the receptors for these AHL molecules. After surpassing a threshold of AHL molecules, LuxR-AHL complexes are responsible for gene transcription and subsequent effects of quorum sensing.[264]

2.4.1.1 LuxI enzyme functions

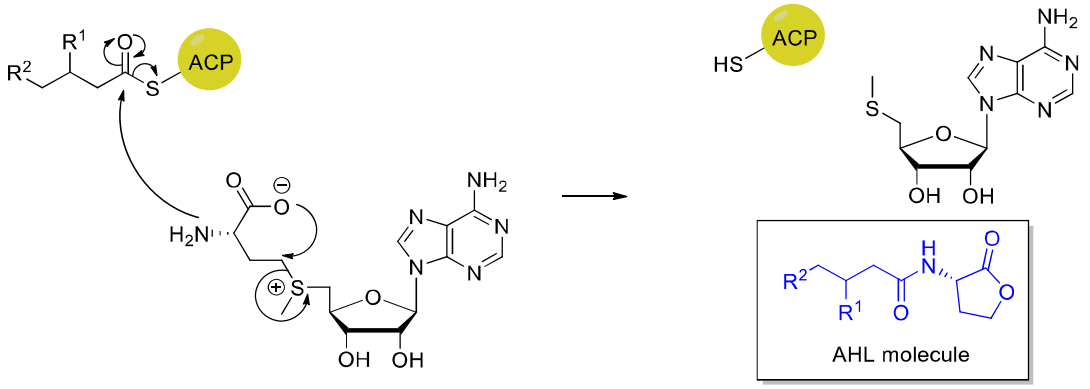


Fig. 2-11 *General mechanism of AHL biosynthesis.*

LuxI-type enzymes bind to acyl-loaded ACP proteins and use *S*-adenosylmethionine (SAM) as substrates.[265,266] They couple specific acyl-ACPs to SAM to form an amide bond between the acyl side chain of acyl-ACP and the amino group of the homocysteine moiety of SAM. AHL autoinducer molecules are formed through subsequent lactonization of the produced amide as shown in **Fig. 2-11**. This mechanism has been shown for LuxI from *V. fischeri* and enzymes from other Gram-negative bacteria.[265,267] The advanced research on the biochemistry of these reactions suggests conservation among all LuxI homologues, with the only difference being the acyl side chains of specific AHL molecules.[255,256,258,268,269] For this to work, LuxI homologues must have specificity for differently loaded acyl-ACP proteins. How this is realized on a structural level and the hierarchy and timing of components binding in the quaternary complex comprised of ACP, acyl-CoA, SAM and AHL synthase is not well studied yet.

2.4.1.2 LuxR receptor functions

The receiving site of AHL molecules, LuxR receptors, will not be part of this work and will therefore be discussed only briefly. Typically, these receptors are built from two major

components: An N-terminal domain responsible for AHL binding and a C-terminal domain involved in DNA binding for signal transmission.[270–272] While the AHL binding domain must be variable, the DNA binding domain of all LuxR homologues has a highly conserved helix-turn-helix motif that binds DNA promoter elements called "lux boxes".[273–275]

2.4.2 EsaI

Pantoea stewartii uses an AHL synthase called EsaI to create autoinducer molecules for quorum sensing. It has been shown that EsaI is mainly responsible for the cell density-dependent production of exopolysaccharide (EPS) and increased virulence of the pathogen.[276] EsaI utilizes 3-oxo-hexanoyl-ACP as a substrate to produce 3-oxo-hexanoyl-homoserine lactone (3-oxo-C6-AHL). Its structure has the same fold as N-acetyltransferases[269], enzymes involved in the transfer of acetyl-CoA to arylamines as part of phase II metabolism of most mammalian species. While the production of small amounts of slightly different AHL molecules has been shown, EsaI displays selectivity to the production of 3-oxo-C6-AHL over 3-oxo-C8-AHL or C6-AHL.[269]

2.4.3 BmaI

BmaI is the AHL synthase for the pathogen *Burkholderia mallei*, which has been shown to be involved in the development of Glanders disease.[277] There are two isoforms involved in quorum sensing, BmaI1 and BmaI3, which produce C8-AHL and 3-hydroxy-C8-AHL, respectively.[278,279] It is generally believed that BmaI, as well as other LuxI-type AHL synthases, achieves substrate specificity through a combination of modulating acyl-ACP pool supplies and specific substrate recognition of AHL synthases.[280] How this is achieved on a structural level remains mysterious and is still not well understood.

2.5 The glucocorticoid induced leucine zipper (GILZ)

Glucocorticoids are common drugs to treat inflammatory diseases or autoimmune diseases[281], but they have distinct disadvantages. Serious side effects include increased blood pressure, water retention, weight gain[282] and severe complications such as osteoporosis[283], risk of infections[284] or gastrointestinal bleeding[285]. Drugs that replace glucocorticoids and have less serious side effects are therefore urgently needed. The glucocorticoid induced leucine zipper (GILZ) is a small protein whose expression is induced by glucocorticoid treatment.[286,287] It is believed that anti-inflammatory effects exerted by glucocorticoids are mainly caused by GILZ induction, as GILZ-deficient mice show attenuated effects.[288]

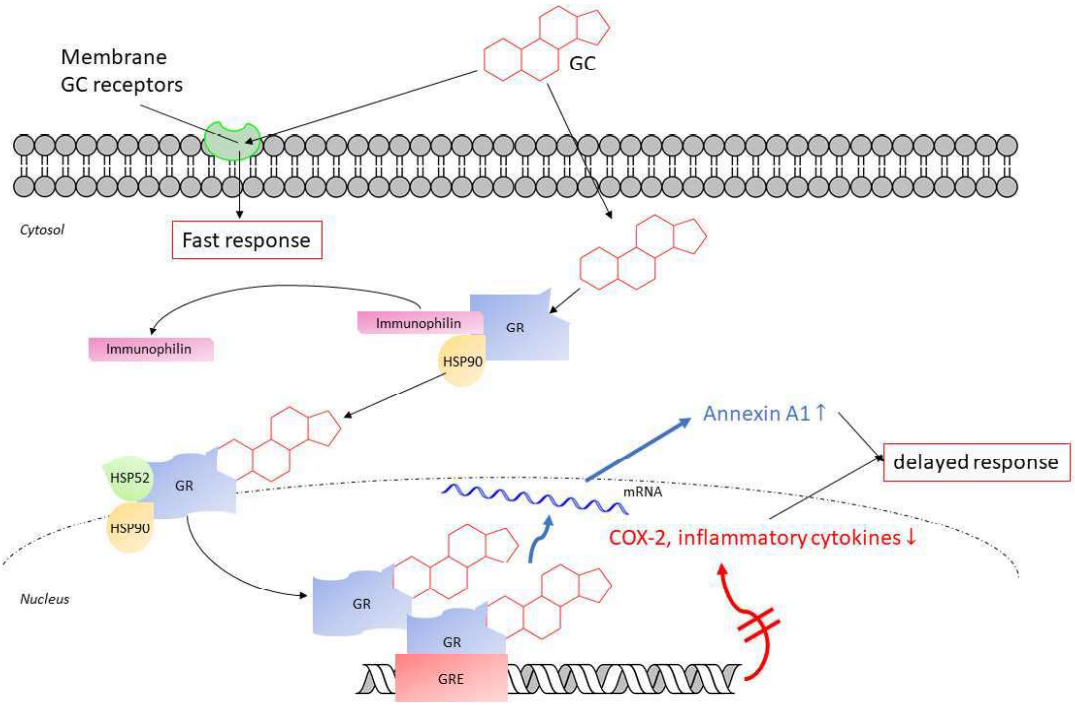


Fig. 2-12 *Schematic depiction of effects induced by glucocorticoids (GCs). Fast responses are mediated by pathways induced by GC binding to membrane bound glucocorticoid receptors (GRs). Delayed responses are caused by diffusion of GCs and pathways induced by GC binding to intracellular GRs.*

2.5.1 Induction of GILZ by glucocorticoids

Glucocorticoids, apart from being used as drugs, are hormones produced by the human body to regulate homeostasis and regulate transcriptional activity of a variety of different genes.[289,290] These genes affect inflammatory responses and the immune system. Most effects exerted by glucocorticoids (GCs) are mediated by their interaction with the glucocorticoid receptor (GR, see **Fig. 2-12**).[291] GRs are kept in the cytosol in protein complexes called receptosomes and dissociate as a result of conformational changes upon GC binding.[292] These receptosomes include stabilizing proteins such as heat shock protein 90 (HSP90) and immunophilin. GC/GR complexes translocate to the nucleus, mediated by the dynein trafficking pathway, where GRs dimerize and interact with coactivators and other transcription factors to bind DNA sequences called glucocorticoid recognition elements (GREs). Transcription of GC-sensitive genes is either repressed or stimulated.[291] Most anti-inflammatory effects are mediated by gene inhibition, such as cyclooxygenase 2 (COX-2) or cytokine repression, except for Annexin A1, which is upregulated.[293] Non-genomic, fast responses to GCs have been reported, but are poorly understood.[294]

GILZ was first discovered to be induced by treatment with dexamethasone, a synthetic GC derivative.[295] It has been shown that GILZ is one of the first GC-induced gene targets. As such, in a drug-free context, GILZ is produced because of environmental changes such as stressful events (infection, depression etc.).[296,297]

2.5.2 Anti-inflammatory effects of GILZ

Both anti-inflammatory and immune-regulating responses induced by GCs have been shown to be mediated by GILZ.[.286,287] GILZ is involved in the regulation of T cell apoptosis: By inhibiting the expression of Fas and FasL, T cells are protected from induced apoptosis by the activation of GILZ.[.284] Furthermore, GILZ associates with nuclear factor-κB (NF-κB), c-Fos, and c-Jun and inhibits their transcriptional activities.[299,300] It also inhibits targets downstream of the Raf and Ras pathway[301], as well as forkhead box O3 (FoxO3)[302]. On the other hand, GILZ can activate transforming growth factor-β (TGF-β) signaling.[287,296,303] Taken together, these actions drive T cell regulation. T_{H1} response and IL-2 secretion is downregulated while T_{H2} responses are

upregulated, mediated by the interaction between GILZ and NF-κB.[299] Through TGF-β, T_{reg} are developed which ultimately inhibit T cell activation. T_{H17} differentiation and IL-17 secretion are also repressed.[303] B cells are regulated negatively by GILZ. Via elevated NF-κB activity in *gilz* deficient mice, B lymphocytosis was observed, suggesting that GILZ might play a role in B cell disorders.[288,304]

2.5.3 GILZ interaction partners and their role in inflammation

Several interaction partners of GILZ have been identified (overview in **Fig. 2-13**).[292] As can be seen, human GILZ is a 134 amino acid long protein that can be divided into four functional domains: The N-terminal domain (N-Ter, amino acids 1-60), which was shown to bind Raf-1.[301] The TGF-β stimulated clone box (TSC) is a short fragment stretching over 15 amino acids from 61-75 which interacts with Ras.[305] Both interactions inhibit the PI3K/Akt and MAPK/ERK pathways, leading to reduced inflammation. The leucine zipper domain (LZ, amino acids 76-97) is involved in GILZ dimerization[299] (discussed in detail later) but, unlike other leucine zipper domains, has not been shown to interact with DNA and function as a transcription factor. Finally, the C-terminal domain (Pro-Glu-rich

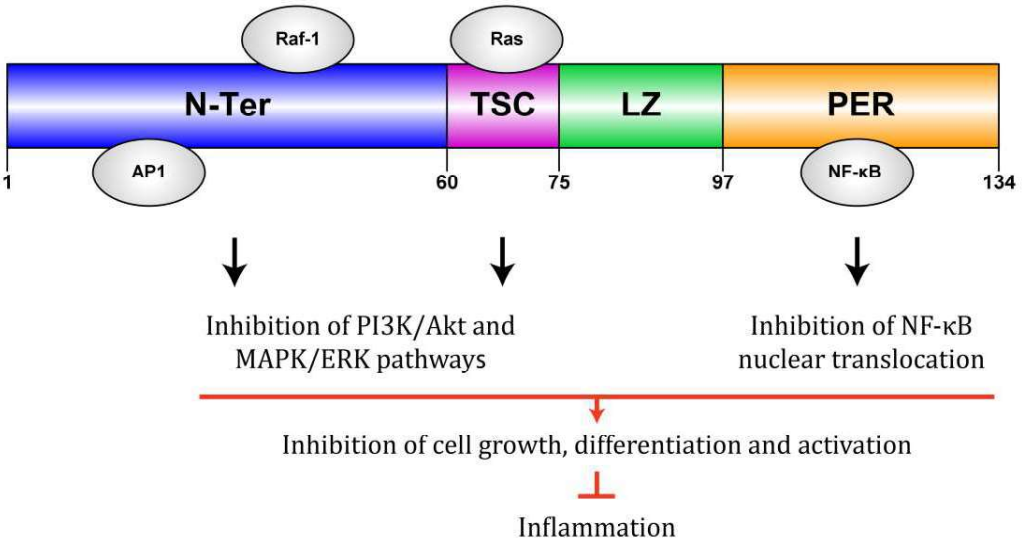


Fig. 2-13 *Domains of GILZ and identified interaction partners. Effects of these interactions and their role in inflammation are shown with arrows. Adapted from O. Bereshchenko, G. Migliorati, S. Bruscoli, C. Riccardi,* Glucocorticoid-Induced Leucine Zipper: A Novel Anti-inflammatory Molecule. Front. Pharmacol. ***2019***, *10, 308.*

region, PER, amino acids 98-134) was shown to harbor the NF-κB binding site and negatively regulate NF-κB nuclear translocation, leading to decreased inflammation.[299,306] Multiple other binding partners of GILZ have been identified (including AP1, SUMO1, MDM2 etc.) but their binding regions have not yet been identified.[292,298,300]

Of particular interest for this work is the interaction of GILZ with histone deacetylase (HDAC) 1 and 2 found by BRUSCOLI et al.[307] The authors showed that HDAC1 co-immunoprecipitates with MyoD (a transcription factor that regulates the differentiation of fibroblasts to myoblasts), but only in the presence of GILZ. RNAi-induced GILZ knockdown depleted the interaction between HDAC1 and MyoD. Furthermore, the authors showed that GILZ-Myc was able to interact with MyoD, HDAC1-FLAG and HDAC2-FLAG. This interaction was determined to be of importance during HDAC1/MyoD-mediated inhibition of myogenin expression. This data suggests the involvement of GILZ in GC-induced antimyogenic effects during anti-inflammatory treatments of myopathies.[307]

2.5.4 Structure of GILZ

Although GILZ is a small protein, its structure has not been solved to date. A three-dimensional model has been created by MARCO et al., using DIP, the sleep-inducing peptide immunoreactive peptide, as a template.[308] The proposed structure of murine GILZ (which shares 97 % identity with human GILZ) consists of a short helix followed by two β-sheet strands in the N-terminal domain and TSC box, a long helix in the leucine zipper domain and a disordered C-terminal PER-domain. Later, a dimerization event through the leucine zipper domain, forming the active coiled-coil dimeric structure show in **Fig. 2-14 C** was proposed.[299] The authors showed, through mutational analysis, that the leucine zipper domain was necessary for dimerization and monomeric GILZ was incapable of binding NF-κB. However, in the assay used by the authors, dimerization was shown through interaction of GILZ mutants with wild type GST-GILZ, so oligomerization beyond dimerization cannot be excluded. In particular, it was shown that only mutants in which all four heptad leucines in the leucine zipper domain (L76, L83, L90 and L97)

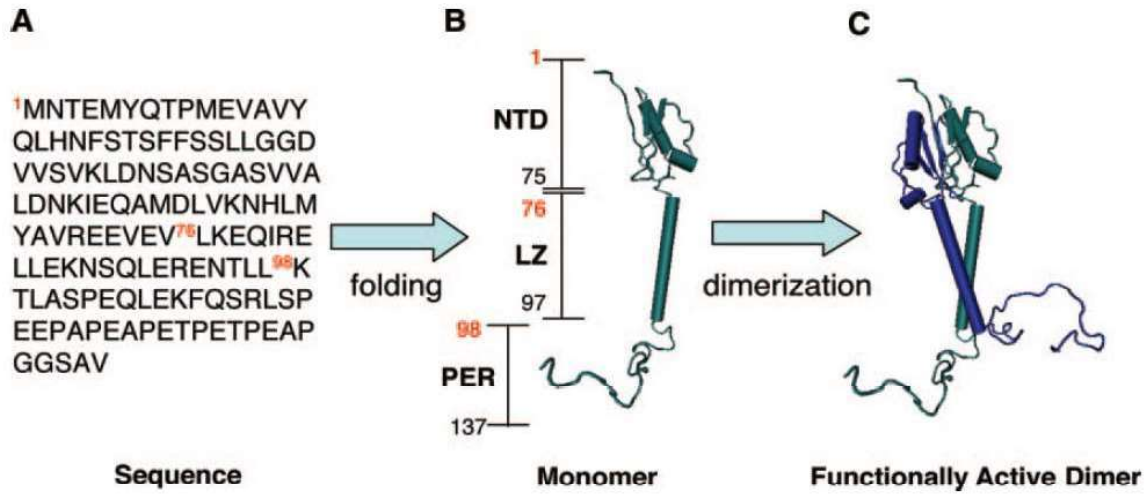


Fig. 2-14 A. *Sequence of murine GILZ.* ***B.*** *Monomeric 3D model of GILZ, calculated in E. Ayroldi, G. Migliorati, L. Cannarile, R. Moraca, D.V. Delfino, C. Riccardi.* CD2 rescues T cells from T-cell receptor/CD3 apoptosis: a role for the Fas/Fas-L system. Blood. ***1997***, 89, *3717-3726.* ***C.*** *Functionally active coiled-coil dimer of GILZ. Taken from B. D. Marco, M. Massetti, S. Bruscoli, A. Macchiarulo, R. D. Virgilio, E. Velardi, V. Donato, G. Migliorati, C. Riccardi,* Glucocorticoid-induced leucine zipper (GILZ)/NF- B interaction: role of GILZ homo-dimerization and C-terminal domain. Nucleic Acids Res. ***2006***, 35, *517–528.*

were mutated to alanines was incapable of oligomerization. Furthermore, single point mutation of N87 to D87 rendered GILZ unable to oligomerize, suggesting the importance of this residue.[299]

2.6 Keap1

Keap1 (Kelch-like ECH-associated protein 1) is an essential regulation protein for the activity of the transcription factor Nrf2 (Nuclear factor erythroid 2-related factor 2). Nrf2 plays an important role in the induction of cytoprotective genes as a response to toxification due to stressors such as electrophilic xenobiotics or reactive oxygen species (ROS).[309–311] Under normal conditions, Nrf2 levels are kept low by constant degradation mediated by Keap1. Keap1, as an adaptor protein of a Cullin 3 (Cul3)-dependent E3-ubiquitin ligase, facilitates ubiquitination and subsequent proteasomal degradation of Nrf2.[312,313] In the presence of electrophiles or ROS, Keap1 is modified directly at cysteine residues, which decreases its E3 ligase activity.[314] The Nrf2-Keap1 axis is implicated in

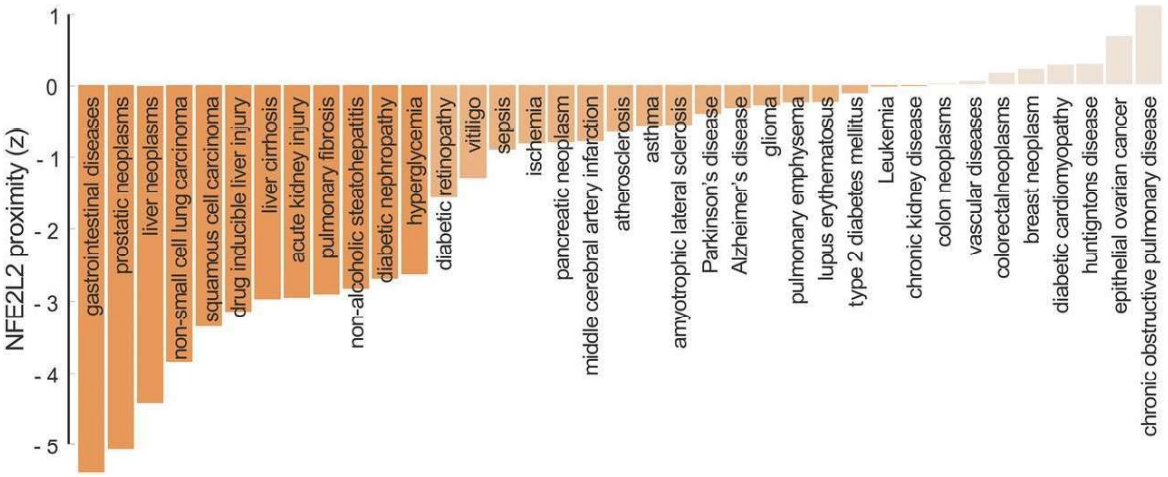


Fig. 2-15 *Bar graph showing the proximity of NRF2 gene (*NFE2L2*) to known disease genes that participate in NRF2-related pathophenotypes. For any disease, proximity is calculated by the distance from NRF2 to the closest known disease gene and then compared to average distances between randomly selected proteins in the interactome. Negative z-scores indicate that the distance is lower than expected by chance. Taken from A. Cuadrado, G. Manda, A. Hassan, M. J. Alcaraz, C. Barbas, A. Daiber, P. Ghezzi, R. León, M. G. López, B. Oliva, et al.,* Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach. Pharmacol. Rev. ***2018***, 70, *348–383.*

a number of diseases, including cancers such as skin, liver and lung neoplasms[315–318], metabolic and cardiovascular diseases such as diabetes, hyperglycemia and atherosclerosis[319–322], respiratory diseases such as asthma and pulmonary fibrosis[323–325] as well as neurodegenerative conditions such as Alzheimer's disease or Parkinson's disease[326–328]. The interactome-based proximity of Nrf2 to genes related to various diseases is shown in **Fig. 2-15.**[329]

2.6.1 Structural basis of the Keap1-Nrf2 interaction

Keap1 is a member of the BTB-Kelch family, which comprises over 50 proteins, each of which assemble with Cullin 3 (Cul3) and Rbx1 to form Cullin–RING (really interesting new gene) ligases.[330] Its domain architecture is depicted in **Fig. 2-16.** It contains an N-terminal BTB domain, which is required for homodimerization and contributes to Cul3 interaction.[331] The 3-box motif in the central intervening region (IVR domain) further provides Cul3 interaction.[330] The C-terminal Kelch domain is required for substrate interaction (i.e. Nrf2) and binds to the ETGE and DLG motifs of Nrf2.[332] After identification of the binding interface between Keap1 and Nrf2, high resolution crystal structures of the complex have been solved.[333,334] The Kelch domain consists of six Kelch repeats which form a β-propeller structure (see **Fig. 2-17A**, taken from[335]). Each propeller blade comprises four antiparallel β strands, with a βA strand forming the central core. The fold of each Kelch repeat is maintained by a small number of conserved residues, while the overall sequence is variable to allow for substrate specificity. The substrate binding site is created by the DA loop connecting β strands A and D to form a shallow pocket.

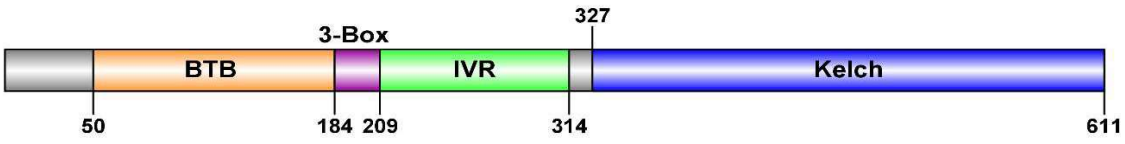


Fig. 2-16 *Domain architecture of Keap1.*

In 2013, the first cysteine-independent small molecule inhibitor of the protein-protein interaction was identified in a high throughput screening and structurally analyzed a year later.[336,337] The same binding site occupied by the Nrf2 peptide is addressed by the

small molecule inhibitor (S,R,S)-1a, suggesting replacement of Nrf2 and subsequent activation of the transcription factor.[338] Other Keap1 inhibitors mainly act as covalent binders of the BTB domain (bardoxolone[339]) or likely the IVR domain[314].

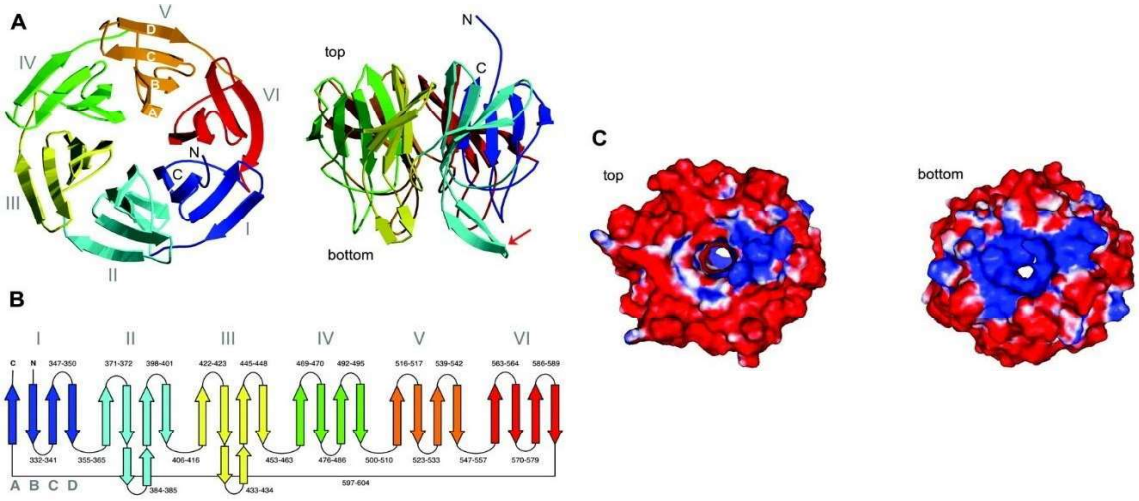


Fig. 2-17 *Structure of the Keap1 Kelch domain.* ***A.*** *Ribbon diagram showing propeller blades from different orientations.* ***B.*** *Topology diagram of the Keap1 Kelch domain.* ***C.*** *Electrostatic surface of the Keap1 Kelch domain calculated by PYMOL. Adapted from X. Li, D. Zhang, M. Hannink, L. J. Beamer,* Crystal Structure of the Kelch Domain of Human Keap1. J. Biol. Chem. ***2004***, 279, *54750–54758.*

2.6.2 Challenges of the structural investigation of Keap1 complexes

Solving the structure of Keap1 in complex with its small molecule inhibitors has long been hampered by the inability of complex formation through ligand soaking. After closer inspection of crystal structures of human Keap1, it became obvious that binding sites are blocked by either crystal-crystal contacts (in the apo form) or by the Nrf2 peptide itself.[340] A mutational analysis in 2013 by HÖRER et al. found a double mutant (E540A/E542A) which crystallizes under different conditions and creates a new crystal form. This double mutant was therefore deemed suitable for soaking experiments with small molecule inhibitors of Keap1.[340]

3 Aim of this Work

3.1 Challenging drug targets

This work is subdivided into several different projects, all of which include novel potential drug targets with challenging aspects regarding their druggability and/or structural analysis. MraY is an integral bacterial membrane protein and potential target for new antibiotics. Even though the protein's structure with and without inhibitors has been solved[76–79], there are still questions unanswered due to the enzyme's dynamics and structural flexibility. Furthermore, exploration of the so-far unstudied interactome of MraY as well as expanding on the knowledge of self-resistance mechanisms employed by *Streptomyces sp.* will be the focus of this work. The nuclear factor of activated T-cells (NFAT) is a transcription factor with mostly intrinsically disordered domains. In particular its N-terminal transactivation domain and regulatory domain harbor multiple interaction and/or phosphorylation sites and are therefore of interest for potential drug development. Another protein investigated in this work is the master regulator of eukaryotic protein translation, eIF4E. eIF4E, the mRNA cap binding protein, undergoes phosphorylation at a single serine residue in position 209. This post translational modification enhances translation of oncogenes and is associated with tumor progression and reduced survival rates in patients with malignant melanoma.[341] It is therefore important to understand the function of this phosphorylation in order to develop effective cancer treatments. A very well researched protein investigated in this work is the acyl carrier protein (ACP) of *E. coli*. However, this thesis will focus on the complexes that form with different acyl homoserine lactone synthases (AHL synthases) of Gram-negative bacteria during the formation of acyl homoserine lactones used for quorum sensing. Gaining insights into these dynamic interactions will help the development of new antibiotic agents. Furthermore, the structure and interactome of the glucocorticoid induced leucine zipper (GILZ) will be discussed. This work aims to gain understanding in how dimerization or possible oligomerization of the protein contributes to its interaction with HDAC7, which was found to be a novel interaction

partner of GILZ. Lastly, a challenging protein-protein interaction of the regulatory adaptor protein Kelch-like ECH-associated protein 1 (Keap1) with the transcription factor Nrf2 will be targeted in an approach to expand the chemical space with computational screenings of over 1 billion compounds to find effective molecules that tackle Nrf2-related diseases.

3.2 Project 1: Bacterial Translocase I – MraY

In this project, several techniques will be applied to study MraY-ligand interactions. Biophysical techniques such as isothermal titration calorimetry (ITC) and microscale thermophoresis (MST) will be used to study ligand-target interactions and determine binding constants as well as thermodynamic parameters. The development of a detergent-free, lipid bilayer environment (i.e. nanodiscs) will be performed. Combined, these results are envisioned to complete the picture of MraY inhibition and help with the development of novel nucleoside antibiotics for potential clinical use.

Mounting evidence suggests the involvement of MraY in a multienzyme complex participating in peptidoglycan synthesis.[84] A potential high-order complex of MraY, the preceding ligase MurF, the flippase MurG and possibly other proteins such as MreD, RodA and FtsW could be used by bacteria to diminish the possible diffusion of intermediate products to the cytoplasm by a controlled channeling network. Understanding the structure of such a complex holds the great potential of interfering with bacteria's organization of enzymatic pathways, thus creating an alternative route for antibiotic development. Noteworthy, none of the published crystal structures were able to resolve loop A, which connects the transmembrane helices TM1 and TM2 of MraY.[76–79] This suggests the possibility of a stabilizing structure, which is missing in the crystallization experiment, such as a cytosolic protein. This work aims to prove an interaction between MraY and its preceding soluble enzyme MurF, which might ultimately lead to solving the structure of this protein-protein complex using X-ray crystallography and cryo-EM.

Also, by investigating MraY isoforms of muraymycin producers (*Streptomyces sp.*), the possibility of self-resistance by single point mutation will be discussed. After structural information about MraY and MraY bound to ligands had become available[76–79], the

involvement of ligand-binding motifs helped to understand MraY inhibition immensely. Of particular interest for this work is the very well conserved "HHH"-motif found in most MraY isoforms except for *Streptomyces sp.* Especially since it was shown by CHUNG et al. that in particular the first histidine needs to be flexible to participate in the binding of muraymycin D2 as an inhibitor[77], the observed replacement to glutamine in muraymycin producers suggests the involvement of this alteration in a self-resistance mechanism.

3.3 Project 2: Nuclear factor of activated T-cells (NFAT)

This work aims to target two of the three intrinsically disordered regions of the nuclear factor of activated T-cells. The N-terminal transactivation domain and its implications in transcriptional activity of NFAT will be subject of investigation, including efforts to understand the role of phosphorylation. Since not much is known about the relevance of individual phosphorylation sites, this project will address the first reported phosphor-serine in the sequence of NFAT: S23. This residue was reported to be phosphorylated by p38 MAPK[342], but the effects are poorly understood. To elucidate the consequence and function of this post-translational modification, this work aims to use solution-NMR accompanied by other techniques to study this effect *in vitro* and in cells.

Furthermore, the heavily phosphorylated regulatory domain of NFAT will be discussed as a model protein to develop novel NMR techniques for the study of intrinsically disordered proteins and to elucidate translocation mechanisms of the transcription factor. As a starting point, the phosphorylation sites addressed by protein kinase A (PKA) will be in the center of research. This kinase was chosen because it engages in the first phosphorylation event in NFATs regulatory domain (S255) and addresses only one other serine within the sequence (S225). The relevance of this phosphorylation has been studied before in NFAT2 and was shown to influence nuclear export, although the effects are linked to PKAs priming ability for GSK-3.[144,343] The influence of the second phosphorylation site (S225) is still unclear.

Lastly, it will be probed if NFAT can be used as a model system for the development of functional protein manipulations through alterations of the protein's sequence. By

overexpressing fluorescently tagged NFAT mutants in mammalian cells, fluorescence microscopy will be used to investigate the effect of protein modifications on subcellular localization. Various binding sites for the ubiquitously expressed protein phosphatase 2A (PP2A) will be incorporated into a disordered part of full-length NFAT and it will be investigated if NFAT activity can be regulated outside of the natural calcium-calmodulin-calcineurin pathway.

3.4 Project 3: The eukaryotic translation initiation factor eIF4E

This project will, on one hand, address the issue of targeting eIF4E using the PROTAC approach. As introduced before, this method targets proteins for E3-ubuquitin ligase-mediated proteasomal degradation and is an upcoming method to develop drugs against difficult targets. Even though the PROTAC approach has been attempted for eIF4E before[250], low binding affinities and cell permeability of the molecules have impeded the development of effective compounds as of yet. Therefore, this work aims to develop new small molecule-based eIF4E degraders to improve on the efforts made so far. PROTACs will be developed using the small molecule 4EGI-1 and a novel biphenyl inhibitor compound, which targets an internal void cavity of eIF4E. Target molecules are depicted in **Fig. 3-1**.

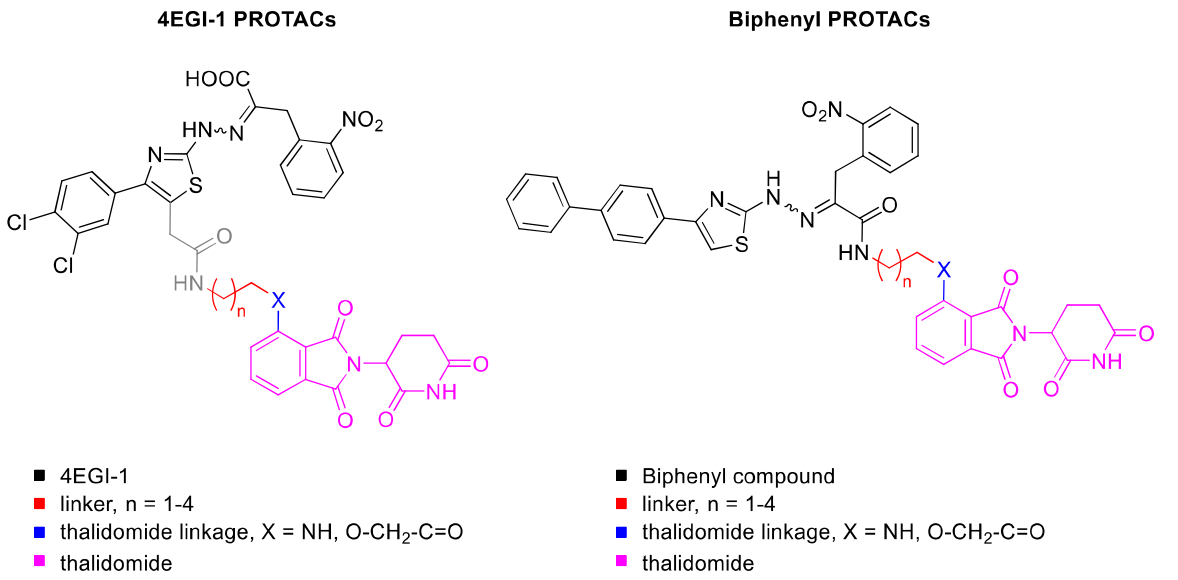


Fig. 3-1 *PROTAC target structures based on 4EGI-1 (left) and the unpublished biphenyl inhibitor (right). Variable regions are highlighted in different colors.*

On the other hand, this work tries to lay the foundations of structural and functional implications of eIF4E phosphorylation by Mnk kinases. For this, human eIf4E, eIF4E S209D (a previously reported phosphomimetic mutant[344–346]) and eIF4E pS209 will be recombinantly expressed and purified. The phosphorylated protein will be produced using an amber codon strategy: Generally, three codons on mRNAs are recognized as stop codons: ochre (UAA), opal (UGA) and amber (UAG). Of these three, the amber codon is used least frequently by *E. coli* (~ 7 %).[347] In order to achieve suppression of termination by the amber codon, complementary amber $tRNA^{CUA}$ can be generated using an orthogonal aminoacyl-tRNA synthetase (aaRS) that only accepts unnatural amino acids (phosphoserine in our case). To achieve stable expression of phosphorylated protein, the *E. coli* strain BL21ΔserB will be used. This strain was developed by JESSE RINEHART and DIETER SÖLL by deleting the gene from *E. coli* BL21 that encodes the phosphoserine phosphatase (serB).[348] The combination of amber codon suppression with an expression system that lacks the phosphoserine phosphatase serB will provide an ideal setup for the large-scale production of phosphorylated eIF4E for X-ray crystallography and NMR studies.

3.5 Project 4: The acyl carrier protein ACP

This project will revolve around the specific recognition of acyl-ACPs by different AHL synthases. To do this, solution-NMR spectroscopy will be employed. Two different acyl-ACP/AHL synthase interactions will be the focus of this work: BmaI1 and EsaI with their native ligands, C8-CoA and 3-oxo-C6-CoA, respectively. In order to study these interactions without substrate turnover, analogues for the individual ACP substrates will be employed (see **Fig. 3-2**).

In these analogues, designed by the group of RAJESH NAGARAJAN who is a collaborator on this project, the labile thioester moiety has been replaced by a non-cleavable thioether group.[349] Holo-ACP (ACP carrying pantetheine) will be recombinantly produced with specific isotope labeling to generate ^{15}N-ACP and ^{15}N-^{13}C-ACP. Apo-ACP (ACP were pantetheine is cleaved to leave unsubstituted Ser36) can be produced by incubation of bacterial lysates with overexpressed holo-ACP at 37 °C, as performed routinely by the NAGARAJAN group according to preexisting protocols with minor modifications.[350,351] Labeled apo-ACP will then be reacted with the inert substrates depicted in **Fig. 3-2** and recombinantly purified surfactin phosphopantetheinyl transferase (Sfp) to generate loaded acyl-ACPs.[352]

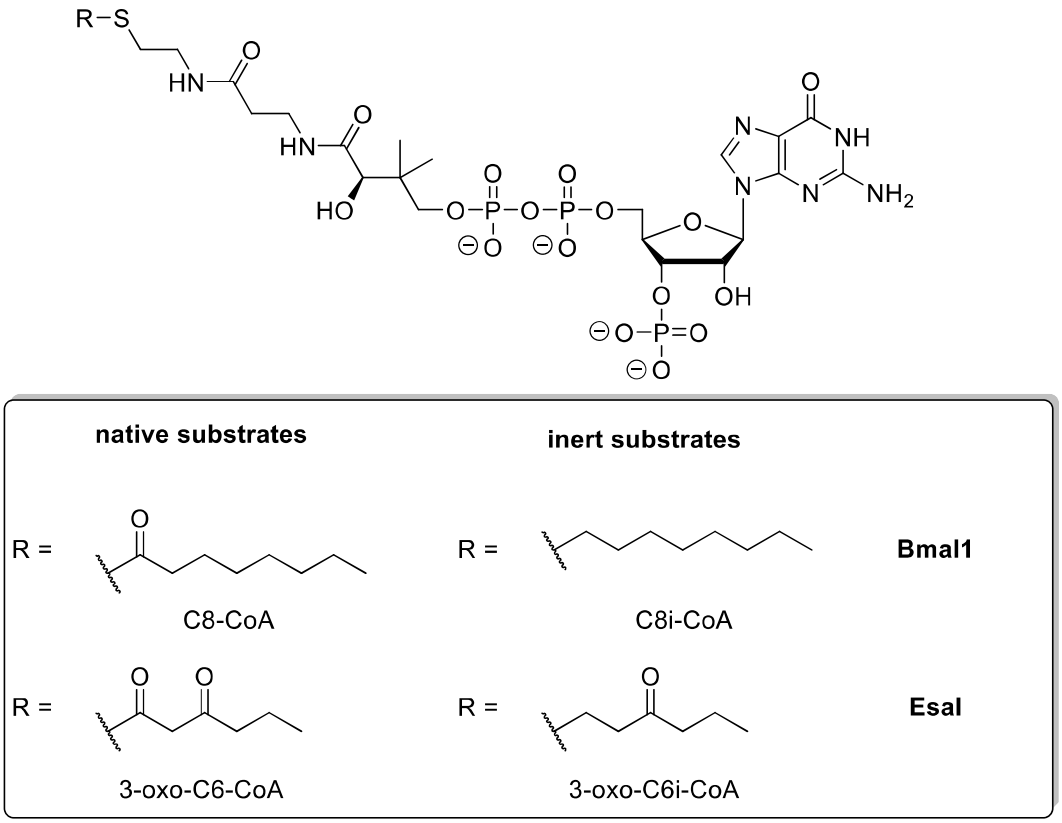


Fig. 3-2 *Structures of ACP-bound substrates of the AHL synthases BmaI1 and EsaI*

^{15}N-^{1}H-HSQC spectra will be recorded for the loaded acyl-ACPs and peaks will be assigned to individual amino acids using triple resonance backbone NMR experiments. The binding interface and conformational changes of acyl-ACPs upon binding of AHL synthases will then be probed by the addition of unlabeled AHL synthases and mapping of the occurring perturbations of chemical shifts in the ACP spectra, as well as relaxation-directed solution-NMR experiments.

3.6 Project 5: The glucocorticoid induced leucine zipper GILZ

To understand how the structure of GILZ affects its interactions, this project aims to solve the structure of GILZ and map novel, poorly understood interaction sites. As preliminary results indicate the oligomeric nature of GILZ[299], NMR fails as a method to study the protein, as its correlation time caused by the high molecular weight will cause enhanced relaxation and therefore low signal intensity. Therefore, X-ray crystallography and cryo-

EM will be attempted to solve the structure of GILZ alone and possibly in complex with interaction partners.

Collaborators have found an interaction between GILZ and HDAC7, a novel interaction which will be investigated in this work in a so-called MAmmalian Protein-Protein Interaction Trap (MAPPIT) screen. In this assay, a signaling-deficient type I cytokine receptor is used (extracellular EpoR fused to the transmembrane domain and cytoplasmic portion of LepR).[353] Ligand binding to this receptor will still activate associated Janus Kinases (JAKs), but because the target tyrosines are mutated, no signal transduction can occur due to lack of phosphorylation. The protein of interest (GILZ) is fused to this receptor, while the potential interaction partner is fused to a different receptor (i.e. glycoprotein 130 receptor [gp130]). Therefore, upon interaction of the two proteins, the JAKs activated by the GILZ-EpoR/LepR complex will phosphorylate tyrosines on gp130, and signal transduction leads to the recruitment of STATs that will activate the transcription of a luciferase reporter gene.[353]

Fragments of both HDAC7 and GILZ will be cloned and assessed for interactions using immunoprecipitation (IP) assays to find the minimal interaction interface between the two proteins.

3.7 Project 6: The Kelch-like ECH-associated protein 1: Keap1

This project will create an effort to increase the number of molecules screened in a computational environment using the Keap1-Nrf2 interaction by an order of magnitude. Using a single processor core, docking 1 billion compounds for 15 s each would take approximately 475 years. Therefore, job splitting onto multiple CPUs will be used as a method to achieve these screening efforts in a matter of days.

Virtual hits will be experimentally validated using orthogonal assays through multiple iterations to assess the hit rate produced by docking and evaluate the suitability of this approach to produce effective binders. Thereby, the chemical space that is possible to screen will be expanded. Currently existing screening methods only allow for 10^6-10^7 molecules to be exploited by virtual screening.[354] However, a higher percentage of the

available chemical space is desirable due to the increase of high affinity hits and decrease of false positives that comes from larger screens.[355]

Four orthogonal methods will be utilized to validate hits obtained from the virtual screen: Fluorescence polarization (FP), surface plasmon resonance (SPR), NMR and bio-layer interferometry (BLI). Fluorescently labeled Nrf2 peptides will be used to assess their association with the Keap1 Kelch domain. The created high molecular weight complex will exhibit slow molecular motions, resulting in a high FP signal. Compounds that are able to compete with the Nrf2 peptide will result in a decrease of FP signal, as fast motions will be in place for the small Nrf2 peptide. SPR methods will be employed with immobilized GST-tagged Keap1 captured on GST sensor chips to evaluate direct binding constants for virtual hits obtained in the screening. In real time, a flow of ligand across the immobilized protein can be utilized to obtain information about binding kinetics. As protein-ligand complexes form, the mass increase on the surface changes the refractive index, which will be registered by the SPR detector. Therefore, using SPR, information about association and dissociation rates of complex formation can be measured directly. Ligand-detected NMR experiments such as saturation-transfer difference (STD) NMR and CPMG (Carr-Purcell-Meiboom-Gill)-based experiments will be used to confirm ligand binding. Labeling methods to specifically express proteins with selectively ^{1}H-^{13}C-labeled methyl groups of Ile, Leu and Val residues (ILV-labeling) in an otherwise deuterated environment allow for the detection of less crowded ^{1}H-^{13}C-TROSY experiments.[356–360] Comparison of chemical shifts perturbations (CSPs) caused by binding of the Nrf2 peptide with those caused by small molecules create the opportunity to filter ligands that bind to the same residues as the Nrf2 peptide. Lastly, biotinylated Nrf2 peptides will be immobilized on biosensor chips and introduced to a solution containing Keap1 preincubated with ligands and interference patterns will be measured in real-time. Binding ligands shift the interference pattern to allow the detection of binding events. Nonspecific binding can be excluded by control experiments using ligand solution in the absence of Keap1.

In addition, an attempt will be made to obtained high resolution crystallography data of Keap1 in complex with the best inhibitors obtained from the virtual screening. For this purpose, Keap1 will be crystallized in the presence and absence of these inhibitors.

Crystals obtained from apo-protein will be subjected to ligand soaking. As ligand soaking will unlikely be successful for wild type Keap1, the double-mutant Keap1 E540A/E542A will be cloned and expressed.

4 Publications

4.1 ^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins

Sandeep Chhabra*, Patrick Fischer*, Koh Takeuchi, Abhinav Dubey, Joshua J. Ziarek, Andras Boeszoermenyi, Daniel Mathieu, Wolfgang Bermel, Norman E. Davey, Gerhard Wagner, and Haribabu Arthanari. (2018) ^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins. **Proc. Natl. Acad. Sci. U. S. A.**, 115 (8) E1710-E1719.

* These authors contributed equally

DOI: https://doi.org/10.1073/pnas.1717560115

URL: www.pnas.org/cgi/doi/10.1073/pnas.1717560115

Summary:

This publication provides a suite of 3D ^{15}N-detected NMR experiments that exploit the narrow line width and wide dispersion of ^{15}N signals in IDPs. Experiments conducted in this research start with ^{1}H magnetization and end in ^{15}N detection to exploit the higher sensitivity of proton excitation and shorter recycling times compared to starting with ^{13}C of ^{15}N excitations. Prior to this publication, an array of ^{13}C-detected experiments for the assignment of IDPs has been developed. Similarly to these experiments, our ^{15}N detected methods provide correlation and connectivity between protein backbone atoms to enable unambiguous assignments. Narrower line widths and increased sensitivity are achieved despite the lower sensitivity of ^{15}N compared to ^{13}C due to slower transverse relaxation rates of nitrogen.

Compared to traditional ^{15}N-^{1}H–HSQC spectra, our 2D-CON correlation experiments provide improved dispersion and therefore better fingerprints for IDPs. Additionally, ^{15}N detected 2D-CON experiments are less sensitive to pH variation, problems occurring from imperfect water suppression and display correlations to proline residues.

Furthermore, this publication shed light on the effects of PKA phosphorylation of the regulatory domain of NFAT and subsequent 14-3-3 binding. Transcriptional activity of NFAT is inhibited by PKA phosphorylation due to a promoted 14-3-3 chaperone association following phosphorylation[361]. To elucidate the mechanism, we performed *in vitro* phosphorylation of $NFAT^{131-294}$ by PKA. Assignments were transferred from ^{15}N-detected 2D-CON experiments to ^{1}H-detected ^{15}N-^{1}H-HSQC spectra. NUS was employed to follow phosphorylation in a time-dependent manner by recording HSCQ spectra over 20 min intervals. PKA phosphorylation resulted in downfield shifts of the ^{1}H resonances of S-225 and S-255. It was noted that S-255 is phosphorylated first and to completion, whereas S-225 remains partially unphosphorylated. When 14-3-3τ was added to a phosphorylated NMR sample of $NFAT_{131-294}$, line broadening was observed from residues 220-300, suggesting binding of 14-3-3 to both phosphorylation sites. Notably, this sequence includes the nuclear localization signal (NLS) 251-253. Only minor changes in the NLS were observed upon the addition of 14-3-3τ to unphosphorylated NFAT, indicating that 14-3-3 engagement is regulated by phosphorylation and 14-3-3 masks the NLS after PKA activation. Hence, the model of 14-3-3 blocking NFAT-importin binding, thus preventing nuclearization, is further strengthened by these results.

Author contribution:

As the author of this thesis, I performed the expression and purification of unlabeled, ^{15}N-labeled and $^{15}N^{13}C$-labeled die together with Sandeep Chhabra. I optimized the experimental conditions for *in vitro* phosphorylation of NMR samples. Assisted by Haribabu Arthanari, I wrote the pulse sequences for the triple resonance NMR experiments and carried out the measurements. I processed and analyzed the NMR data using nmrPipe and CcpNMR-Analysis, respectively, and assigned spectra to corresponding amino acids of unphosphorylated and phosphorylated $NFAT_{131-294}$. Together with Sandeep Chhabra, I transferred assignments from 2D-CON experiments to ^{15}N-^{1}H–HSQC spectra and interpreted effects of phosphorylation and 14-3-3 binding. With the help of Sandeep Chhabra and all other co-authors, I assisted Haribabu Arthanari in writing the manuscript.

^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins

Sandeep Chhabra[a,b,1], Patrick Fischer[a,1], Koh Takeuchi[c], Abhinav Dubey[a,b], Joshua J. Ziarek[a,d], Andras Boeszoermenyi[a,b], Daniel Mathieu[e], Wolfgang Bermel[e], Norman E. Davey[f], Gerhard Wagner[a,2], and Haribabu Arthanari[a,b,2]

[a]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; [b]Department of Cancer Biology, Dana–Farber Cancer Institute, Boston, MA 02215; [c]Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, 135-0064 Tokyo, Japan; [d]Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, IN 47405; [e]Magnetic Resonance Spectroscopy NMR Application, Bruker BioSpin GmbH, 76287 Rheinstetten, Germany; and [f]Conway Institute of Biomolecular and Biomedical Sciences, University College Dublin, Dublin 4, Ireland

Contributed by Gerhard Wagner, January 11, 2018 (sent for review October 10, 2017; reviewed by Richard W. Kriwacki and Peter E. Wright)

Studies over the past decade have highlighted the functional significance of intrinsically disordered proteins (IDPs). Due to conformational heterogeneity and inherent dynamics, structural studies of IDPs have relied mostly on NMR spectroscopy, despite IDPs having characteristics that make them challenging to study using traditional ^{1}H-detected biomolecular NMR techniques. Here, we develop a suite of 3D ^{15}N-detected experiments that take advantage of the slower transverse relaxation property of ^{15}N nuclei, the associated narrower linewidth, and the greater chemical shift dispersion compared with those of ^{1}H and ^{13}C resonances. The six 3D experiments described here start with aliphatic ^{1}H magnetization to take advantage of its higher initial polarization, and are broadly applicable for backbone assignment of proteins that are disordered, dynamic, or have unfavorable amide proton exchange rates. Using these experiments, backbone resonance assignments were completed for the unstructured regulatory domain (residues 131–294) of the human transcription factor nuclear factor of activated T cells (NFATC2), which includes 28 proline residues located in functionally important serine–proline (SP) repeats. The complete assignment of the NFATC2 regulatory domain enabled us to study phosphorylation of NFAT by kinase PKA and phosphorylation-dependent binding of chaperone protein 14-3-3 to NFAT, providing mechanistic insight on how 14-3-3 regulates NFAT nuclear translocation.

15N detection | IDP | NFAT | NMR resonance assignment | nuclear localization

Over the last two decades, proteins or protein segments that lack stable secondary structure have been recognized for their importance. Collectively referred to as intrinsically disordered proteins (IDPs) (1, 2), they orchestrate diverse cellular processes from transcription and translation to self-assembly and signaling (2–8). Consequently, disordered regions occur with greater probability in human proteins associated with cancer (~79%) and signaling (~66%). IDPs show remarkable structural flexibility by either transiently forming secondary structure, and/or adopting stable conformations upon binding (9, 10). Some IDPs sample ordered conformations resembling classical secondary structural elements (9), while other retain dynamics in their bound conformation resulting in "fuzzy" complexes (11). The inherently disordered nature of these proteins combined with their structural plasticity renders structural characterization methods such as X-ray crystallography and cryo-EM ineffective.

Solution-state NMR spectroscopy has been the primary structural tool; however, traditional ^{1}H-detected experiments have several limitations for studying IDPs, especially those with large molecular weight and many repeat sequences. Their conformational heterogeneity, prevalence of repetitive sequences, low fraction of hydrophobic amino acids, and high occurrence of charged amino acids homogenize the chemical complexity of IDP residues and lead to extremely narrow chemical shift dispersion (H_N; ~1 ppm). This is further compounded by a high frequency of functionally critical proline residues (12, 13), which cause "sequential connectivity breaks" in the traditional $^{1}H_N$-detected backbone assignment scheme due to the lack of backbone amide protons in proline. Furthermore, prolines are known to undergo *cis/trans* isomerization in small unstructured peptides (14, 15) and protein loops (16), which dramatically increases the complexity of the NMR spectra by causing neighboring resonances to exist in multiple chemical environments. In addition, the lack of a stable hydrogen-bonding network and secondary/tertiary structure facilitates the rapid exchange of labile amide protons with bulk water, which broadens the resonances. Exchange broadening of amide protons is exacerbated at physiological pH (7.0–7.5) and ambient temperatures (25–37 °C)—compounding challenges for in-cell NMR experiments. To address these challenges, ^{13}C-detected (17) and higher-dimensional experiments (18–20), as well as experiments that utilize long-range ^{13}C couplings (21–23), have been developed. ^{15}N-detected experiments, with distinct advantages and disadvantages in comparison with ^{1}H and ^{13}C experiments,

Significance

Intrinsically disordered proteins (IDPs) have attracted significant attention due to their roles in crucial cellular processes. NMR is the only technique that allows the study of IDPs at atomic-level resolution. However, narrow chemical shift dispersion, rapid exchange with solvent, and high proline content challenge conventional ^{1}H-detected experiments. Here, we report the development of a suite of 3D experiments based on ^{15}N direct detection that harnesses the slow relaxation and the larger chemical shift dispersion of ^{15}N nuclei for complete backbone assignment of IDPs, including proline residues, which are critical to the study of IDPs. Using this approach, we have assigned the regulatory domain of NFATC2 and have identified a likely mechanism by which 14-3-3 proteins regulate NFAT nuclear translocation.

Author contributions: K.T., G.W., and H.A. designed research; S.C., P.F., A.D., A.B., D.M., W.B., and H.A. performed research; S.C., P.F., K.T., A.D., J.J.Z., A.B., D.M., W.B., N.E.D., G.W., and H.A. analyzed data; and S.C., K.T., G.W., and H.A. wrote the paper.

Reviewers: R.W.K., St. Jude Children's Research Hospital; and P.E.W., The Scripps Research Institute.

The authors declare no conflict of interest.

[1]S.C. and P.F. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: gerhard_wagner@hms.harvard.edu or hari@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717560115/-/DCSupplemental.

have also been introduced in recent years for studying IDPs and show comparable sensitivity to their ^{13}C-detected counterparts (24, 25). Furthermore, they are orthogonal to the existing paradigm of ^{1}H and ^{13}C experiments.

Here, we present a suite of six 3D ^{15}N-detected experiments to overcome the challenges of studying IDPs with traditional ^{1}H-detected assignment strategies. These experiments exploit the greater chemical shift dispersion in the ^{15}N dimension relative to that of $^{1}H_N$ ($^{15}N_H \sim 30$ ppm) and combine it with the chemical shift dispersion in the ^{13}C dimension by encoding the correlated ^{13}C frequency in the indirect dimension. This enables direct detection and assignment of all amino acids, including frequently occurring prolines, as these experiments do not rely on amide protons. Although the low gyromagnetic ratio and the associated lower intrinsic sensitivity are disadvantages of ^{15}N detection, the slower transverse relaxation rate of ^{15}N nuclei results in ultra-narrow linewidths and a corresponding increase in signal height and sensitivity. ^{15}N detection also encodes nuclei with the largest dispersion and narrowest intrinsic linewidths in the direct dimension, which allows access to high resolution without sacrificing instrument time. The advantages of ^{15}N detection mean that the 3D experiments developed here can provide essential information necessary for the unambiguous assignment of IDP backbone resonances.

These ^{15}N-detected experiments were used to assign the backbone resonances of human transcription factor nuclear factor of activated T cells (NFATC2) (residues 131–294). In this report, we use NFAT to mean NFATC2, unless otherwise specified. This portion of NFAT is disordered and harbors 28 proline residues (17% of the residues in the region), many of which are located in functionally critical serine–proline (SP)-rich regions. The regulatory region of NFAT (residues 131–294) used in this study harbors phosphorylation sites modified by protein kinase A (PKA). NFAT phosphorylation by PKA is known to create binding sites for 14-3-3 (26); however, the mechanism by which the NFAT–14-3-3 complex regulates transcriptional activity is not understood. Complete backbone assignment of the NFAT (131–294) regulatory domain using ^{15}N-detected experiments enabled us to follow phosphorylation by PKA. Our data demonstrate that, upon PKA phosphorylation, 14-3-3 binds to NFAT and blocks access to the nuclear localization sequence (NLS), which may prevent the nuclear transport protein importin from translocating NFAT to the nucleus, thus preventing transcriptional activation.

Results

Suite of 3D ^{15}N-Detection Experiments. Six separate 3D experiments are reported here that encode the highly resolved ^{15}N chemical shift in the direct dimension. All experiments start from either $^{1}H^{\alpha}$ or ^{1}H-aliphatic spins to maximize sensitivity by starting with a sensitive hydrogen nucleus and to minimize the recovery delay after acquisition (24). The experiments are designed to be performed in a D_2O solvent to take advantage of the slower relaxation of $^{15}N_D$ compared with that of proton-decoupled $^{15}N_H$ at conventional magnetic fields (25, 27) (*SI Appendix*, Fig. S1). All of the experiments are designed in an out-and-stay fashion where frequency encoding is done during magnetization transfer using constant-time evolution. The general magnetization flow in these experiments is shown in Fig. 1. Here, in the pulse sequence names, the lowercase letters denote magnetization transfer without frequency encoding and the uppercase letters represent those transfers with frequency encoding. The hcaCOCAN experiment provides the intraresidue and sequential connectivities of $^{13}C'_{i-1}/^{13}C^{\alpha}_{i-1}$ pairs with $^{15}N_{i-1}$ and $^{15}N_i$ (Fig. 1 and *SI Appendix*, Fig. S2). The hCACON can be used to disambiguate the $^{15}N_{i-1}$ and $^{15}N_i$ resonances by isolating the $^{13}C^{\alpha}_{i-1}$ and $^{15}N_i$ correlations (Fig. 1 and *SI Appendix*, Fig. S3). The hcaCOCAN and hCACON experiments provide information analogous to the ^{1}H-detected HNCA and HNcoCA, respectively. Assignment ambiguities and degeneracies can be further resolved by incorporating $^{13}C^{\beta}$ chemical shifts. We designed experiments with three different variations to exploit the dispersion in the $^{13}C^{\beta}$ chemical shift. First, we designed the hCBCACON experiment (Fig. 1 and *SI Appendix*, Fig. S4), which provides connectivity between the $^{13}C^{\alpha}_{i-1}/^{13}C^{\beta}_{i-1}$ pair and their corresponding $^{15}N_i$ resonance, dispersed by the $^{13}C'_{i-1}$ frequency in the third dimension. This experiment shares two common dimensions (^{15}N and $^{13}C'$) with the hcaCOCAN and hCACON experiments and can be used to break degeneracy in the $^{13}C^{\alpha}$-encoded experiments by providing amino acid-specific $^{13}C^{\beta}$ chemical shift. In addition, we developed the HBHACBCAN experiment (Fig. 1 and *SI Appendix*, Fig. S5), which provides both sequential and intraresidue connectivities between $^{1}H^{\alpha}_{i-1}{}^{13}C^{\alpha}_{i-1}/^{1}H^{\beta}_{i-1}{}^{13}C^{\beta}_{i-1}$ pairs and the corresponding $^{15}N_{i-1}$ and $^{15}N_i$. To unambiguously assign the sequential resonances, we use the HBHACBCAcoN experiment (Fig. 1 and *SI Appendix*, Fig. S6), which provides only the sequential connectivities between $^{1}H^{\alpha}_{i-1}{}^{13}C^{\alpha}_{i-1}/^{1}H^{\beta}_{i-1}{}^{13}C^{\beta}_{i-1}$ pairs and $^{15}N_i$ resonances. These experiments encode the chemical shifts of aliphatic protons in a semiconstant time (SCT) fashion (28) without significant loss of sensitivity. The sixth experiment, the hCcaCON–total correlation spectroscopy (TOCSY), provides amino acid-specific information by correlating side-chain carbon ($^{13}C^{aliph}_{i-1}$) chemical shifts to the $^{15}N_i$ resonances, with dispersion by the $^{13}C'_{i-1}$ frequency in the third dimension (Fig. 1 and *SI Appendix*, Fig. S7).

The hCACON, hCBCACON, and hCcaCON-TOCSY experiments follow similar magnetization pathways. Initial proton magnetization is transferred to the aliphatic carbons by a simple INEPT step (29) followed by the encoding of the $^{13}C^{\alpha}/^{13}C$-aliphatic carbon frequencies in a constant-time (CT) fashion (30, 31). The CT period eliminates $^{1}J_{CC}$ homonuclear as well as $^{1}J_{CH}$ and $^{1}J_{CN}$ heteronuclear couplings. The $^{13}C^{\alpha}/^{13}C^{\beta}$-aliphatic evolution period is also concatenated with coherence transfer to $^{13}C'$. The $^{13}C'$ coherence is encoded in a CT manner while being transferred to the $^{15}N_i$ spins using an INEPT sequence. The antiphase C'_zN_x coherence is refocused and detected as ^{15}N in-phase coherence. An adiabatic WURST pulse (32) with MLEV-16 supercycle (33) is used to achieve broadband decoupling of $^{13}C'$ and $^{13}C^{\alpha}$ spins, and ^{2}H is decoupled using a WALTZ-16 composite pulse decoupling scheme (34). The sequential hcaCOCAN experiment uses a slightly modified strategy. The initial $^{1}H^{\alpha}$-to-$^{13}C^{\alpha}$ INEPT sequence is followed by a second INEPT transfer to $^{13}C'$ spins. These two transfer steps are used primarily to gain sensitivity, and there is no frequency encoding during these transfer steps. The antiphase $C^{\alpha}_zC'_x$ coherence evolved in the transfer step is not refocused until used in the $^{13}C'$ to $^{13}C^{\alpha}$ back-transfer step. The $^{13}C'$ frequency is then encoded in a real-time fashion, followed by a $^{13}C'$-to-$^{13}C^{\alpha}$ INEPT back-transfer sequence. The $^{13}C^{\alpha}$ chemical shift is encoded as the second indirect dimension with a CT block while the $^{13}C^{\alpha}$ coherence is transferred to $^{15}N_{i-1}/^{15}N_i$. Finally, the antiphase $C^{\alpha}_zN_x$ coherence is refocused and detected as ^{15}N in-phase coherence with simultaneous ^{13}C and ^{2}H decoupling.

In the HBHACBCAN and HBHACBCAcoN experiments, $^{1}H^{\alpha}_{i-1}/^{1}H^{\beta}_{i-1}$ chemical shifts are encoded in a SCT fashion while the coherence is INEPT transferred from the aliphatic protons to the aliphatic carbons to minimize sensitivity loss due to relaxation. The $^{13}C^{\alpha}/^{13}C^{\beta}$-aliphatic carbon frequencies are then encoded in a CT fashion. In the HBHACBCAN experiment, the coherences are transferred to both $^{15}N_{i-1}/^{15}N_i$ providing sequential and intraresidue connectivity, whereas the HBHACBCAcoN experiment transfers the magnetization to C', which results in connectivity only to the $^{15}N_i$ resonance. ^{15}N is detected after refocusing the antiphase $C^{\alpha}_zN_x$ or C'_zN_x coherences to in-phase with simultaneous ^{13}C and ^{2}H decoupling as described above. Combining the HBHACBCAN and the HBHACBCAcoN enables

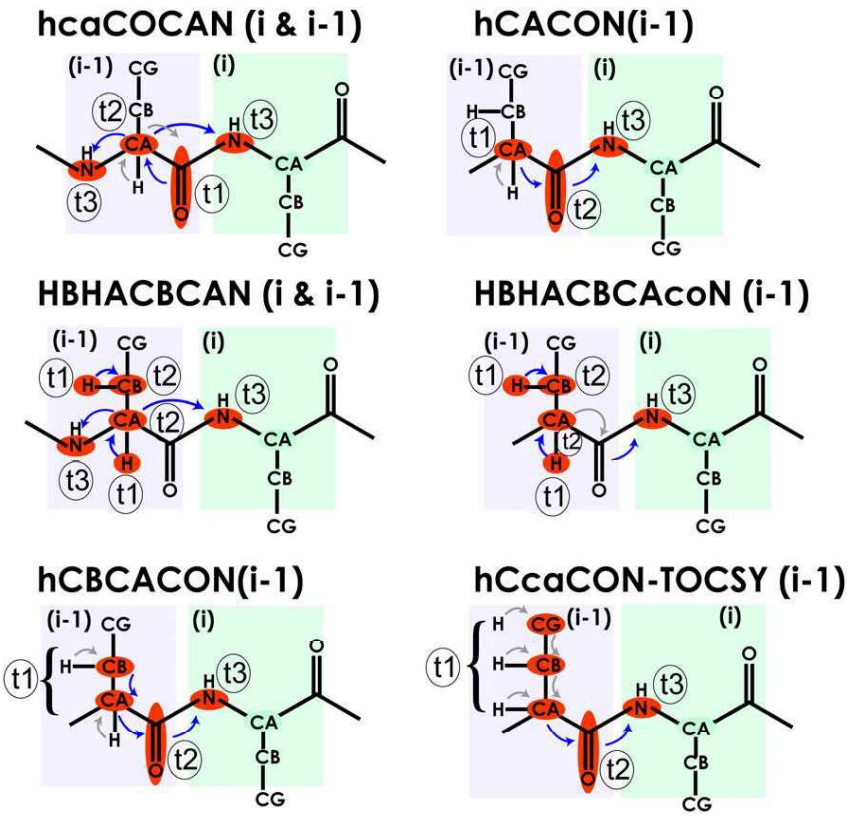


Fig. 1. Magnetization flow in the panel of ^{15}N-detected experiments. Atoms that are frequency encoded in each experiment are colored in red. t_1 and t_2 denote the atoms that are frequency encoded in the two indirect dimensions, and t_3 is ^{15}N nuclei that is encoded in the direct dimension. Blue arrows indicate frequency labeling during magnetization transfer, and gray arrows indicate magnetization transfer without frequency encoding. Relevant atoms of a given residue i and those of the preceding residue $i-1$ are highlighted in light green and light purple boxes, respectively.

the unambiguous assignment of $^{13}C^{\alpha}/^{13}C^{\beta}$ with their backbone ^{15}N resonances. Representative 2D planes from each of these 3D experiments are shown in *SI Appendix* along with a table describing acquisition parameters (*SI Appendix*, Figs. S8–S14 and Table S1).

Complete Backbone Assignment of NFAT Regulatory Domain. To demonstrate the utility of our ^{15}N-detection experiments, we assigned the disordered N-terminal regulatory domain of the human transcription factor NFAT ($NFAT_{131–294}$). NFAT is ~1,000 aa in length, and only the DNA binding domain (residues 392–678 in human NFAT) is structured (*SI Appendix*, Fig. S15*A*). The N-terminal regulatory domain of NFAT ($NFAT_{131–294}$), which harbors NLS (residues 251–253), is predicted to be intrinsically disordered (*SI Appendix*, Fig. S15*B*) and displays narrow $^{1}H_N$ dispersion (*SI Appendix*, Fig. S16). This regulatory region of NFAT harbors 18 phosphorylation sites and is targeted by five different kinases (35). In its phosphorylated form, NFAT resides inactive in the cytoplasm. Immunoreceptor and receptor tyrosine kinase (RTK) activation results in the subsequent activation of calcineurin, which dephosphorylates NFAT and induces its translocation to the nucleus, where it initiates transcription of key immune response factors (e.g., IL2, TNFα, INF-γ) (36–38). $NFAT_{131–294}$ contains 28 prolines (~17%), as part of the SP-rich regions, which present a major challenge when assigning residues by conventional ^{1}H-detected methods (*SI Appendix*, Fig. S15*C*), thus slowing determination of the mechanism of NFAT cytoplasmic sequestration.

The suite of six ^{15}N-detected experiments were applied to $NFAT_{131–294}$. Nonuniform sampling (NUS) was used to efficiently acquire high-resolution spectra, with 12% of the Nyquist grid sampled in the two indirect dimensions (39). A striking feature of all of these experiments is the presence of proline ^{15}N resonances at around 138 ppm. Well-separated from all other amino acid types, these "proline islands" can clearly be seen in the overlay of the hCACON and hcaCOCAN spectra (*SI Appendix*, Fig. S8). Moreover, the hCACON and hcaCOCAN experiments display linewidths that are extremely narrow in the ^{15}N dimension, allowing convenient assignment from 2D planes (CAN) alone. These proline islands are created from their unique ^{15}N frequency coupled with their upfield $^{13}C^{\alpha}$ frequency and can facilitate assignment through stretches of amino acids interspaced with prolines (Fig. 2). In this case, we were able to obtain complete ^{15}N, $^{13}C^{\alpha}$, and $^{13}C'$ resonance assignments from the 3D hCACON and hcaCOCAN experiments alone. In situations when chemical shift ambiguity and degeneracy stymie the assignment process, the information from the above two experiments can be combined with $^{13}C^{\beta}$ or ^{13}C-aliphatic chemical shifts obtained from the hCBCACON, HBHACBCAN, HBHACBCAcoN, and hCcaCON-TOCSY experiments. The established assignments of NFAT are best represented in the CON plane of the hCACON experiment (Fig. 3), where there is one resonance for every amino acid in the primary sequence, making it similar to a ^{1}H-^{15}N–heteronuclear single-quantum coherence (HSQC). It should be noted, however, that the CON planes also include proline resonances, which would be absent in a traditional ^{1}H-^{15}N–HSQC (Fig. 3).

Transfer of Assignments to ^{1}H-Detected Experiments. The resonance assignments obtained using ^{15}N detection can be subsequently transferred to ^{1}H-detected experiments such as a ^{1}H-^{15}N–HSQC. This is accomplished by transferring the ^{15}N and $^{13}C'$ assignments from a ^{15}N-detected hCACON experiment to the CON plane of a ^{1}H-detected HNCO experiment and then subsequently correlating

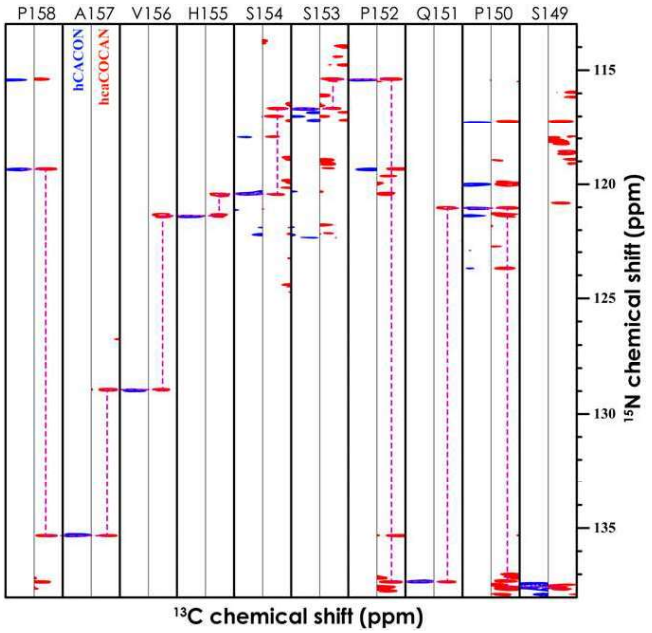


Fig. 2. An example of sequential connectivity through proline residues using ^{15}N-detection experiments for NFAT$_{131–294}$. Strip plot showing resonances from hCACON (blue) and hcaCOCAN (red) spectra. The experimental time was 3.5 d per experiment on a ~1 mM sample. NUS was used to collect 12% of the Nyquist grid in the two indirect dimensions (see *SI Appendix*, Table S1 for detailed acquisition parameters).

to the amide proton resonances. Users must account for the isotope shift of experiments conducted in D_2O vs. H_2O. This shift is typically 0.08 and 0.75 ppm in the $^{13}C'$ and ^{15}N dimensions, respectively, and is not sensitive to amino acid type (40, 41). A comparison of the CON planes from ^{1}H- and ^{15}N-detected experiments for NFAT$_{131–294}$ is shown in Fig. 4. The majority of peaks overlap after correction of the deuterium isotope shift, but a small fraction of the peaks shifted and were assigned by visual inspection. The difference in peak position between the ^{15}N-detected and ^{1}H-detected experiments could be due to partial hydrogen bonding and requires further investigation.

2D-CON Correlation Provides a Better Spectral "Fingerprint" for IDPs. The 2D-CON correlation provides a more representative fingerprint for IDPs compared with the ^{1}H-^{15}N–HSQC. The signals in the 2D-CON are better dispersed than those in the ^{1}H-^{15}N–HSQC for IDPs (18, 42). The CON correlations are less sensitive to exchange broadening even at higher pH conditions, and the proline residues are visible. In addition, there are no problems originating from imperfect water suppression, something occasionally encountered in ^{1}H-detected ^{1}H-^{15}N–HSQC experiments. The CON plane can be recorded either using ^{15}N-detected or ^{13}C-detected experiments [hcaCON for ^{15}N-detected and hcacoNCO[in-phase anti-phase (IPAP)] for ^{13}C-detected experiments, respectively] (43). It should also be noted that having a proline residue right after a phosphorylation site is a rather common feature in IDPs. The phosphorylation of serine–proline/threonine–proline (SP/TP) sequences is often associated with the regulation of functional modules in IDPs, including the control of subcellular localization through activation or inhibition of NLS or nuclear export signal (NES) motifs. The effect of phosphorylation on neighboring proline residues cannot be defined in ^{1}H-detected ^{1}H-^{15}N–HSQC experiments; however, the effect would be evident in a 2D-CON plane. It should be noted that resonances corresponding to the side chains of asparagine and glutamine would also appear in the ^{15}N-detected CON spectrum at around 35 ppm. These can be avoided by using a selective C^{α} pulse that does not excite these side-chain resonances (C^{γ} of glutamine and C^{β} of asparagine).

Sensitivity Comparison of ^{15}N- and ^{13}C-Detected 2D-CON. A one-to-one comparison of sensitivity between the ^{15}N- and ^{13}C-detected methods is not straightforward due to the differences in inherent probe design (^{13}C-optimized vs. ^{15}N-optimized TXO probes) and differences in the style of frequency encoding (CT/SCT vs. real time evolution) in the pulse sequence design. Theoretical estimates indicate that the ^{15}N-detected experiment has an advantage in sensitivity (in time-equivalent experiments) compared with its ^{13}C-detected counterpart *SI Appendix, Comparison of the Signal Height in ^{15}N- vs. ^{13}C-Detected CON Experiments*). To directly compare the sensitivities of the ^{15}N- vs. ^{13}C-detected CON experiments, we modified the ^{15}N-detected CON experiment to encode the C' chemical shift in a semiconstant time fashion, which allowed the acquisition of ^{13}C and ^{15}N experiments with the same resolution in both the direct and indirect dimensions.

To compare the relative sensitivities of the ^{15}N- and ^{13}C-detected CON experiments, we recorded time-equivalent ^{15}N-detected hcaCON and ^{13}C-detected hcacoNCO-IPAP experiments (*SI Appendix*, Figs. S13 and S14 and Table S2). The spectral width was set to 40 ppm (~3,422 Hz) in the ^{15}N dimension to include the proline resonances and 10 ppm (~2,012 Hz) in the $^{13}C'$ dimension (*SI Appendix*, Table S2). These spectral widths are typical for IDPs. The nucleus with the larger spectral width and slower relaxation (^{15}N) requires a larger number of points to be recorded, giving the ^{15}N-detected experiment an immediate advantage over the ^{13}C-detected experiment in which each point in the ^{15}N dimension needs to be recorded as a separate free induction decay (FID). In contrast, in the ^{15}N-detection experiment, high resolution in the ^{15}N dimension can be achieved with a negligible increase in experimental time. We observe that overall the peak heights are larger in the ^{15}N-detected experiment than in the ^{13}C-detected experiment when the noise levels of the spectra are set to be equal. We also observe broader linewidths in the $^{13}C'$ dimension in the ^{13}C-detected CON plane, despite the $^{13}C'$ evolution times being equal in both experiments. One probable reason for this difference is the use of an IPAP scheme in the ^{13}C-detected experiment. This requires that the in-phase and anti-phase signals be recorded as two different spectra, in an interleaved fashion. The final spectrum is obtained by shifting these two spectra in frequency space by a value equivalent to the scalar coupling between C'-Ca (~55 Hz) and then adding them together. At this high resolution, even a small discrepancy between the chosen value of the scalar coupling and the actual scalar coupling will lead to broadening. Another advantage of ^{15}N-detected experiment is that it does not require an IPAP scheme, as the ^{13}C-detected experiment does, to remove scalar couplings.

Phosphorylation of NFAT by PKA and Subsequent Binding by 14-3-3. PKA phosphorylation of NFAT promotes its association with the 14-3-3 chaperone protein, which results in the inhibition of NFAT transcriptional activity (26). The proximity of the 14-3-3 binding site to the NLS of NFAT has led to speculation that 14-3-3 binding could block importin binding to the NLS, thus preventing translocation of NFAT to the nucleus, and inhibiting its activity (26). However, mechanistic details of PKA-dependent NFAT regulation have yet to be elucidated. To understand the importance of phosphorylation by PKA, we performed an in vitro phosphorylation of NFAT by PKA. To follow the effect of phosphorylation on the NFAT regulatory domain, backbone assignments were transferred from the CON plane of a ^{15}N-detected experiment to a conventional ^{1}H-^{15}N–HSQC.

The kinase reaction was performed in an NMR tube by adding unlabeled PKA to 50 μM NFAT. Phosphorylation of NFAT was followed as a function of time by ^{1}H-detected ^{1}H-^{15}N–HSQC

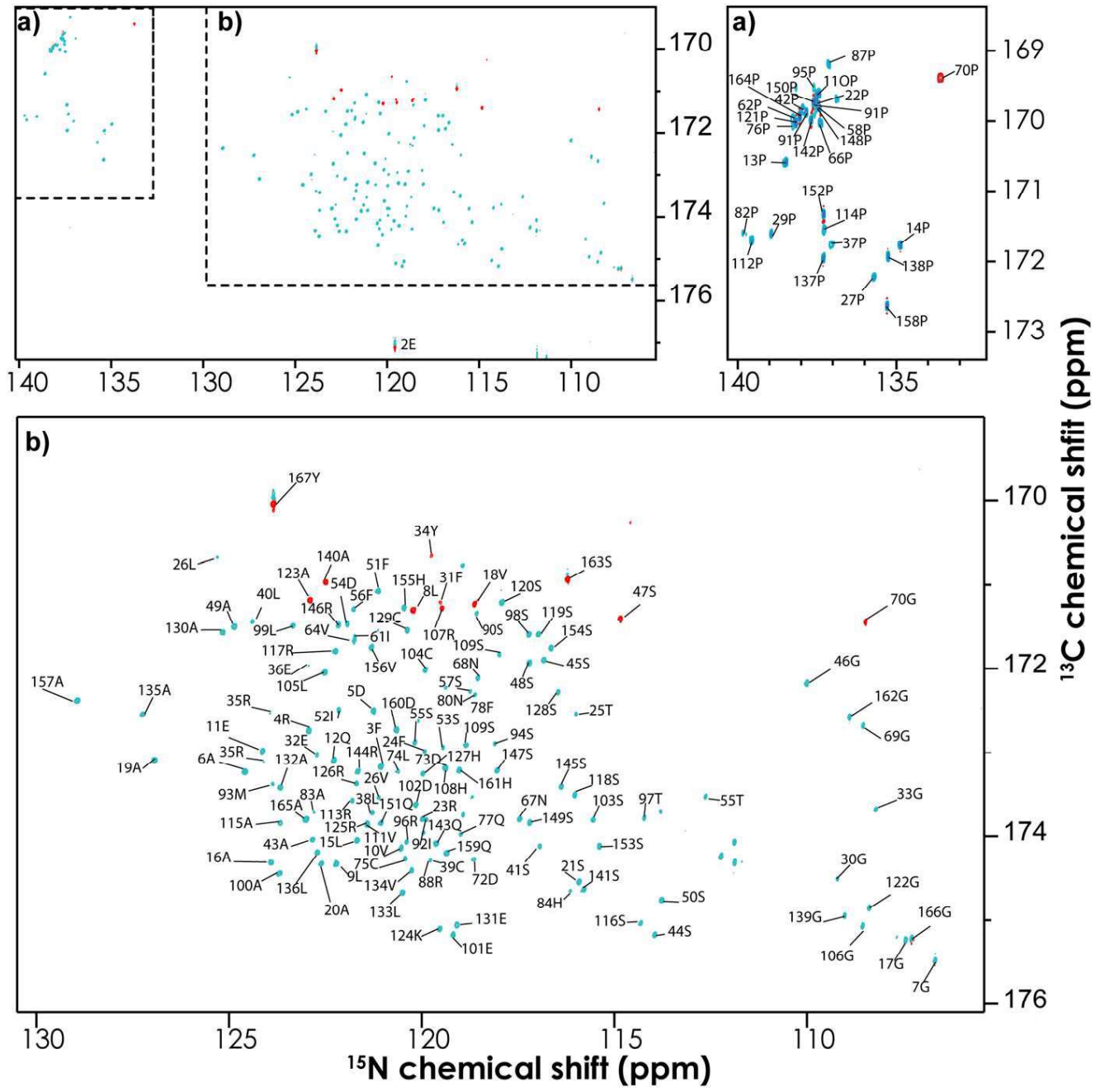


Fig. 3. A 2D-CON plane of the 3D hCOCAN experiment showing the assignments for the regulatory region of NFAT$_{131–294}$. Residues preceding glycines have negative intensity (red) due to the constant-time nature of the $^{13}C^{\alpha}$ encoding. *Insets A* and *B* are expanded to display the resonance assignments.

experiments. NUS was employed to collect each ^{1}H-^{15}N-HSQC at 20-min intervals. PKA phosphorylation resulted in downfield shifts of the ^{1}H resonances of two serine residues that are assigned to S-225 and S-255. We could readily assign phosphoserine resonances by following the disappearance of the parent serine or nearby residues and the concomitant appearance of a new signal representing the phosphorylated residue (Fig. 4 and *SI Appendix*, Fig. S17). The phosphoserine assignments were later confirmed by 2D (CAN) planes of the ^{15}N-detected hCACON and hcaCOCAN experiments acquired on the phosphorylated sample (*SI Appendix*, Fig. S18). It is interesting to note that the rates of phosphorylation of the two serine residues vary significantly. Under the current experimental conditions, S-255 is phosphorylated first and completely, whereas phosphorylation of S-225 is incomplete with a subpopulation of S-225 remaining unphosphorylated (Fig. 4*D* and *SI Appendix*, Fig. S17). An overlay of the ^{1}H-^{15}N–HSQC spectra of the control and phosphorylated samples is shown in Fig. 4*C*, and the phosphoserine resonances are highlighted. Interestingly, some of the unphosphorylated serine residues have weaker intensities in the ^{1}H-detected ^{1}H-^{15}N–HSQC spectrum and the corresponding ^{1}H-detected 3D experiments, as had been reported previously in IDPs (44, 45). However, in the ^{15}N-detected experiments, the intensities of all of the serine residues are comparable with those of the other amino acids.

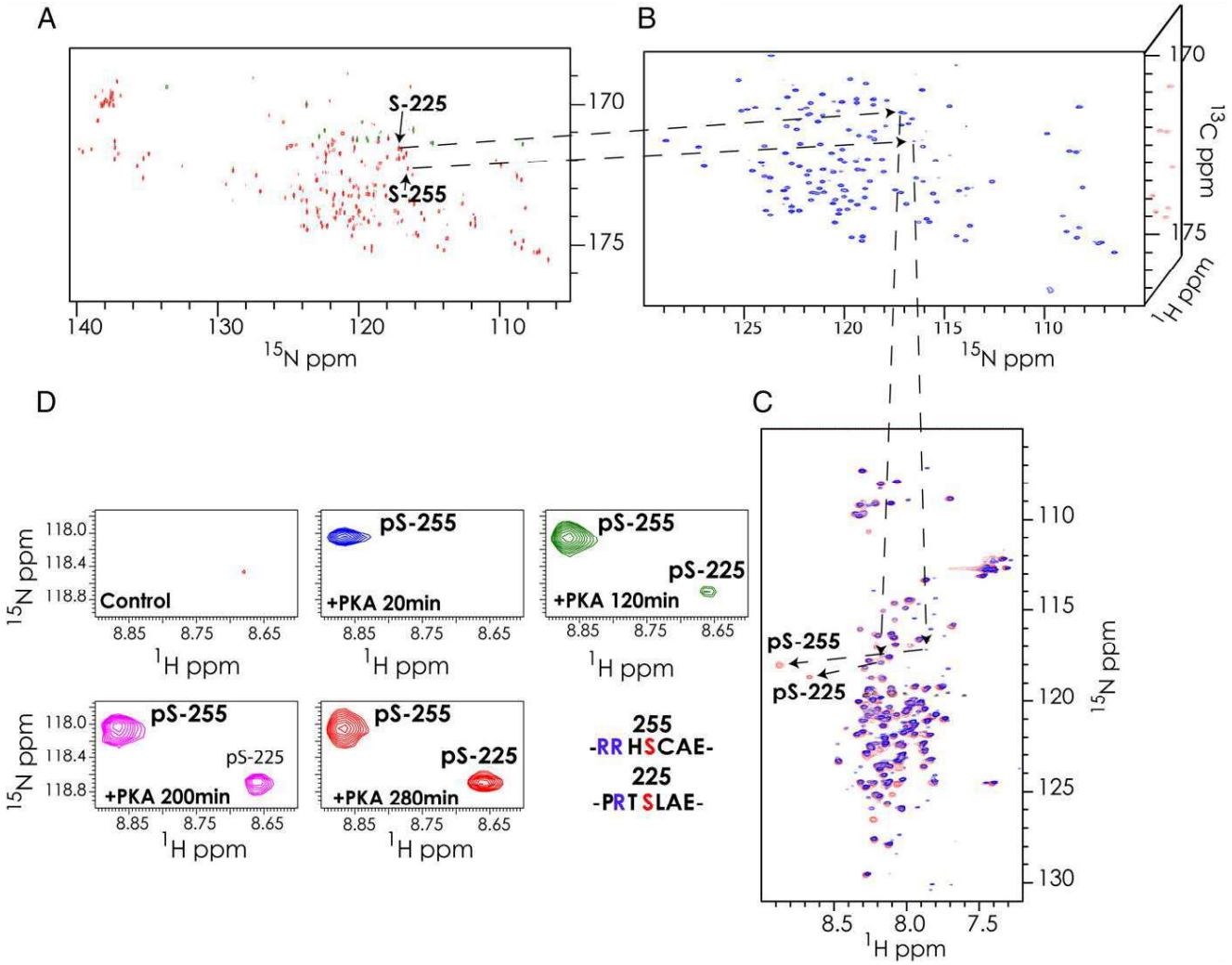


Fig. 4. Transfer of assignments from ^{15}N-detected to ^{1}H-detected experiments. (*A*) A 2D-CON plane from the ^{15}N-detected hCACON experiment, and (*B*) a 2D-CON plane from the ^{1}H-detected 3D-HNCO experiment of the regulatory region of $NFAT_{131–294}$. Assignments of serine-225 and serine-255 resonances obtained in the ^{15}N-detected experiments were transferred to the CON plane of the ^{1}H-detected HNCO by matching the nitrogen and carbonyl frequencies common to both experiments, thus allowing assignment of the amide ^{1}H resonances. Once the ^{1}H-^{15}N resonances are assigned, conventional ^{1}H-^{15}N–HSQC experiments can be used to follow interactions and modifications such as phosphorylation. (*C*) NFAT is phosphorylated in vitro by the kinase PKA. An overlay of a ^{1}H-detected ^{15}N-HSQC spectrum of the control sample (blue) and the phosphorylated sample (red) is shown. The two phosphorylated residues are highlighted. (*D*) Time course of the phosphorylation events on NFAT by the kinase PKA. The region of the ^{1}H-^{15}N–HSQC spectrum harboring the phosphoserine resonances is highlighted. Five panels corresponding to different time periods following initiation of phosphorylation in addition to the control spectrum are shown. Each spectrum was recorded for 20 min, and the data in each spectrum represent the average of events in each 20-min interval. The amino acid sequences of the seven-residue region centered on phosphoserine residues 225 and 255 are shown.

PKA is a basophilic kinase and has a strong preference for basic residues (R/K) two ($i - 2$) and three ($i - 3$) residues before (N-terminal) the phosphorylation site i (46). Although PKA is a serine/threonine kinase, it strongly prefers serine for phosphorylation (47). Therefore, PKA will preferentially target sites with the consensus (R/K)XS or (R/K)XXS for modification. In fact, many of the key substrates of PKA have two basic residues and conform to the consensus (R/K)(R/K)XS (47). A proline at position $i + 1$ is largely incompatible with PKA modification, which distinguishes the modification targets of PKA from that of proline-directed kinases (48). It has been postulated that GSK3, a kinase that phosphorylates NFAT after transcriptional activation and promotes NFAT nuclear export (49), requires prior phosphorylation by a priming kinase such as PKA (50).

The regulatory region of NFAT ($NFAT_{131–294}$) used in this study harbors 10 possible phosphorylation sites for a basophilic kinase, namely, T152, S217, S225, S236, S243, S246, S247, S255, S274, and S276. However, only S255 and S225 are actually modified by PKA. S255 is a strong PKA site with two basic residues (Arg) at the $i - 2$ and $i - 3$ positions and dominant phosphorylation kinetics in the in vitro PKA phosphorylation assay. Phosphorylation of this site by PKA has been previously observed in NFAT2 and has been shown to control nuclear import (49, 50). S225, which harbors only one basic residue (Arg) at position $i - 2$, is phosphorylated by PKA after S255 (Fig. 4*D*). This order of phosphorylation is expected as S255 conforms more closely to the optimal PKA phosphorylation motif. S255 and the arginines/lysines at positions $i - 2$ and $i - 3$ are conserved in all four isoforms of NFAT (Fig. 5*A*). However, S225 and the arginine at position $i - 2$ are conserved in isoforms 1, 2, and 3, but not in 4. The biological significance of this difference is not known.

Five out of the eight remaining unmodified sites in NFAT (S217, S236, S247, S274, and S276) have a proline at the $i + 1$

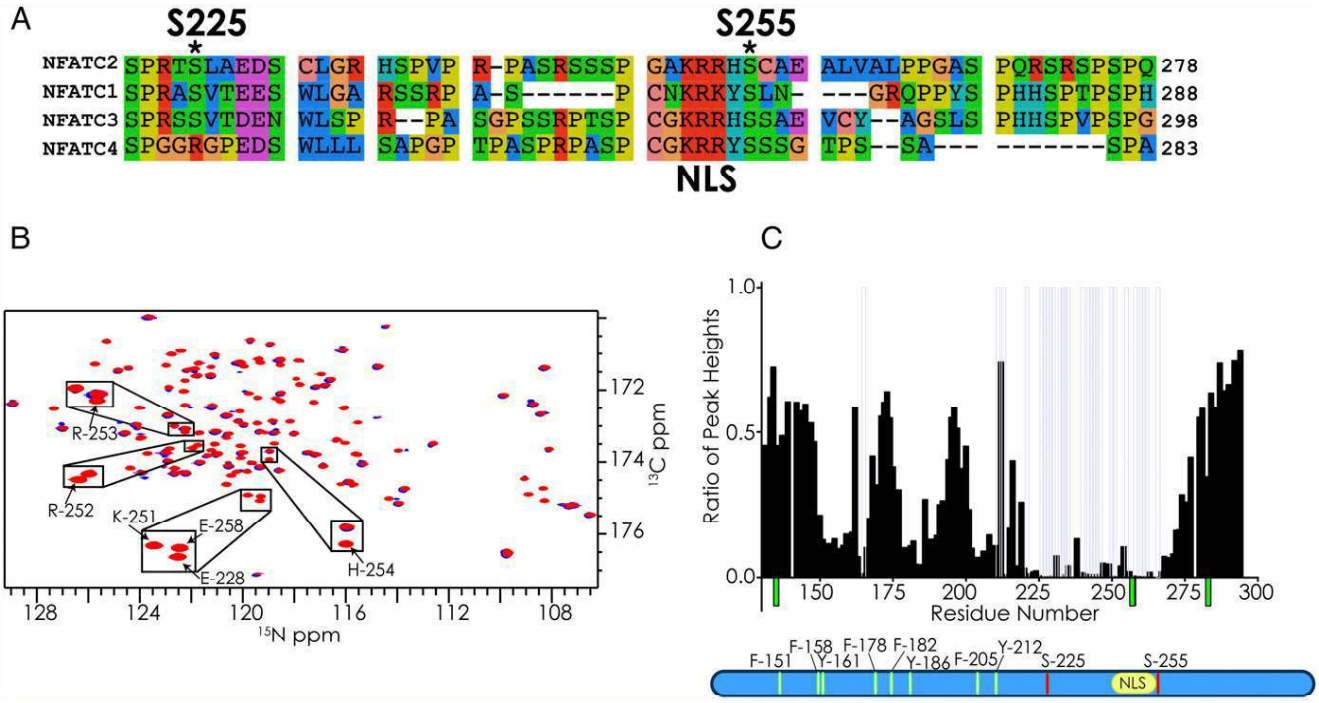


Fig. 5. (*A*) Sequence alignment of the region harboring the NLS and PKA phosphorylation sites for all four isoforms of NFAT. (*B*) Overlay of the 2D-CON planes (from ^{1}H-detected 3D-HNCO experiments) of PKA-phosphorylated NFAT alone (red) and PKA-phosphorylated NFAT with the addition of unlabeled 14-3-3 in the molar ratio 1:1 (blue). Resonances corresponding to the residues of the NLS are highlighted. (*C*) The ratios of peak heights of PKA-phosphorylated NFAT with and without the addition of 14-3-3. Transparent gray bars indicate the residues whose resonances are broadened to below five times the noise level of the spectrum. Green bars represent residues that were not assigned in the PKA-phosphorylated sample. The PKA phosphorylation site (red), NLS (yellow), and the aromatic amino acids (green) are shown in the schematic below.

position after the serine residue, which is disfavored by PKA. These serine sites would be preferred by DYRK or MAP kinases, additional kinases that are known to phosphorylate NFAT. T152 is a suboptimal PKA substrate due to the presence of threonine at the phosphorylation site (i) instead of serine. Residues S243 and S246 have arginine residues at position $i - 2$; but they are not phosphorylated by PKA. This may be due to the presence of prolines at positions other than the $i + 1$ position, potentially making these serines unsuitable for PKA modification.

PKA Regulated NFAT–14-3-3 Interaction and NLS Masking. 14-3-3 is a dimeric phosphodependent chaperone that regulates an array of diverse pathways (51, 52). 14-3-3 binding is a common regulatory mechanism to modulate accessibility to functional modules in disordered regions (53), allowing 14-3-3 to act as a regulatory switch that decodes the cell state-dependent phosphorylation pattern generated by active kinases.

There are numerous experimentally characterized examples of 14-3-3 regulating key cellular proteins including cystic fibrosis transmembrane conductance regulator (CFTR) (54), Bcl2-associated agonist of cell death (BAD) (55), and RAF proto-oncogene serine/threonine-protein kinase (RAF1) (56). Often, 14-3-3 acts as a switch to control the localization of a protein by modulating access to subcellular localization signals. Examples of this include 14-3-3 binding to Forkhead box protein O3 (FOXO3) (57), Yes-associated protein (YAP) (58), carbohydrate-response element-binding protein (ChREBP) (59), and protein Mdm4 (60). The classical consensus for 14-3-3 binding is RXXS$_p$XP, which overlaps with the phosphorylation motif consensus of basophilic kinases. However, there is no simple rule that can describe all instances of 14-3-3 engagement (61). Furthermore, 14-3-3 targets can harbor a single or multiple 14-3-3 binding sites (59, 61). As 14-3-3 is a dimer, containing two antiparallel binding pockets (61), multivalent interactions are stronger due to avidity effects (62). Often the 14-3-3 binding sites of a protein are tuned for optimal binding strength corresponding to the desired biological function. This overall binding is mediated by strong and weak sites, based on consensus match, and often associates with a correspondingly strong or weak binding affinity (63).

When the τ isoform of 14-3-3 was added to PKA-phosphorylated NFAT, we observe broadening of resonances from residues 220–300, suggesting that 14-3-3 engages both phosphorylation sites (Fig. 5 and *SI Appendix*, Figs. S19 and S20). Since we were concentration limited in the amount of phosphorylated NFAT (0.2 mM), where the kinase PKA was the limiting factor, we recorded a 3D ^{1}H-detected HNCO to extract the 2D CON plane. The assignments for 2D CON plane (derived from ^{1}H-detected HNCO) were transferred from the ^{15}N-detected experiments as described above. These results validate the hypothesis that the activity of dimeric 14-3-3 can result from the binding of two sites on one target (64). Furthermore, we see moderate broadening of resonances near aromatic residues in NFAT, suggesting that these residues have additional interaction with 14-3-3 once the phosphorylated consensus sequence is engaged. Of note, the region that broadened upon 14-3-3 binding includes the NLS (251–253) (Fig. 5 *B* and *C*), implying that the NLS of NFAT is masked by phosphorylation-dependent binding of 14-3-3. In contrast, when 14-3-3 was added to unphosphorylated NFAT, only minor changes were observed in the resonances corresponding to the NLS, indicating that engagement of NFAT at this site by 14-3-3 is primarily driven by phosphorylation (*SI Appendix*, Figs. S19*A* and S20). It should be noted that the NMR results do indicate that 14-3-3 can bind to unphosphorylated NFAT at the concentrations and under the conditions used here. However, the primary binding interface of 14-3-3 to unphosphorylated NFAT is the region corresponding

to residues 184–204, which harbor three SP repeats and not the NLS (*SI Appendix*, Fig. S20). It has been previously shown that 14-3-3 can bind to proteins in a phosphorylation-independent manner (65–67) and that its specificity increases upon phosphorylation of its substrates (26). These results further support the model that NLS masking by 14-3-3 blocks importin binding, thus preventing the translocation of NFAT to the nucleus and the initiation of NFAT-dependent transcription.

Discussion and Conclusions

IDPs display larger chemical shift dispersion in the ^{13}C and ^{15}N dimensions than in the ^{1}H dimension (68, 69). The advantage of obtaining higher resolution outweighs the intrinsic low sensitivity of ^{13}C- and ^{15}N-detected experiments. Here, we present a suite of 3D ^{15}N-detected experiments that leverage the narrow linewidths and wide dispersion of ^{15}N signals in IDPs. These experiments start with aliphatic ^{1}H magnetization and end in ^{15}N detection with the indirect dimension encoded in a constant-time fashion during magnetization transfer. Starting the experiment with ^{1}H magnetization has two distinct advantages: (*i*) higher sensitivity and (*ii*) shorter recycling times (compared with an experiment that starts with ^{13}C or ^{15}N magnetization). In the ^{15}N-detected experiments described here, there are no residual couplings in the ^{13}C or ^{15}N dimensions, which provides narrow lines in the 3D spectra. In traditional "out-and-back"–style HNCA or HNCACB experiments, the $^{13}C^{\alpha}/^{13}C^{\beta}$-aliphatic carbon resonances are limited to an approximate digital resolution of 38 Hz because of the $^{1}J_{13C\text{-}13C}$ couplings. Here, the $^{13}C^{\alpha}/^{13}C^{\beta}$ coherences are evolved in a CT fashion removing the couplings. The amide protons in IDPs have very narrow dispersion, but in the HBHACBCAN and HBHACBCAcoN experiments we encode both $^{1}H^{\beta}$ and $^{1}H^{\alpha}$ resonances, providing better resolving power. Since the experiments are done using 100% D_2O as the solvent, we minimize broadening due to chemical exchange between $^{15}N_H$ and $^{15}N_D$ species. Furthermore, water suppression elements are not required in ^{15}N-detection experiments, and this allows the encoding of all $^{1}H^{\alpha}$ resonances. This is in contrast to ^{1}H-detected experiments performed in H_2O solvent, where the $^{1}H^{\alpha}$ resonances near to and/or overlapping with the water frequency are affected by the various water suppression schemes needed in these experiments.

An array of ^{13}C-detected experiments (17, 18, 23, 42) have been developed especially for IDPs as an alternative to traditional ^{1}H-detected methods. The ^{15}N-detected experiments described here provide correlations and connectivities between the various backbone nuclei similar to those measured by the ^{13}C-detected experiments. Despite the reduced intrinsic sensitivity of ^{15}N compared with ^{13}C due to its lower gyromagnetic ratio [$(10/4)^{1.5} = 3.9$], the slower transverse relaxation rate of ^{15}N results in narrower line widths with increased signal height (25). For example, assuming a local rotational correlation time of 8 ns at 500 MHz in D_2O, the ^{15}N-detected hCACON signal height is calculated to be slightly better than that of its ^{13}C-detected hCAcoNCO counterpart (i.e., identical $^{13}C^{\alpha}_{i}$-$^{13}C'_{i}$-$^{15}N_{i+1}$ correlations; *SI Appendix*, *Comparison of the Signal Height in ^{15}N- vs. ^{13}C-Detected Experiments*). This is primarily due to the 3.8 times slower transverse relaxation rate of $^{15}N_D$ (3.7 s^{-1}) compared with that of $^{13}C'_N$ (14.2 s^{-1}). The relative sensitivity gain of ^{15}N detection over ^{13}C-detected experiments should increase with magnet field strength due to the stronger chemical shift anisotropy of $^{13}C'_N$ spins (*SI Appendix*, Fig. S1). In addition, the increase in rotational correlation time due to transiently forming secondary structures or adoption of stable conformations upon binding would be advantageous to the ^{15}N-detected experiments as the transverse relaxation rate of $^{15}N_D$ is less sensitive to molecular weight. Direct ^{13}C-detection experiments have the additional disadvantage of one-bond ^{13}C-^{13}C scalar couplings that increase spectral crowding and diminish signal intensity. Although these problems can be managed by a judicious choice of detected carbon type (e.g., $^{13}C'$ vs. $^{13}C^{\alpha}$), isotope labeling scheme, and experimental design (e.g., IPAP, S^3E), these solutions also increase the likelihood of imperfect signal recovery, especially for $^{13}C^{\alpha}$ (as reviewed in ref. 27). Here, we report broader resonances for $^{13}C'$ in the ^{13}C-detected experiments compared with those in the ^{15}N-detected hcaCON experiment.

Adapting other optimization methods in pulse sequence design will further enhance the sensitivity of these ^{15}N-detection experiments. For example, CT periods can be replaced with $^{13}C^{\alpha}$ selective pulses to minimize the relaxation losses during $^{13}C^{\alpha}$ encoding. This would improve the sensitivity for most amino acids, except those whose $^{13}C^{\alpha}$ chemical shifts are close to $^{13}C^{\beta}$ resonances, such as glycine, serine, and threonine. The use of the accelerated ^{1}H-longitudinal relaxation of $^{1}H^{\alpha}$ spins (^{1}H-flip) would enable the hCACON and hcaCOCAN experiments to be executed with shorter recycling delays (70). The ^{15}N-detected 3D experiments described here can be easily converted to 4D experiments by encoding the $^{1}H^{\alpha}$ or ^{1}H-aliphatic chemical shift during the initial INEPT transfer in a SCT fashion. It should be noted that most of the 3D experiments presented here can also be easily converted to high-resolution 2D planes with minimal to no loss in sensitivity, as the frequency encoding in the indirect dimension is incorporated into the magnetization transfer step.

The broadening of the ^{15}N signals that are detected in the direct dimension by the weak two- and three-bond scalar couplings (1–3 Hz) to $^{1}H^{\alpha}$ and $^{1}H^{\beta}$ protons (71) can be removed by ^{1}H decoupling during acquisition. Narrowband ^{1}H decoupling can be employed, targeting the $^{1}H^{\alpha}$ and $^{1}H^{\beta}$ resonances, and thereby reducing the required decoupling power. Although modern cryogenic probes are capable of dissipating heat from simultaneous ^{13}C, ^{1}H, and ^{2}H decoupling, careful attention must be paid to the decoupling power levels and acquisition times. The experiments here were recorded at sample concentrations of 0.5–1 mM on a Bruker TXO cryoprobe where the ^{13}C and ^{15}N coils are positioned closest to the sample. The sensitivity of the ^{15}N-detected experiments will further improve with improvements in coil design in the TXO probes optimized for ^{15}N nuclei. Although the ^{15}N-detected experiments are not as sensitive as the ^{1}H-detected experiments, which can be performed at lower sample concentrations, the advantages in resolution and ability to observe prolines in the ^{15}N-detected experiments offer distinct advantages for characterizing IDPs. The ^{15}N-detected experiments described here can be readily combined with data from ^{1}H- and ^{13}C-detected experiments, especially those that were developed for characterizing IDPs.

Although the ^{15}N-detected experiments described here were applied to the intrinsically disordered region of $NFAT_{131–294}$, the application is not limited to IDPs. The relaxation rates of ^{15}N in deuterated solvent is comparable to that of the ^{15}N transverse relaxation optimized spectroscopy (TROSY) at low field strengths. Thus, these methods can be applied to structured globular proteins. The recent application of TROSY selection to ^{15}N detection would extend these experiments to larger molecular weights and higher magnetic fields (71, 72). Pulse sequence designs, like the ones presented here, that enable direct ^{15}N detection with TROSY selection (where the TROSY component of $^{15}N_H$ resonance is detected) would facilitate studies of high–molecular-weight proteins, and systems marred by spectral overlap and/or incomplete hydrogen backexchange.

Our newly developed ^{15}N-detected 3D experiments have shed light on the mechanism of the interaction of 14-3-3 with NFAT and its regulation of NFAT through NLS masking. This example illustrates how ^{15}N- and ^{13}C-detected multidimensional experiments open avenues of structural investigation into IDPs. Greater understanding of disordered regions, at atomic resolution, will allow us to dissect the functional and mechanistic details of how IDPs orchestrate cellular functions.

Materials and Methods

Expression and Purification of NFAT Regulatory Domain. The regulatory domain of human NFATC2 (residues 131–294), $NFAT_{131–294}$, was expressed as a fusion protein with an N-terminal His-GB1 solubility tag in the *Escherichia coli* strain BL21 (DE3). Cells were grown at 37 °C to an OD_{600} of 0.8. Then the temperature was dropped to 20 °C, and expression was induced by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were grown for an additional 15 h at 20 °C before harvesting. For [U-^{15}N^{13}C]-$NFAT_{131–294}$, the bacterial culture was grown using M9 media containing 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 mM $MgSO_4$, and 0.1 mM $CaCl_2$ in H_2O supplemented with 2 g/L ^{13}C-glucose and 1 g/L of $^{15}NH_4Cl$. To obtain [U-^{15}N]-$NFAT_{131–294}$, 3 g/L of ^{12}C-glucose was used in an otherwise-identical M9 media.

The harvested cells were resuspended in 40 mL of 50 mM Tris·HCl (pH 8.0), 350 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol (β-ME). The suspended cells were then disrupted by sonication, and the insoluble fraction was removed by centrifugation for 40 min at 32,914 × *g*. The protein was initially purified by gravity-flow affinity chromatography using 5 mL (10 mL of a 50% slurry) of Ni-NTA resin (Qiagen). After washing the resin with 40 mL of 50 mM Tris·HCl (pH 8.0), 350 mM NaCl, 40 mM imidazole, and 5 mM β-ME, the protein was eluted in an identical buffer containing 350 mM imidazole. The elution fraction was dialyzed against a buffer containing 30 mM Na_2HPO_4 (pH 6.7), NaCl (150 mM), and DTT (5 mM), and the His-GB1 solubility tag was removed using TEV protease. The digested NFAT and His-GB1 were separated and further purified using size exclusion chromatography (Superdex 75 10/300 GL; GE Healthcare Life Sciences).

Expression and Purification of 14-3-3. Human 14-3-3 (τ isoform) was expressed in *E. coli* as an N-terminally His-tagged protein with a TEV cleavage site using the pNIC28 plasmid (Addgene ID 38931). The τ isoform of the protein 14-3-3 was used in all of the studies described in this manuscript and is referred to as 14-3-3. Cells were grown at 37 °C to an OD_{600} of 0.8. Then the temperature was dropped to 20 °C, and the expression was induced by 1 mM IPTG. Cells were grown for additional 15 h at 20 °C before harvesting. The harvested cells were resuspended in 40 mL of 50 mM Tris·HCl (pH 8.0), 350 mM NaCl, 10 mM imidazole, and 5 mM β-ME. The suspended cells were then disrupted by sonication, and the insoluble fraction was removed by centrifugation for 40 min at 32,914 × *g* rpm. The protein was initially purified by gravity-flow affinity chromatography using 5 mL (10 mL of a 50% slurry) of Ni-NTA resin (Qiagen). After washing the resin with 40 mL of 50 mM Tris·HCl (pH 8.0), 350 mM NaCl, 40 mM imidazole, and 5 mM β-ME, the protein was eluted in an identical buffer containing 350 mM imidazole. The eluted protein was concentrated to a 5-mL volume and further purified using size exclusion chromatography in kinase reaction buffer [50 mM Mes (pH 6.7), 140 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, and 5 mM DTT].

NMR Spectroscopy. All 3D spectra were collected at 288 K using ∼1 mM protein in 30 mM Na_2HPO_4 (pH 6.7), 150 mM NaCl, and 5 mM DTT dissolved in 100% D_2O. For backbone resonance assignment, ^{15}N-detected NMR spectra (hCACON, hCBCACON, hCcaCON-TOCSY) were acquired on a Bruker Avance I 500 spectrometer equipped with a triple-resonance cryogenic probe (TXO) in which the carbon and nitrogen channels are both on the inner coil and detected with cryogenically cooled preamplifiers. The hcaCOCAN, hCACON (2D planes for ^{13}C-detection comparison), HBHACBCAN, and HBHACBCAcoN experiments, and the ^{13}C-detected hacaCON (2D plane) were recorded on a Bruker Avance III 800 MHz spectrometer with a TXO-style cryogenically cooled probe. The spectral widths for the $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\alpha}/^{13}C^{\beta}$, C-aliphatic, and $^{1}H^{\alpha}/^{1}H^{\beta}$ were set to 20, 34, 68, 68, and 5 ppm, respectively. The hCcaCON-TOCSY experiment employed the flip-flop spectroscopy (73) mixing sequence with the mixing time set to 18 ms. All 3D spectra were acquired using NUS with 12% sampling of the Nyquist grid using Poisson Gap Sampling (74). The Nyquist grid for the hCACON and hcaCOCAN experiments consisted of 48/64 complex points in the $^{13}C^{\alpha}/^{13}C^{\alpha}$-$^{13}C^{\beta}$ dimension and 32/64 complex points in the $^{13}C'$ dimension. For the hCcaCON-TOCSY experiment, the grid consisted of 64 complex points in the C-aliphatic dimension and 32 complex points in the $^{13}C'$ dimension. The NUS spectra were reconstructed and processed with hmsIST (39). Processed spectra were analyzed with CcpNmr-Analysis (version 2.4.1) (75). The average experimental time was 3.5–4 d per 3D experiment (see *SI Appendix*, Table S1 for detailed acquisition parameters).

In Vitro Phosphorylation and 14-3-3 Binding Experiments. The 2D NMR experiments were performed on a Varian (DD2 700; Agilent) spectrometer equipped with a cryogenically cooled probe, and the spectrum was recorded at 288 K. The kinase PKA was purchased from Signalchem. The phosphorylation reaction was performed with a sample containing 0.05 mM ^{15}N-labeled NFAT with the addition of 2 μg of PKA in kinase reaction buffer [50 mM Mes (pH 6.7), 140 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 2 mM ATP, and 5 mM DTT]. First, a control experiment of unphosphorylated NFAT in the same kinase reaction buffer was recorded followed by addition of PKA, and phosphorylation was monitored using 2D ^{1}H-^{15}N–HSQC experiments. In the ^{1}H-detected HNCO experiment recorded for obtaining the CON plane (shown in Fig. 5 and *SI Appendix*, Fig. S20), the concentration of the PKA-phosphorylated NFAT and the control NFAT sample was held at 0.2 mM and 14-3-3 was added in a 1:1 molar ratio. All spectra were processed using nmrPipe and analyzed with CcpNmr-Analysis (version 2.4.1) (75).

ACKNOWLEDGMENTS. We thank Kendra E. Leigh for the insightful discussions regarding the work presented here. This work was supported by NIH Grants GM047467 and AI03758 (to G.W.). S.C. acknowledges National Health and Medical Research Council Australia for the C. J. Martin Fellowship. H.A. acknowledges funding from Claudia Adams Barr Program for Innovative Cancer Research. J.J.Z. was supported by NIH Grant K99 GM115814. A.B. thanks Fonds zur Förderung der Wissenschaftlichen Forschung (Project J3872-B21) for support. Maintenance of some of the instruments used for this research was supported by NIH Grant EB002026. Funding was also provided by Japan Science and Technology Agency, Precursory Research for Embryonic Science and Technology (Grant JPMJPR14L5, to K.T.).

1. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
2. Tompa P (2011) Unstructural biology coming of age. *Curr Opin Struct Biol* 21:419–425.
3. Babu MM, Kriwacki RW, Pappu RV (2012) Structural biology. Versatility from protein disorder. *Science* 337:1460–1461.
4. Calabretta S, Richard S (2015) Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem Sci* 40:662–672.
5. Flock T, Weatheritt RJ, Latysheva NS, Babu MM (2014) Controlling entropy to tune the functions of intrinsically disordered regions. *Curr Opin Struct Biol* 26:62–72.
6. Oldfield CJ, Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* 83:553–584.
7. Uversky VN, et al. (2014) Pathological unfoldomics of uncontrolled chaos: Intrinsically disordered proteins and human diseases. *Chem Rev* 114:6844–6879.
8. Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16:18–29.
9. Tamiola K, Acar B, Mulder FA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 132:18000–18003.
10. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
11. Fuxreiter M, Tompa P (2012) Fuzzy complexes: A more stochastic view of protein function. *Adv Exp Med Biol* 725:1–14.
12. Davis RJ (1993) The mitogen-activated protein kinase signal transduction pathway. *J Biol Chem* 268:14553–14556.
13. Li H, Rao A, Hogan PG (2011) Interaction of calcineurin with substrates and targeting proteins. *Trends Cell Biol* 21:91–103.
14. Grathwohl C, Wüthrich K (1976) Nmr studies of the molecular conformations in the linear oligopeptides H-(L-Ala)$_n$-L-Pro-OH. *Biopolymers* 15:2043–2057.
15. Grathwohl C, Wüthrich K (1976) The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides. *Biopolymers* 15:2025–2041.
16. Ivanov D, Stone JR, Maki JL, Collins T, Wagner G (2005) Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. *Mol Cell* 17:137–143.
17. Dziekański P, Grudziąż K, Jarvoll P, Koźmiński W, Zawadzka-Kazimierczuk A (2015) ^{13}C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing. *J Biomol NMR* 62:179–190.
18. Bermel W, et al. (2012) Speeding up sequence specific assignment of IDPs. *J Biomol NMR* 53:293–301.
19. Bermel W, et al. (2013) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR* 57:353–361.
20. Nováček J, et al. (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR* 50:1–11.
21. Bermel W, et al. (2006) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J Am Chem Soc* 128:3918–3919.
22. Takeuchi K, Gal M, Takahashi H, Shimada I, Wagner G (2011) HNCA-TOCSY-CANH experiments with alternate ^{13}C-^{12}C labeling: A set of 3D experiment with unique supra-sequential information for mainchain resonance assignment. *J Biomol NMR* 49:17–26.
23. Yoshimura Y, Kulminskaya NV, Mulder FA (2015) Easy and unambiguous sequential assignments of intrinsically disordered proteins by correlating the backbone ^{15}N or $^{13}C'$ chemical shifts of multiple contiguous residues in highly resolved 3D spectra. *J Biomol NMR* 61:109–121.
24. Gal M, Edmonds KA, Milbradt AG, Takeuchi K, Wagner G (2011) Speeding up direct ^{15}N detection: hCaN 2D NMR experiment. *J Biomol NMR* 51:497–504.

25. Takeuchi K, Heffron G, Sun ZY, Frueh DP, Wagner G (2010) Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *J Biomol NMR* 47:271–282.
26. Chow CW, Davis RJ (2000) Integration of calcium and cyclic AMP signaling pathways by 14-3-3. *Mol Cell Biol* 20:702–712.
27. Takeuchi K, Gal M, Shimada I, Wagner G (2012) Low-[γ] nuclei detection experiments for biomolecular NMR. *Recent Developments in Biomolecular NMR* (Royal Society of Chemistry, London), pp 25–52.
28. Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J Biomol NMR* 3:185–204.
29. Morris GA, Freeman R (1979) Enhancement of nuclear magnetic resonance signals by polarization transfer. *J Am Chem Soc* 101:760–762.
30. Santoro J, King GC (1992) A constant-time 2D overbodenhausen experiment for inverse correlation of isotopically enriched species. *J Magn Reson (1969)* 97:202–207.
31. Vuister GW, Bax A (1992) Resolution enhancement and spectral editing of uniformly ^{13}C-enriched proteins by homonuclear broadband ^{13}C decoupling. *J Magn Reson (1969)* 98:428–435.
32. Kupce E, Freeman R (1995) Adiabatic pulses for wideband inversion and broadband decoupling. *J Magn Reson A* 115:273–276.
33. Jacobs JWM, Van Os JWM, Veeman WS (1983) Broadband heteronuclear decoupling. *J Magn Reson (1969)* 51:56–66.
34. Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broadband decoupling: WALTZ-16. *J Magn Reson (1969)* 52:335–338.
35. Okamura H, et al. (2000) Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity. *Mol Cell* 6:539–550.
36. Hogan PG, Chen L, Nardone J, Rao A (2003) Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev* 17:2205–2232.
37. Macian F (2005) NFAT proteins: Key regulators of T-cell development and function. *Nat Rev Immunol* 5:472–484.
38. Rao A, Luo C, Hogan PG (1997) Transcription factors of the NFAT family: Regulation and function. *Annu Rev Immunol* 15:707–747.
39. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. *J Biomol NMR* 52:315–327.
40. Abildgaard J, Hansen PE, Manalo MN, LiWang A (2009) Deuterium isotope effects on ^{15}N backbone chemical shifts in proteins. *J Biomol NMR* 44:119–126.
41. Henry GD, Weiner JH, Sykes BD (1987) Backbone dynamics of a model membrane protein: Measurement of individual amide hydrogen-exchange rates in detergent-solubilized M13 coat protein using ^{13}C NMR hydrogen/deuterium isotope shifts. *Biochemistry* 26:3626–3634.
42. Sahu D, Bastidas M, Showalter SA (2014) Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Anal Biochem* 449:17–25.
43. Bermel W, Bertini I, Felli IC, Peruzzini R, Pierattelli R (2010) Exclusively heteronuclear NMR experiments to obtain structural and dynamic information on proteins. *ChemPhysChem* 11:689–695.
44. Iakoucheva LM, et al. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32:1037–1049.
45. Theillet FX, et al. (2012) Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR* 54:217–236.
46. Fujii K, et al. (2004) Kinase peptide specificity: Improved determination and relevance to protein phosphorylation. *Proc Natl Acad Sci USA* 101:13744–13749.
47. Shabb JB (2001) Physiological substrates of cAMP-dependent protein kinase. *Chem Rev* 101:2381–2411.
48. Zhu G, et al. (2005) Exceptional disfavor for proline at the P + 1 position among AGC and CAMK kinases establishes reciprocal specificity between them and the proline-directed kinases. *J Biol Chem* 280:10743–10748.
49. Beals CR, Sheridan CM, Turck CW, Gardner P, Crabtree GR (1997) Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* 275:1930–1934.
50. Sheridan CM, Heist EK, Beals CR, Crabtree GR, Gardner P (2002) Protein kinase A negatively modulates the nuclear accumulation of NF-ATc1 by priming for subsequent phosphorylation by glycogen synthase kinase-3. *J Biol Chem* 277: 48664–48676.
51. Aitken A (2006) 14-3-3 proteins: A historic overview. *Semin Cancer Biol* 16:162–172.
52. Mackintosh C (2004) Dynamic interactions between 14-3-3 proteins and phosphoproteins regulate diverse cellular processes. *Biochem J* 381:329–342.
53. Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE (2013) The switches.ELM resource: A compendium of conditional regulatory interaction interfaces. *Sci Signal* 6: rs7.
54. Stevers LM, et al. (2016) Characterization and small-molecule stabilization of the multisite tandem binding between 14-3-3 and the R domain of CFTR. *Proc Natl Acad Sci USA* 113:E1152–E1161.
55. Datta SR, et al. (2000) 14-3-3 proteins and survival kinases cooperate to inactivate BAD by BH3 domain phosphorylation. *Mol Cell* 6:41–51.
56. Molzan M, et al. (2010) Impaired binding of 14-3-3 to C-RAF in Noonan syndrome suggests new approaches in diseases with increased Ras signaling. *Mol Cell Biol* 30: 4698–4711.
57. Brunet A, et al. (2002) 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. *J Cell Biol* 156:817–828.
58. Basu S, Totty NF, Irwin MS, Sudol M, Downward J (2003) Akt phosphorylates the Yes-associated protein, YAP, to induce interaction with 14-3-3 and attenuation of p73-mediated apoptosis. *Mol Cell* 11:11–23.
59. Sato S, et al. (2016) Metabolite regulation of nuclear localization of carbohydrate-response element-binding protein (ChREBP): Role of AMP as an allosteric inhibitor. *J Biol Chem* 291:10515–10527.
60. Pereg Y, et al. (2006) Differential roles of ATM- and Chk2-mediated phosphorylations of Hdmx in response to DNA damage. *Mol Cell Biol* 26:6819–6831.
61. Dinkel H, et al. (2016) ELM 2016–Data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res* 44:D294–D300.
62. Yaffe MB, et al. (1997) The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell* 91:961–971.
63. Yaffe MB (2002) How do 14-3-3 proteins work? Gatekeeper phosphorylation and the molecular anvil hypothesis. *FEBS Lett* 513:53–57.
64. Shikano S, Coblitz B, Wu M, Li M (2006) 14-3-3 proteins: Regulation of endoplasmic reticulum localization and surface expression of membrane proteins. *Trends Cell Biol* 16:370–375.
65. Bonet R, Vakonakis I, Campbell ID (2013) Characterization of 14-3-3-ζ interactions with integrin tails. *J Mol Biol* 425:3060–3072.
66. Fuglsang AT, et al. (1999) Binding of 14-3-3 protein to the plasma membrane H^+-ATPase AHA2 involves the three C-terminal residues Tyr^{946}-Thr-Val and requires phosphorylation of Thr^{947}. *J Biol Chem* 274:36774–36780.
67. Waterman MJ, Stavridi ES, Waterman JL, Halazonetis TD (1998) ATM-dependent activation of p53 involves dephosphorylation and association with 14-3-3 proteins. *Nat Genet* 19:175–178.
68. Csizmok V, Felli IC, Tompa P, Banci L, Bertini I (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. *J Am Chem Soc* 130:16873–16879.
69. O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating ^{1}H chemical shift determination into ^{13}C-direct detected spectroscopy of intrinsically disordered proteins in solution. *J Magn Reson* 200:354–358.
70. Bermel W, Bertini I, Felli IC, Pierattelli R (2009) Speeding up ^{13}C direct detection biomolecular NMR spectroscopy. *J Am Chem Soc* 131:15339–15345.
71. Takeuchi K, Arthanari H, Imai M, Wagner G, Shimada I (2016) Nitrogen-detected TROSY yields comparable sensitivity to proton-detected TROSY for non-deuterated, large proteins under physiological salt conditions. *J Biomol NMR* 64:143–151.
72. Takeuchi K, Arthanari H, Shimada I, Wagner G (2015) Nitrogen detected TROSY at high field yields high resolution and sensitivity for protein NMR. *J Biomol NMR* 63: 323–331.
73. Kadkhodaie M, Rivas O, Tan M, Mohebbi A, Shaka AJ (1991) Broadband homonuclear cross polarization using flip-flop spectroscopy. *J Magn Reson (1969)* 91:437–443.
74. Hyberts SG, Takeuchi K, Wagner G (2010) Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc* 132:2145–2147.
75. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59:687–696.

SI Appendix

^{15}N-Detection Harnesses the Slow Relaxation Property of Nitrogen: Delivering Enhanced Resolution for Intrinsically Disordered Proteins

Sandeep Chhabra[a,b†], Patrick Fischer[a†], Koh Takeuchi[c], Abhinav Dubey[a,b], Joshua J. Ziarek[a,f], Andras* Boeszoermenyi[a,b], *Daniel Mathieu[d], Wolfgang Bermel[d], Norman E. Davey[e] Gerhard Wagner[a,1] and Haribabu Arthanari[a,b,1]

---

[a] Haribabu Arthanari, Gerhard Wagner, Sandeep Chhabra, Patrick Fischer, Joshua Ziarek, Andras Boeszoermenyi
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, USA.

[b] Haribabu Arthanari, Sandeep Chhabra, Andras Boeszoermenyi, Abhinav Dubey
Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA.

[c] Koh Takeuchi
Molecular Profiling Research Center for Drug Discovery (Molprof), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064 (Japan)

[d] Daniel Mathieu, Wolfgang Bermel
Bruker BioSpin GmbH, Silberstreifen 4, 76287, Rheinstetten, Germany.

[e] Norman E Davey
Conway Institute of Biomolecular and Biomedical Sciences, University College Dublin, Ireland.

[f] Joshua Ziarek
Molecular and Cellular Biochemistry Department, Indiana University, Bloomington

[†] These authors contributed equally to this work.
Supporting Information Appendix for this article is given via a link at the end of the document.

[1]To whom correspondence may be addressed.
E-mail: hari@hms.harvard.edu or gerhard_wagner@hms.harvard.edu

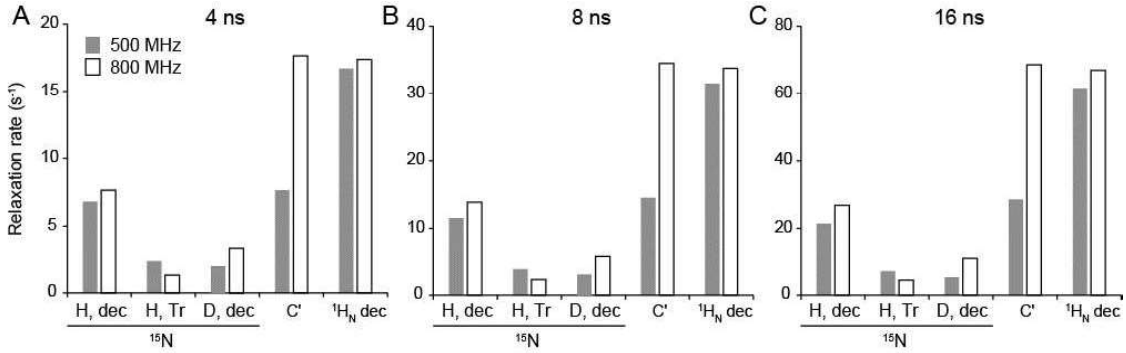


Fig. S1. Transverse relaxation rates of main-chain nuclei of a uniformly ^{15}N-^{13}C-labeled protein with rotational correlation times of (A) 4 ns, (B) 8 ns, and (C) 16 ns. The transverse relaxation rates for proteins at magnetic fields corresponding to proton frequencies of 500 MHz and 800 MHz are shown with filled and open bars, respectively. Dipole-dipole (DD) interactions with directly bonded nuclei and distal nuclei within the distances defined below were used to mimic the typical proton density in a protein. Details of the calculations and the actual values used are included below.

Estimation of transverse relaxation rates

The transverse relaxation rates were calculated based on the standard equations shown below:

$$R_{2I} = \sum_j^n p_{IS_j}^2 \left[(4J(0) + 3(J(\omega_I) + J\left(|\omega_I - \omega_{S_j}|\right) + 6J(\omega_I) + 6J\left(\omega_I + \omega_{S_j}\right)\right] + \delta_I^2[(4J(0) + 3(J(\omega_I)] \quad (1)$$

where

$$p_{IS_j} = -(\frac{\mu_0}{2\sqrt{2}})\frac{\gamma_I \gamma_{S_j} \hbar}{r_{IS_j}^3} \quad (2)$$

$$\delta_I = -(\frac{\mu_0}{3\sqrt{2}})\gamma_I B_0 \Delta\sigma_I \quad (3)$$

$\Delta\sigma_I$ is the difference between the axial and the perpendicular principal components of the axially symmetric chemical-shift tensors. Spectrum densities were estimated using equation 4, shown below:

$$J(\omega) = 2\tau_c / (5 \text{ x } (1+ (\tau_c \omega)^2) \quad (4)$$

where the values for ω are the Larmor frequencies of the ^{1}H or ^{15}N spins, or the sum or difference of the two.

Dipole-dipole interactions with directly bonded nuclei and nuclei falling within defined distances were used to mimic the typical proton density in a protein. The distance cut-off criteria used were < 4.5 Å for ^{1}H-^{1}H pairs, < 3.3 Å for ^{1}H-^{13}C pairs, and < 2.1 Å for ^{1}H-

^{15}N pairs. Distances typical for an α-helical region were used for these estimates. Chemical shift anisotropies were also included in the calculations ($\Delta\sigma^{15}N_H$: -160ppm, $\Delta\sigma^{1}H_N$: 16 ppm, $\Delta\sigma^{13}C\alpha$: 2.2 ppm, $\Delta\sigma^{13}C'$: 170 ppm).

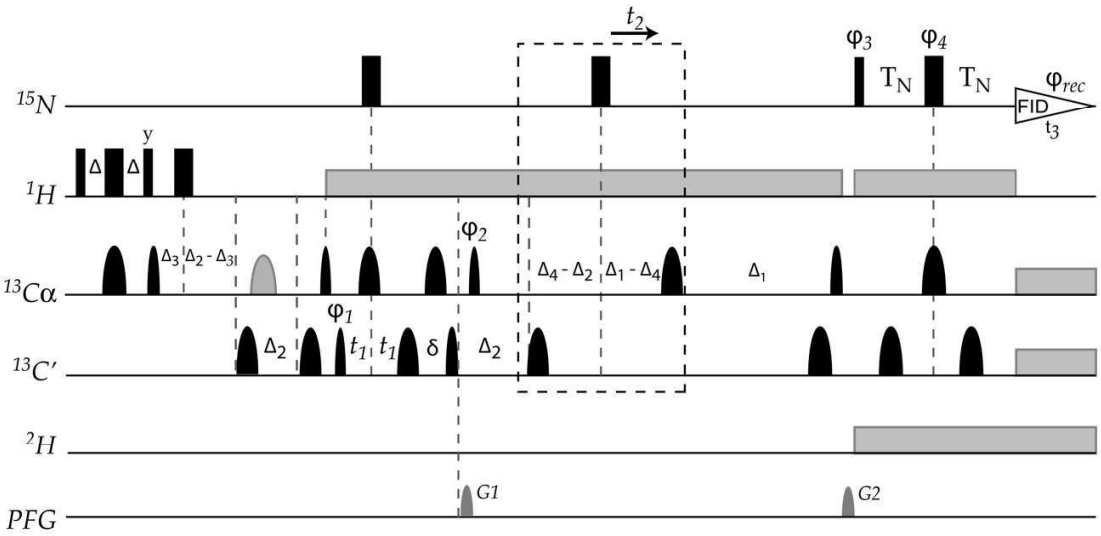


Fig. S2. hcaCOCAN pulse sequence. Narrow and wide, rectangular, black bars indicate non-selective π/2 and π pulses, respectively. Narrow and wide, black, semi-elliptical shapes represent π/2 and π Gaussian cascade pulses selective for $C_{\alpha/\beta}$ or C' carbon nuclei frequencies (Q5 / 320 μs and Q3 / 256 μs, respectively, for a 500 MHz spectrometer, and Q5 / 240 μs and Q3 / 192 μs, respectively, for an 800 MHz spectrometer). The wide, gray, semi-elliptical shape represents a π Gaussian cascade pulse selective for the C_α carbon nuclei frequency (Q3 / 750 μs). All pulses are applied along the x-axis unless otherwise indicated. The delays are $\Delta = 1/4J_{H\alpha C\alpha} = 1.7$ ms, $\Delta_1 = 1/2J_{C\alpha C\beta} = 14.3$ ms, $\Delta_2 = 1/4J_{C_\alpha C'} = 4.5$ ms, $\Delta_3 = 1/6J_{H\alpha C\alpha} = 1.1$ ms, $\Delta_4 = 1/4J_{NC'} = 12.4$ ms, and $T_N = 1/4J_{C\alpha N} = 12.4$ ms. The phase cycle employed is φ_1 = (x, -x), φ_2 = (x, x, x, x, x, x, x, x, -x, -x, -x, -x, -x, -x, -x, -x), φ_3 = (x, x, -x, -x), φ_4 = (x, x, x, x, -x, -x, -x, -x), and φ_{rec} = (x, -x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, x, -x). Phase sensitive spectra in the indirect $C_{aliphatic}$ and C′ dimensions (t1 and t2) are obtained by incrementing the phases φ_1 and φ_2 in a States-TPPI manner. The recycling delay is set to 1 s. Two sine-shaped pulsed field gradients (PFG) are applied along the z-axis for 1.0 ms with maximum intensities of G1 = 10.6 G/cm and G2 = 17.5 G/cm. Proton and deuterium decoupling are achieved by application of a WALTZ-16 sequence (3.3 kHz and 1 kHz, respectively). Adiabatic WURST decoupling was applied for broadband carbon decoupling, centered between C_α and C' frequencies at 115 ppm (2 ms, 30 kHz bandwidth, MLEV-16 super cycle).

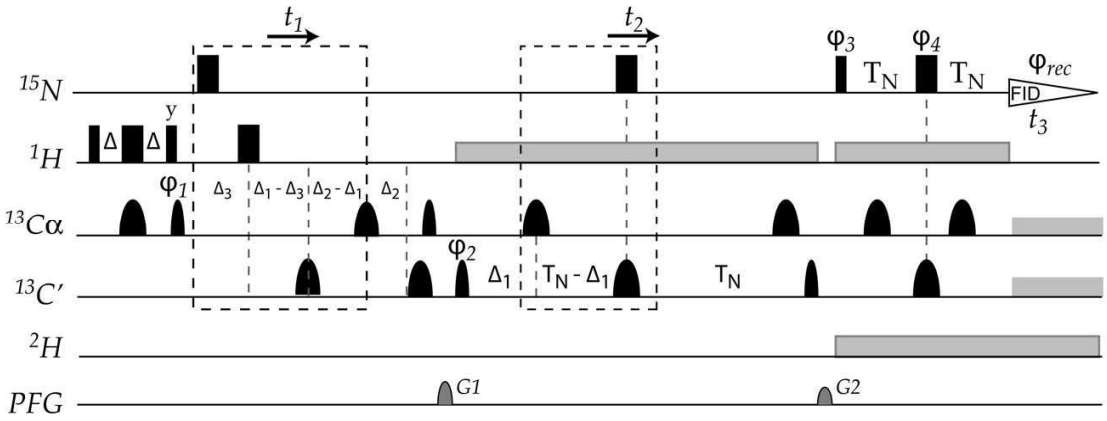


Fig. S3. hCACON pulse sequence. Narrow and wide, rectangular, black bars indicate non-selective π/2 and π pulses, respectively. Narrow and wide, black, semi-elliptical shapes represent π/2 and π Gaussian cascade pulses selective for $C_{aliphatic}$ or C′ carbon nuclei frequencies (Q5 / 320 μs and Q3 / 256 μs, respectively, for a 500 MHz spectrometer, and Q5 / 240 μs and Q3 / 192 μs, respectively, for an 800 MHz spectrometer). All pulses are applied along the x-axis unless otherwise indicated. The delays are $\Delta = 1/4J_{H\alpha C\alpha} = 1.7$ ms, $\Delta_1 = 1/4J_{C\alpha C'} = 4.5$ ms, $\Delta_2 = 1/2J_{C\alpha C\beta} = 14.3$ ms, $\Delta_3 = 1/6J_{H\alpha C\alpha} = 1.1$ ms, and $T_N = 1/4J_{NC'} = 12.4$ ms. The phase cycle employed is φ_1 = (x, -x), φ_2 = (x, x, x, x, x, x, x, x, -x, -x, -x, -x, -x, -x, -x, -x), φ_3 = (x, x, -x, -x), φ_4 = (x, x, x, x, -x, -x, -x, -x), and φ_{rec} = (x, -x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, x, -x). Phase sensitive spectra in the indirect Cα and C′ dimensions (t1 and t2) are obtained by incrementing the phases φ_1 and φ_2 in a States-TPPI manner. The recycling delay is set to 1 s. Two sine-shaped pulsed field gradients (PFG) were applied along the z-axis for 1.0 ms with maximum intensities of G1 = 10.6 G/cm and G2 = 17.5 G/cm. Proton and deuterium decoupling are achieved by application of a WALTZ-16 sequence (3.3 kHz and 1 kHz, respectively). Adiabatic WURST decoupling was applied for broadband carbon decoupling, centered between C_α and C' frequencies at 115 ppm (2 ms, 30 kHz bandwidth, MLEV-16 super cycle).

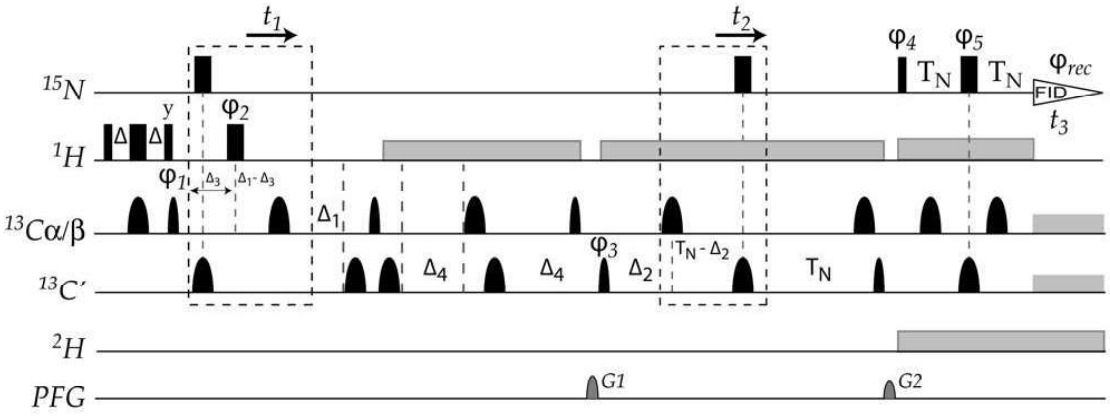


Fig. S4. hCBCACON pulse sequence. Narrow and wide, rectangular, black bars indicate non-selective π/2 and π pulses, respectively. Narrow and wide, black, semi-elliptical shapes represent π/2 and π Gaussian cascade pulses selective for $C_{\alpha/\beta}$ or C′ carbon nuclei frequencies (Q5 / 320 μs and Q3 / 256 μs, respectively, for a 500 MHz spectrometer, and Q5 / 240 μs and Q3 / 192 μs, respectively, for an 800 MHz spectrometer). All pulses are applied along the x-axis unless otherwise indicated. The delays are $\Delta = 1/4J_{H\alpha C\alpha} = 1.7$ ms, $\Delta_1 = 1/8J_{C\alpha C\beta} = 3.6$ ms, $\Delta_2 = 1/4J_{C\alpha C'} = 4.4$ ms, $\Delta_3 = 1/6J_{H\alpha C\alpha} = 1.1$ ms, $\Delta_4 = 1/8J_{C\alpha C\beta} = 1/4J_{C\alpha C'} = 3.6$ ms and $T_N = 1/4J_{NC'} = 12.4$ ms. The phase cycle employed is φ_1 = (x, -x), φ_2 = (x, x, -x, -x), φ_3 = (x, x, x, x, x, x, x, x, -x, -x, -x, -x, -x, -x, -x, -x), φ_4 = (x, x, -x, -x), φ_5 = (x, x, x, x, -x, -x, -x, -x), and φ_{rec} = (x, -x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, x, -x). Phase sensitive spectra in the indirect $C_{\alpha/\beta}$ and C′ dimensions (t1 and t2) are obtained by incrementing the phases φ_1 and φ_3 in a States-TPPI manner. The recycling delay is set to 1 s. Two sine-shaped pulsed field gradients (PFG) are applied along the z-axis for 1.0 ms with maximum intensities of G1 = 10.6 G/cm and G2 = 17.5 G/cm. Proton and deuterium decoupling are achieved by application of a WALTZ-16 sequence (3.3 kHz and 1 kHz, respectively). Adiabatic WURST decoupling is applied for broadband carbon decoupling, centered between C_α and C' frequencies at 115 ppm (2 ms, 30 kHz band width, MLEV-16 super cycle).

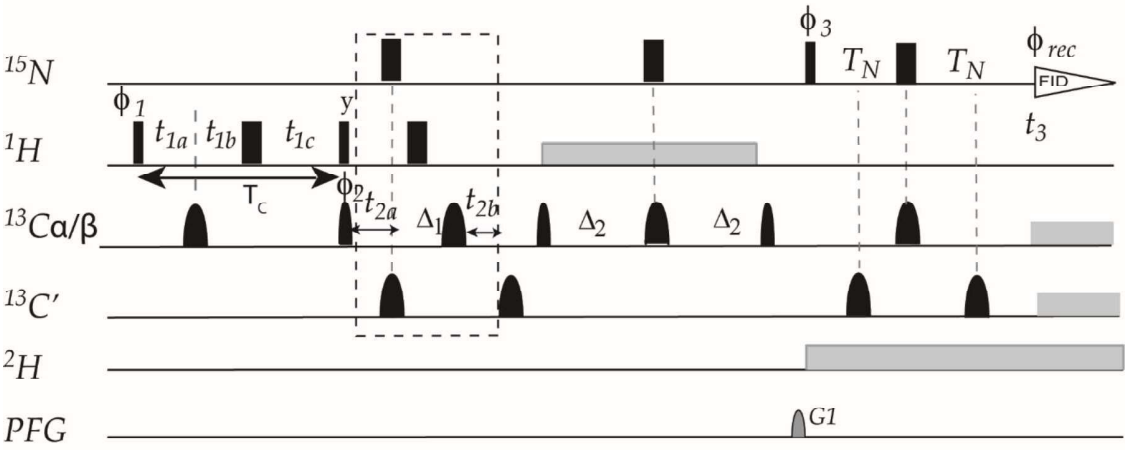


Fig. S5. HBHACBCAN pulse sequence. Narrow and wide, rectangular, black bars indicate non-selective π/2 and π pulses, respectively. Narrow and wide, black, semi-elliptical shapes represent π/2 and π Gaussian cascade pulses selective for $^{13}C_{\alpha/\beta}$ or $^{13}C'$ carbon nuclei frequencies (Q5 / 240 μs and Q3 / 192 μs, respectively, for an 800 MHz spectrometer). All pulses are applied along the x-axis unless otherwise indicated. The chemical shifts of $^{1}H_{\alpha/\beta}$ are evolved in a semi-constant time manner with initial values of $t_{1a} = 1/4J_{H\alpha C\alpha} = 1.7$ ms, and $t_{1b} = t_{1a} + t_{1c}$ (3 μs). The chemical shifts of $^{13}C_{\alpha/\beta}$ are evolved in a constant time manner with an initial value of $t_{2b} = 1/8J_{C\alpha C\beta} = 3.6$ ms. The delays are $\Delta_1 = 1/8J_{C\alpha C\beta} = 3.6$ ms, $\Delta_2 = 1/4J_{C\alpha C'} = 4.5$ ms, and $T_N = 1/4J_{C\alpha N} = 12.4$ ms. The phase cycle employed is φ_1 = (x, x, -x, -x), φ_2 = (x,-x), φ_3 = (x, x, -x, -x), φ_4 = (x, x, x, x, x, x, x, x, -x, -x, -x, -x, -x, -x, -x, -x), φ_5 = (x, x, -x, -x), φ_6 = (x, x, x, x, -x, -x, -x, -x), and φ_{rec} = (x, -x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, x, -x). Phase sensitive spectra in the indirect $^{1}H_{\alpha/\beta}$ and $^{13}C_{\alpha/\beta}$ dimensions (t_1 and t_2) are obtained by incrementing the phases φ_1 and φ_2 in a States-TPPI manner. The recycling delay is set to 1.5 s. Two sine-shaped pulsed field gradients (PFG) are applied along the z-axis for 1.0 ms with maximum intensities of G1 = 10.6 G/cm and G2 = 17.5 G/cm. Proton and deuterium decoupling are achieved by application of a WALTZ-16 sequence (3.3 kHz and 1 kHz, respectively). Adiabatic WURST decoupling was applied for broadband carbon decoupling, centered between C_α and C' frequencies at 115 ppm (2 ms, 30 kHz band width, MLEV-16 super cycle).

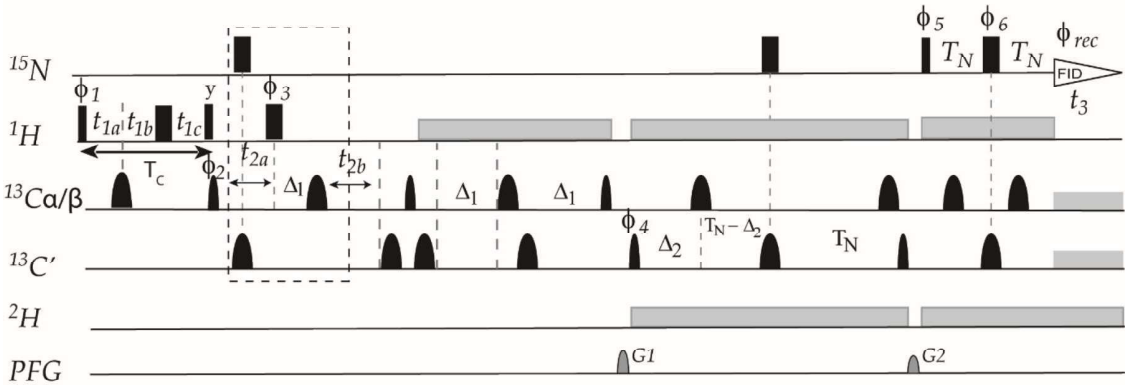


Fig. S6. HBHACBCAcoN pulse sequence. Narrow and wide, rectangular, black bars indicate non-selective π/2 and π pulses, respectively. Narrow and wide, black, semi-elliptical shapes represent π/2 and π Gaussian cascade pulses selective for $^{13}C_{\alpha/\beta}$ or $^{13}C'$ carbon nuclei frequencies (Q5 / 240 μs and Q3 / 192 μs, respectively, for an 800 MHz spectrometer). All pulses are applied along the x-axis unless otherwise indicated. The chemical shifts for $H_{\alpha/\beta}$ are evolved in a semi-constant time manner with initial values of $t_{1a} = 1/4J_{H\alpha C\alpha} = 1.7$ ms, and $t_{1b} = t_{1a} + t_{1c}$ (3 μs). The chemical shifts for $^{13}C_{\alpha/\beta}$ are evolved in a constant time manner with an initial value of $t_{2b} = 1/8J_{C\alpha C\beta} = 3.6$ ms. The delays are $\Delta_1 = 1/8J_{C\alpha C\beta} = 3.6$ ms, $\Delta_2 = 1/4J_{C\alpha C'} = 4.5$ ms, and $T_N = 1/4J_{C\alpha N} = 12.4$ ms. The phase cycle employed is φ_1 = (x, x, -x, -x), φ_2 = (x,-x), φ_3 = (x, x, -x, -x), φ_4 = (x, x, x, x, x, x, x, x, -x, -x, -x, -x, -x, -x, -x, -x), φ_5 = (x, x, -x, -x), φ_6 = (x, x, x, x, -x, -x, -x, -x), and φ_{rec} = (x, -x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, x, -x). Phase sensitive spectra in the indirect $^1H_{\alpha/\beta}$ and $^{13}C_{\alpha/\beta}$ dimensions (t_1 and t_2) are obtained by incrementing the phases φ_1 and φ_2 in a States-TPPI manner. The recycling delay is set to 1.5 s. Two sine-shaped pulsed field gradients (PFG) are applied along the z-axis for 1.0 ms with maximum intensities of G1 = 10.6 G/cm and G2 = 17.5 G/cm. Proton and deuterium decoupling are achieved by application of a WALTZ-16 sequence (3.3 kHz and 1 kHz, respectively). Adiabatic WURST decoupling was applied for broadband carbon decoupling, centered between $^{13}C_\alpha$ and $^{13}C'$ frequencies at 115 ppm (2 ms, 30 kHz band width, MLEV-16 super cycle).

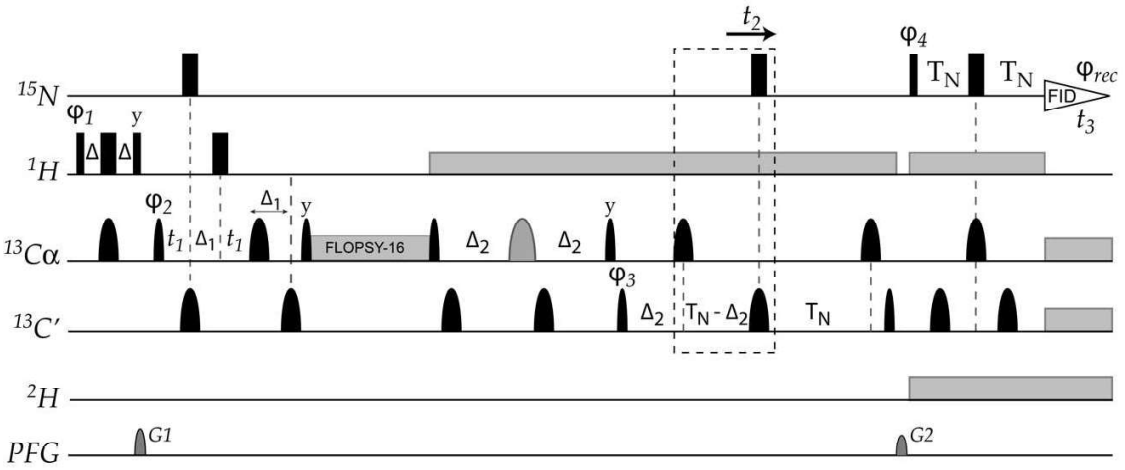


Fig. S7. hCcaCON-TOCSY pulse sequence. Narrow and wide, rectangular, black bars indicate non-selective π/2 and π pulses, respectively. Narrow and wide, black, semi-elliptical shapes represent π/2 and π Gaussian cascades pulses selective for $C_{\alpha/\beta}$ or C′ carbon nuclei frequencies (Q5 / 320 μs and Q3 / 256 μs, respectively, for a 500 MHz spectrometer, and Q5 / 240 μs and Q3 / 192 μs, respectively, for an 800 MHz spectrometer). The wide, gray, semi-elliptical shape represents a π Gaussian cascade pulse selective the C_α carbon nuclei frequency (Q3 / 750 μs on an 800 MHz spectrometer). All pulses are applied along the x-axis unless otherwise indicated. The delays are $\Delta = 1/4J_{H\alpha C\alpha} = 1.7$ ms, $\Delta_1 = 1/6J_{H\alpha C\alpha} = 1.1$ ms, $\Delta_2 = 1/4J_{C\alpha C'} = 4.4$ ms, and $T_N = 1/4J_{NC'} = 12.4$ ms. The phase cycle employed is φ_1 = (x, x, x, x, x, x, x, x, -x, -x, -x, -x, -x, -x, -x, -x), φ_2 = (x, x, -x, -x), φ_3 = (x, x, x, x, -x, -x, -x, -x), φ_4 = (x, -x), and φ_{rec} = (x, -x, -x, x, x, -x, -x, x, -x, x, -x, x, x, -x, -x, x, x, -x). Phase sensitive spectra in the indirect $C_{aliphatic}$ and C′ dimensions (t1 and t2) are obtained by incrementing the phases $\varphi_2$and φ_3 in a States-TPPI manner. The recycling delay is set to 1 s. Two sine-shaped pulsed field gradients (PFG) are applied along the z-axis for 1.0 ms with maximum intensities of G1 = 10.6 G/cm and G2 = 17.5 G/cm. Proton and deuterium decoupling are achieved by application of a WALTZ-16 sequence (3.3 kHz and 1 kHz, respectively). Adiabatic WURST decoupling was applied for broadband carbon decoupling, centered between C_α and C' frequencies at 115 ppm (2 ms, 30 kHz band width, MLEV-16 super cycle). An 18 ms FLOPSY-16 sequence was applied for spin locking (3.2 kHz).

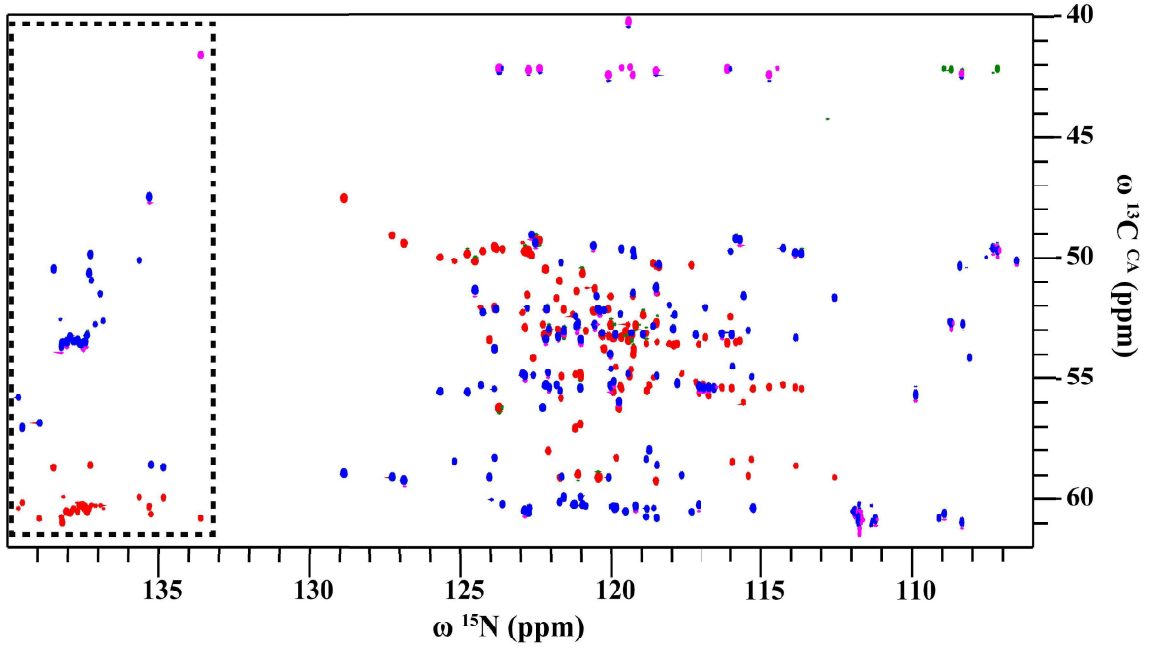


Fig. S8. Overlay of 2D $^{13}C^{\alpha}$-^{15}N planes from a 3D hcaCOCAN experiment (red (positive)/green negative)) and a 3D hCACON (blue (positive)/pink (negative)) experiment. Resonances from prolines residues (the "proline island") are highlighted in the dotted box.

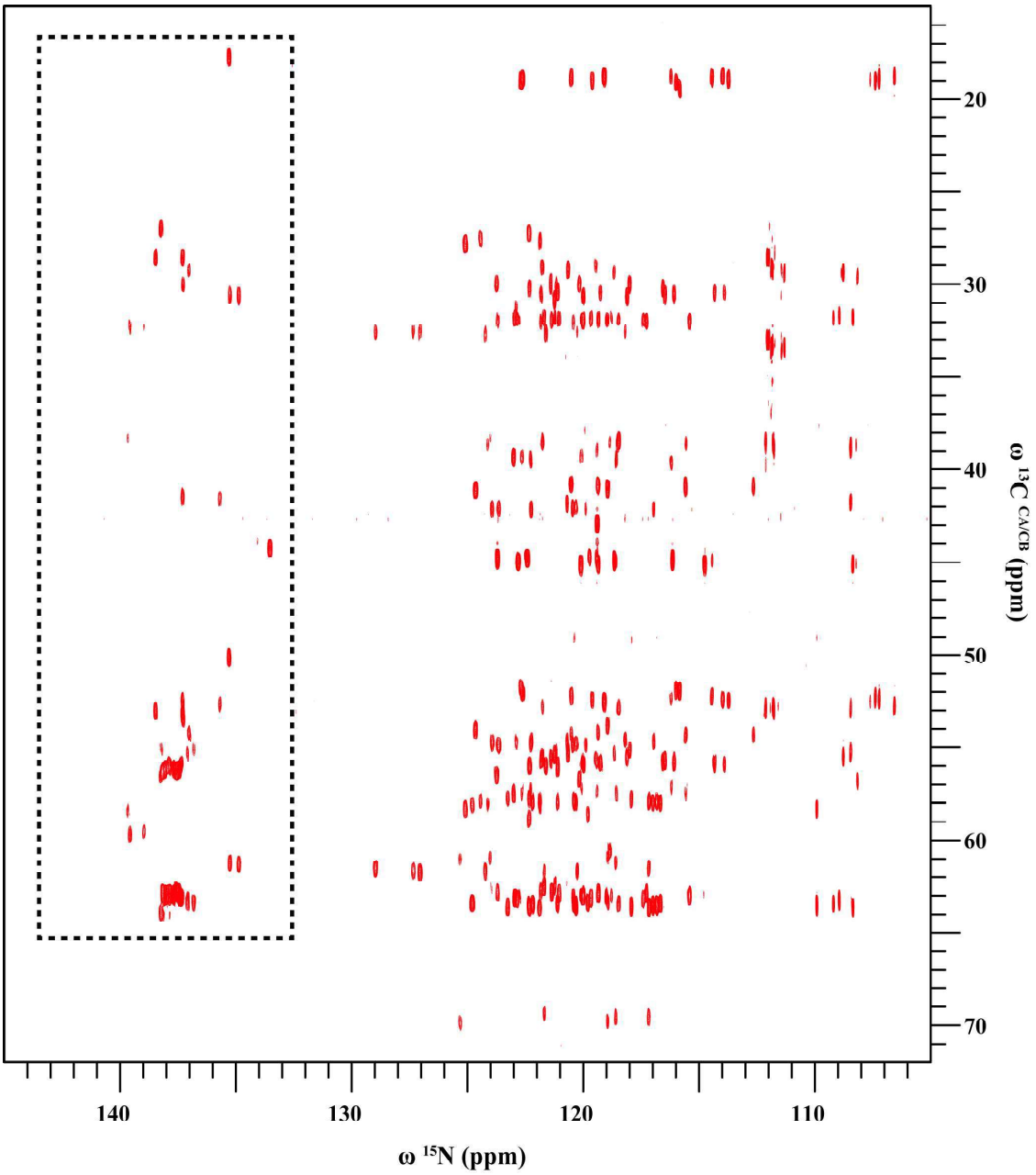


Fig. S9. A 2D $^{13}C^{\alpha\beta}$-^{15}N plane of the 3D hCBCACON spectrum of NFAT (residues 131-294). Resonances from prolines residues (the "proline island") are highlighted in the dotted box.

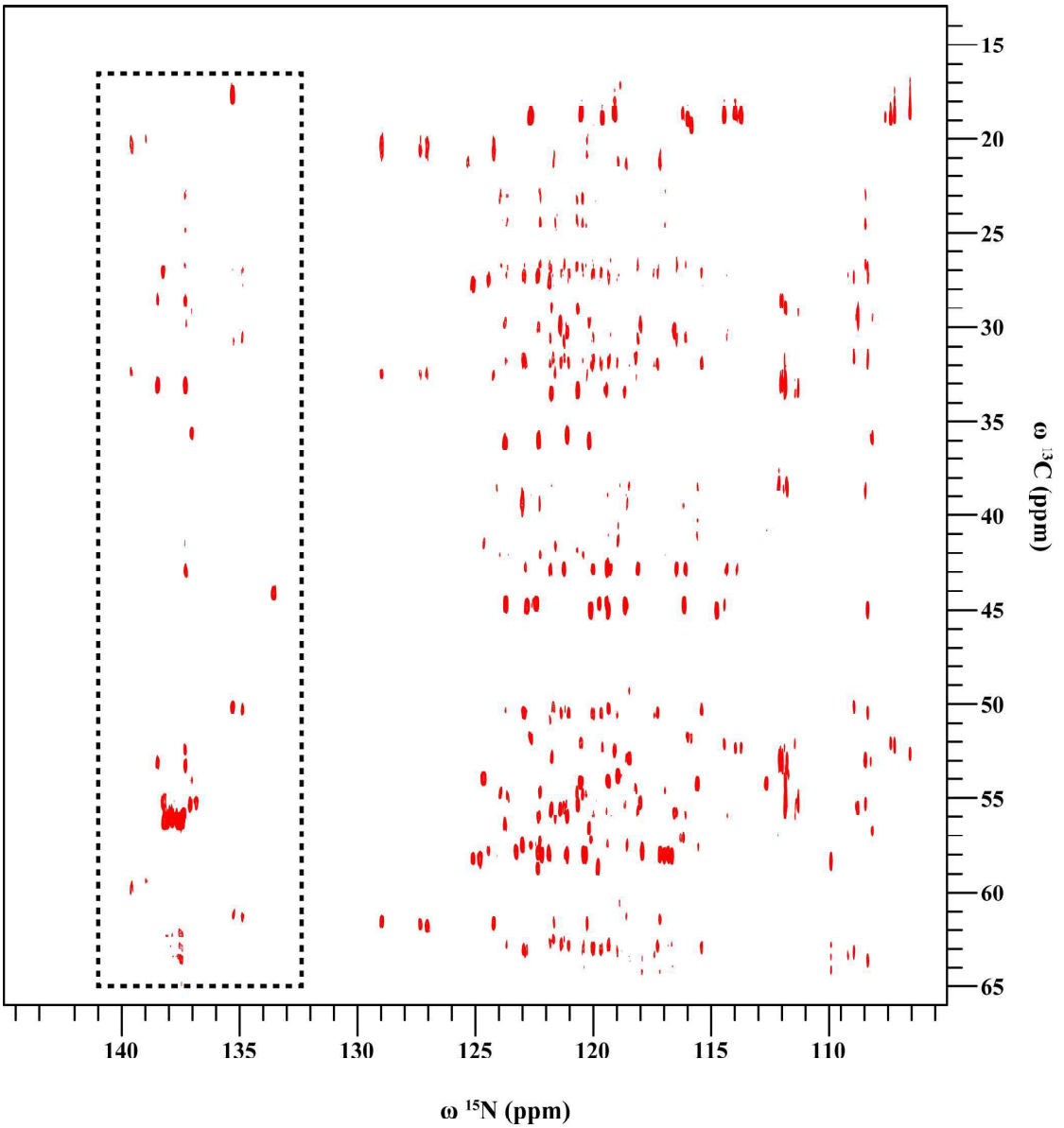


Fig. S10. A 2D $^{13}C^{aliphatic}$-^{15}N plane of the 3D hCcaCON-TOCSY spectrum of NFAT (residues 131-294). Resonances from prolines residues (the "proline island") are highlighted by the dotted box.

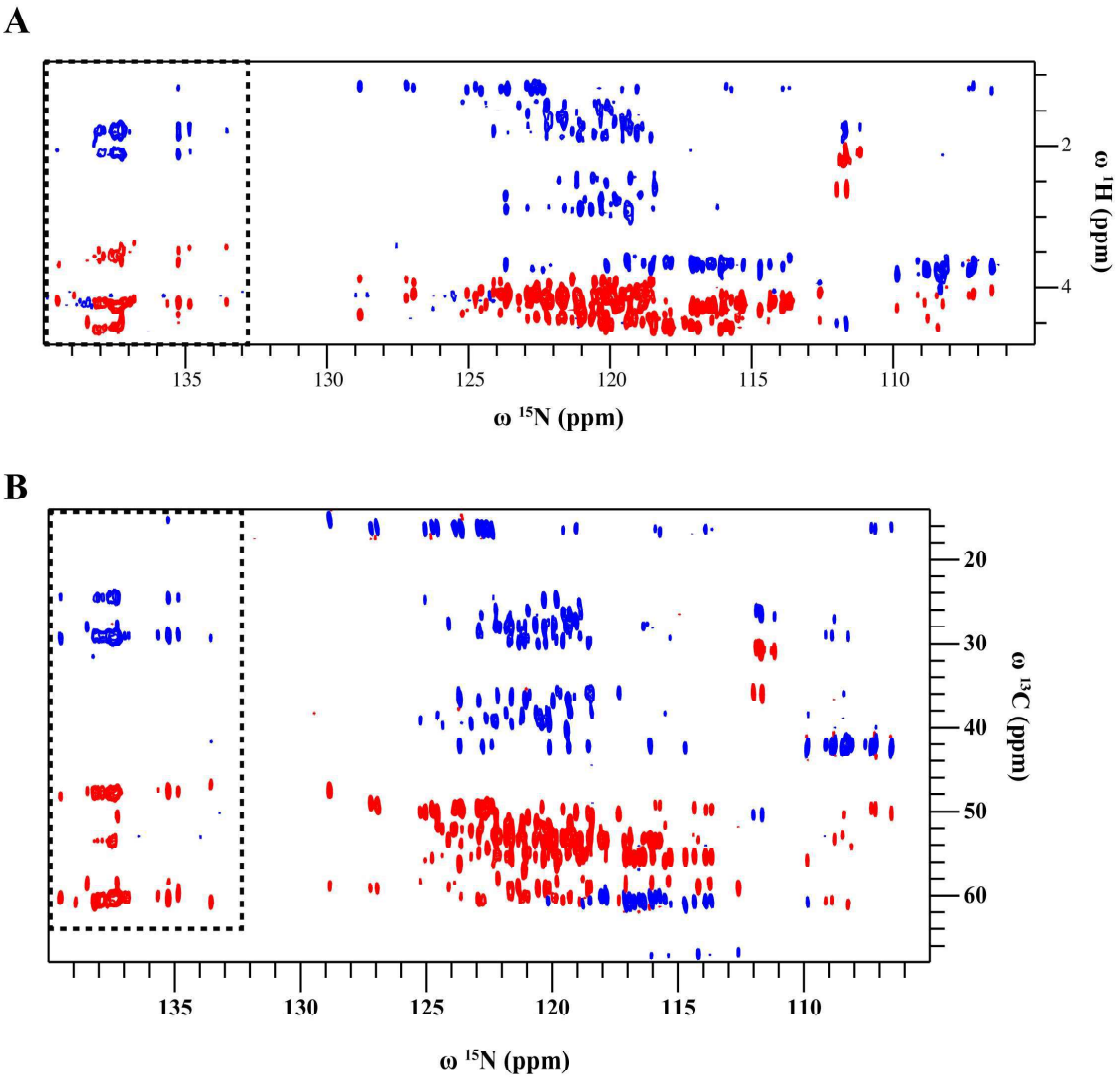


Fig. S11. (A) A 2D $^{1}H^{\alpha\beta}$-^{15}N plane of a 3D HBHACACBN. Red peaks are from $^{1}H^{\alpha}$ and blue peaks are from $^{1}H^{\beta}$ (B) A 2D $^{13}C^{\alpha\beta}$-^{15}N plane of a 3D HBHACBCAN spectrum of NFAT (residues 131-294). Red peaks are from $^{13}C^{\alpha}$ and blue peaks are from $^{13}C^{\beta}$. Resonances from prolines residues (the "proline island") are highlighted in both panels by the dotted boxes.

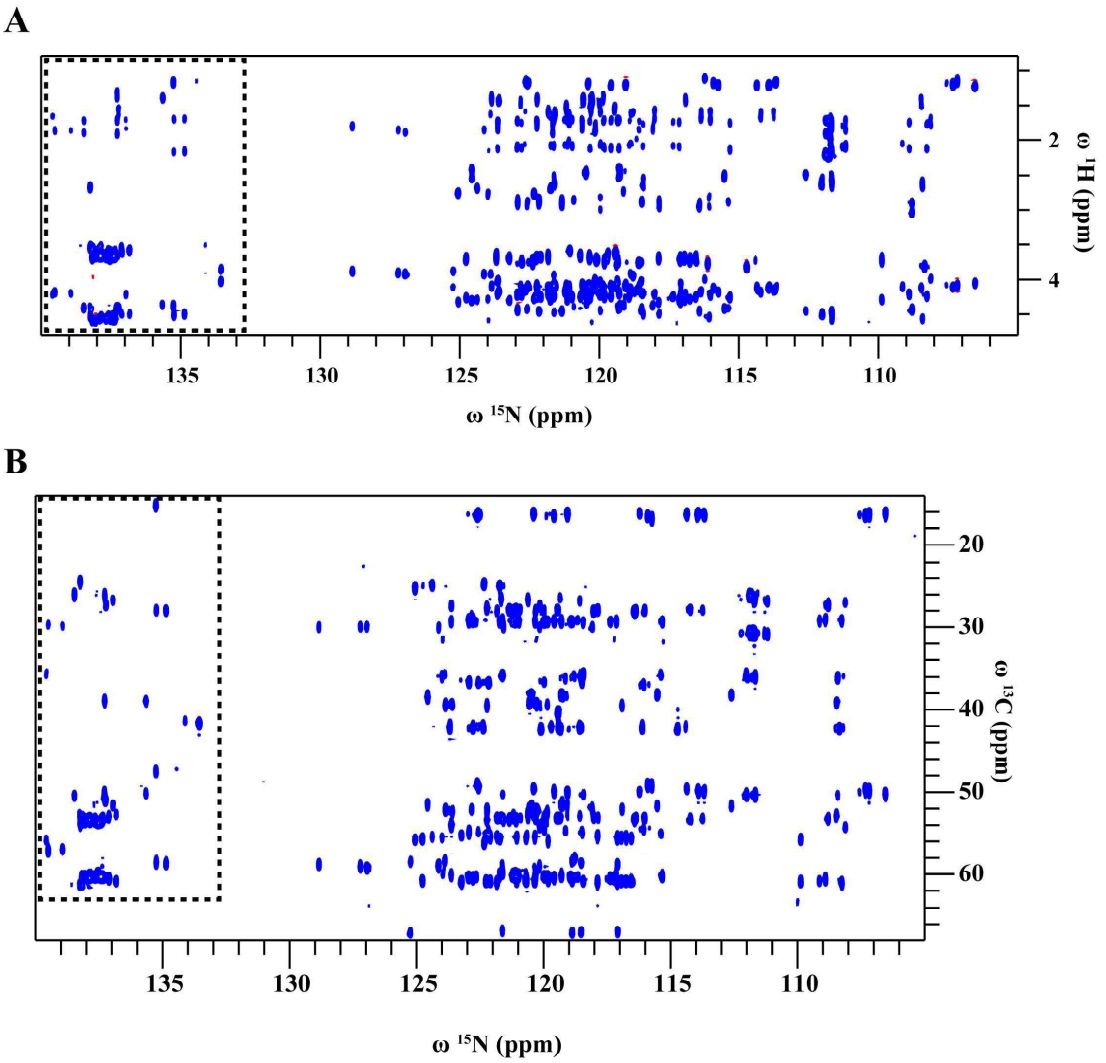


Fig. S12. (A) A 2D $^{1}H^{\alpha\beta}$-^{15}N plane, and (B) a 2D $^{13}C^{\alpha\beta}$-^{15}N plane of the 3D HBHACBCAcoN spectrum of NFAT (residues 131-294). Resonances from prolines residues (the "proline island") are highlighted in both panels by the dotted boxes.

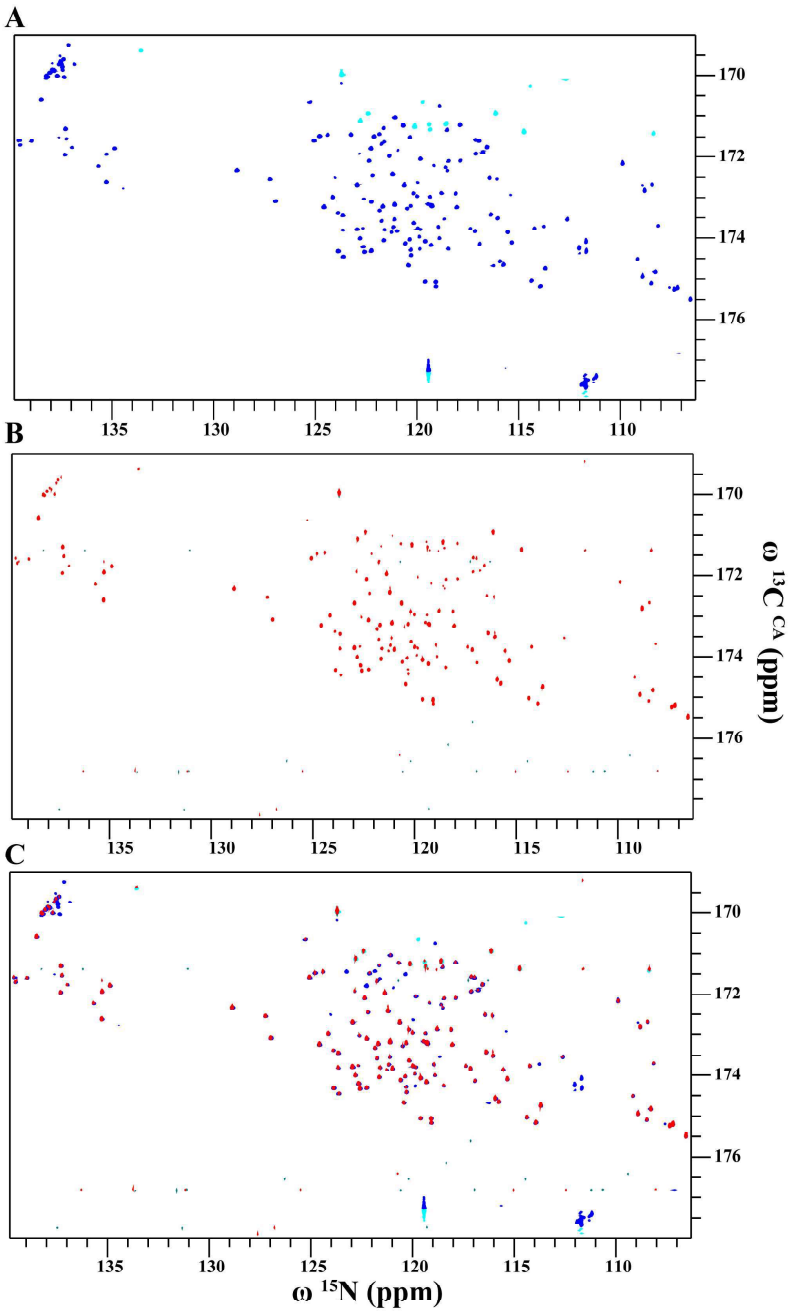


Fig. S13. Time-equivalent 2D ^{13}C'-^{15}N (CON) planes from (A) a ^{15}N-detected hcaCON, and (B) a ^{13}C-detected hcacoNCO. (C) An overlay of the ^{15}N-detected CON (blue (positive)/cyan (negative)) and ^{13}C-detected hcacoNCO (red) experiments. In the ^{15}N-detected experiment, the residues preceding glycines are in negative phase and are shown in cyan.

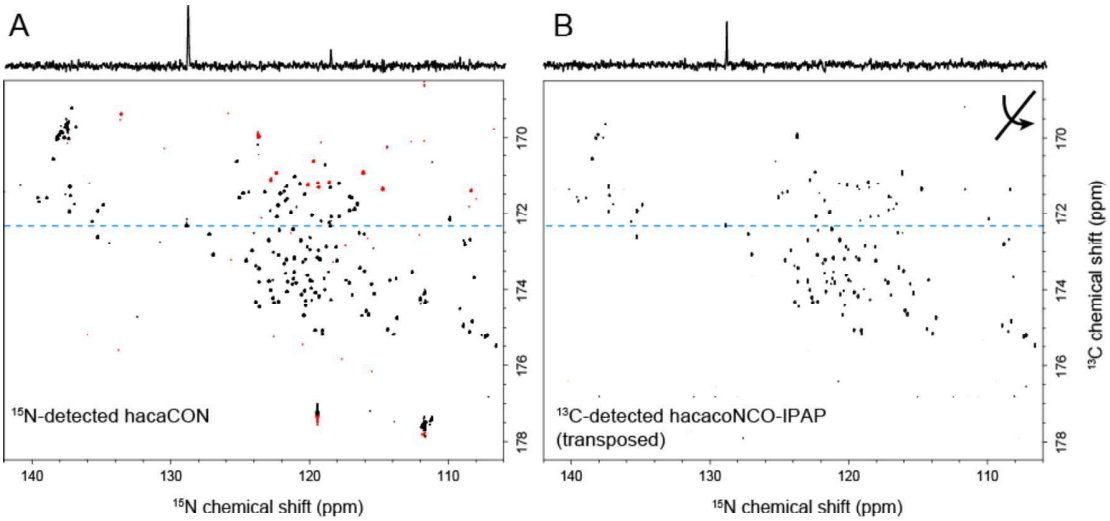


Fig. S14. Comparison of time-equivalent (A) ^{15}N-detected hcaCON, and (B) ^{13}C-detected hcacoNCO-IPAP experiments. The horizontal slice showing the peak height at the indicated frequency (cyan dashed line) is shown above each spectrum. Both spectra are plotted with the vertical threshold adjusted to the same noise level. A detailed description of the acquisition parameters is provided in Table 2.

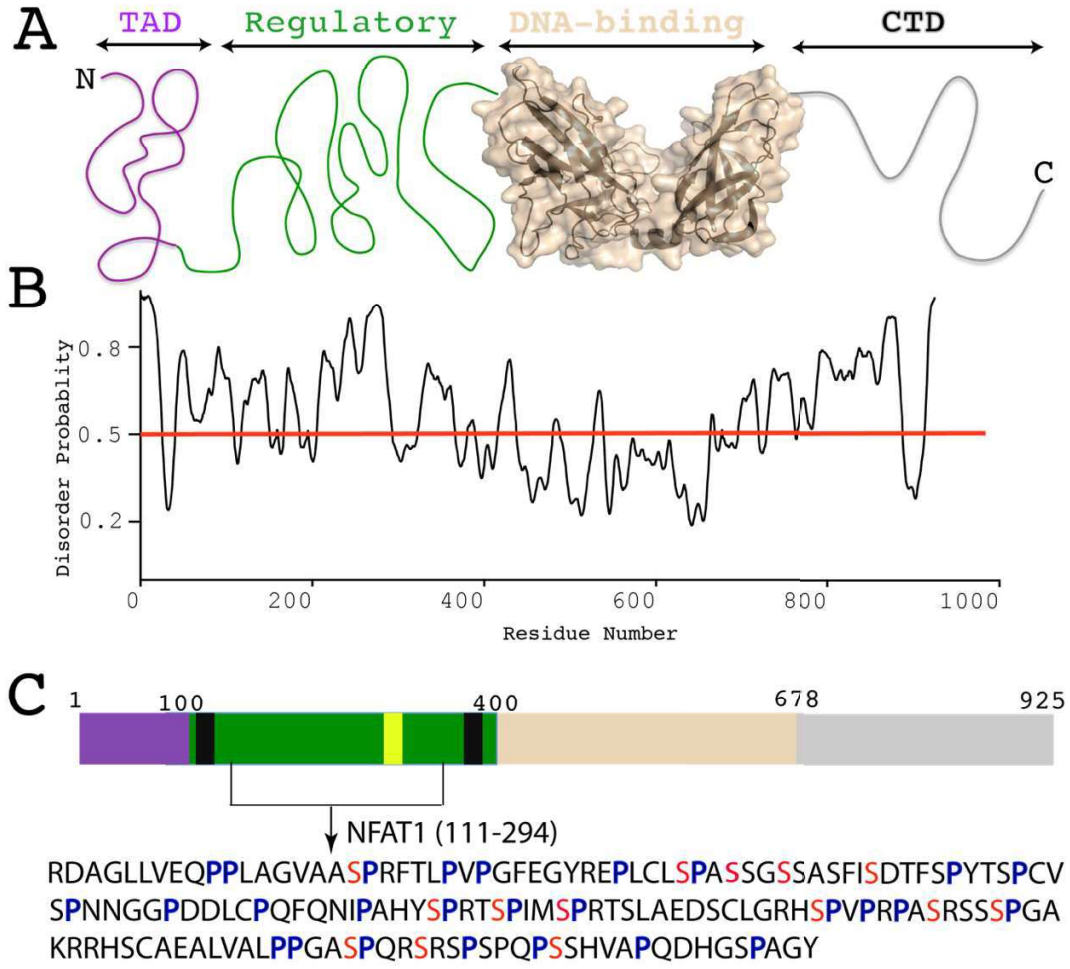


Fig. S15. Domain architecture of NFAT1. (A) A schematic representation of the functional domains of NFAT1. Purple: transactivation domain; green: regulatory domain; beige: DNA binding domain (REL homology region); grey: C-Terminal regulatory domain. (B) Intrinsic disorder prediction of NFAT1 using the Psipred server (http://bioinf.cs.ucl.ac.uk/psipred/). NFAT1 is predicted to be unstructured except for the DNA binding domain. (C) A block representation of the NFAT1 primary sequence annotated with residue numbers and domain boundaries based on functional and structural studies. The two black bars represent the calcineurin binding sites. The yellow bar represents the proposed nuclear localization signal (NLS). The sequence of the NFAT1 regulatory domain (residues 131-294) used in this study is shown below the schematic. The serine residues that have been reported to be phosphorylated are colored red. The proline residues, invisible in traditional ^{1}H-detected backbone experiments, but visible in ^{15}N-detected experiments, are colored blue.

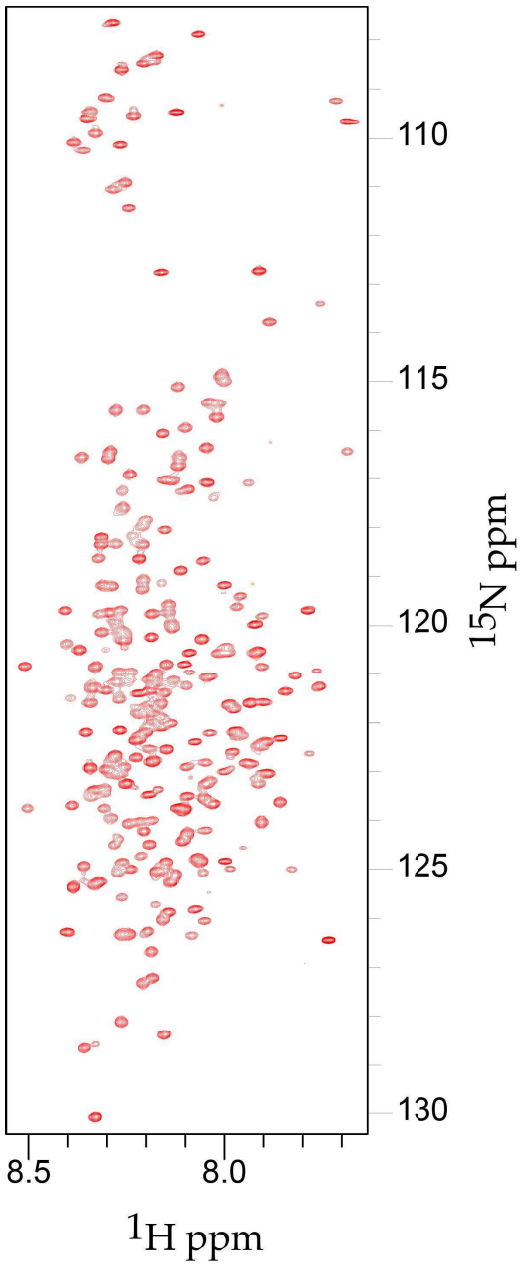


Fig. S16. ^{1}H-^{15}N-HSQC spectrum of NFAT$_{131\text{-}294}$. The narrow chemical shift dispersion in the ^{1}H dimension confirms the disordered nature of this section of NFAT.

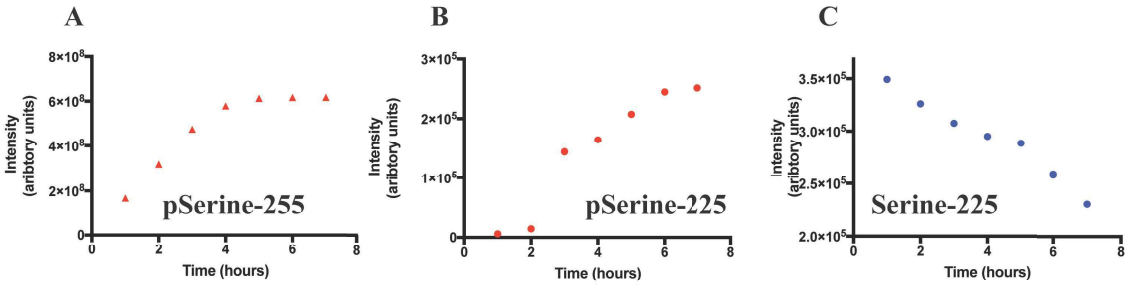


Fig. S17. Following phosphorylation of NFAT by PKA using ^{1}H-^{15}N-HSQC spectroscopy. (A) Build-up in the intensity of the resonance corresponding to phospho-serine-255 as a function of time following addition of PKA. (B) Build-up in the intensity of the resonance corresponding to phospho-serine-225 as a function of time following addition of PKA. (C) Reduction in the intensity of the resonance corresponding to unphosphorylated serine-225 as a function of time after addition of PKA. The resonance corresponding to Ser-255 is weak (near to the level of the noise) in the unphosphorylated spectrum, and therefore the reduction in intensities could not be plotted. The assignments of the phospho-serine resonances were confirmed using ^{15}N-detected experiments on a phosphorylated NFAT sample.

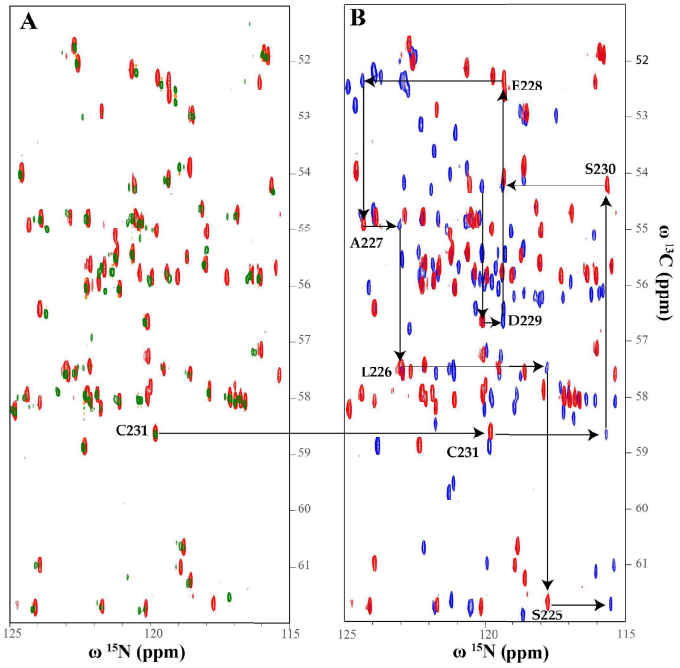


Fig. S18. Confirming the assignment of the phospho-serine residues in NFAT upon phosphorylation by PKA. (A) Overlay of the 2D ^{15}N-detected CAN planes of unphosphorylated NFAT (green) and PKA-phosphorylated NFAT (red). Using the assignment of unphosphorylated NFAT and the expectation that the resonances near the phosphorylation site will experience chemical shift perturbations, we identified a resonance near the phosphorylation site that does not move upon phosphorylation, C231. (B) Overlay of the 2D ^{15}N-detected CAN planes of the hCACON spectrum (red) and the hcaCOCAN (blue) spectrum of phosphorylated NFAT. Here we start from the known (unperturbed) assignment of unphosphorylated NFAT and proceed to assign the resonances that have moved upon phosphorylation. It should be noted that the CAN plane of the hcaCOCAN has both the i and i-1 resonances. These assignments were then transferred to the ^{1}H-^{15}N HSQC using a ^{1}H-detected HNCO experiment. The assignments match those obtained by following the disappearance and appearance of the resonances in a ^{1}H-^{15}N HSQC. NMR experiments were performed on $NFAT_{131-294}$ at a concentration of ~ 0.5 mM after the addition of ~5ug PKA.

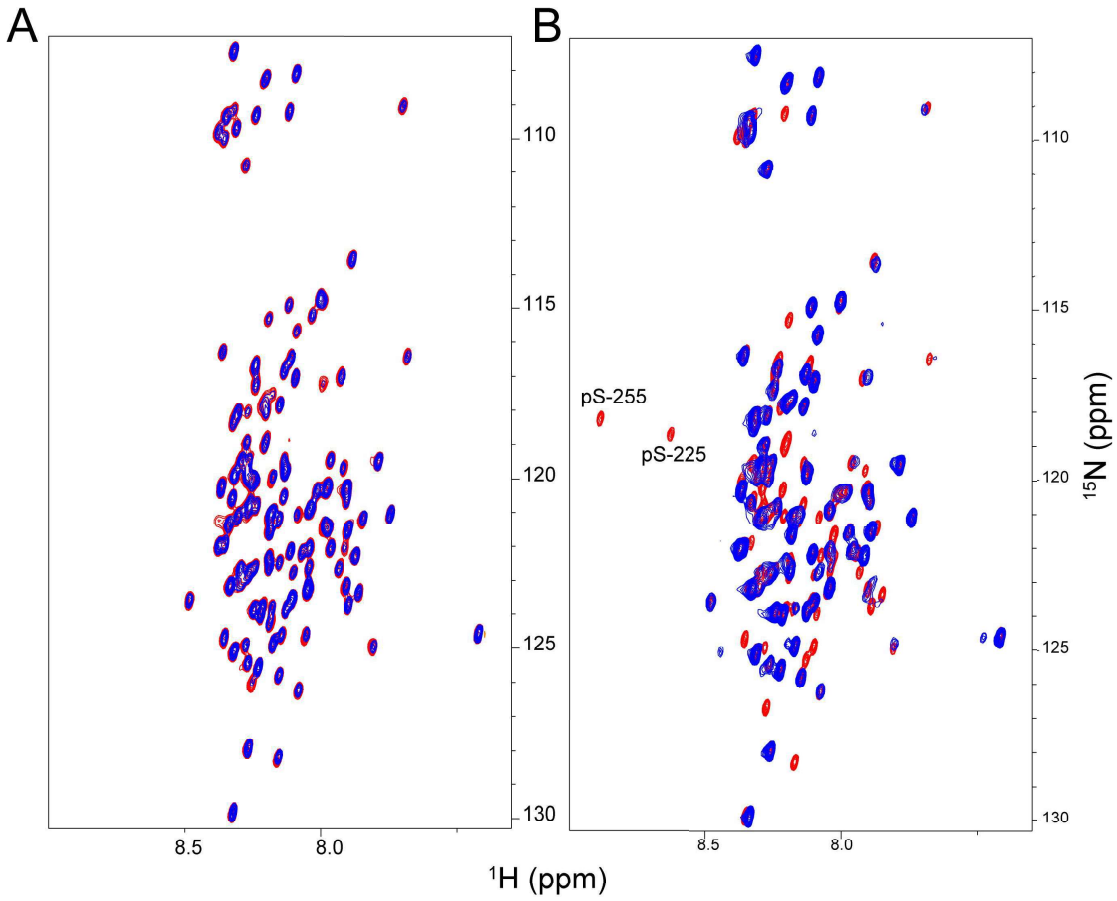


Fig. S19. Interaction of 14-3-3 with unphosphorylated and phosphorylated NFAT. (A) Overlay of the ^{1}H-detected ^{15}N-HSQC spectrum of unphosphorylated NFAT alone (red) and in the presence of 14-3-3 (blue) at a molar ratio of 1:1. (B) Overlay of a ^{1}H-detected ^{15}N-HSQC spectrum of PKA-phosphorylated NFAT alone (red) and in the presence of 14-3-3 (blue) at a molar ratio of 1:1. The two serine residues of NFAT phosphorylated by PKA are labeled.

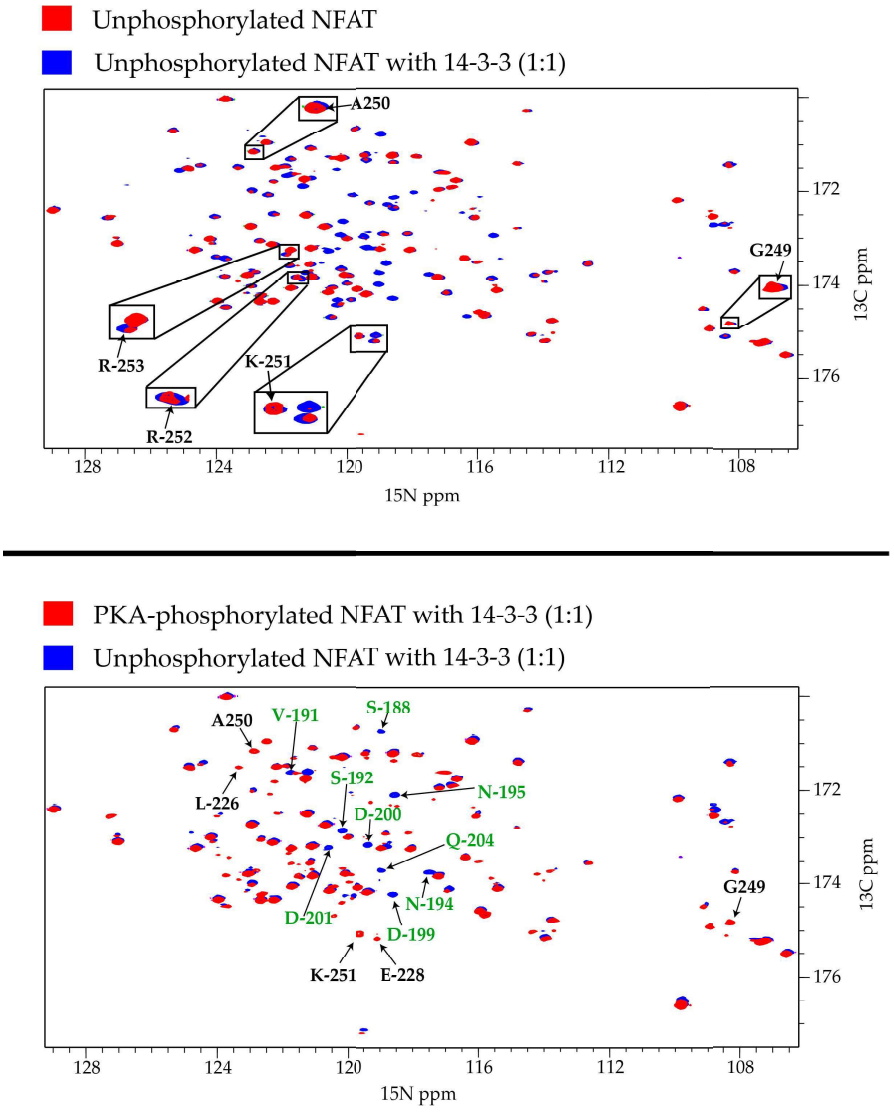


Fig. S20. Interaction of 14-3-3 with unphosphorylated and PKA-phosphorylated NFAT. Top) Overlay of CON plane from a ^{1}H-detected 3D-HNCO of unphosphorylated NFAT (red) and unphosphorylated NFAT with 14-3-3 protein (blue). The resonances corresponding to the residues near the NLS are highlighted (G-249, A-250, K-251, R-252, R-253). These resonances do not experience significant broadening compared to the case of PKA-phosphorylated NFAT binding to 14-3-3 (Fig. 5). Bottom) Overlay of CON plane from a ^{1}H-detected 3D-HNCO of PKA-phosphorylated NFAT with 14-3-3 protein (red) and unphosphorylated NFAT with 14-3-3 protein (blue). Resonances corresponding to the residues near the NLS of NFAT (G-249, A-250, K-251) and resonances near phospho-serine 225 (L226, E228) are highlighted in black. These resonances are only broadened when PKA-phosphorylated NFAT binds to 14-3-3 and not when unphosphorylated NFAT binds to 14-3-3. Interestingly, when 14-3-3 binds to NFAT it shows broadening of resonances near the residues 184-204. These resonances are highlighted with green labels. These data clearly indicate that phosphorylation increase the specificity of 14-3-3 for the NLS of NFAT.

Comparison of the Signal Height in ^{15}N- vs ^{13}C-detected CON Experiments.
The signal height of a multidimensional experiment, S_H, can be described as a product of the initial coherence, C_0, with the product of the remaining coherence ratio in the j_{th} coherence transfer step, P_j, the 1.5th power of gyromagnetic ratio (γ) of the detected nuclei, γ_o, and the relaxation rate of the observing coherences, R_o, as shown in the following equation,

$$S_H \propto C_0 \times \prod P_j \times \gamma_o^{1.5} / R_o \qquad (5)$$

In the ^{15}N-detected hcaCON and the ^{13}C-detected hcacoNCO experiments, the initial coherence and the coherence transfer steps up to the C^{α} to CO transfer are identical. After this transfer, the ^{15}N-detected hcaCON experiment has a SCT CO evolution period, which is concatenated with the CO to N transfer, followed by a $J_{C'N}$ refocusing period where the $^{15}N_D$ coherence is in transverse plane. In constrast, in the ^{13}C-detected hcacoNCO experiment, there is a CO to N transfer and an indirect, real-time (RT) nitrogen evolution period, followed by a $J_{C'N}$ refocusing period where the ^{13}C' coherence is in the transverse plane. When calculating relaxation loss in a SCT evolution, the constant time (CT) part is concatenated with the coherence transfer period, and thus only the RT evolution period after the CT period should be considered with the scaling defined by the ratio of the CT and RT evolution periods. Thus the relative sensitivity of the hcaCON experiment over the counterpart hcacoNCO experiment can be calculated as

$$S_H(\text{hcaCON}) / S_H(\text{hcacoNCO}) = [P(\text{SCT indirect CO evolution}) \times P(^{15}N_D \text{ refocus period}) / [P(\text{RT indirect nitrogen evolution}) \times P(^{13}\text{C' refocus period})] \times [\gamma_O(^{15}N_D)/ \gamma_O(^{13}\text{C'})]^{1.5} / [R(^{15}N_D)/ R(^{13}\text{C'})] \qquad (6)$$

Based on the transverse relaxation rate of $^{15}N_D$ (3.6 s^{-1}) and ^{13}C' (18.8 s^{-1}), assuming a local rotational correlation time of 4.4 ns at 800 MHz in D_2O, $P(^{15}N_D$ refocus period) and $P(^{13}$C' refocus period) can be calculated as

$$P(^{15}N_D \text{ refocus period}) = \sin(\pi \Delta N \times J_{C'N}) \times \exp(- R(^{15}N_D) \times \Delta N) = 0.91 \qquad (7)$$
$$P(^{13}\text{C' refocus period}) = \sin(\pi \Delta C \times J_{C'N}) \times \exp(- R(^{13}\text{C'}) \times \Delta C) = 0.63 \qquad (8)$$

where, ΔN and ΔC are refocusing delays (24.8 ms) and $J_{C'N}$ is set to 16 Hz. The nitrogen transverse relaxation value used here is closely matched with the mean relaxation rate of NFAT at 800 MHz (3.6 ± 0.5 s^{-1}). Thus, without considering relaxation loss during the nitrogen evolution, the ratio of the peak heights is

$$[S_H(\text{hcaCON}) / S_H(\text{hcacoNCO})] = [P(^{15}N_D \text{ refocus period}) / P(^{13}\text{C' refocus period})] \times [\gamma_O(^{15}N_D)/ \gamma_O(^{13}\text{C'})]^{1.5} / [R(^{15}N_D)/ R(^{13}\text{C'})] = 1.93.$$

In addition, the sensitivity of both the ^{15}N-detected and ^{13}C-detected experiments will be affected by relaxation losses during the indirect evolution time period. The ^{15}N-detected hcaCON will be affected by relaxation losses during the SCT CO evolution in the indirect dimension. The relaxation losses during the SCT evolution can be factored into two parts: 1) relaxation loss during the CT portion of the evolution time period, where the relaxation

will stay constant, at a magnitude corresponding to the CT delay (25 ms) and 2) relaxation losses during the RT part of the evolution, where there will be an exponential decay. The sensitivity of the ^{13}C-detected hcacoNCO experiment would decay exponentially during the RT evolution, which encompasses the indirect ^{15}N evolution time period. In the experiment that we carried out, the maximum evolution time period in the indirect carbon (^{15}N-detected) and nitrogen (^{13}C-detected) dimensions were set to 250 ms. In the ^{15}N-detected experiment this corresponds to 25 ms of CT evolution and 225 ms of RT evolution time periods, in the indirect dimension, for the CO nuclei, making the overall signal height 0.306 times smaller compared to a case where there was no relaxation during the indirect evolution time period. The ^{13}C-detected hcacoNCO experiment would have a 250 ms RT evolution for the ^{15}N nuclei, in the indirect dimension, making the overall signal height 0.660 times smaller, compared to a case where there was no relaxation during the indirect evolution time period. Including the relaxation loss from the indirect encoding time periods, the relative intensity of the ^{15}N-detected hcaCON experiment compared to the ^{13}C-detected hcacoNCO experiment, { S_H(hCACON)/S_H(hCAcoNCO) }, would be 0.895 (1.93 × (0.306/0.660)), in an experimental set with the same number of scans for the ^{15}N and ^{13}C detected experiments.

However in the time- and resolution-equivalent experimental setup, the ^{15}N-detected hcaCON experiment had 52 scans whereas the ^{13}C-detected hcacoNCO experiment had 32 scans, due to the larger number of points (to accommodate the larger ^{15}N spectral width) in the indirect dimension for the ^{13}C-detected experiment. Thus including the factor, corresponding to the difference in the number of scans (52 for ^{15}N-detceted vs. 32 for ^{13}C-detected), the relative signal height between a ^{15}N-detected hcaCON experiment compared to the ^{13}C-detected hcacoNCO experiment in a time- and resolution-equivalent experimental setup { S_H(hCACON)/S_H(hCAcoNCO) }, would be 1.45 (0.895 × (52/32)). This advantage stems from the fact the nuclei with the largest dispersion and slowest relaxation in IDPs, i.e. the ^{15}N, is encoded in the direct dimension.

Table S1. Acquisition parameters for the 3D ^{15}N-detected experiments performed on a ^{13}C-^{15}N-labeled NFAT sample dissolved in 99.9% D_2O.

Experiments	HBHACBCAN	HBHACBCAcoN	hCACON	hcaCOCAN	hCBCACON	hCcaCON-TOCSY
Complex Points (Real + Imaginary)	1536 x 128 x 120 ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	1536 x 128 x 92 ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	1536 x 64 x 96 ^{15}N x $^{13}C'$ x $^{13}C_{alpha}$	1280 x 128 x 128 ^{15}N x $^{13}C_{alpha}$ x $^{13}C'$	1536 x 64 x 96 ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$	1536 x 64 x 128 ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$
Magnetic Field Strength (in terms of ^{1}H frequency, MHz)	800	800	500	800	500	500
Scans	128	128	224	128	224	256
Spectral Width	50 x 66 x 5 (ppm) ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	50 x 66 x 5 (ppm) ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	42 x 20 x 34 (ppm) ^{15}N x $^{13}C'$ x $^{13}C_{alpha}$	45 x 25 x 18 (ppm) ^{15}N x $^{13}C_{alpha}$ x $^{13}C'$	42 x 20 x 68 (ppm) ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$	42 x 20 x 68 (ppm) ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$
Spectral Width	4058 x 13280 x 4001 (Hz) 15N x 13C$_{aliph}$ x 1H$_{aliph}$	4058 x 13280 x 4001 (Hz) ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	2126 x 2513 x 4272(Hz) ^{15}N x $^{13}C'$ x $^{13}C_{alpha}$	3650 x 5031 x 3622 (Hz) ^{15}N x $^{13}C_{alpha}$ x $^{13}C'$	2126 x 2513 x 8544 (Hz) ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$	2126 x 2513 x 8544 (Hz) ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$
Acquisition time (ms)	189.2, 4.82, 14.99 ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	189.2, 5.3,38.1 ^{15}N x $^{13}C_{aliph}$ x $^{1}H_{aliph}$	361, 12.7, 11.23 ^{15}N x $^{13}C'$ x $^{13}C_{alpha}$	175, 12.7, 17.6 ^{15}N x $^{13}C_{alpha}$ x $^{13}C'$	361, 12.7, 5.6 ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$	361, 12.7, 7.5 ^{15}N x $^{13}C'$ x $^{13}C_{aliph}$
Experiment time	~ 4 days (12% NUS)	~ 4 days (12% NUS)	~ 3.5 days (12% NUS)	~ 3.5 days (12% NUS)	~ 3.5 days (12% NUS)	~4 days (12% NUS)
Inter-scan delay	1.5 s	1.5 s	1.5 s	1.0 s	1.5 s	1.5 s

Note: The 3D ^{15}N-detected NMR experiments were performed on $NFAT_{131-294}$ at a concentration of ~1 mM in 30 mM sodium phosphate, pH 6.7, 150 mM sodium chloride, 2 mM DTT, and 0.01% NaN_3.

Table S2. Acquisition parameters for the time-equivalent 2D ^{13}C-^{15}N planes from ^{15}N-detected hcaCON and ^{13}C-detected hcacoNCO experiments performed on a ^{13}C-^{15}N-NFAT sample dissolved in 99.9% D_2O. Experiments were recorded on a Bruker Avance III 800 MHz spectrometer with a TXO-style cryogenically cooled probe.

Experiments	^{15}N-hcaCON	^{13}C-hcacoNCO
Scans	52	16 (gets added up to 32 after split and combine)
Spectral width in the direct dimension	3243.945 Hz (40 ppm)	2012.882 (~10 ppm)
Spectral width in the indirect dimension	2012.45 Hz (~10 ppm)	3243.741 (40 ppm)
Number of points in the direct dimension (acquisition time)	1620 (250 ms)	1006 (250 ms)
Number of points the indirect dimension (acquisition time)	1006 (250 ms; 25 ms constant time, 225 ms real time)	1622 (250 ms; real time)
Experimental time	25 hours 23 min	25 hours 10 min

Note: NMR experiments comparing the ^{15}N- and ^{13}C-detected CON planes were performed on $NFAT_{131-294}$ at a concentration of ~1 mM in 30 mM sodium phosphate, pH 6.7, 150 mM sodium chloride, 2 mM DTT, and 0.01% NaN_3.

4.2 Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics

Stefan Koppermann, Zheng Cui, Patrick D. Fischer, Xiachang Wang, Jannine Ludwig, Jon S. Thorson, Steven G. Van Lanen, and Christian Ducho. (2018). Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics. **ChemMedChem**, 13 (8), 779-784.

DOI: https://doi.org/10.1002/cmdc.201700793

URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.201700793

Summary:

This publication provides the first detailed structure-activity relationship (SAR) study of naturally occurring muraymycins with their target protein MraY. A detailed SAR is given by comparison of inhibitory effects from muraymycin subgroups A-D. In conjunction with antibacterial activities and modelling data derived from an X-ray crystal structure, the influence of muraymycin variations was interpreted. Muraymycin families are divided based on the modification of the central 3-hydroxy-L-leucine moiety in the peptide chain. A-type muraymycins are identified by 3-*O*-acylation with ω-functionalized fatty acids terminating with a guanidinium functionality. Muraymycins from the B series are 3-*O*-acylated with unfunctionalized, branched fatty acids. C-type muraymycins lack 3-*O*-acylation, while congeners from the D series contain unmodified 3-hydroxy-L-leucine. Using a fluorescence-based *in vitro* activity assay, inhibitory information was generated for members of each of the muraymycin families. The assays were performed with MraY from different species and using different purification methods. Overexpressed MraY from *Staphylococcus aureus* was used as a mixture of crude *E. coli* membranes from overexpressing cells. Following an adapted protocol from Chung et al., MraY from *Aquifex aeolicus* was purified in detergent micelles and used to compare inhibitory potencies generated from crude membranes with purified protein. To assess this information more effectively, *Clostridium bolteae* MraY was used both as overexpressed protein in crude membranes as well as purified in detergent micelles according to Hakulinen et al. With overexpressed *Staphylococcus aureus* MraY in crude

membranes, all tested muraymycins inhibited the protein in the picomolar range. Due to similar activities of muraymycins of types A, B and C, it was concluded that the fatty acid moiety absent in C series muraymycins does not significantly enhance *in vitro* inhibition of MraY. However, the missing modification of the 3-hydroxy-L-leucine moiety in the D series led to a loss of inhibition, thus suggesting the beneficial role of a β-substituent in this amino acid side chain. It was also concluded that the absence of the 5'-aminoribosyl moiety in a tested synthetic analogue furnished a ca. 6000-fold loss of inhibitory effect. Generally, inhibitory effects of muraymycins against solubilized protein were much smaller, even though the general ranking of inhibitors was retained. It can be concluded that different sources of MraY activity may provide different inhibitory potencies depending on the used homologue and mainly the preparation method used to obtain the enzyme.

Author contribution:

As the author of this thesis, I performed the expression and purification of detergent-solubilized MraY from *Aquifex aeolicus* and *Clostridium bolteae*. I assisted Stefan Koppermann in preparing crude membranes of overexpressed MraY from *Staphylococcus aureus* and *Clostridium bolteae* and helped analyzing and discussing the inhibition data.

Published in final edited form as:
ChemMedChem. 2018 April 23; 13(8): 779–784. doi:10.1002/cmdc.201700793.

Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics

Stefan Koppermann[a], **Zheng Cui**[b], **Patrick D. Fischer**[a], **Xiachang Wang**[c],[d], **Jannine Ludwig**[a], **Jon S. Thorson**[b],[d], **Steven G. Van Lanen**[b], and **Christian Ducho**[a]

[a]Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Campus C2 3, 66123 Saarbrücken (Germany)

[b]Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 789 S. Limestone Street, Lexington, KY 40536 (USA)

[c]Jiangsu Key Laboratory for Functional Substance of Chinese Medicine, School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023 (P. R. China)

[d]Center for Pharmaceutical Research and Innovation, College of Pharmacy, University of Kentucky, 789 S. Limestone Street, Lexington, KY 40536 (USA)

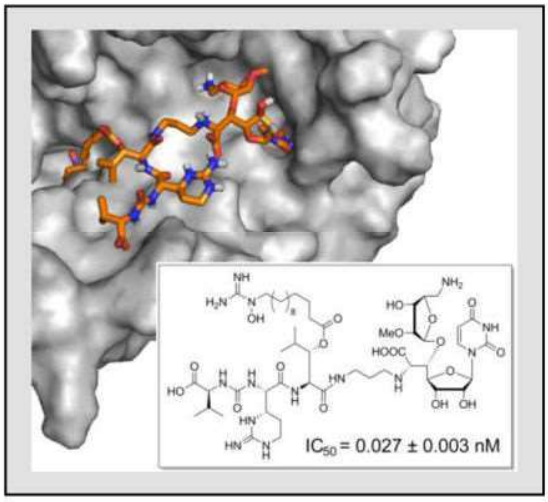
Abstract

Muraymycins are a subclass of antimicrobially active uridine-derived natural products. Biological data on several muraymycin analogues have already been reported, including some inhibitory *in vitro* activities towards their target protein, the bacterial membrane enzyme MraY. However, a structure-activity relationship (SAR) study on naturally occurring muraymycins based on such *in vitro* data has been missing so far. In this work, we report a detailed SAR investigation on representatives of the four muraymycin subgroups A–D using a fluorescence-based *in vitro* MraY assay. For some muraymycins, inhibition of MraY with IC_{50} values in the low pM range was observed. These inhibitory potencies were compared with antibacterial activities and were correlated to modelling data derived from a previously reported X-ray crystal structure of MraY in complex with a muraymycin inhibitor. Overall, these results will pave the way for the development of muraymycin analogues with optimized properties as antibacterial drug candidates.

Natural products revisited

Naturally occurring muraymycin nucleoside antibiotics were studied for inhibitory activities towards their target protein, the bacterial membrane enzyme MraY. An *in vitro* MraY assay furnished SAR data on several representative muraymycins. These inhibitory potencies were correlated to antibacterial activities and to modelling data, thus providing detailed insights into the target interaction of muraymycins.

Correspondence to: Christian Ducho.



Keywords

antibiotics; natural products; nucleosides; activity assays; structure-activity relationship

Resistance against clinically used antibiotics continues to emerge and represents a significant challenge for human healthcare.[1] Therefore, research on novel antimicrobial agents with new or as of yet unexploited modes of action is highly relevant. The bacterial membrane protein MraY (translocase I) represents such a potential new target.[2] MraY is an enzyme involved in the intracellular membrane-associated stages of cell wall (peptidoglycan) biosynthesis. It catalyzes the transformation of the cytosolic precursor UDP-Mur*N*Ac-pentapeptide ('Park's nucleotide') **1** with the membrane anchor undecaprenyl phosphate, yielding the membrane-bound biosynthetic intermediate lipid I **2** (Scheme 1).[3,4]

Methods for the overexpression of MraY, including cell-free approaches, have been established.[5] Following a previously described topology model,[6a] the first ligand-free X-ray crystal structure of MraY (from the extremophile *Aquifex aeolicus*) confirmed the enzyme to be an integral membrane protein with ten transmembrane helices and five cytoplasmic loops forming the active site.[6b] Mechanistic studies on MraY including mutagenesis have not been fully conclusive.[7]

Naturally occurring uridine derivatives ('nucleoside antibiotics', e.g., muraymycins, caprazamycins, liposidomycins, capuramycins and mureidomycins) and their analogues are potent inhibitors of MraY and therefore attractive candidate structures for antimicrobial drug development.[8] Muraymycins (e.g., compounds **3–8**, Table 1) were discovered as a collection of 19 structurally related, secondary metabolites from *Streptomyces* sp. in 2002.[9] The muraymycin scaffold consists of a (5′*S*,6′*S*)-glycyluridine (GlyU) core structure, which is connected to a peptide moiety by an alkyl linker. In most muraymycins, this core is 5′-*O*-aminoribosylated to generate an unusual disaccharide. Muraymycins have been categorized into four subgroups A–D, based on the chemical modification of the central L-leucine, with groups A–C containing (3*S*)-3-hydroxy-L-leucine at this position. A-type

muraymycins (e.g., A1 **3**) are 3-*O*-acylated with ω-functionalized fatty acids terminating with a guanidinium functionality. B-series congeners (e.g., B2 **4** and B6 **5**) are 3-*O*-acylated with unfunctionalized, branched fatty acids. C-type muraymycins (e.g., C1 **6**) completely lack the 3-*O*-acyl unit. Finally, members of the D-series (e.g., D1 **7** and D2 **8**) contain an unmodified L-leucine residue at the central position (Table 1). First insights into the biosynthesis of these antibacterial natural products have already been obtained and are summarized elsewhere.[8c]

Several semi-synthetic and synthetic analogues of muraymycins as well as their biological activities have already been reported, thus furnishing some structure-activity relationship (SAR) data mostly for simplified muraymycin variants.[10] Further insights into the biological action of muraymycins were recently obtained from the X-ray co-crystal structure of MraY from *Aquifex aeolicus* in complex with the natural product muraymycin D2.[11] In many of the aforementioned SAR investigations on muraymycin analogues, biological potencies were mainly evaluated on the basis of antibacterial activities (i.e., minimal inhibitory concentrations (MIC)). However, antibacterial activity is the overall result of a compound's influx into the bacterial cell, its target interaction, its intracellular stability and potential efflux phenomena.

A significantly more informative approach to study the SAR of MraY-inhibiting antibiotics would be to include an investigation into their inhibitory potencies towards MraY using an *in vitro* enzyme assay. This way, MIC values would provide additional information on other factors influencing antibacterial activity besides target interaction. Such an approach has only been pursued for some of the previously reported muraymycin analogues.[10] Remarkably, even information on the MraY-inhibiting properties of the parent natural products is scarce. The originally isolated, naturally occurring muraymycins were solely assessed by their inhibition of the formation of lipid II (i.e., the biosynthetic intermediate following lipid I **2**) and peptidoglycan in coupled assays, thus only indirectly supporting MraY inhibition.[9a,b] These assays revealed inhibition of lipid II and peptidoglycan formation at concentrations as low as 0.27 μg/mL, being equivalent to the low nM range. However, quantitative information on direct MraY inhibition remains unknown.

We therefore aimed to isolate and test the MraY inhibition of the naturally produced muraymycins **3–8** as selected representatives of the four subclasses A–D (see Table 1). Using modified fermentation conditions based partly on previously reported methods,[9a,b] *Streptomyces* sp. (NRRL30473 for isolation of muraymycins A1 **3**, B2 **4** and B6 **5**; NRRL30475 for isolation of muraymycins D1 **7** and D2 **8**; NRRL30477 for isolation of muraymycin C1 **6**) were cultivated for production. Methanol extracts of the mycelium and the water phase were subjected to size exclusion chromatography and semi-preparative HPLC to furnish muraymycins **3–8**. For the *in vitro* evaluation of MraY inhibition, we employed a fluorescence-based MraY assay originally reported by Bugg et al.[12] In this assay, a fluorescence-labelled derivative of Park's nucleotide **1** (dansylated in the L-lysine side chain, Scheme 1) is employed and furnishes, upon incubation with MraY and undecaprenyl phosphate, a dansylated derivative of lipid I **2**. This product shows stronger fluorescence intensity than the dansylated substrate, thus leading to an increase of fluorescence over time as a measure of MraY activity. We have recently reported the

preparation of the dansylated substrate by total synthesis and a modified version of this assay with lower substrate concentration.[12d]

All isolated natural products **3–8** were initially tested for their inhibitory potency towards two different preparations of MraY: (i) MraY from *S. aureus*, overexpressed in *E. coli* and used as crude membranes from overexpressing cells;[5a,10f,12d] and (ii) MraY from *Aquifex aeolicus*, overexpressed in *E. coli* and purified to homogeneity in the presence of a non-denaturing detergent (modified protocol from ref.[6b,11a]). Attempts to extract and purify the MraY homologue from *S. aureus* from the crude membranes led to denaturation and therefore afforded inactive protein. Activity assays were performed in triplicates, providing (after data fitting) IC_{50} values for inhibition with standard deviations (Table 1). We had previously reported the synthetic 5′-defunctionalized (i.e., 5′-deoxy) muraymycin analogue **9** (Table 1),[10f] which was also included in this study.

Assays with MraY from *S. aureus* in crude membranes revealed all tested naturally occurring muraymycins **3–8** to be picomolar inhibitors of the enzyme (Table 1). For this source of MraY activity, muraymycins A1 **3**, B2 **4**, B6 **5** and C1 **6** were most active with IC_{50} values in the range of 10–27 pM. It was therefore concluded that the fatty acyl motif found in muraymycins of the A- and B-series apparently does not contribute to MraY inhibition as it is absent in congener C1 **6**. This finding supports a previous hypothesis that the fatty acyl unit might instead contribute primarily to the cellular uptake rather than target interaction.[10d,13] On the other hand, muraymycin D1 **7** was a ca. 30-fold less active MraY inhibitor than C1 **6**, although it only differs in the missing β-hydroxy group in the leucine moiety. This indicated a general beneficial role of such a β-substituent in the leucine unit for MraY inhibition, irrespective of *O*-acylation in this position. Inhibitory activities of muraymycins D1 **7** and D2 **8** against MraY from *S. aureus* were nearly identical, thus demonstrating that the 2-*O*-methyl motif found in the aminoribose unit of several naturally occurring muraymycins does not contribute to inhibition. The validity of these inhibitory activities was confirmed using the commercially available nucleoside antibiotic tunicamycin as a reference MraY inhibitor (see Supporting Information).

In contrast to the activities of natural products **3**–**8** in the pM range, synthetic 5′-defunctionalized analogue **9** showed an IC_{50} value of 95 ± 19 nM against MraY from *S. aureus* (Table 1). This value was slightly lower than previously reported[10f] due to differences in the assay protocol, i.e., lower substrate concentration. A comparison of **9** with the most closely related natural product C1 **6** revealed a ca. 6000-fold loss in activity due to the absence of the aminoribose unit, thus further establishing that the aminoribose motif mediates a key interaction with MraY. This has also been derived from the X-ray co-crystal structure of MraY from *Aquifex aeolicus* in complex with muraymycin D2 **8**.[11] Nonetheless, **9** was still a fairly strong MraY inhibitor, thus making it a suitable, structurally simplified model system for SAR studies, for instance, on variations in the peptide unit.

We then tested for inhibition of solubilized and purified MraY from *Aquifex aeolicus* (Table 1). Unexpectedly, inhibitory potencies were generally worse than with the *S. aureus* protein, with IC_{50} values of muraymycins **3–6** being up to ca. 22-fold higher. For muraymycins D1 **7** and D2 **8**, IC_{50} values were in the nM range, demonstrating that inhibitory activities were ca.

200-fold and 120-fold, respectively, lower than for the protein preparation from *S. aureus*. Synthetic analogue **9** was only active at μM concentrations. Overall, the order of activities within the compound series **3–9** was similar irrespective of the source of MraY activity. However, structural variations which were disadvantageous for MraY inhibition had a much more pronounced effect with the solubilized *Aquifex aeolicus* protein. For instance, the inhibitory activity of muraymycin D1 **7** was ca. 280-fold lower than the one of C1 **6** (instead of 30-fold for the *S. aureus* protein, *vide supra*), and the formal removal of the aminoribose unit in **9** (relative to **6**) even furnished more than five orders of magnitude loss in potency.

It was unclear if this pronounced decrease in inhibitory activities was a result from the solubilization of the *Aquifex aeolicus* protein, or if it was related to the intrinsic properties of this MraY homologue. As *Aquifex aeolicus* is an extremophile, the latter could not be ruled out, even if sequence alignment of the *S. aureus* and *Aquifex aeolicus* proteins suggested otherwise (Figure S1, Supporting Information). We therefore decided to overexpress a third MraY homologue, i.e., MraY from *Clostridium bolteae*. This MraY homologue was recently used to obtain a co-crystal structure with tunicamycin[14] and has three relevant features: (i) it does not originate from an extremophile, but rather from a bacterial class with human pathogenicity; (ii) it can be solubilized and purified (like the *Aquifex aeolicus* protein); (iii) in contrast to the MraY homologue from *Aquifex aeolicus*, it is not modified with a maltose-binding protein (MBP) purification tag for overexpression. The MBP tag solubilizes MraY and therefore precludes its membrane insertion. Hence, the activity of the *C. bolteae* protein can alternatively also be applied in the form of non-purified crude membranes from overexpressing cells, as it is the case for the MraY homologue from *S. aureus*.

Inhibition data of **3–9** with solubilized and purified MraY from *C. bolteae* were very similar to activities observed with the solubilized *Aquifex aeolicus* protein (Table 1). In contrast, IC_{50} values of muraymycins **3–5** with MraY from *C. bolteae* in crude membranes were nearly identical to values obtained with the *S. aureus* protein in crude membranes. The IC_{50} value of muraymycin C1 **6** with MraY from *C. bolteae* in crude membranes was at least closer to the *S. aureus* than to the *Aquifex aeolicus* inhibitory activity. For all other inhibitors **7–9**, IC_{50} values with MraY from *C. bolteae* in crude membranes was in between the *S. aureus* and the *Aquifex aeolicus* values. We therefore concluded that inhibitory potencies may differ for MraY homologues from different bacteria, in a way that structural variations of the inhibitors can have more or less pronounced effects. This can obviously even occur in the case of high sequence homology, as it is displayed by the MraY homologues selected for this study (Figure S1, Supporting Information). However, relative tendencies within a series of inhibitors are notably retained. On the other hand, the preparation of the protein for *in vitro* MraY assays appears to play a significant role. While solubilization with non-denaturing detergents in principle retains MraY activity, it leads to differences in MraY inhibition. The latter apparently depends on the native lipid environment provided by cellular membranes. The comparison of the X-ray co-crystal structure of MraY from *Aquifex aeolicus* in complex with muraymycin D2 **8**[11] with the previously reported structure of the apo-enzyme[6b] reveals that MraY undergoes a large conformational change upon inhibitor binding. Such conformational transitions may be different in a lipid

membrane environment than in detergent micelles, which could potentially explain the observed differences in inhibition data.

In order to correlate *in vitro* activities for MraY inhibition with antibacterial properties, we then studied inhibition of bacterial growth by muraymycins **3–9** (Table 1). Antibacterial activities of muraymycins against a range of pathogens have been reported before,[8c,9a,b] so we mainly aimed to confirm some of the previously reported data and to comparatively study activities against an efflux-deficient bacterial strain. In our hands, muraymycin A1 **3** was the only congener with notable activity against *S. aureus*, which was in good agreement with previously published data.[9a,b] This result was remarkable though as muraymycins **3–6** all showed very similar inhibitory potencies towards MraY, thus indicating that the unique activity of **3** is correlated to cellular access to the target. We have reported a model system to study the membrane-interacting properties of the ω-functionalized fatty acid moiety of A-series muraymycins such as **3**, indicating both an efficient membrane accumulation and penetration mediated by the guanidinium-containing structural motif.[13] This present study encourages further investigations on the membrane-penetrating properties of *O*-acylated muraymycins of the A- and B-groups.

Natural products **3–8** were not active against the efflux-competent *E. coli* strain DH5α, but showed strong activities against the efflux-deficient *E. coli* Δ*tolC* mutant (MIC 1–2 μg/mL, Table 1). This demonstrated that efflux also plays a significant role in the antibacterial activity of muraymycins. Remarkably, differences of **3–8** in MraY inhibition did not furnish differences in their inhibition of the growth of *E. coli* Δ*tolC*. For instance, muraymycins C1 **6** and D1 **7** showed a ca. 30- to 430-fold difference in MraY inhibition (crude membranes only), but identical MIC values (1 μg/mL). As both compounds are very similar in size and polarity, it is highly unlikely that they differ in cellular uptake, and efflux is hampered in the *E. coli* Δ*tolC* mutant. We therefore conclude that an increase in inhibitory activity towards MraY beyond a certain threshold does not necessarily lead to increased antibacterial activity. Contrastingly, synthetic analogue **9** differed from natural products **3–8** as it inhibited the growth of both the *E. coli* strain DH5α and the Δ*tolC* mutant, mostly likely because it was less prone to cellular efflux. Taken together, the obtained antibacterial data suggest that antibacterial activities of muraymycins are largely influenced by cellular uptake and efflux phenomena and are less dependent on pharmacodynamic properties.

The recently reported X-ray co-crystal structure of MraY from *Aquifex aeolicus* in complex with muraymycin D2 **8**[11] provided the first structural insights into the target interactions of muraymycins. With exact inhibition data of compounds **3**–**9** in hand, we were interested if the relative inhibitory potencies of muraymycins could be correlated to structural aspects by *in silico* modelling, based on the co-crystal structure. Attempts to dock muraymycin ligands into this structure of MraY (after deletion of **8** from the complex) failed due to the pronounced conformational flexibility of muraymycins (which hampered their energy optimization in solution) and of the enzyme. We therefore proceeded by changing the structure of ligand **8** within the complex into the structure of other muraymycins, and the energy of the respective ligand-protein complex was then minimized *in silico*. This procedure furnished calculated structures of other muraymycin-MraY complexes (Figure S2, Supporting Information). The structure of muraymycin A1 **3** in complex with MraY is a

representative example (Figure 1). An overlay of the calculated structure of **3** and the reported structure of muraymycin D2 **8** in complex with MraY revealed a very similar orientation of the ligands. The ω-functionalized fatty acid moiety of muraymycin A1 **3**, which is absent in **8**, is oriented towards a hydrophobic cleft of the protein (Figure 1A). Within this hydrophobic cleft, there are no apparent key interactions of the fatty acyl unit with MraY though (Figure 1B).

In conclusion, we report a detailed study on the inhibition of the bacterial membrane enzyme MraY by naturally occurring muraymycin nucleoside antibiotics. Several muraymycins proved to be extremely potent MraY inhibitors with IC_{50} values in the low pM range. The tested compounds provided SAR data on the roles of the aminoribose unit (mediates a key interaction), its 2-*O*-methylation (not significant for MraY inhibition) and the fatty acid unit (probably mainly required for uptake-related effects). Different sources of MraY activity revealed (i) that inhibitory potencies may differ for MraY homologues from different bacteria, but that relative SAR trends seem to be retained, and (ii) that the preparation of the protein (crude membranes vs. solubilization with detergents) significantly influences inhibitory data from *in vitro* MraY assays. Antibacterial activities do not fully correlate to SAR data for MraY inhibition, indicating that cellular uptake as well as efflux play important roles. Overall, our results have strong implications for future work. First, the obtained SAR data will contribute to the design of novel muraymycin analogues. Second, MraY assays should be performed with protein preparations providing a lipid environment similar to the native one for MraY, i.e., crude membranes or lipid bilayer nanodiscs. Work along this line is ongoing in our laboratories.

Experimental Section

Full experimental details are disclosed in the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Deutsche Forschungsgemeinschaft (DFG, grant DU 1095/5-1) for financial support. P.D.F. is grateful for a doctoral fellowship of the Fonds der Chemischen Industrie (FCI). We thank Professor Seok-Yong Lee (Duke University Medical Center) for providing us with the plasmid for the overexpression of MraY from *Aquifex aeolicus*.

References

1. a) Taubes G. Science. 2008; 321:356–361. [PubMed: 18635788] b) Cooper MA, Shlaes D. Nature. 2011; 472:32. [PubMed: 21475175]
2. a) Dini C. Curr Top Med Chem. 2005; 5:1221–1236. [PubMed: 16305528] b) Bugg TDH, Lloyd AJ, Roper DI. Infect Disorders Drug Targets. 2006; 6:85–106.
3. a) Struve WG, Neuhaus FC. Biochem Biophys Res Commun. 1965; 18:6–12. [PubMed: 14265759] b) Anderson JS, Matsuhashi M, Haskin MA, Strominger JL. Proc Natl Acad Sci USA. 1965; 53:881–889. [PubMed: 14324547] c) Heydanek MG Jr, Struve WG, Neuhaus FC. Biochemistry. 1969; 8:1214–1221. [PubMed: 5781013] d) Ikeda M, Wachi M, Jung HK, Ishino F, Matsuhashi M. J

Bacteriol. 1991; 173:1021–1026. [PubMed: 1846850] e) Boyle DS, Donachie WD. J Bacteriol. 1998; 180:6429–6432. [PubMed: 9829961]

4. Vollmer W, Blanot D, De Pedro MA. FEMS Microbiol Rev. 2008; 32:149–167. [PubMed: 18194336]
5. a) Bouhss A, Crouvoisier M, Blanot D, Mengin-Lecreulx D. J Biol Chem. 2004; 279:29974–29980. [PubMed: 15131133] b) Ma Y, Münch D, Schneider T, Sahl HG, Bouhss A, Ghoshdastider U, Wang J, Dötsch V, Wang X, Bernhard F. J Biol Chem. 2011; 286:38844–38853. [PubMed: 21937437] c) Henrich E, Ma Y, Engels I, Münch D, Otten C, Schneider T, Henrichfreise B, Sahl HG, Dötsch V, Bernhard F. J Biol Chem. 2016; 291:2535–2546. [PubMed: 26620564]
6. a) Bouhss A, Mengin-Lecreulx D, Le Beller D, van Heijenoort J. Mol Microbiol. 1999; 34:576–585. [PubMed: 10564498] b) Chung BC, Zhao J, Gillespie RA, Kwon DY, Guan Z, Hong J, Zhou P, Lee SY. Science. 2013; 341:1012–1016. [PubMed: 23990562]
7. a) Lloyd AJ, Brandish PE, Gilbey AM, Bugg TDH. J Bacteriol. 2004; 186:1747–1757. [PubMed: 14996806] b) Al-Dabbagh B, Henry X, Ghachi ME, Auger G, Blanot D, Parquet C, Mengin-Lecreulx D, Bouhss A. Biochemistry. 2008; 47:8919–8928. [PubMed: 18672909] c) Al-Dabbagh B, Olatunji S, Crouvoisier M, El Ghachi M, Blanot D, Mengin-Lecreulx D, Bouhss A. Biochimie. 2016; 127:249–257. [PubMed: 27312048]
8. a) Winn M, Goss RJM, Kimura KI, Bugg TDH. Nat Prod Rep. 2010; 27:279–304. [PubMed: 20111805] b) Ichikawa S, Yamaguchi M, Matsuda A. Curr Med Chem. 2015; 22:3951–3979. [PubMed: 26282943] c) Wiegmann D, Koppermann S, Wirth M, Niro G, Leyerer K, Ducho C. Beilstein J Org Chem. 2016; 12:769–795. [PubMed: 27340469] d) Bugg TDH, Rodolis MT, Mihalyi A, Jamshidi S. Bioorg Med Chem. 2016; 24:6340–6347. [PubMed: 27021004]
9. a) McDonald LA, Barbieri LR, Carter GT, Lenoy E, Lotvin J, Petersen PJ, Siegel MM, Singh G, Williamson RT. J Am Chem Soc. 2002; 124:10260–10261. [PubMed: 12197711] b) Carter, GT., Lotvin, JA., McDonald, LA. WO. 02/085310 A2. 2002. c) Cui Z, Wang X, Koppermann S, Thorson JS, Ducho C, Van Lanen SG. J Nat Prod. 2018 submitted.
10. a) Lin YI, Li Z, Francisco GD, McDonald LA, Davis RA, Singh G, Yang Y, Mansour TS. Bioorg Med Chem Lett. 2002; 12:2341–2344. [PubMed: 12161129] b) Yamashita A, Norton E, Petersen PJ, Rasmussen BA, Singh G, Yang Y, Mansour TS, Ho DM. Bioorg Med Chem Lett. 2003; 13:3345–3350. [PubMed: 12951123] c) Tanino T, Ichikawa S, Al-Dabbagh B, Bouhss A, Oyama H, Matsuda A. ACS Med Chem Lett. 2010; 1:258–262. [PubMed: 24900205] d) Tanino T, Al-Dabbagh B, Mengin-Lecreulx D, Bouhss A, Oyama H, Ichikawa S, Matsuda A. J Med Chem. 2011; 54:8421–8439. [PubMed: 22085339] e) Takeoka Y, Tanino T, Sekiguchi M, Yonezawa S, Sakagami M, Takahashi F, Togame H, Tanaka Y, Takemoto H, Ichikawa S, Matsuda A. ACS Med Chem Lett. 2014; 5:556–560. [PubMed: 24900879] f) Spork AP, Büschleb M, Ries O, Wiegmann D, Boettcher S, Mihalyi A, Bugg TDH, Ducho C. Chem Eur J. 2014; 20:15292–15297. [PubMed: 25318977] g) Mitachi K, Aleiwi BA, Schneider CM, Siricilla S, Kurosu M. J Am Chem Soc. 2016; 138:12975–12980. [PubMed: 27617631]
11. a) Chung BC, Mashalidis EH, Tanino T, Kim M, Matsuda A, Hong J, Ichikawa S, Lee SY. Nature. 2016; 533:557–560. [PubMed: 27088606] b) Koppermann S, Ducho C. Angew Chem. 2016; 128:11896–11898.Angew Chem Int Ed. 2016; 55:11722–11724.
12. a) Brandish PE, Burnham MK, Lonsdale JT, Southgate R, Inukai M, Bugg TDH. J Biol Chem. 1996; 271:7609–7614. [PubMed: 8631795] b) Brandish PE, Kimura KI, Inukai M, Southgate R, Lonsdale JT, Bugg TDH. Antimicrob Agents Chemother. 1996; 40:1640–1644. [PubMed: 8807054] c) Stachyra T, Dini C, Ferrari P, Bouhss A, van Heijenoort J, Mengin-Lecreulx D, Blanot D, Biton J, Le Beller D. Antimicrob Agents Chemother. 2004; 48:897–902. [PubMed: 14982781] d) Wohnig S, Spork AP, Koppermann S, Mieskes G, Gisch N, Jahn R, Ducho C. Chem Eur J. 2016; 22:17813–17819. [PubMed: 27791327]
13. Ries O, Carnarius C, Steinem C, Ducho C. Med Chem Commun. 2015; 6:879–886.
14. Hakulinen JK, Hering J, Brändén G, Chen H, Snijder A, Ek M, Johansson P. Nat Chem Biol. 2017; 13:265–267. [PubMed: 28068312]

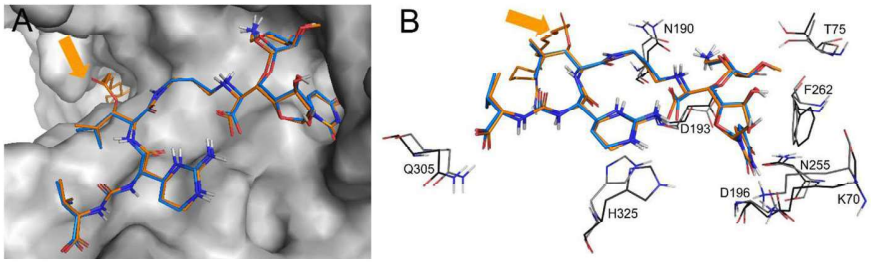
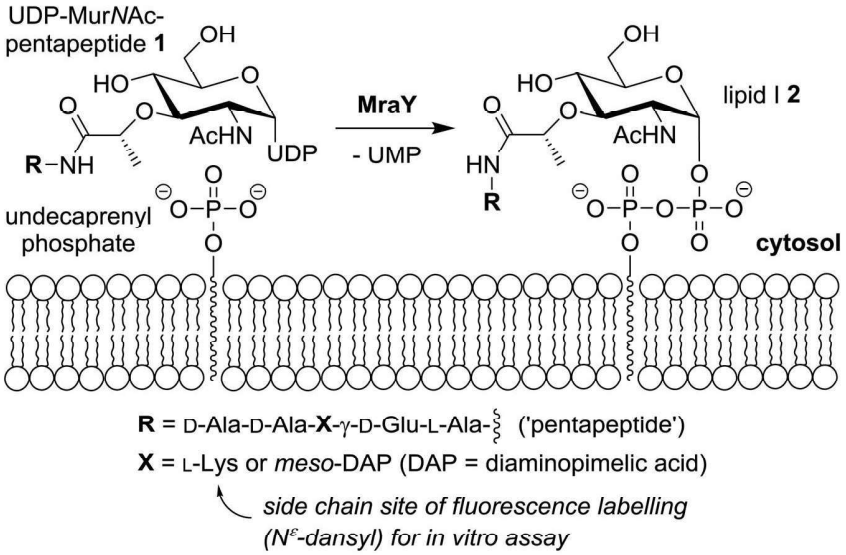


Figure 1.

A: Overlay of muraymycins A1 **3** (hydrocarbon scaffold in orange) and D2 **8** (hydrocarbon scaffold in light blue) in complex with MraY from *Aquifex aeolicus* (grey), with the position of **8** derived from the previously reported X-ray co-crystal structure[11] and the position of **3** derived from *in silico* modelling. The orange arrow indicates the position of the fatty acyl side chain of **3**, which points into a hydrophobic cleft of the protein. **B:** Overlay of muraymycins A1 **3** (orange) and D2 **8** (light blue) from the aforementioned complexes with MraY, showing parts of MraY (thin stick representations) which mediate key interactions (light grey for **3**, dark grey for **8**). The orange arrow indicates the position of the fatty acyl side chain of **3**, which does not show specific key interactions with the protein.

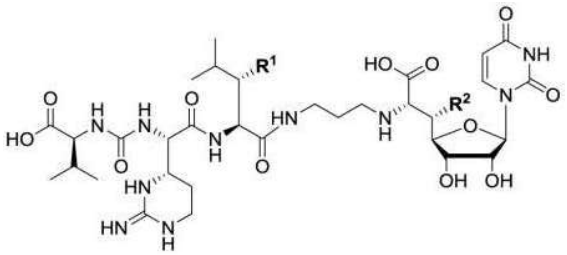


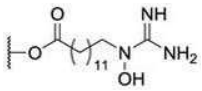
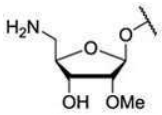
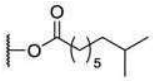
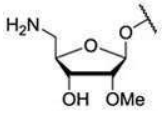
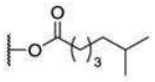
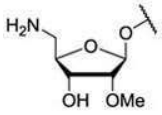
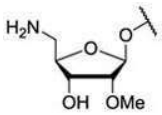
Scheme 1.

MraY-catalyzed reaction of Park's nucleotide **1** towards lipid I **2** (undecaprenyl phosphate represented schematically; UDP = uridine diphosphate; UMP = uridine monophosphate). The exact composition of the pentapeptide moiety can vary in different bacteria.[4]

Table 1

Inhibition data (from *in vitro* MraY assays) and antibacterial activities of naturally occurring muraymycins **3–8** as well as of synthetic analogue **9**.



muraymycin	R^1	R^2	IC_{50} (MraY *S. aureus*) [nM][a]	IC_{50} (MraY *Aquifex aeolicus*) [nM][b]	IC_{50} (MraY *C. bolteae* solubil./crude) [nM][c]	MIC (*S. aureus*) [μg/mL][d]	MIC (*E. coli* Δ*tolC*/DH5α) [μg/mL][e]
A1 **3**			0.027 ± 0.003	0.11 ± 0.02	0.16 ± 0.04 0.017 ± 0.004	18	2 > 32
B2 **4**			0.010 ± 0.002	0.11 ± 0.01	0.14 ± 0.01 0.0083 ± 0.0006	> 32	1 > 32
B6 **5**			0.021 ± 0.002	0.19 ± 0.02	0.38 ± 0.03 0.036 ± 0.004	> 32	1 > 32
C1 **6**	OH		0.016 ± 0.002	0.35 ± 0.04	0.45 ± 0.17 0.093 ± 0.020	> 32	1 > 32

ChemMedChem. Author manuscript; available in PMC 2019 April 23.

muraymycin	R^1	R^2	IC_{50} (MraY *S. aureus*) [nM][a]	IC_{50} (MraY *Aquifex aeolicus*) [nM][b]	IC_{50} (MraY *C. bolteae* solubil./crude) [nM][c]	MIC (*S. aureus*) [μg/mL][d]	MIC (*E. coli* Δ*tolC*/DH5α) [μg/mL][e]
						> 80	1
							> 32
D1 **7**	H		0.48 ± 0.13	99 ± 30	380 ± 110 40 ± 6		
						> 80	1
							> 32
D2 **8**	H		0.39 ± 0.11	46 ± 13	200 ± 30 25 ± 4		
analogue **9**	OH	H	95 ± 19	(50% @100 μM)[f]	58% @100 μM)[f] 860 ± 300	> 50	50 15

[a]Fluorescence-based *in vitro* assay with MraY from *S. aureus* in crude membranes, IC_{50} ± SD.

[b]Fluorescence-based *in vitro* assay with MraY from *Aquifex aeolicus* in solubilized and purified form, IC_{50} ± SD (except where indicated).

[c]Fluorescence-based *in vitro* assay with MraY from *C. bolteae*, first value: IC_{50} ± SD (except where indicated) for protein in solubilized and purified form, second value: IC_{50} ± SD for protein in crude membranes.

[d]Minimal inhibitory concentration against growth of *S. aureus* subsp. *aureus* (Newman strain).

[e]Minimal inhibitory concentration against growth of *E. coli*, first value: efflux-deficient Δ*tolC* mutant, second value: DH5α strain.

[f]Percentage of inhibition at the given concentration of the inhibitor.

CHEMMEDCHEM
Chemistry & Drug Discovery

Supporting Information

Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics

Stefan Koppermann,[a] Zheng Cui,[b] Patrick D. Fischer,[a] Xiachang Wang,[c, d] Jannine Ludwig,[a] Jon S. Thorson,[b, d] Steven G. Van Lanen,[b] and Christian Ducho*[a]

cmdc_201700793_sm_miscellaneous_information.pdf

Table of contents

Additional figures

```
CLUSTAL O(1.2.4) multiple sequence alignment

A.aeolicus      MLYQLALLLKDYWFAFNVLKYITFRSFTAVLIAFFLTLVLSPSFINRLRKIQRLFGGYVR
E.coli          MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKL--SFGQVVR
S.aureus        ------------------------MIFVYALLALVITFVLVPVLIPTLKRM--KFGQSIR
C.bolteae       ----------------------MIHETILAIIIAFAISALLCPIIIPFLHKL--KFGQQVR
                                       :   : *: ::  : * :*  *:::   **  :*

A.aeolicus      EYTPESHEVKKYTPTMGGIVILIVVTLSTLLLM--RWDIKYTWVVLLSFLSFGTIGFWDD
E.coli          NDGPESHFSKRGTPTMGGIMILTAIVISVLLWA--YPSNPYVWCVLVVLVGYGVIGFVDD
S.aureus        EEGPQSHMKKTGTPTMGGLTFLLSIVITSLVAIIFVDQANPIILLLFVTIGFGLIGFIDD
C.bolteae       DDGPESHLKKQGTPTMGGLIILSSIIITSVF---YIPSYPKIIPVLFVTVGFGIIGFLDD
                :  *:**  *  ******: :*  : :: :.      .      :*.  :.:* *** **

A.aeolicus      YVKLKNKKGISIKTKFLLQVLSASLISV-LIYYW-------ADIDTILYFPFFK----EL
E.coli          YRKVVRKDTKGLIARWKYFWMSVIALGVAFALYL----AGKDTPATQLVVPFFK----DV
S.aureus        YIIVVKKNNQGLTSKQKFLAQIG----IAIIFFVLSNVFHLVNFSTSIHIPFTNVA----
C.bolteae       YIKIVMKRSEGLKPMQKLVGQFI----ITGIFAWY--LLNSGEVGTDMLIPFTGGFDGGS
                *  :  *   .:               :                 * : .**

A.aeolicus      YVDLGVLYLPFAVFVIVGSANAVNLTDGLDGLAIGPAMTTATALGVVAYAVGHSKIAQYL
E.coli          MPQLGLFYILLAYFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYL
S.aureus        -IPLSFAYVIFIVFWQVGFSNAVNLTDGLDGLATGLSIIGFTMYAIMSFVLGETAI----
C.bolteae       FLSLGIFFVPALFFIMLGTDNGVNFTDGLDGLCTSVTILVATFLTIVAIG-EDMGI----
                   *.. ::    *  :*  *.**:*******.   ::       :::    .  :

A.aeolicus      NIPYVPYAGELTVFCFALVGAGLGFLWFNSFPAQMFMGDVGSLSIGASLATVALLTKSEF
E.coli          HIPYLRHAGELVIVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEF
S.aureus        -----------GIFCIIMLFALLGFLPYNINPAKVFMGDTGSLALGGIFATISIMLNQEL
C.bolteae       -----------SPITGAVVGSLLGFLLFNVYPAKVFMGDTGSLALGGFVAASCYMMRMPL
                            .   :: : **** :*  **::****.***::*. ..  . : .  :

A.aeolicus      IFAVAAGVFVFETISVILQIIYFRWTGGKRLFKRAPFHHHLELNGLPEPKIVVRMWIISI
E.coli          LLVIMGGVFVVETLSVILQVGSFKLR-GQRIFRMAPIHHHYELKGWPEPRVIVRFWIISL
S.aureus        SLIFIGLVFVIETLSVMLQVASFKLT-GKRIFKMSPIHHHFELIGWSEWKVVTVFWAVGL
C.bolteae       FIPVIGLIYLVEVLSVIIQVTYFKRTGGKRIFKMAPIHHHFELCGWSETRVVAVFAIVTA
                 : . . :::.*.:**::*:  *:   *:*:*: :*:*** ** *  * :::. :  :

A.aeolicus      LLAIIAISMLKLR
E.coli          MLVLIGLATLKVR
S.aureus        ISGLIGLWIGVH-
C.bolteae       ILCMVA-YLGL--
                :  ::.

Percent Identity  Matrix - created by Clustal2.1

     1: A.aeolicus  100.00   47.75   36.57   36.66
     2: E.coli       47.75  100.00   43.13   42.99
     3: S.aureus     36.57   43.13  100.00   46.33
     4: C.bolteae    36.66   42.99   46.33  100.00
```

Figure S1. Sequence alignment (using CLUSTAL O)[S1-S3] and percent identity matrix of MraY homologues from *Aquifex aeolicus* VF5, *E. coli* K12, *S. aureus* MRSA252 and *C. bolteae* A0A9, demonstrating a pronounced similarity of the four MraY congeners. Residues involved in key interactions with muraymycin D2 **8** (according to the previously reported X-ray co-crystal structure,[S4] also see Figure 1) are highlighted in yellow and are conserved in all investigated homologues.

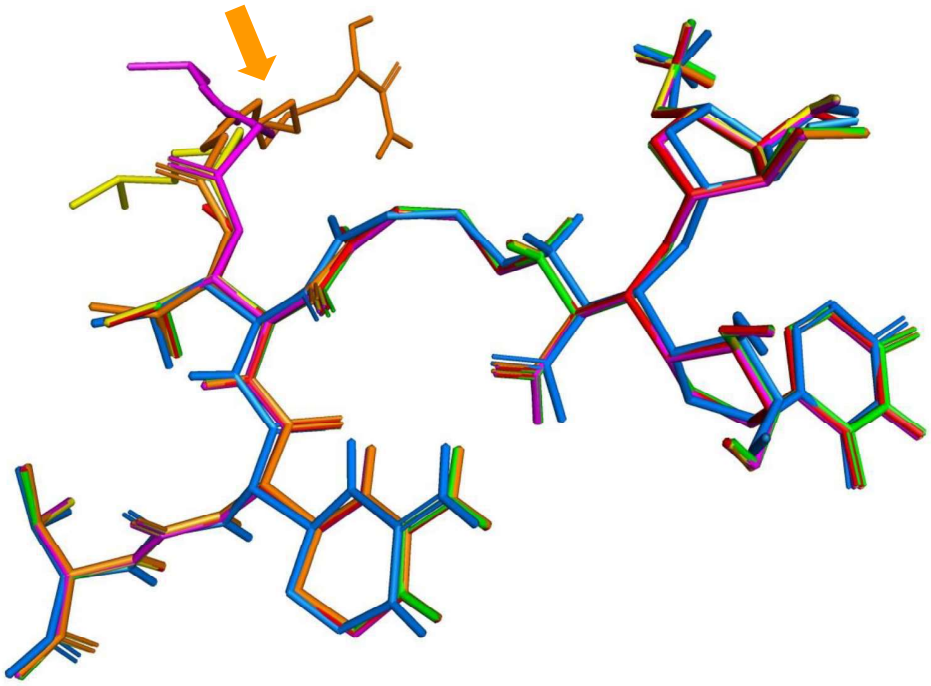


Figure S2. Overlay of muraymycin structures (stick representations; orange: muraymycin A1 **3**, magenta: B2 **4**, yellow: B6 **5**, red: C1 **6**, green: D1 **7**, light blue: D2 **8**) from their respective complexes with MraY (obtained by *in silico* modelling for **3-7** and from the previously reported X-ray co-crystal structure for **8**[S4]). The *in silico* calculations suggest a very similar binding mode for all muraymycins, but some flexibility in the orientation of the fatty acyl moiety (for **3-5**, indicated by orange arrow). This is in agreement with the finding that the fatty acyl moiety probably does not undergo specific key interactions with the protein, but mainly rather unspecific hydrophobic interactions (also see Figure 1).

Experimental section

General methods: NMR data were recorded at 400 MHz for ^{1}H and 100 MHz for ^{13}C with a Varian Inova NMR spectrometer (Agilent, Santa Clara, CA). HR-ESI-MS spectra were acquired with an AB SCIEX Triple TOF 5600 System (AB Sciex, Framingham, MA, USA). Semi-preparative HPLC was performed with a Waters 600 controller and pump (Milford, MA) equipped with a 996 diode array detector, 717 plus autosampler, and an Apollo C_{18} column (250 x 10 mm, 5 μm). All solvents used for the isolation of muraymycins were minimally of ACS grade and purchased from Pharmco-AAPER (Brookfield, CT).

Production and isolation of muraymycins: *Streptomyces* sp. (NRRL30473 for isolation of muraymycins A1 **3**, B2 **4** and B6 **5**; NRRL30475 for isolation of muraymycins D1 **7** and D2 **8**; NRRL30477 for isolation of muraymycin C1 **6**) were cultivated in 250 mL Erlenmeyer flasks containing 50 mL TSBG (Tryptic soy broth supplemented with 20 g/L D-glucose) at 30 °C on a rotary shaker (250 rpm) for 72 h. The seed cultures were used to inoculate flasks (250 mL) containing 50 mL PM-1 medium.[S5] PM-1 was composed of D-glucose 2%, soluble starch 1%, pressed yeast 0.9%, peptone 0.5%, meat extract 0.5%, NaCl 0.5%, $CaCO_3$ 0.3% and CB-442 (NOF Co., Ltd.) 0.01% (pH 7.4, before sterilization). The fermentation was continued on a rotary shaker (210 rpm) at 23 °C for 7 d. All culture flasks were then combined and centrifuged at 5000 rpm for 15 min to separate the mycelium and water phase. The mycelial-cake portion was extracted with MeOH by sonication, and the organic phase was evaporated to obtain a brown crude extract. Amberlite XAD-16 resin (Sigma) 3% was added to the water phase and stirred for 12 h. The resin was washed with water until the effluent became colorless, and then eluted with MeOH. The MeOH extract was concentrated under reduced pressure to obtain the crude extract. Components of the mycelium crude extract

and water phase crude extract were subjected to a Sephadex LH-20 (25-100 μm, GE Healthcare) column, and MeOH was used to elute compounds at a flow rate of 2 mL/min. The major fraction was further purified by using semi-preparative HPLC to obtain all muraymycins. A linear gradient was applied from water (A) to 0.025% TFA in MeCN (B) in the following manner (beginning time and ending time with linear increase to % B): 0 min, 10% B; 25 min, 22% B; 55 min, 55% B; 55-60 min 10% B; flow rate 3.5 mL/min. All muraymycins were analyzed by NMR and MS (Figures S5 to S35, *vide infra*).

(+)-HR-ESI-MS data:

muraymycin A1 **3**: *m/z* 1215.6599 $[M + H]^+$ (calcd for $C_{52}H_{90}N_{14}O_{19}$ 1215.6579).

muraymycin B2 **4**: *m/z* 1100.5833 $[M + H]^+$ (calcd for $C_{48}H_{81}N_{11}O_{18}$ 1100.5834).

muraymycin B6 **5**: *m/z* 1072.5553 $[M + H]^+$ (calcd for $C_{46}H_{77}N_{11}O_{18}$ 1072.5521).

muraymycin C1 **6**: *m/z* 946.4475 $[M + H]^+$ (calcd for $C_{38}H_{63}N_{11}O_{17}$ 946.4476).

muraymycin D1 **7**: *m/z* 930.4573 $[M + H]^+$ (calcd for $C_{38}H_{63}N_{11}O_{16}$ 930.4527).

muraymycin D2 **8**: *m/z* 916.4368 $[M + H]^+$ (calcd for $C_{37}H_{61}N_{11}O_{16}$ 916.4371).

Overexpression of MraY from *S. aureus*: The overexpression of MraY from *S. aureus* in *E. coli* was performed as described before.[S6]

Overexpression of MraY from *Aquifex aeolicus*: For the overexpression of MraY from *Aquifex aeolicus* in *E. coli*, the plasmid previously reported by Lee and co-workers[S4,S7] was used (kindly provided by Professor Seok-Yong Lee, Duke University Medical Center). This plasmid contained the codon-optimized *mraY* gene from *Aquifex aeolicus* encoding a fusion protein with decahistidine maltose-binding protein (His-MBP) and a PreScission protease cleavage site for removal of the tag. The plasmid was transformed into C41 (DE3) *E. coli* cells and MraY was expressed in terrific broth (TB) medium containing 35 μg/mL

kanamycin. Protein expression was induced at OD_{600} ~1.0 at 37 °C for 3 h using 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation at 5750 x g. All further steps were carried out at 4 °C. The cells were resuspended in 10-20 mL resuspension buffer per liter culture (resuspension buffer: 50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 2 mM β-mercaptoethanol (BME), lysozyme, DNAse I and one cOmplete™ EDTA-free Protease Inhibitor Cocktail tablet) and then lysed by sonication (30% amplitude, 3/10 s pulse, 7/10 s pause for 15 min). *n*-Dodecyl-β-D-maltoside (DDM, 30 mM) was added and the resultant lysate was stirred for 2 h. After centrifugation at 30 000 x g for 45 min, the cleared lysate was loaded onto pre-equilibrated TALON Superflow™ resin (~0.5 mL/L culture), imidazole was added to a final concentration of 7 mM and the mixture was rotated at 4 °C overnight. Subsequently, the resin was washed with 20 column volumes of washing buffer (50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 2 mM BME, 10 mM imidazole, 1 mM DDM) and eluted with 10-20 mL elution buffer (50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 2 mM BME, 250 mM imidazole, 1 mM DDM). Dithiothreitol (DTT, final concentration of 1 mM), ethylenediaminetetraacetic acid (EDTA, final concentration of 1 mM) and PreScission protease (ratio to His-MBP-MraY 1:100) were added and the solution was incubated at 4 °C overnight. The cleaved protein was purified by size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg in a buffer containing 20 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 2 mM DTT and 5 mM *n*-decyl-β-D-maltopyranoside (DM). The resultant solution was employed as a source of solubilized MraY from *Aquifex aeolicus* for *in vitro* MraY assays.

Overexpression of MraY from *C. bolteae*: For the overexpression of MraY from *C. bolteae* in *E. coli*, we constructed a plasmid similar to the previously reported one by Ek, Johansson and co-workers[S8] (synthetic plasmid purchased from Eurofins Genomics). This $pET26b^+$

plasmid contained the codon-optimized *mraY* gene from *C. bolteae* with an N-terminal PelB signal sequence and a C-terminal decahistidine tag. The plasmid was transformed into C41 (DE3) *E. coli* cells and MraY was expressed in TB medium containing 25 μg/mL kanamycin. Protein expression was induced at OD_{600} ~1.0 at 19 °C for 19 h using 0.4 mM IPTG. Cells were harvested by centrifugation at 5750 x g. All further steps were carried out at 4 °C. The cells were resuspended in 10-20 mL resuspension buffer per liter culture (resuspension buffer: 50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 2 mM BME, lysozyme, DNAse I and one cOmplete™ EDTA-free Protease Inhibitor Cocktail tablet) and then lysed by sonication (30% amplitude, 3/10 s pulse, 7/10 s pause for 15 min). The cell debris was removed by centrifugation at 4900 x g for 15 min. Membranes were isolated by centrifugation at 138 000 x g for 1 h. For the crude membrane preparation of MraY from *C. bolteae*, the resultant membrane pellet was resuspended in 0.9 mL/L culture buffer (50 mM TRIS-HCl (pH 7.5), 1 mM $MgCl_2$, 2 mM BME), flash frozen in liquid nitrogen and stored at -80 °C. This MraY-containing crude membrane preparation was used for MraY activity assays without further purification. For the preparation of solubilized and purified MraY from *C. bolteae*, the aforementioned membrane pellet was resuspended in ~10 mL/L culture membrane buffer (40 mM TRIS-HCl (pH 7.5), 300 mM NaCl, 2 mM BME, 10% (v/v) glycerol), DM (final concentration of 25 mM) was added and the mixture was stirred for 2 h. Insoluble parts were pelleted by centrifugation at 180 000 x g for 1 h. The soluble fraction was diluted 1:1 in membrane buffer and the pH was adjusted to 8.0 using 1 M NaOH. This solution was loaded onto pre-equilibrated TALON Superflow™ resin (~0.5 mL/L culture), washed with washing buffer (50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 1 mM $MgCl_2$, 2 mM BME, 10 mM imidazole, 5 mM DM, 10% (v/v) glycerol) and eluted with 10-20 mL elution buffer (50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 2 mM BME, 140 mM imidazole, 3 mM DM). The protein was purified by size exclusion chromatography on a pre-equilibrated

HiLoad 16/600 Superdex 200 pg in a buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl, 2 mM DTT and 3 mM DM. The resultant solution was employed as a source of solubilized MraY from *C. bolteae* for *in vitro* MraY assays.

Fluorescence-based MraY assay: *In vitro* MraY assays were performed using our previously reported adapted version[S6] of Bugg's fluorescence-based method.[S9-S11] Fluorescence intensity over time was measured at λ_{ex} = 355 nm and λ_{em} = 520 nm (BMG Labtech POLARstar Omega, 384-well plate format). Each well contained a total volume of 20 μL with 100 mM TRIS-HCl buffer (pH 7.5), 200 mM KCl, 10 mM $MgCl_2$, 0.1% Triton X-100, 0-5% DMSO, 50 mM undecaprenyl phosphate, 7.5 mM dansylated Park's nucleotide (synthetic or semi-synthetic),[S6] a protein preparation (*vide infra*) and the potential inhibitor at various concentrations. The amount of DMSO in the assay mixture depended on the solubility of the inhibitor, and inhibitor-free control assays with different DMSO content (up to 5%) showed no change in MraY activity. For MraY from *S. aureus*, 1 μL of a crude membrane preparation with a total protein concentration of 1.0 mg/mL was used and the reaction was initiated by the addition of the protein preparation. For solubilized MraY from *Aquifex aeolicus* and *C. bolteae*, 1 μL of a solution of purified protein was used and was preincubated with the assay mixture (including the inhibitor, but except the dansylated Park's nucleotide) at rt for 30 min. The reaction was then initiated by the addition of Park's nucleotide. Concentrations of the proteins were 0.1 mg/mL for solubilized MraY from *Aquifex aeolicus* and 0.04 mg/mL for solubilized MraY from *C. bolteae*, respectively. For MraY from *C. bolteae* as a crude membrane preparation, the procedure was identical to the protocol for MraY from *S. aureus* (*vide supra*). All protein preparations showed comparable MraY activity in the absence of inhibitors. MraY activity at a certain inhibitor concentration was determined by a linear fit of the fluorescence intensity curve from 0 to 2 min. This measure of

enzymatic activity was plotted against the logarithmic inhibitor concentration and fitted with a sigmoidal fit using the formula shown below, thus furnishing IC_{50} values (selected examples are provided in Figures S3 and S4).

$$y = A1 + \frac{(A2 - A1)}{1 + 10^{\log(x_0 - x)\cdot p}}$$

The aforementioned preincubation period for solubilized MraY (*i.e.*, 30 min. preincubation of the protein with the respective inhibitor, *vide supra*) was found to be necessary in order to obtain robust IC_{50} values, *i.e.*, values independent of a further extension of the preincubation period. This was probably due to sterically hindered access to the active site of MraY when it was in complex with solubilizing detergents, thus leading to a slower formation of the protein-ligand equilibrium. In contrast, IC_{50} values with the crude membrane preparations were independent of a preincubation period.

The validity of the inhibitory activities was confirmed using the commercially available nucleoside antibiotic tunicamycin (mixture of compounds with different fatty acyl units, obtained from Sigma-Aldrich) as a reference MraY inhibitor. Under our assay conditions and with MraY from *S. aureus* in crude membranes, this tunicamycin mixture inhibited MraY with $IC_{50} = 13.6 \pm 1.2$ ng/mL. If one assumes the four main components of the tunicamycin mixture to be present in equimolar amounts, this corresponds to an estimated IC_{50} value of ~16 nM. This result was in the same order of magnitude as previously reported IC_{50} values for MraY inhibition by tunicamycin, which were obtained with a different, FRET-based assay method.[S12]

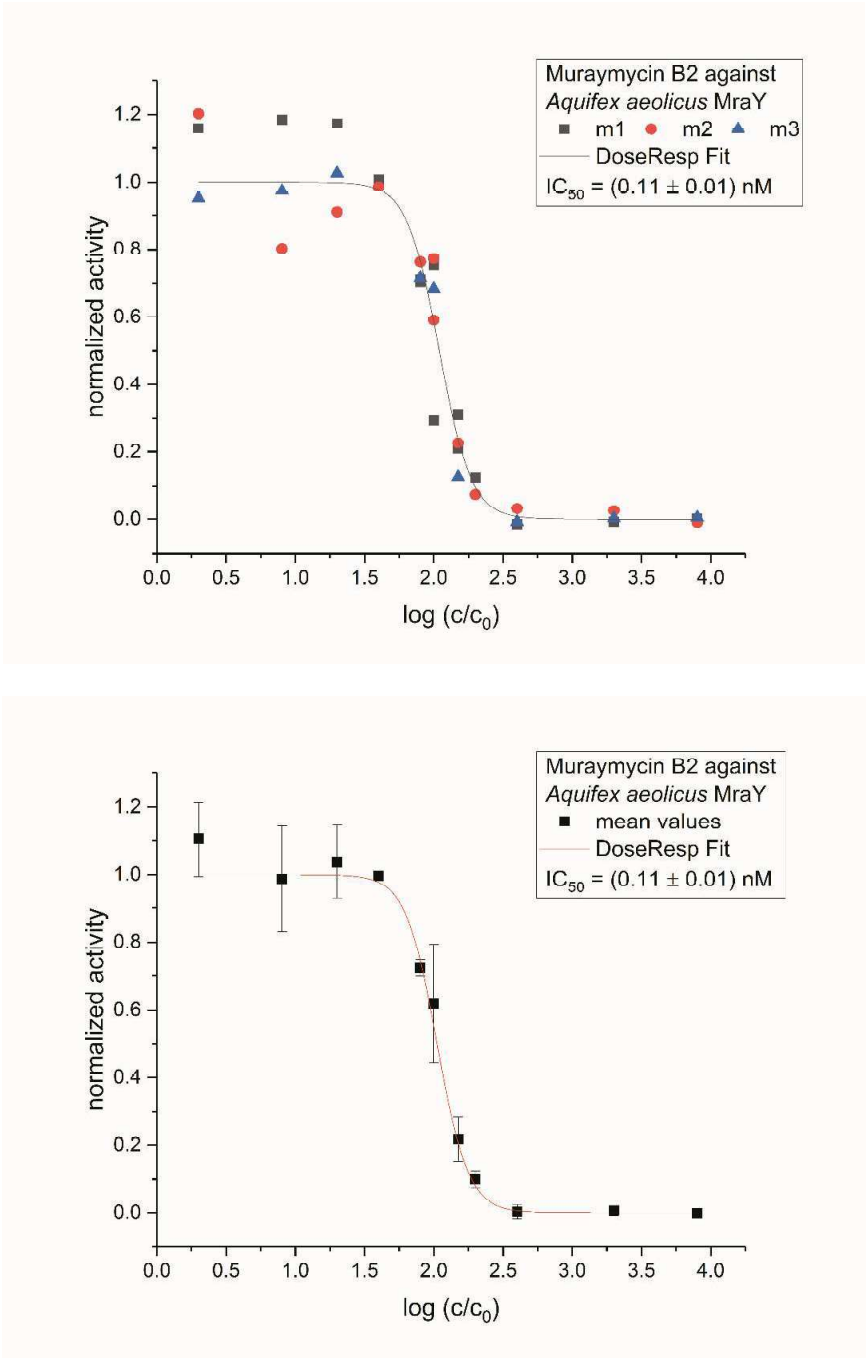


Figure S3. MraY assay using solubilized and purified MraY from *Aquifex aeolicus* and muraymycin B2 **4** as inhibitor, with the fitted curve for MraY activity (top: original triplicate data, shown as individual series of measurements; bottom: averaged data from the triplicate measurement).

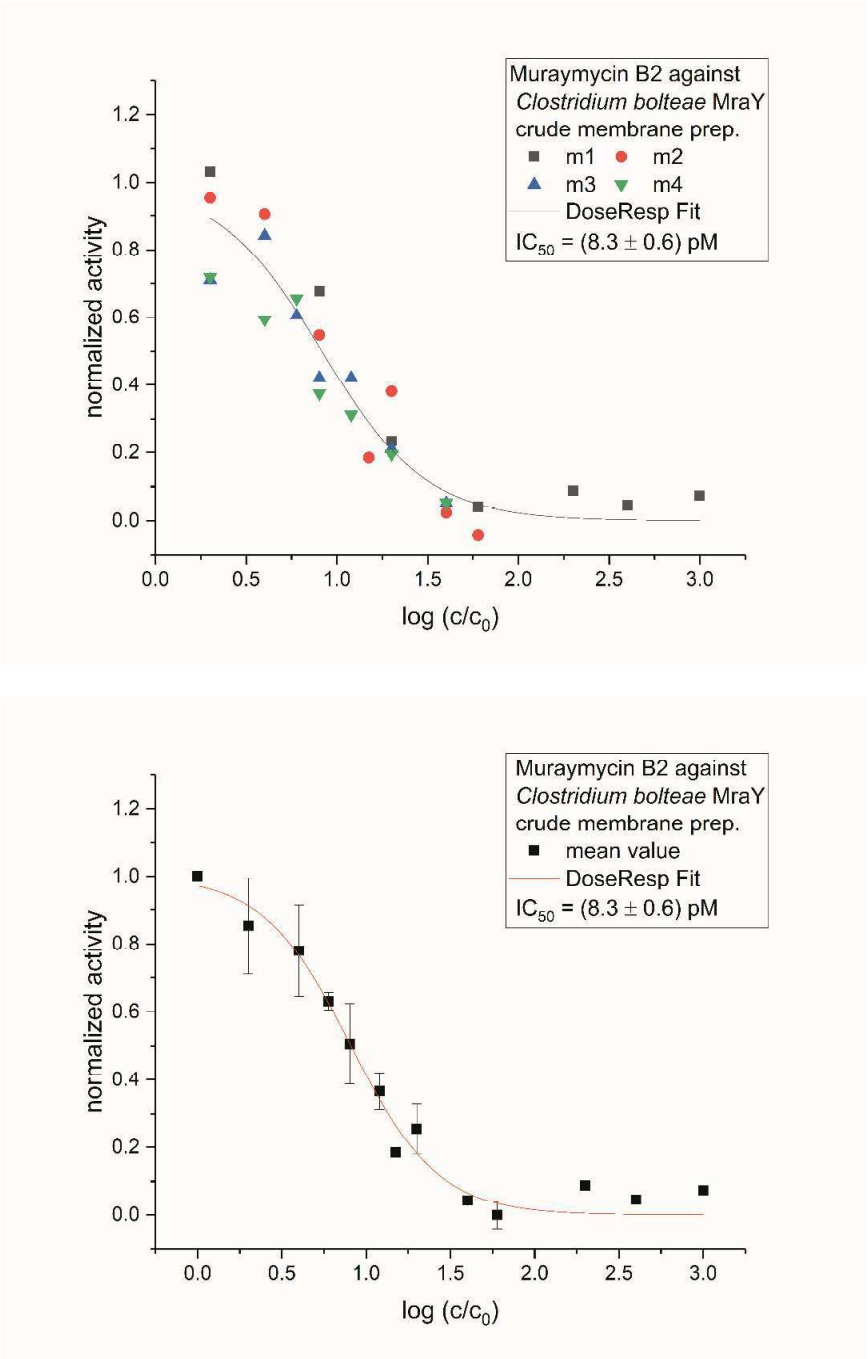


Figure S4. MraY assay using MraY from *C. bolteae* as crude membrane preparation and muraymycin B2 **4** as inhibitor, with the fitted curve for MraY activity (top: original triplicate data, shown as individual series of measurements; bottom: averaged data from the triplicate measurement).

Antibacterial activities: Antibacterial activities were determined as described before.[S13]

Molecular modelling: All procedures for *in silico* modelling were performed using the Molecular Operating Environment (MOE) software package (version 2010, Chemical Computing Group). The previously reported X-ray co-crystal structure of MraY from *Aquifex aeolicus* in complex with muraymycin D2 **8** (PDB 5CKR)[S4] was used as a starting point. The structure of the ligand muraymycin D2 **8** was changed into other muraymycins (A1 **3**, B2 **4**, B6 **5**, C1 **6**, D1 **7**), and energy minimization was performed using the AMBER99 force field, R-Field solvatization and standard parameters with ligand and protein free to move.

NMR and MS spectra of isolated muraymycins

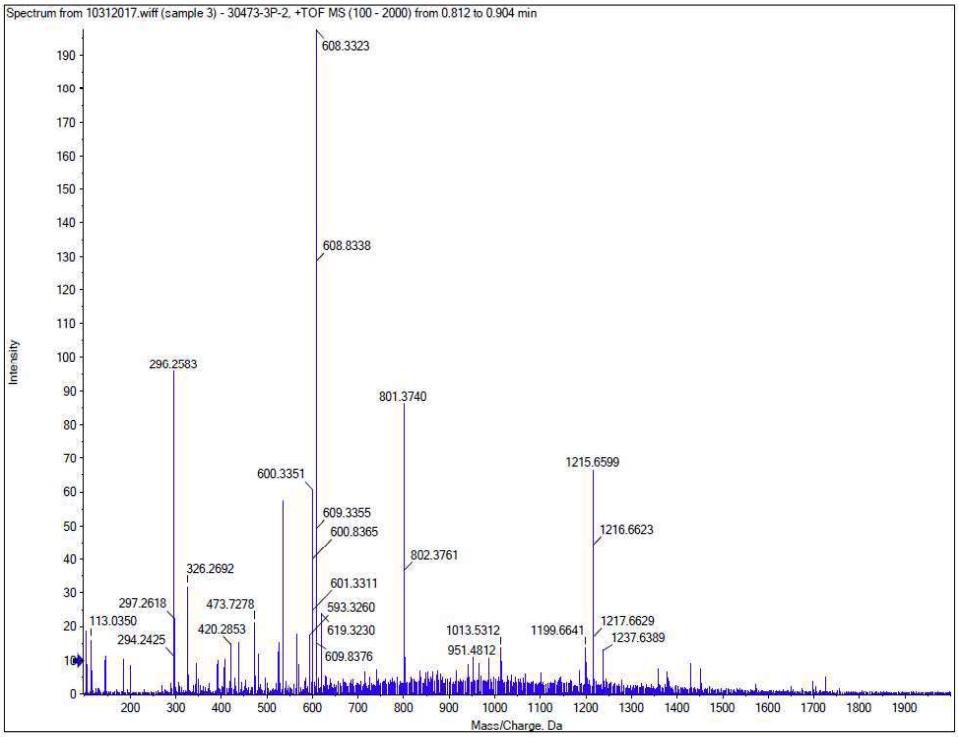


Figure S5. (+)-HR-ESI-MS (positive mode) of muraymycin A1 **3**.

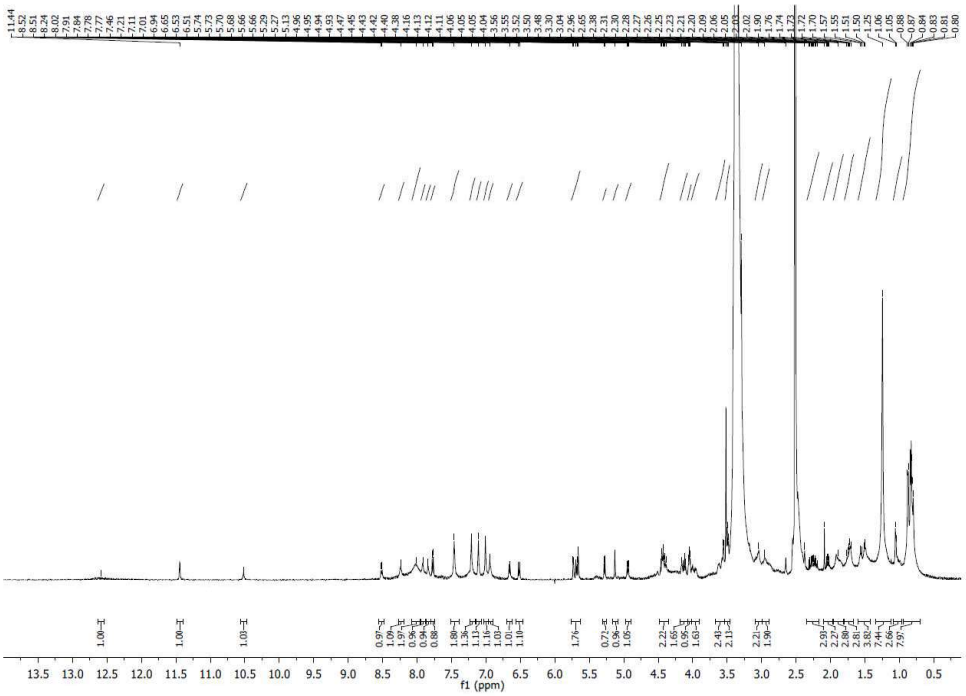


Figure S6. ^{1}H NMR spectrum (400 MHz, DMSO-d_6) of muraymycin A1 **3**.

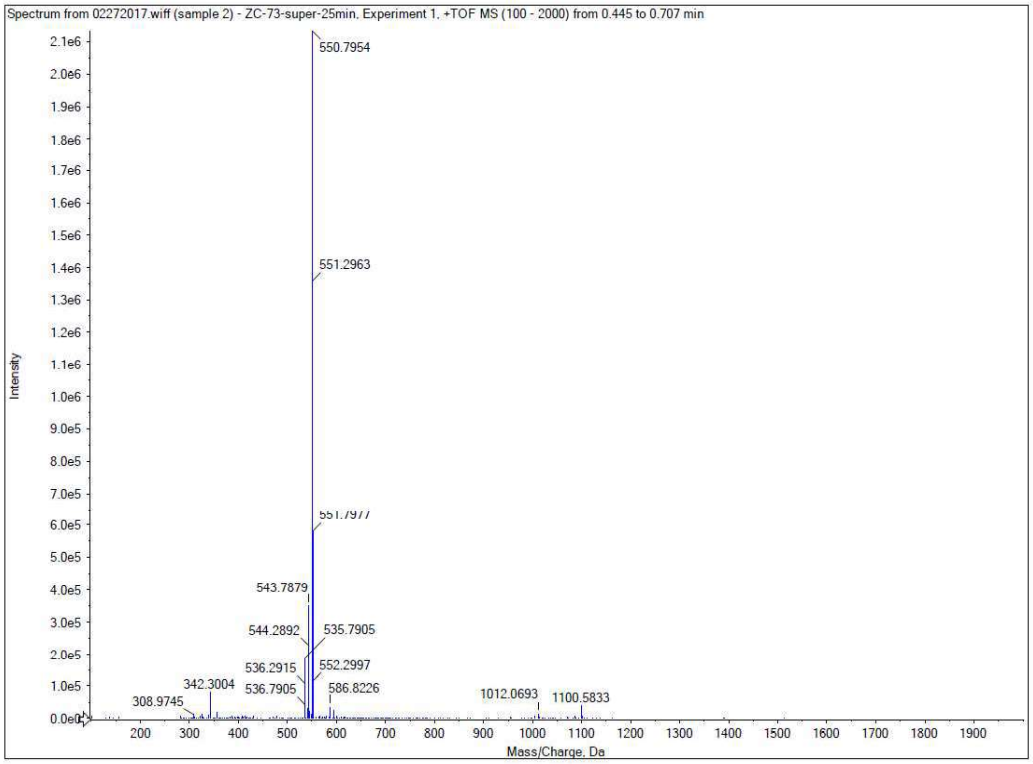


Figure S7. (+)-HR-ESI-MS (positive mode) of muraymycin B2 **4**.

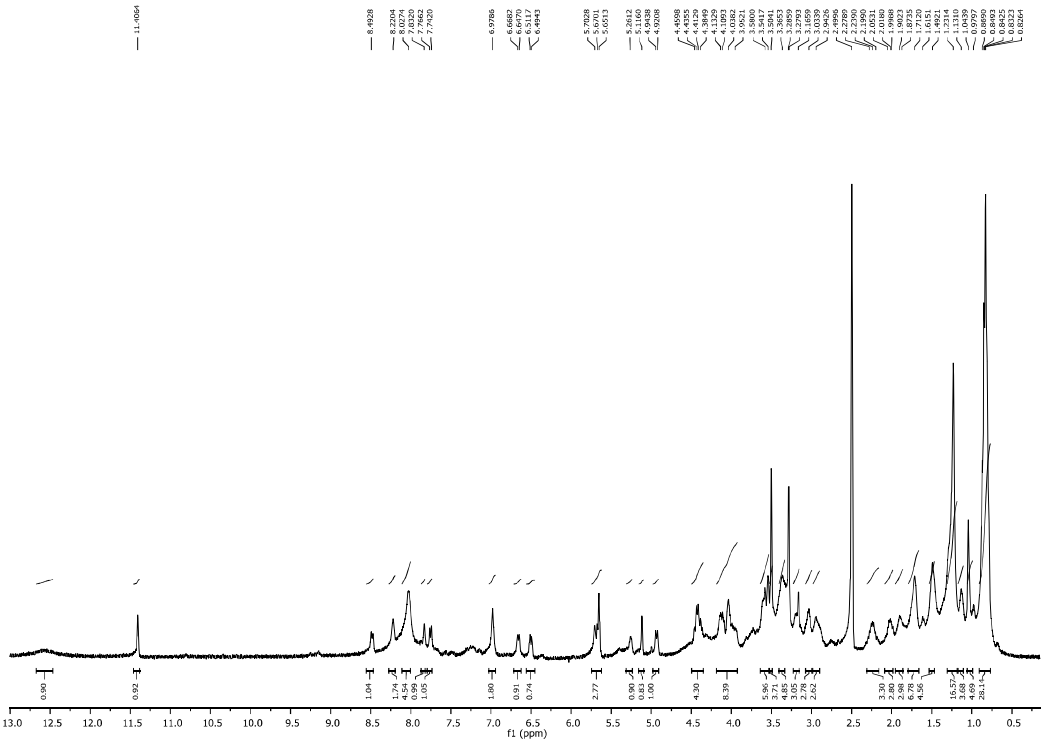


Figure S8. ^{1}H NMR spectrum (400 MHz, DMSO-d_6) of muraymycin B2 **4**.

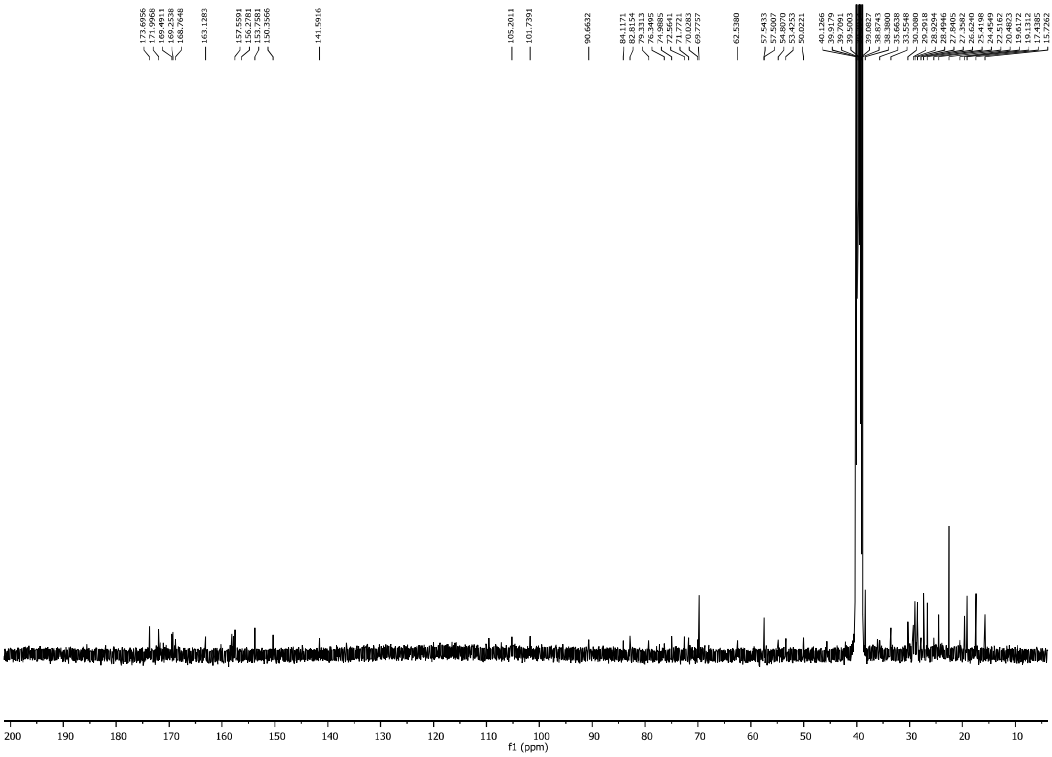


Figure S9. ^{13}C NMR spectrum (100 MHz, DMSO-d_6) of muraymycin B2 **4**.

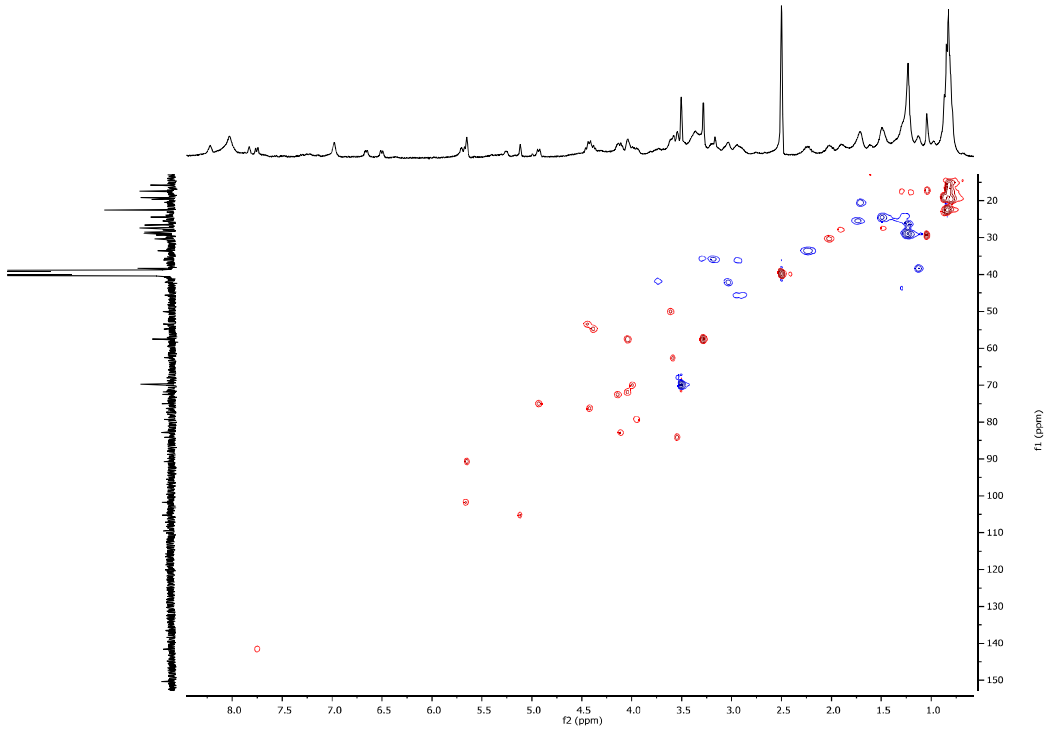


Figure S10. HSQC spectrum (400 MHz, DMSO-d_6) of muraymycin B2 **4**.

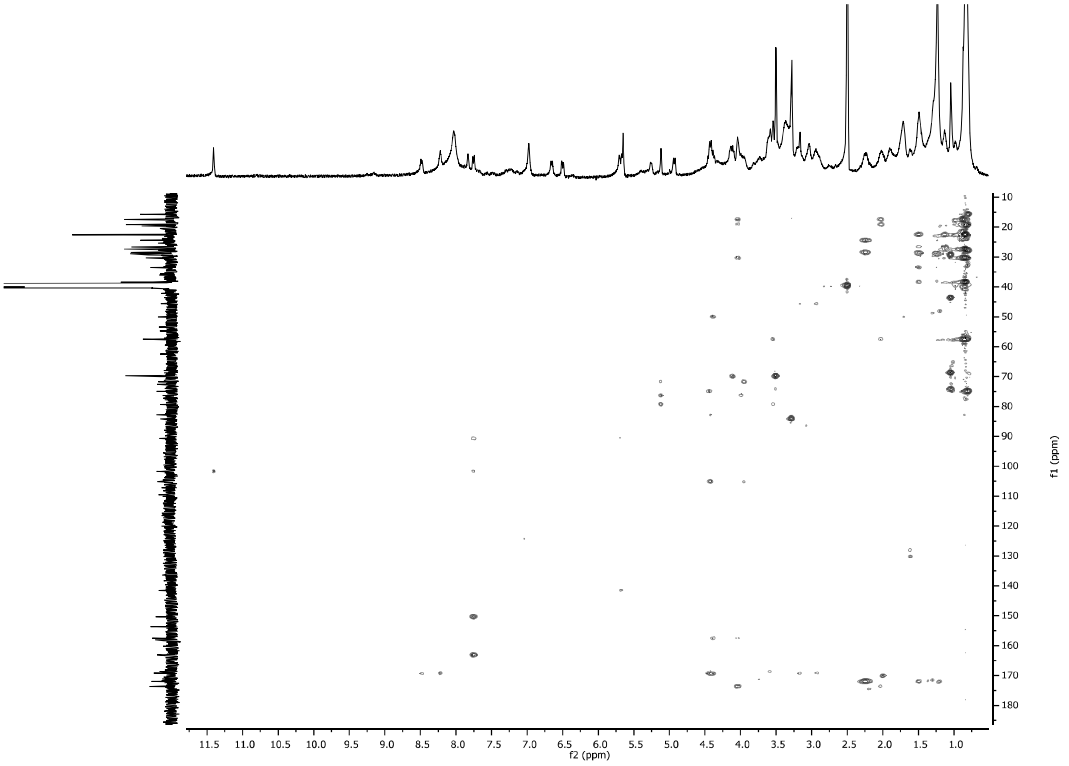


Figure S11. HMBC spectrum (400 MHz, DMSO-d_6) of muraymycin B2 **4**.

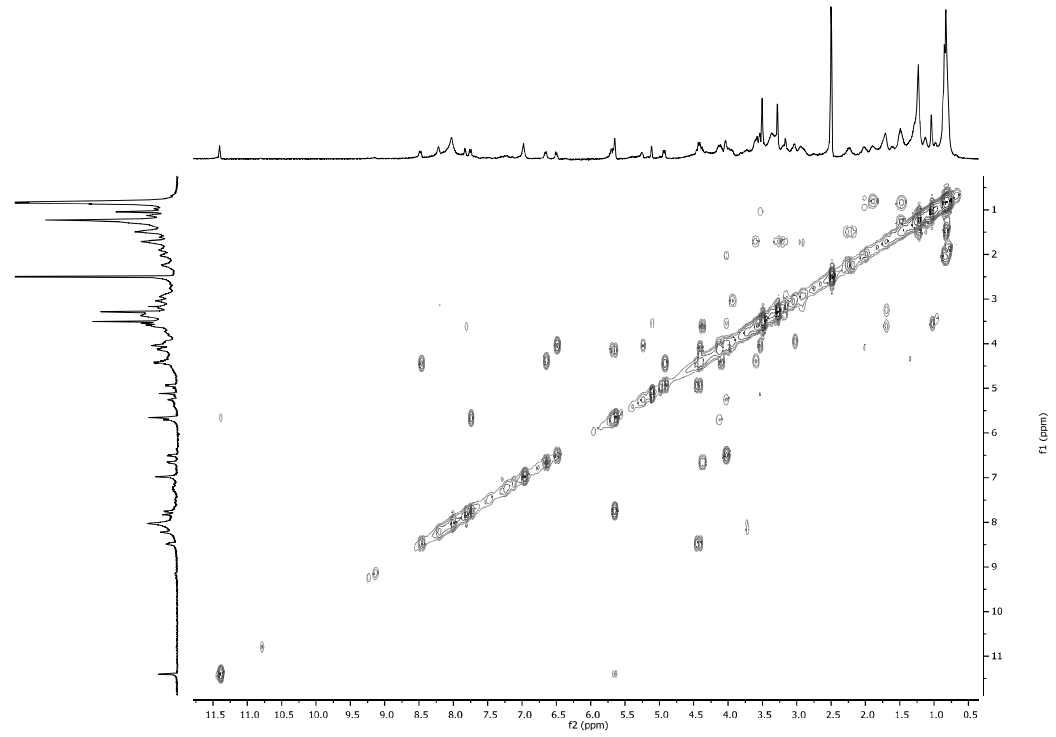


Figure S12. ^{1}H,^{1}H-COSY spectrum (400 MHz, DMSO-d_6) of muraymycin B2 **4**.

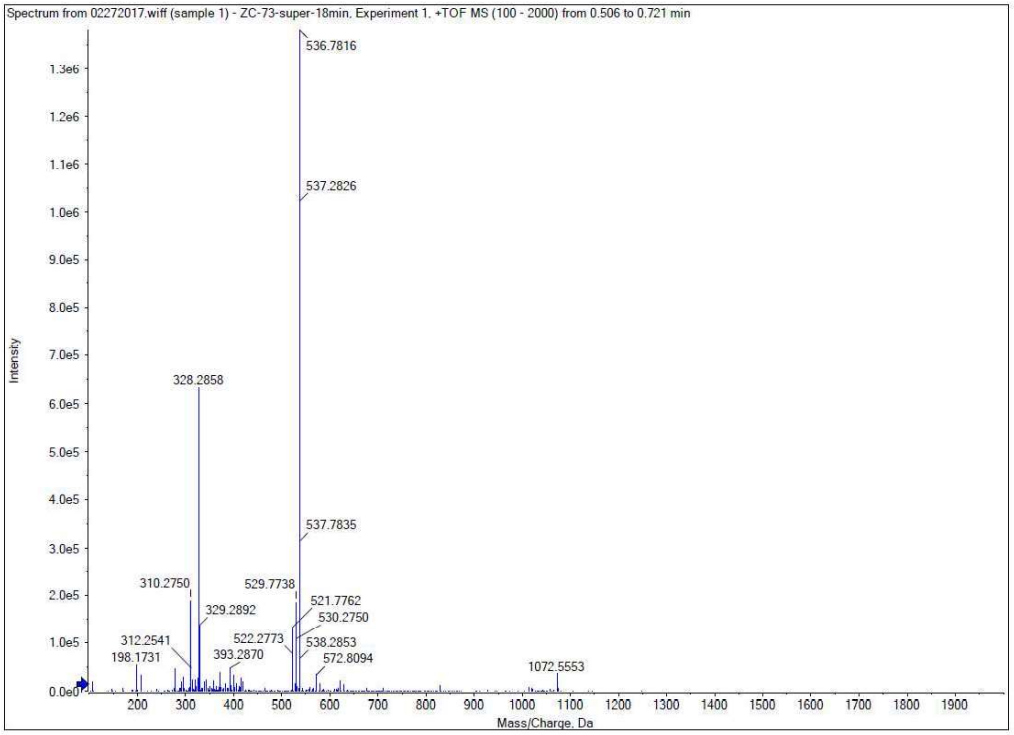


Figure S13. (+)-HR-ESI-MS (positive mode) of muraymycin B6 **5**.

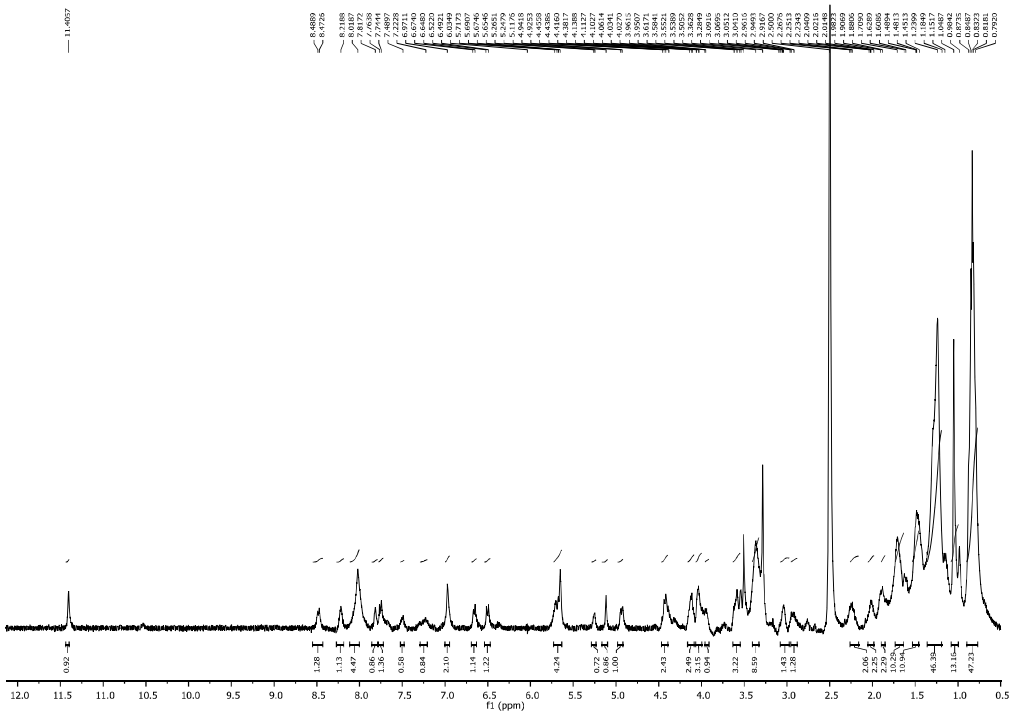


Figure S14. ^{1}H NMR spectrum (400 MHz, DMSO-d_6) of muraymycin B6 **5**.

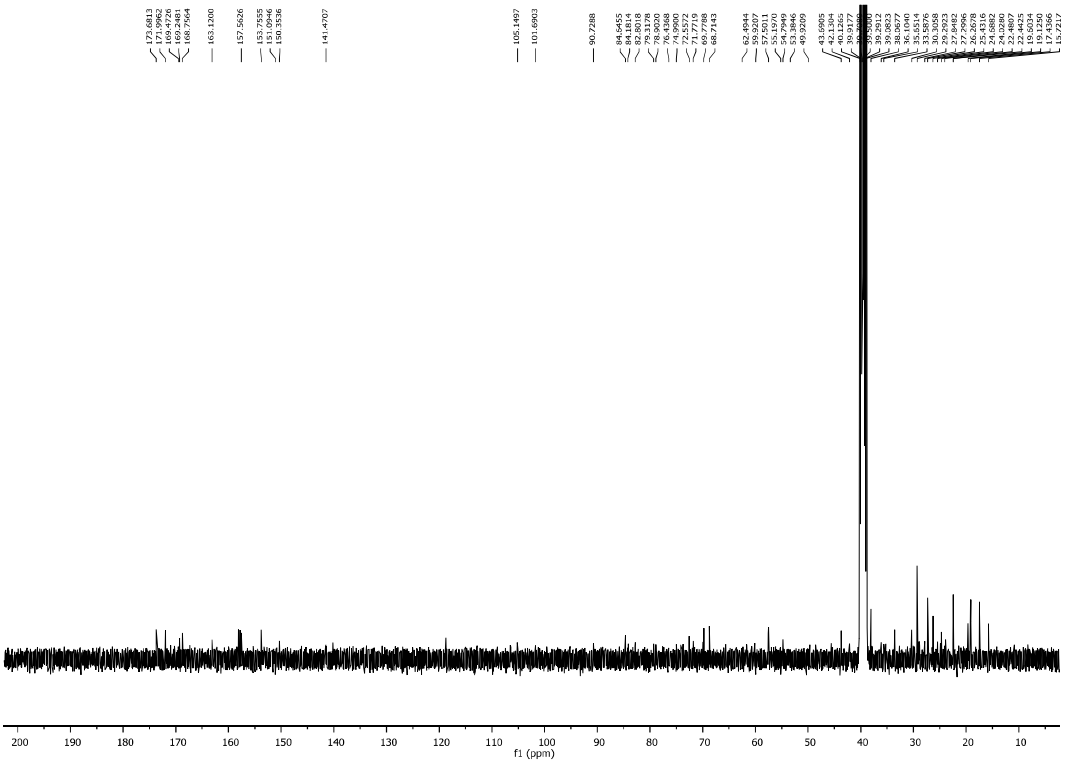


Figure S15. ^{13}C NMR spectrum (100 MHz, DMSO-d_6) of muraymycin B6 **5**.

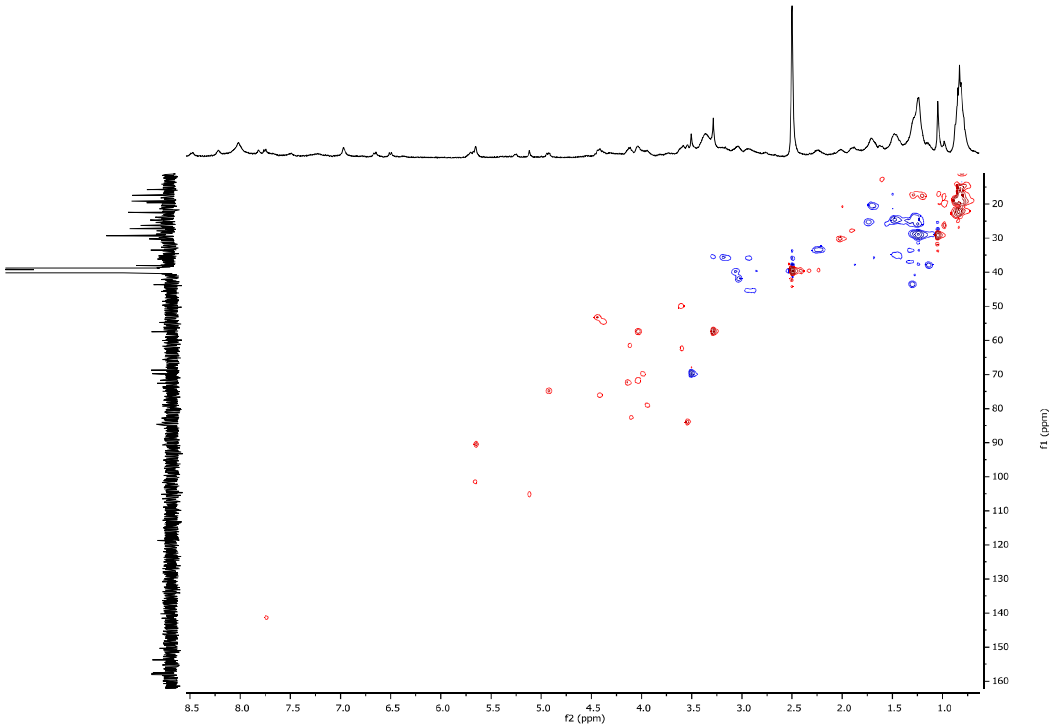


Figure S16. HSQC spectrum (400 MHz, DMSO-d_6) of muraymycin B6 **5**.

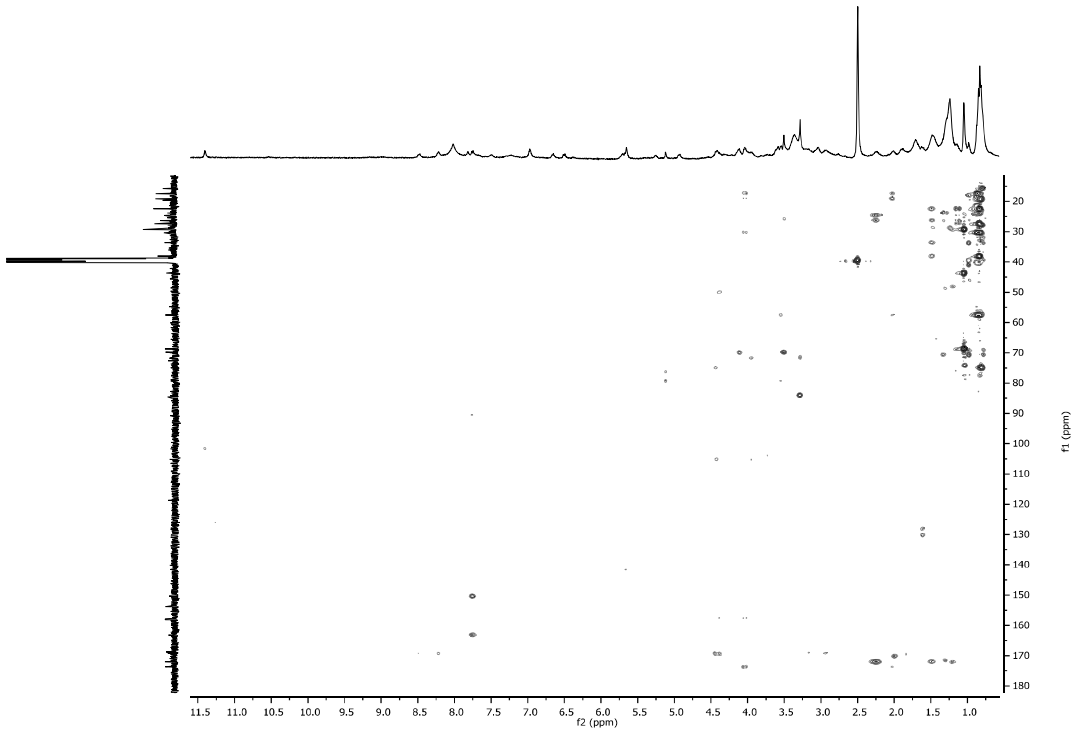


Figure S17. HMBC spectrum (400 MHz, DMSO-d_6) of muraymycin B6 **5**.

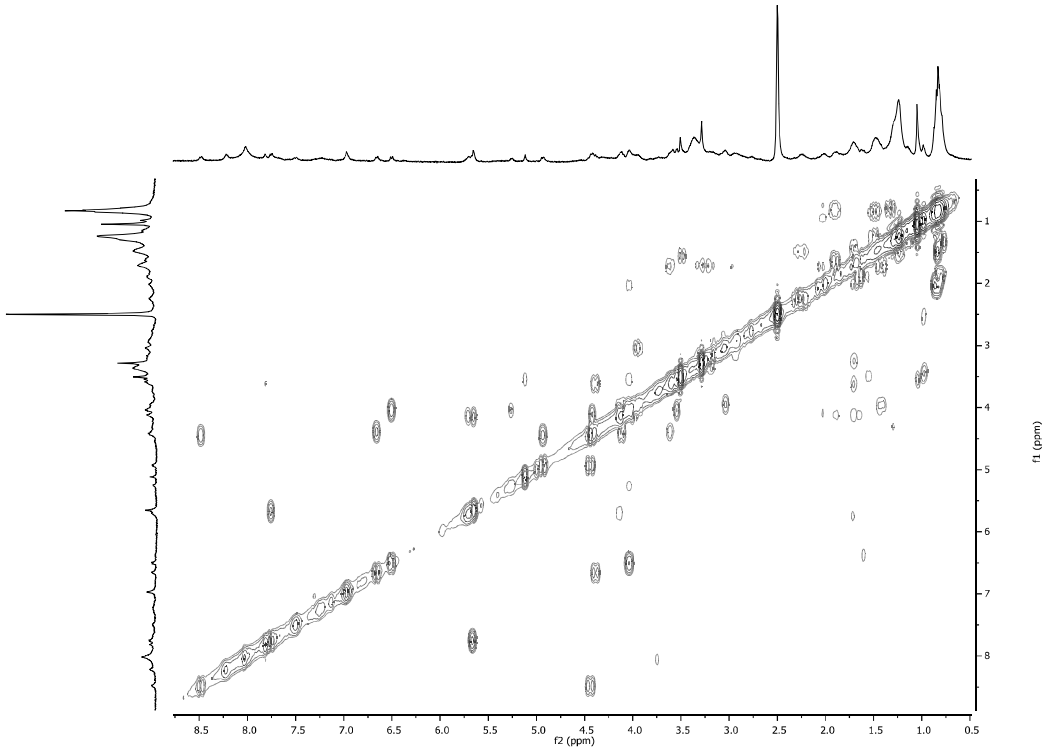


Figure S18. ^{1}H,^{1}H-COSY spectrum (400 MHz, DMSO-d_6) of muraymycin B6 **5**.

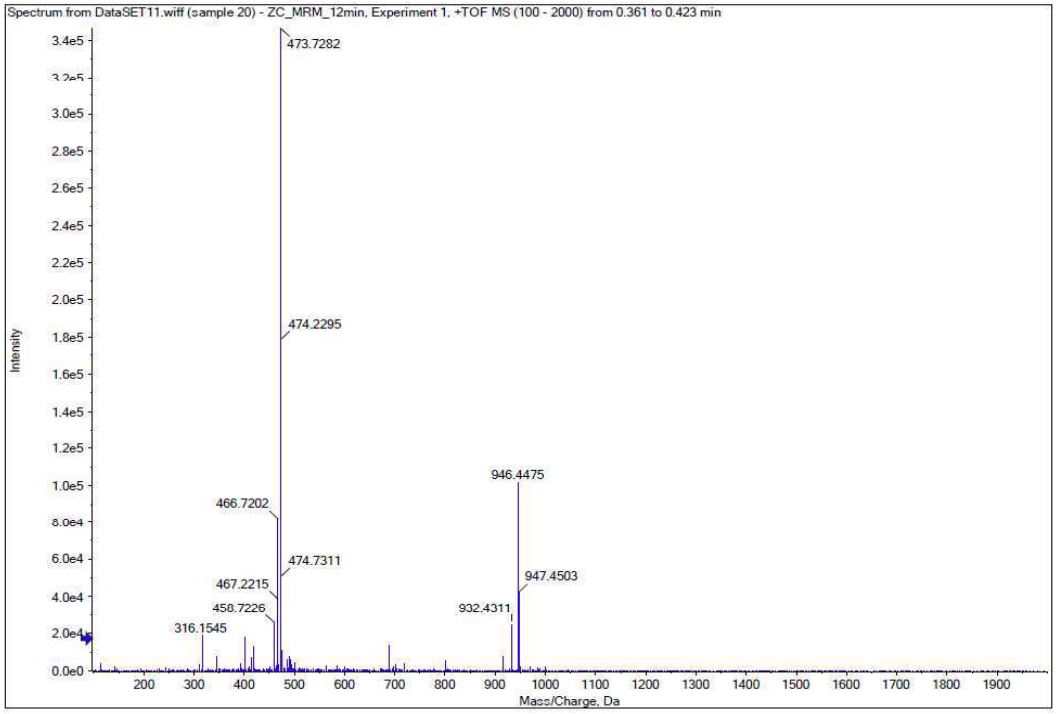


Figure S19. (+)-HR-ESI-MS (positive mode) of muraymycin C1 **6**.

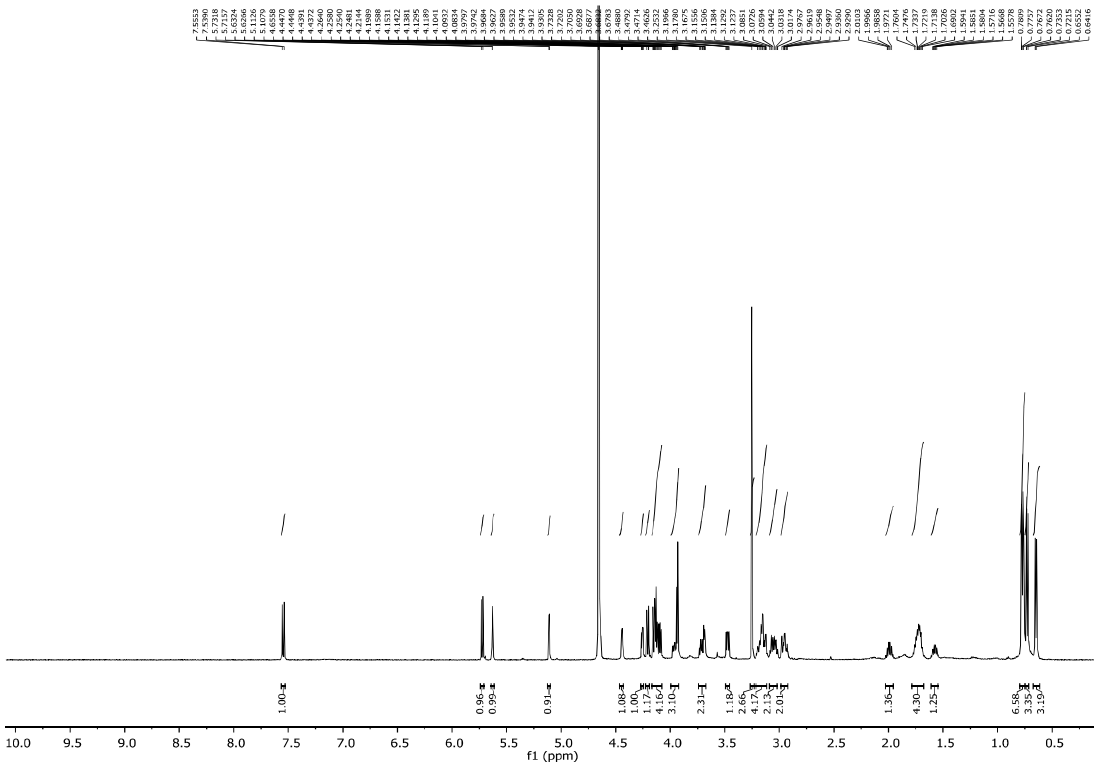


Figure S20. ^{1}H NMR spectrum (400 MHz, D_2O) of muraymycin C1 **6**.

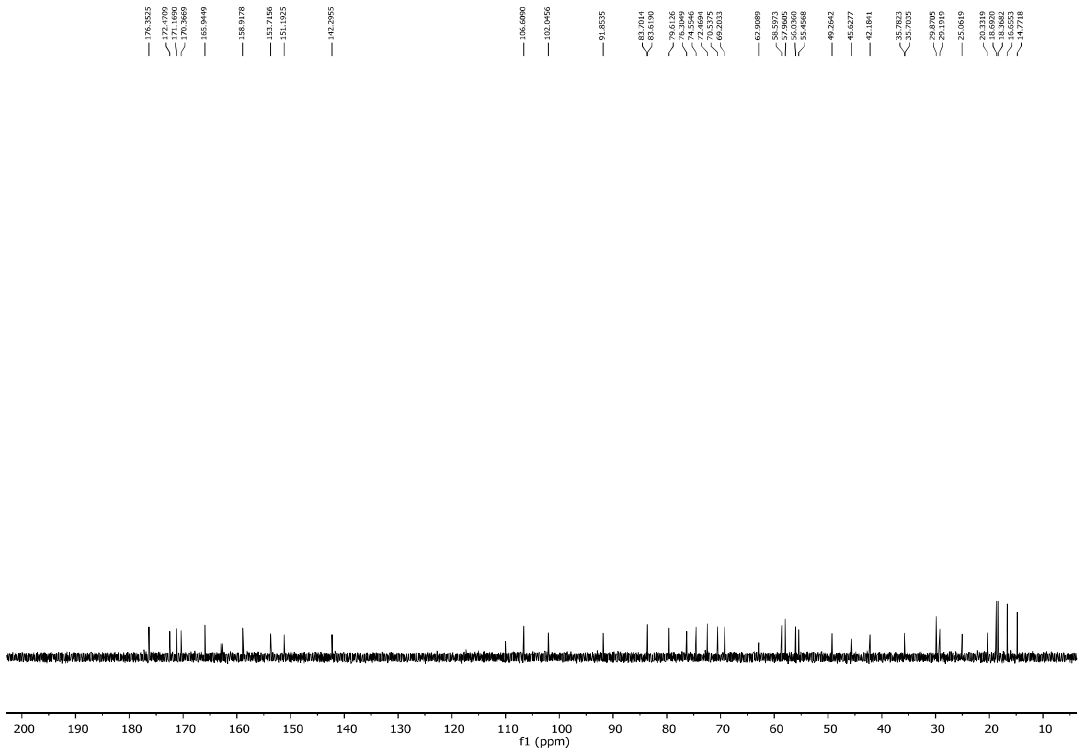


Figure S21. ^{13}C NMR spectrum (100 MHz, D_2O) of muraymycin C1 **6**.

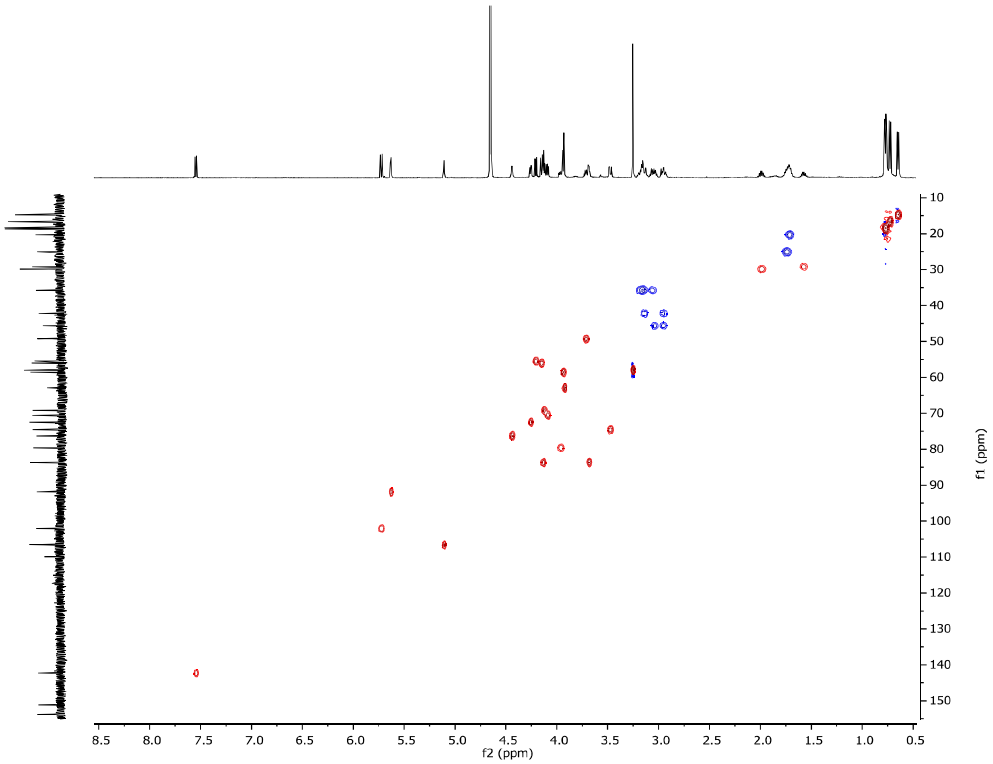


Figure S22. HSQC spectrum (400 MHz, D_2O) of muraymycin C1 **6**.

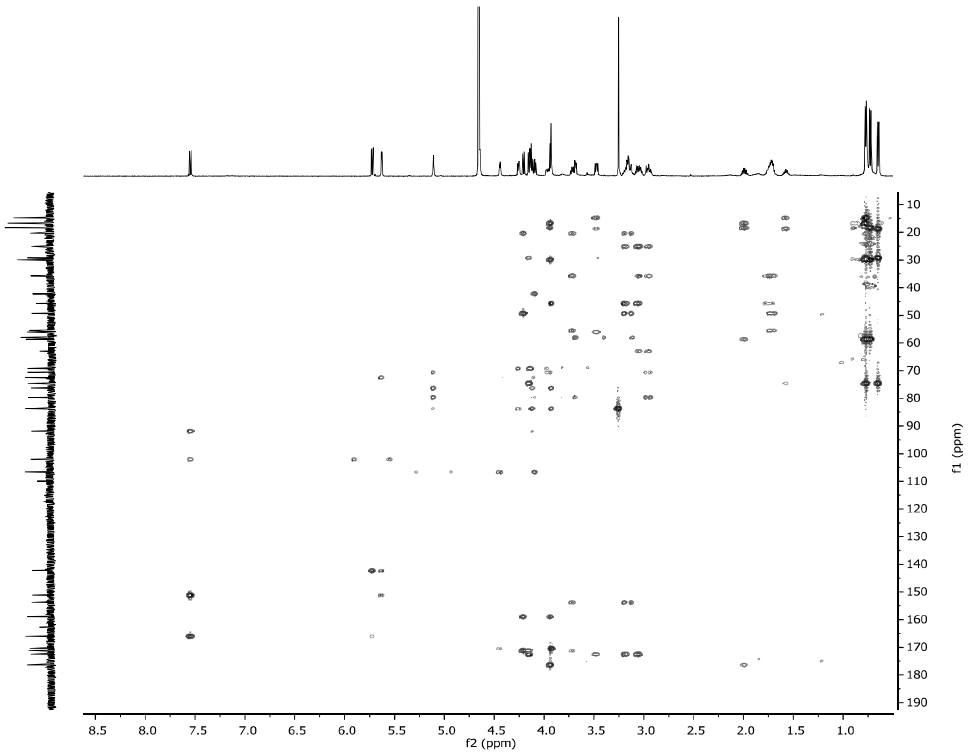


Figure S23. HMBC spectrum (400 MHz, D_2O) of muraymycin C1 **6**.

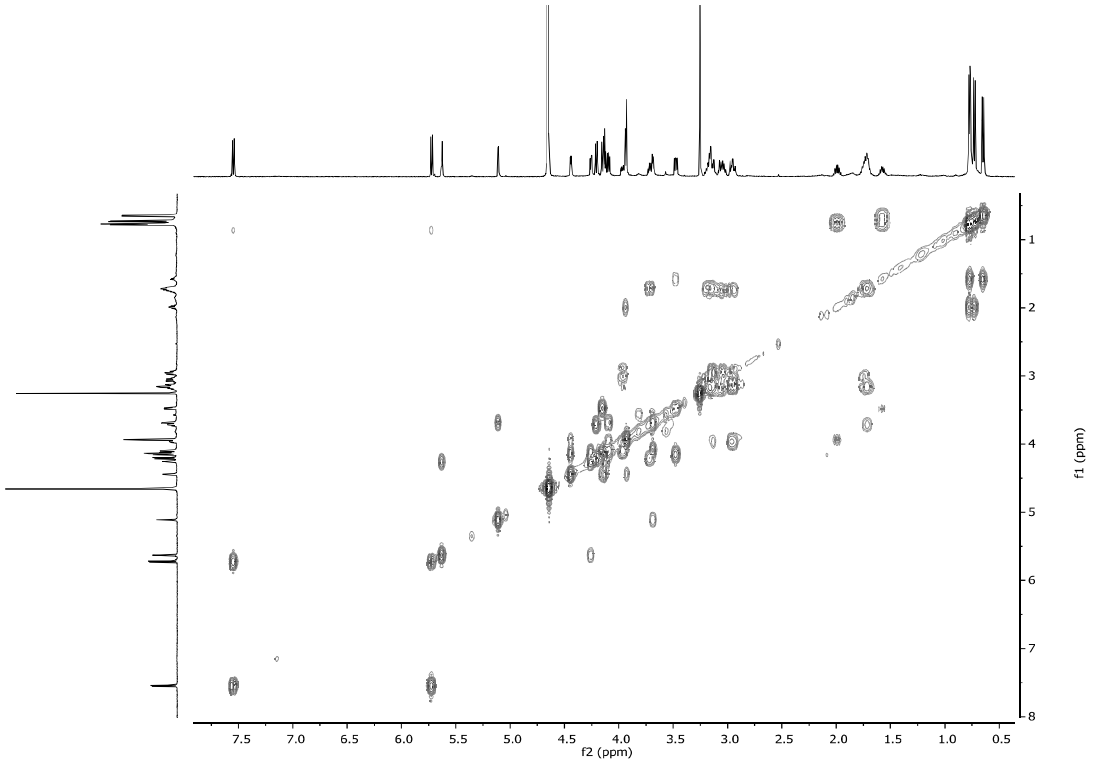


Figure S24. $^1H,^1H$-COSY spectrum (400 MHz, D_2O) of muraymycin C1 **6**.

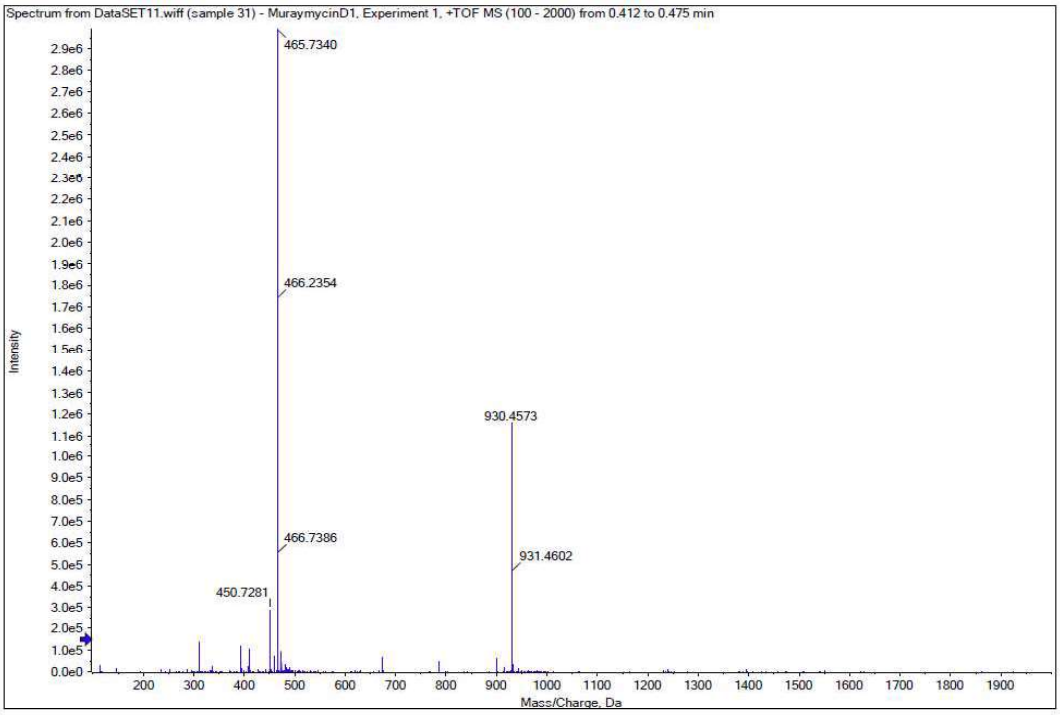


Figure S25. (+)-HR-ESI-MS (positive mode) of muraymycin D1 **7**.

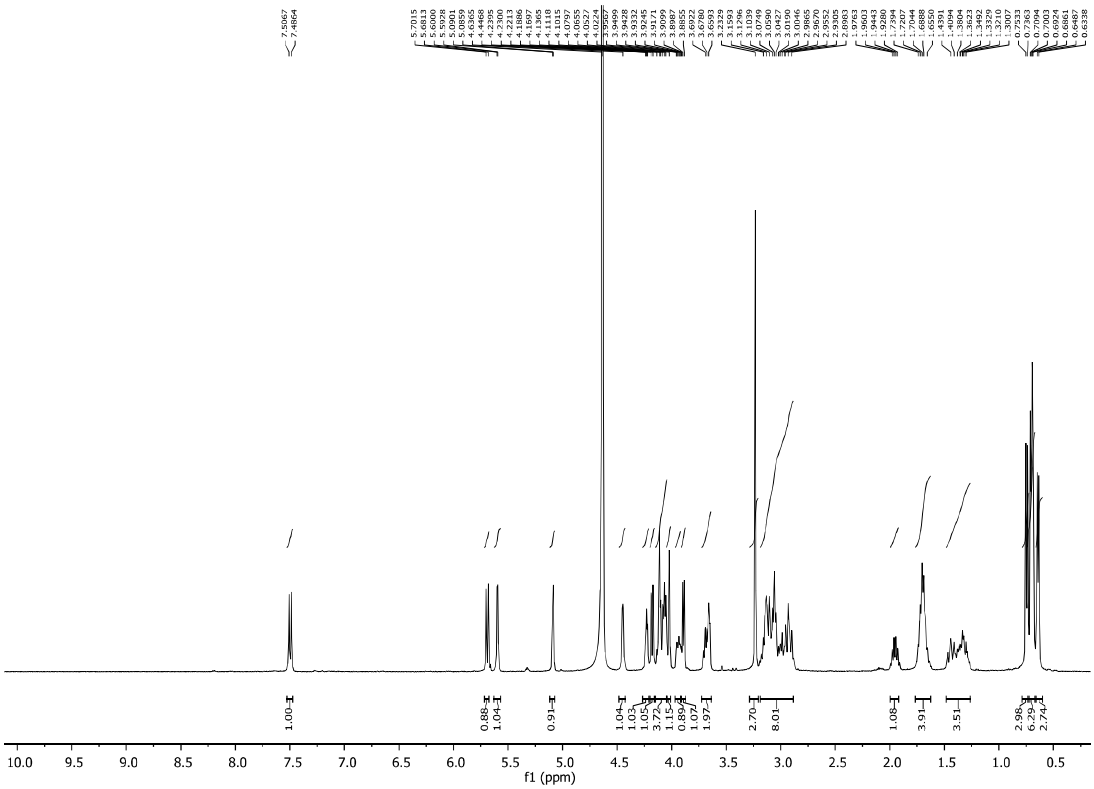


Figure S26. ^{1}H NMR spectrum (400 MHz, D_2O) of muraymycin D1 **7**.

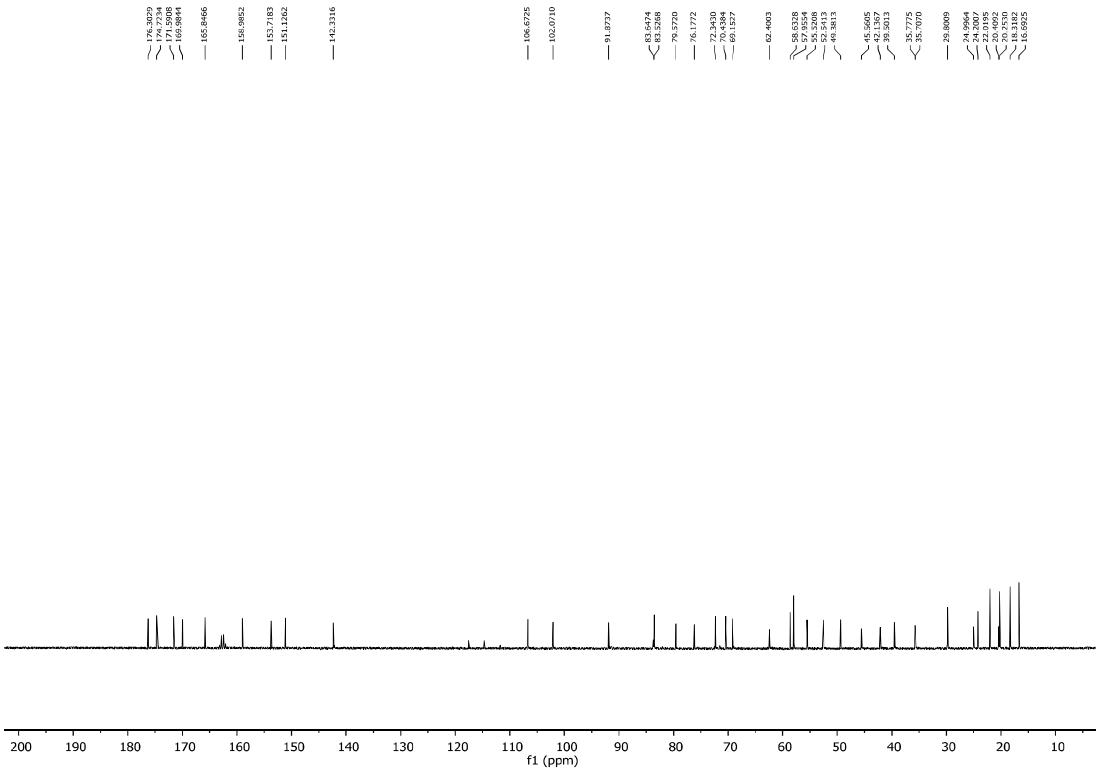


Figure S27. ^{13}C NMR spectrum (100 MHz, D_2O) of muraymycin D1 **7**.

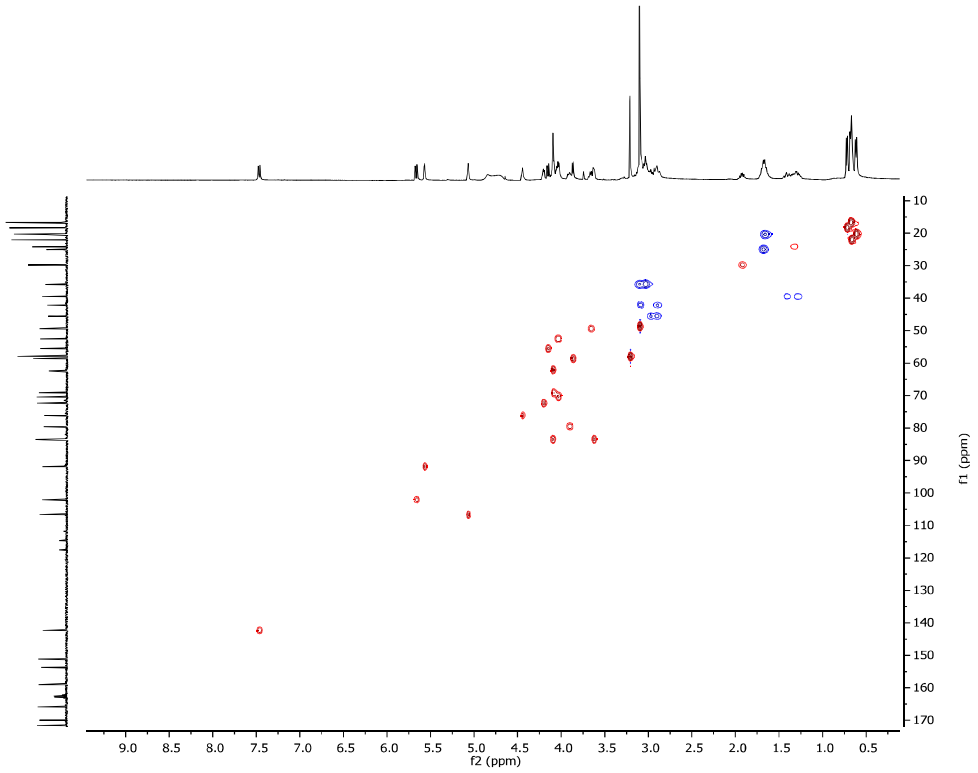


Figure S28. HSQC spectrum (400 MHz, D_2O) of muraymycin D1 **7**.

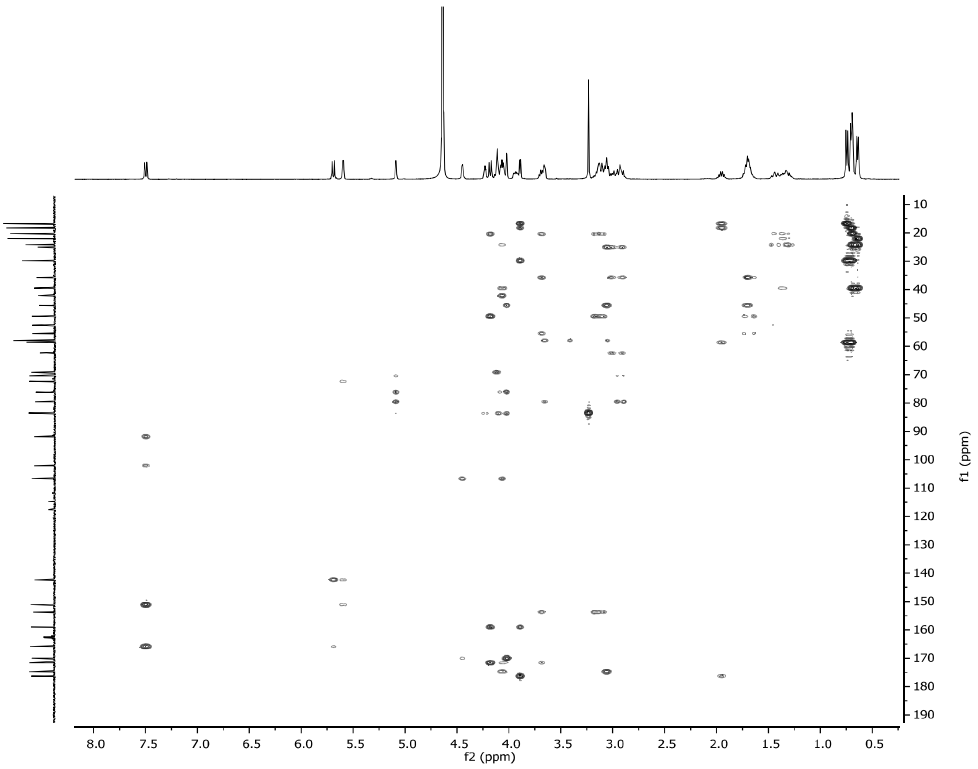


Figure S29. HMBC spectrum (400 MHz, D_2O) of muraymycin D1 **7**.

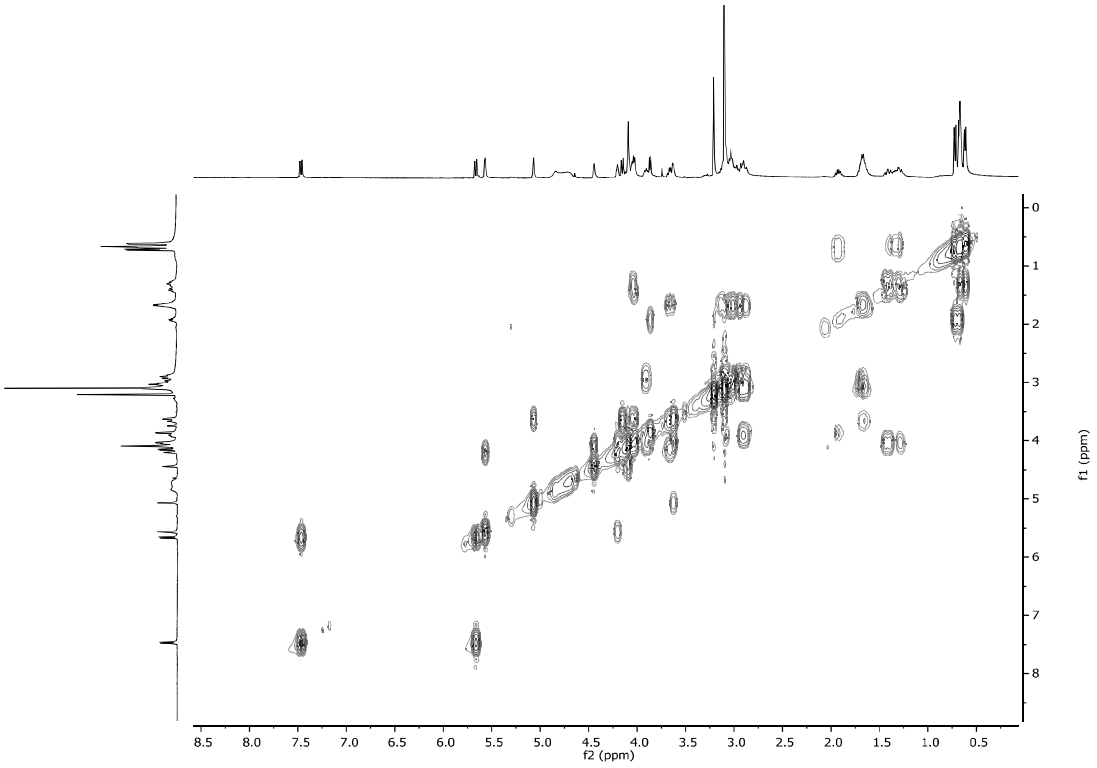


Figure S30. $^1H,^1H$-COSY spectrum (400 MHz, D_2O) of muraymycin D1 **7**.

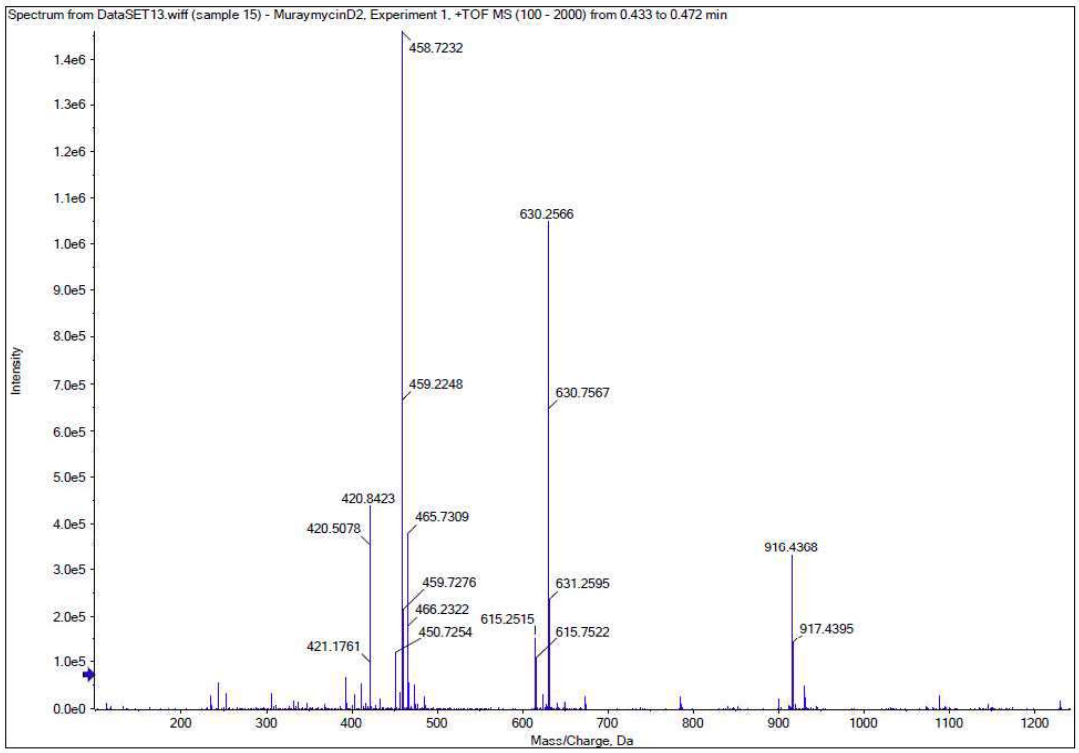


Figure S31. (+)-HR-ESI-MS (positive mode) of muraymycin D2 **8**.

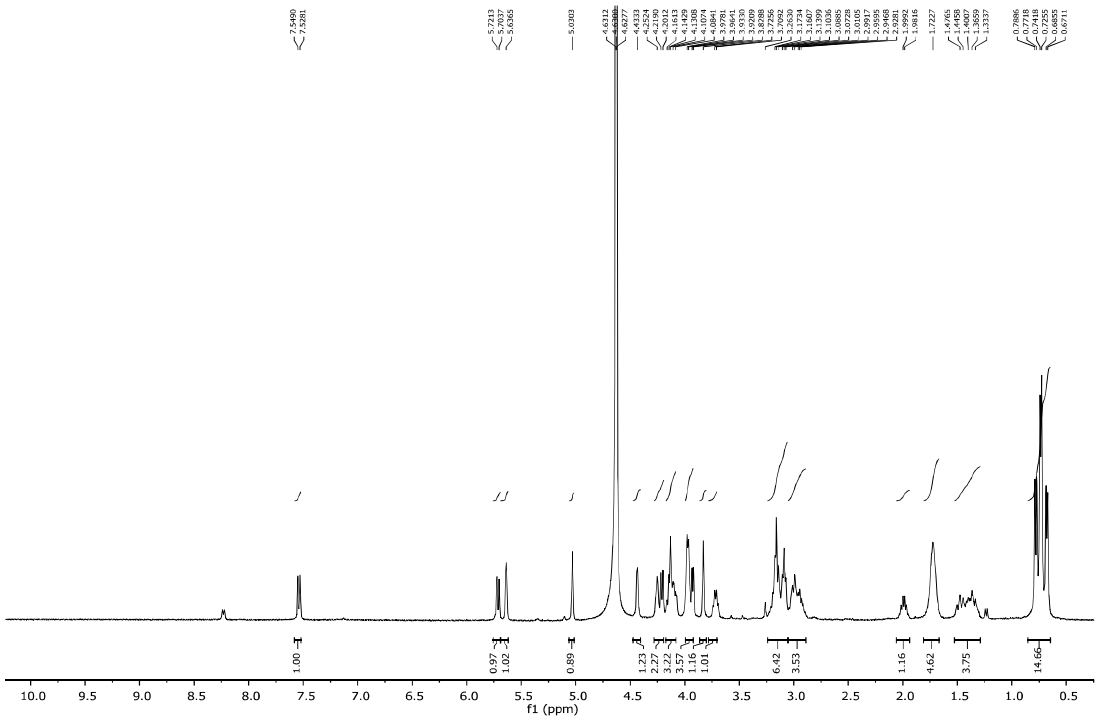


Figure S32. ^{1}H NMR spectrum (400 MHz, D_2O) of muraymycin D2 **8**.

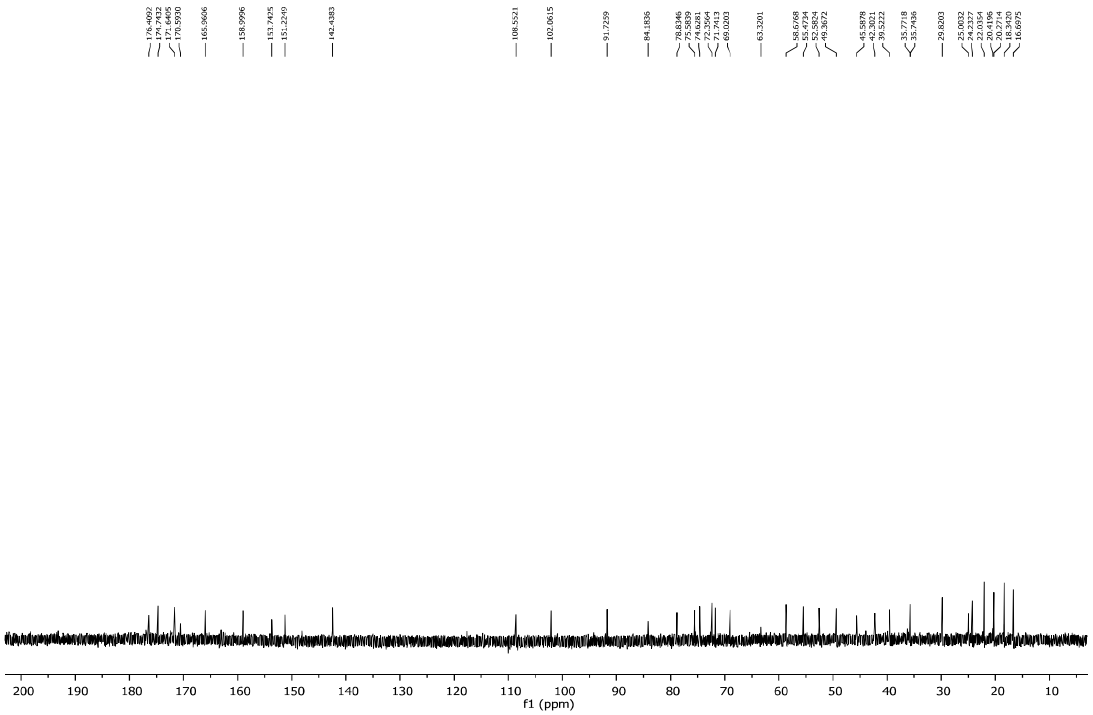


Figure S33. ^{13}C NMR spectrum (100 MHz, D_2O) of muraymycin D2 **8**.

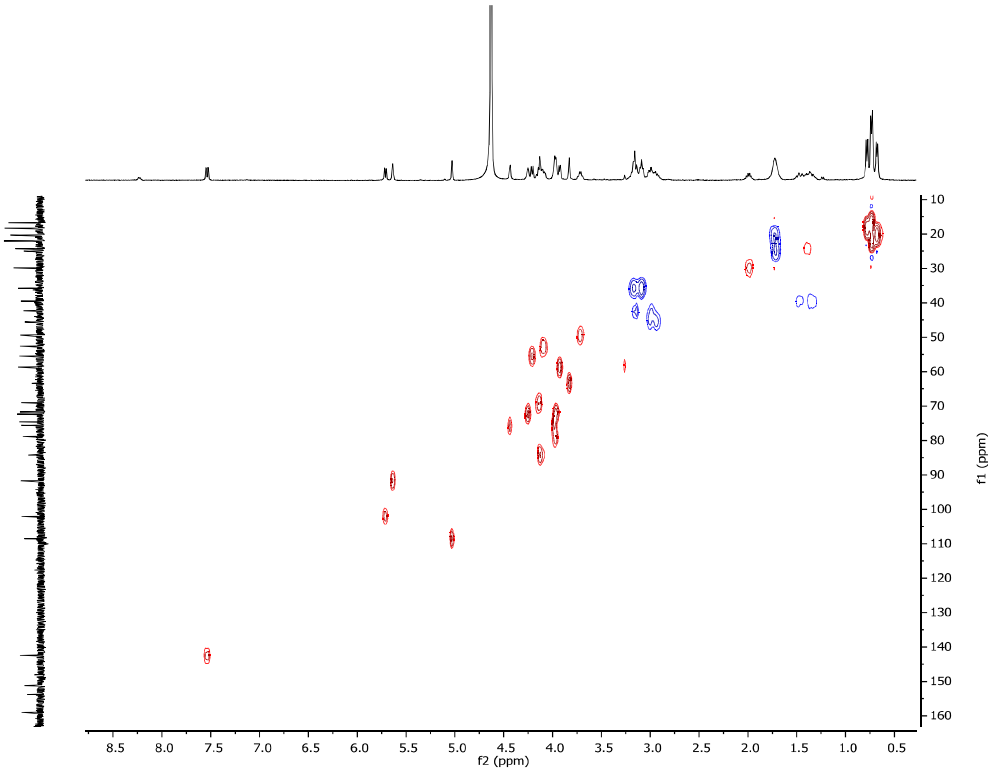


Figure S34. HSQC spectrum (400 MHz, D_2O) of muraymycin D2 **8**.

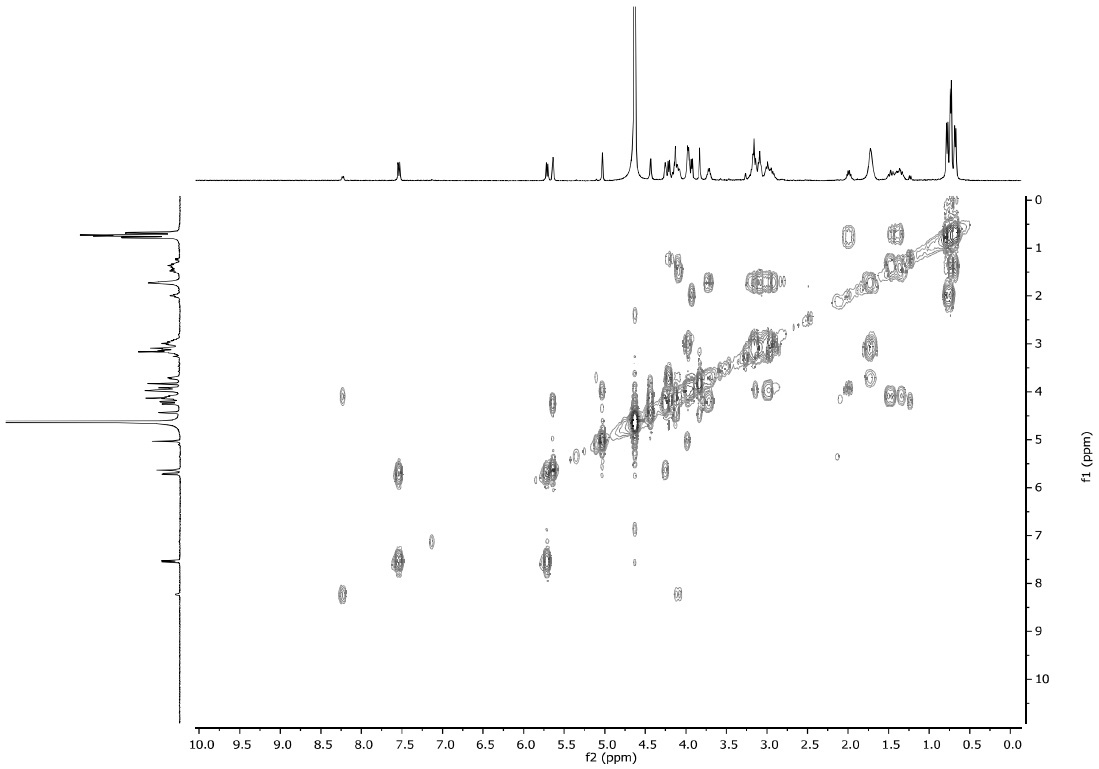


Figure S35. $^{1}H,^{1}H$-COSY spectrum (400 MHz, D_2O) of muraymycin D2 **8**.

References

S1 F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Mol. Syst. Biol.* **2011**, *7*, 539.

S2 W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso, R. Lopez, *Nucleic Acids Res.* **2015**, *43*, W580-584.

S3 H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, R. Lopez, *Nucleic Acids Res.* **2013**, *41*, W597-600.

S4 B. C. Chung, E. H. Mashalidis, T. Tanino, M. Kim, A. Matsuda, J. Hong, S. Ichikawa, S.-Y. Lee, *Nature* **2016**, *533*, 557-560.

S5 Y. Muramatsu, M. Arai, Y. Sakaida, Y. Takamatsu, S. Miyakoshi, M. Inukai, *J. Antibiot.* **2006**, *59*, 601-606.

S6 S. Wohnig, A. P. Spork, S. Koppermann, G. Mieskes, N. Gisch, R. Jahn, C. Ducho, *Chem. Eur. J.* **2016**, *22*, 17813-17819.

S7 B. C. Chung, J. Zhao, R. A. Gillespie, D.-Y. Kwon, Z. Guan, J. Hong, P. Zhou, S.-Y. Lee, *Science* **2013**, *341*, 1012-1016.

S8 J. K Hakulinen, J. Hering, G. Brändén, H. Chen, A. Snijder, M. Ek, P. Johansson, *Nat. Chem. Biol.* **2017**, *13*, 265-267.

S9 P. E. Brandish, M. K. Burnham, J. T. Lonsdale, R. Southgate, M. Inukai, T. D. H. Bugg, *J. Biol. Chem.* **1996**, *271*, 7609-7614.

S10 P. E. Brandish, K.-I. Kimura, M. Inukai, R. Southgate, J. T. Lonsdale, T. D. H. Bugg, *Antimicrob. Agents Chemother.* **1996**, *40*, 1640-1644.

S11 T. Stachyra, C. Dini, P. Ferrari, A. Bouhss, J. van Heijenoort, D. Mengin-Lecreulx, D. Blanot, J. Biton, D. Le Beller, *Antimicrob. Agents Chemother.* **2004**, *48*, 897-902.

S12 A. B. Shapiro, H. Jahić, N. Gao, L. Hajec, O. Rivin, *J. Biomol. Screening* **2012**, *17*, 662-672.

S13 A. P. Spork, M. Büschleb, O. Ries, D. Wiegmann, S. Boettcher, A. Mihalyi, T. D. H. Bugg, C. Ducho, *Chem. Eur. J.* **2014**, *20*, 15292-15297.

4.3 An open-source drug discovery platform enables ultra-large virtual screens

Christoph Gorgulla, Andras Boeszoermenyi, Zi-Fu Wang, Patrick D. Fischer, Paul Coote, Krishna M. Padmanabha Das, Yehor S. Malets, Dmytro S. Radchenko, Yurii S. Moroz, David A. Scott, Konstantin Fackeldey, Moritz Hoffmann, Iryna Iavniuk, Gerhard Wagner, Haribabu Arthanari. (2020). An open-source drug discovery platform enables ultra-large virtual screens, **Nature.** 580, 663-668.

DOI: 10.1038/s41586-020-2117-z

URL: https://www.nature.com/articles/s41586-020-2117-z

Summary:

This publication reports the development and introduction of an open-source platform that enables virtual screens on the largest possible scale so far. As of now, large-scale structure-based virtual screenings in a flexible and accessible manner on computer clusters had not been possible, despite the existence of substantial databases. The open-source program introduced in this publication, VirtualFlow, is the first highly automated platform to prepare and screen a variety of large molecule databases, using the most powerful available docking programs. Over 1 billion commercially available molecules were used in a screen against the kelch-like ECH-associated protein 1 (Keap1). Using this method, a novel nanomolar inhibitor was discovered (iKeap1, K_D = 114 nM) and experimentally verified. We were also able to identify a set of structurally diverse molecules that bind Keap1 with sub-micromolar affinity and are either capable of binding the protein in a deep tunnel binding pocket or replace its interaction partner nuclear factor erythroid-derived 2-related factor 2 (NRF2). Experimental validation of hits was performed using surface plasmon resonance (SPR), fluorescence polarization (FP), bio-layer interferometry (BLI) and nuclear magnetic resonance (NMR) techniques such as saturation transfer difference (STD) as a ligand-detected method and ^{1}H-^{13}C HMQC experiments with Keap1 selectively ^{1}H and ^{13}C labeled at the methyl groups of isoleucine,

leucine, and valine residues in otherwise perdeuterated background as a protein-detected method.

Author contributions:

As the author of this thesis, I designed and carried out the FP and NMR experiments, together with Haribabu Arthanari, Andras Boeszoermenyi, Zi-Fu Wang and Krishna M. Padmanabha Das. I expressed and purified GST-tagged Keap1 and cleaved Keap1 using natural abundance isotopes as well as site-selective ^{1}H-^{13}C-methyl labeled Keap1 at isoleucine, leucine and valine residues. With the help of Andras Boeszoermenyi, I carried out quality control of the purchased molecules from the used databases by NMR spectroscopy and liquid chromatography-mass spectrometry (LC-MS). I assisted Haribabu Arthanari, Christoph Gorgulla and Andras Boeszoermenyi in preparing the manuscript and writing the paper.

An open-source drug discovery platform enables ultra-large virtual screens

https://doi.org/10.1038/s41586-020-2117-z

Received: 5 March 2019

Accepted: 27 February 2020

Published online: 9 March 2020

Check for updates

Christoph Gorgulla[1,2,3]✉, Andras Boeszoermenyi[1,3,11], Zi-Fu Wang[1,11], Patrick D. Fischer[1,3,4], Paul W. Coote[1,3], Krishna M. Padmanabha Das[1,3], Yehor S. Malets[5,6], Dmytro S. Radchenko[5,6], Yurii S. Moroz[6,7], David A. Scott[1,3], Konstantin Fackeldey[8,9], Moritz Hoffmann[10], Iryna Iavniuk[5], Gerhard Wagner[1] & Haribabu Arthanari[1,3]✉

On average, an approved drug currently costs US$2–3 billion and takes more than 10 years to develop[1]. In part, this is due to expensive and time-consuming wet-laboratory experiments, poor initial hit compounds and the high attrition rates in the (pre-)clinical phases. Structure-based virtual screening has the potential to mitigate these problems. With structure-based virtual screening, the quality of the hits improves with the number of compounds screened[2]. However, despite the fact that large databases of compounds exist, the ability to carry out large-scale structure-based virtual screening on computer clusters in an accessible, efficient and flexible manner has remained difficult. Here we describe VirtualFlow, a highly automated and versatile open-source platform with perfect scaling behaviour that is able to prepare and efficiently screen ultra-large libraries of compounds. VirtualFlow is able to use a variety of the most powerful docking programs. Using VirtualFlow, we prepared one of the largest and freely available ready-to-dock ligand libraries, with more than 1.4 billion commercially available molecules. To demonstrate the power of VirtualFlow, we screened more than 1 billion compounds and identified a set of structurally diverse molecules that bind to KEAP1 with submicromolar affinity. One of the lead inhibitors (iKeap1) engages KEAP1 with nanomolar affinity (dissociation constant (K_d) = 114 nM) and disrupts the interaction between KEAP1 and the transcription factor NRF2. This illustrates the potential of VirtualFlow to access vast regions of the chemical space and identify molecules that bind with high affinity to target proteins.

Repeated optimization of lead compounds and late-stage failure of drug candidates are the primary causes of the longer development times and increased costs of drug development. Improving the quality of the initial lead compounds would minimize these lead optimization cycles and result in drug candidates that enter (pre-)clinical phases with greater specificity and higher affinity. Virtual screening to identify molecules that bind to a specified site on a receptor protein has become an important part of the drug discovery pipeline[2–5].

Current virtual screening paradigms routinely sample only a small fraction, on the order of 10^6–10^7 molecules, of the total chemical space of small organic compounds that are suitable for drug discovery, which have been estimated to encompass more than 10^{60} molecules[6].

However, the scale of a virtual screen is of central importance because the more compounds that are screened, the lower the rate of false positives and the more favourable the quality of the lead compounds (for example, molecules that bind with higher affinity). It was recently shown experimentally that ultra-large scale screening improves the rate of true positives[2]. Here we derived a probabilistic model of the true-positive rate as a function of the number of compounds screened; analysis of our ultra-large screen confirms that the docking score of the highest-scoring compounds improves with the scale. Increasing the scale of a virtual screen can improve the quality of initial hits in two distinct ways: by identifying hits with tighter binding affinity, which can result in lowered dosages and fewer off-target effects and by discovering compounds with more favourable pharmacokinetic and/or less inherent cytotoxic properties.

To increase the number of compounds evaluated in a virtual screen by orders of magnitude and make it accessible to any researcher, there is a need for a platform that can integrate all of the tasks in the virtual screening process. Such a platform should ideally scale linearly with the number of CPUs, efficiently handle billions of files, minimize input and output load, run robustly (for example, skip incorrectly encoded ligands, resist temporary input and output problems and resume after unexpected termination), run on any type of computing cluster

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Harvard University, Boston, MA, USA. [2]Department of Physics, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, USA. [3]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbrücken, Germany. [5]Enamine, Kyiv, Ukraine. [6]National Taras Shevchenko University of Kyiv, Kyiv, Ukraine. [7]Chemspace, Kyiv, Ukraine. [8]Zuse Institute Berlin, Berlin, Germany. [9]Institute of Mathematics, Technical University Berlin, Berlin, Germany. [10]Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany. [11]These authors contributed equally: Andras Boeszoermenyi, Zi-Fu Wang. ✉e-mail: cgorgulla@g.harvard.edu; hari@hms.harvard.edu

(including cloud platforms) and be user friendly and easy to use for non-computational scientists. Furthermore, to provide flexibility, a structure-based virtual screening platform should be able to interface with a variety of docking programs, support both rigid and flexible receptor docking, test multiple docking scenarios in a single workflow, allow for consensus and ensemble docking, and carry out multiple replicas of the same docking scenario. Lastly, to democratize access, facilitate widespread use and catalyse further development, such a platform would need to be open source.

With these requirements in mind, we designed VirtualFlow, an open-source platform that is able to screen chemical space on a large scale. Screening 1 billion compounds on a single processor core, with an average docking time of 15 s per ligand, would take approximately 475 years. By contrast, VirtualFlow can dock 1 billion compounds in approximately 2 weeks when leveraging 10,000 CPU cores simultaneously. Such high-performance computing facilities are available to researchers through several potential sources, including computer clusters of local institutes, national super-computing centres or cloud computing platforms.

Targeting KEAP1 using VirtualFlow

To test the advantages of ultra-large-scale in silico screening and the performance of the VirtualFlow platform, we decided to target the challenging and therapeutically relevant protein–protein interaction between nuclear factor erythroid-derived 2-related factor 2 (NRF2) and Kelch-like ECH-associated protein 1 (KEAP1). NRF2 is a master regulator of cellular resistance to oxidative stress and cellular repair[7]. Under unstressed conditions, NRF2 is sequestered by KEAP1—an E3 ubiquitin ligase substrate adaptor—and targeted for degradation[8]. However, upon oxidative stress, reactive oxidants dissociate NRF2 from KEAP1 and NRF2 translocates to the nucleus to activate its transcriptional program of approximately 250 genes[9]. The NRF2–KEAP1 pathway is critical in protecting the cell under oxidative stress and inflammation and is implicated in a number of diseases[10]. There are ten drugs that target KEAP1 that are in clinical trials and nine more that are at the preclinical stage[10]. Using VirtualFlow, we screened approximately 1.3 billion compounds (around 1 billion compounds from the Enamine REAL library and about 330 million compounds from the ZINC library) against the NRF2 interaction interface on KEAP1. First, however, we describe the salient features of VirtualFlow and its scalability.

Characteristic features of VirtualFlow

One of the key features of VirtualFlow is its linear scaling behaviour ($O(N)$, where N is the number of cores) with respect to the number of CPUs and nodes used. VirtualFlow can run on computer clusters operated with any of the major resource managers (SLURM (https://slurm.schedmd.com), Moab/TORQUE (http://www.adaptivecomputing.com), PBS (http://www.pbspro.org), LSF (https://www.ibm.com/us-en/marketplace/hpc-workload-management) and SGE (http://gridscheduler.sourceforge.net)), and compatibility with additional job schedulers can be easily added. Thus VirtualFlow is also ideally configured for cloud computing platforms such as Amazon's Web Services (AWS), Microsoft's Azure and Google's Cloud Platform (GCP). VirtualFlow is able to run autonomously from the first to the last ligand in the screening pipeline, a feature that is facilitated by automatic submission of new batch system jobs. The workflow can be monitored and controlled during runtime. The VirtualFlow package consists of two applications that work seamlessly together: the VFLP (VirtuaFlow for Ligand Preparation) module—which prepares small molecules for screening—and the VFVS (VirtualFlow for Virtual Screening) module, which executes the virtual screening procedures (Fig. 1). The separation of ligand preparation and virtual screening is desirable because the same ready-to-dock ligand library can be used in any number of VFVS virtual screens.

The VFLP module

VFLP prepares ligand databases by converting them from the SMILES format to any desired target format (for example, the PDBQT format, which is required by many of the AutoDock-based docking programs). VFLP uses the JChem package of ChemAxon as well as Open Babel to desalt ligands, neutralize them, generate (one or multiple) tautomeric states, compute protonation states at specific pH values, calculate three-dimensional coordinates and convert the molecules into desired target formats (Extended Data Fig. 2). The output file formats that are currently supported by VFLP are shown in Supplementary Table 7.

Preparation of the Enamine REAL library

Commercially available compounds constitute the most interesting subset of the chemical space, as these compounds can be readily purchased. One of the largest vendor libraries that is currently available is the REAL library of Enamine, which contains approximately 1.4 billion make-on-demand compounds (as of October 2019, the ZINC 15 database contained 1.46 billion compounds, but only provided 630 million molecules in a ready-to-dock format). We used VFLP to convert the approximately 1.4 billion compounds of the REAL library into PDBQT format (Methods) and have made this library freely available on the VirtualFlow homepage, accessible through a graphical interface (Supplementary Fig. 5). The entire database has a six-dimensional lattice architecture, the general concept of which was modelled after the ZINC 15 database[11], in which each dimension corresponds to a physico-chemical property of the compounds (molecular mass, partition coefficient, number of hydrogen bond donors and acceptors, number of rotatable bonds and the topological polar surface area). The preparation of ligands using VFLP is a one-time effort.

The VFVS module

To set up a virtual screen with VFVS, a set of docking scenarios is specified by the user. Docking scenarios are defined by the choice of the external docking program, the receptor structure and the docking parameters (which include the pre-defined docking surface on the receptor, residues on the receptor that are allowed to be flexible during docking and the rigor of the docking routine). VirtualFlow currently supports the following docking programs: AutoDock Vina[12], QuickVina 2[13], Smina (which includes the Vinardo and AutoDock 4 scoring functions)[14], AutoDockFR[15], QuickVina-W[5], VinaXB[16] and Vina-Carb[17]. By supporting an array of different docking programs, VFVS can be used in a variety of cases by leveraging the unique advantages of each program. VFVS allows the specification of multiple docking scenarios to be carried out for each ligand, enabling consensus docking procedures, as well as ensemble docking procedures[18,19]. VirtualFlow is also amenable to the integration of other docking programs that are currently not a part of this platform.

Scaling behaviour of VFVS

To measure the scaling behaviour of VFVS, we measured the performance on two local clusters, LC1 and LC2. On LC1, we used 18,000 CPU cores of heterogeneous composition (different models of Intel Xeon and AMD Opteron processors), whereas on LC2 we used up to 30,000 Intel Xeon 8268 cores. The scaling behaviour was effectively linear in both cases (that is, $O(N)$) (Extended Data Fig. 3a). These results meet theoretical expectations as there is no direct communication between the processes running in parallel, which is key to perfect scaling behaviour without bounds. The independence of its parallel processes means that VirtualFlow is expected to scale linearly even if millions of cores are used. We also tested the performance of the platform on cloud-based computing systems, including GCP and AWS. On the GCP, we carried

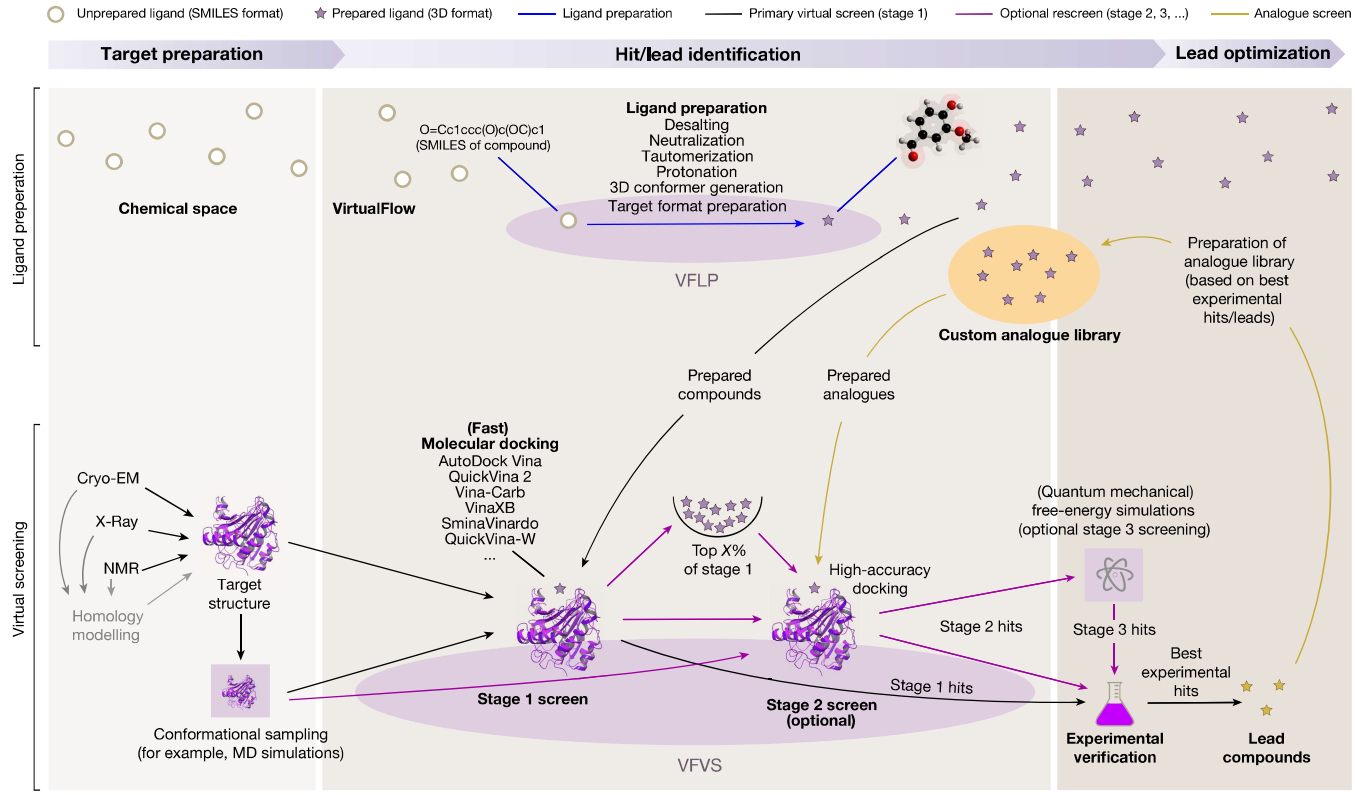


Fig. 1 | Application of VirtualFlow to the drug discovery process. Before the screening can begin, the target structure—which is generally obtained through X-ray, NMR, cryo-electron microscopy (cryo-EM) or homology-modelling studies—needs to be prepared. The preparation step can include molecular dynamics (MD) simulations to obtain one or more relevant conformations of the target protein. Once the structure is prepared, it can be used to identify new hit compounds by virtual screening-based approaches. The two independent modules of VirtualFlow, VFLP and VFVS, are designed to aid virtual screening-based needs. VFLP prepares (blue arrow) the desired chemical space into ready-to-dock ligand libraries, which can subsequently be used by VFVS during the virtual screen. The virtual screen usually consists of a primary virtual screen (stage 1), but can also be implemented using the multi-stage setting (pink arrows), which can include protein side chain flexibility or inclusion of multiple protein conformations (for example, from molecular dynamics simulations or NMR structures). In addition, complimentary software packages can be used during multi-staged screening (for example, to carry out quantum mechanics-based free-energy simulations as a final step) to improve the true hit rate and estimated binding affinities. After the virtual screening procedure, experimental verification can be carried out to identify true binders. Promising molecules (lead compounds) can be further optimized by creating custom analogue libraries and screening them again with VirtualFlow (yellow arrows), followed again by experimental verification of the hits.

out large-scale benchmarks with up to 160,000 CPUs and, despite this massive scaling in CPU volume, VirtualFlow still exhibited linear scaling behaviour (Extended Data Fig. 3a). A typical high-throughput screen, such as the one described in this study, of 1 billion compounds will take around 15 h on the GCP with 160,000 CPUs. The linear scaling behaviour over a large number of CPUs makes VirtualFlow suitable for the highly anticipated exascale computing age.

Multi-staged virtual screens with VFVS

VFVS can also be used to organize virtual screens with multiple stages to substantially increase the quality of the results (Fig. 2a). In the multi-staging approach, several virtual screens are executed in succession. The number of top-scoring compounds that advance from one stage to the next is successively reduced, with concomitant increases in docking accuracy and computational cost.

Using VFVS to screen 1.3 billion ligands

To validate the performance of VFVS, we screened a virtual library of 1.3 billion commercially available compounds (around 330 million compounds from the ZINC 15 database[11] and approximately 1 billion compounds from the Enamine REAL library) against KEAP1. It should be noted that there is some overlap of compounds between the two libraries.

This effort was completed in around 4 weeks, using on average approximately 8,000 cores on a heterogeneous Linux cluster.

To illustrate the benefit of an ultra-large-scale screen, we chose subsets of the ligands (0.1, 1, 10 and 100 million compounds) randomly from the around 1 billion compound screen of the REAL library and considered the scores of the top 50 compounds (Fig. 2b). As the scale of the screen increased, the average docking score increased, thus improving the chances of identifying molecules with higher binding affinity. This in turn leads to higher true hit rates and tighter experimental binding affinities, as predicted by a probabilistic model that we derived (Supplementary Information section D) and that has been experimentally demonstrated previously[2].

To demonstrate the use of VirtualFlow in a multi-staging context, we subjected the top approximately 3 million ranking compounds from the primary virtual screen to a rescoring procedure (Fig. 2a). In stage 2, the 13 residues of KEAP1 at the NRF2 interaction interface were allowed to be flexible. This flexibility accounted for the movement and/or dynamics of the amino acids at the binding interface, which are not captured by a static structure. In the rescoring procedure we used two different docking programs (Smina Vinardo and AutoDock Vina), and two replicas of each docking scenario were carried out to

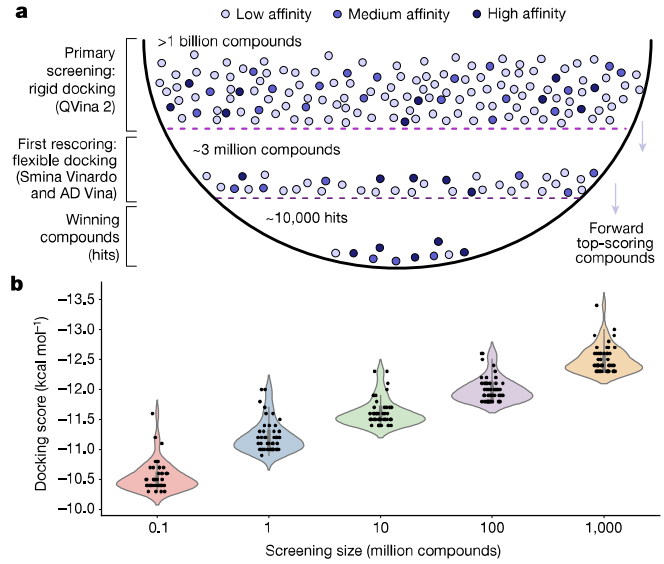


Fig. 2 | Schematic overview of the multi-stage screen and benefits of ultra-large-scale screens. a, In stage 1, approximately 1.3 billion compounds were screened with the fast docking program QuickVina 2 at the lowest accuracy level. In stage 2, 13 residues of the receptor were allowed to be flexible and the top approximately 3 million scoring compounds were rescored with higher accuracy using AutoDock Vina and Smina Vinardo. **b**, Violin plots of the docking scores of the top 50 molecules from virtual screens that targetted KEAP1 with different starting library sizes (0.1, 1, 10, 100 and 1,000 million ligands). To mimic virtual screens of smaller sizes, we randomly chose a subset of the ligands from the around 1 billion compound screen of the REAL library and considered the scores of the top 50 compounds. The docking score is an estimation of the free energy of binding (in kcal mol^{-1}) and, therefore, the more negative the value is, the tighter the hit binds to the target. The distributions show that the docking score of the top 50 compounds improves with the scale of the screen. The procedure was repeated independently five times with similar results.

further increase the conformational space sampled during the docking runs. The necessity of multi-stage screening depends on the target of choice and the computational resources available, but this type of virtual screen is particularly useful in cases in which dynamics at the docking interface is expected to have a marked role.

Experimental validation

From the in silico screen described above, we chose 590 hits for experimental validation. Of these, 492 compounds were from the top 0.03% of the stage 2 screen and 98 compounds were from the top 0.0001% of the stage 1 screen. Hits from stage 1 were ordered to compare the true hit rate between stage 1 and stage 2 hits, in a multi-stage setting. In addition to the ranking by docking score, the choice of these compounds was based on factors such as drug-likeness, availability for procurement, ligand efficiency and chemical diversity. We used four established biophysical methods: fluorescence polarization, surface plasmon resonance (SPR), nuclear magnetic resonance (NMR) and bio-layer interferometry (BLI) to experimentally validate the binding of the VirtualFlow-derived hits to KEAP1. Fluorescence polarization and SPR were initially used in a high-throughput manner (level 1) to detect binding and the compounds identified here were subsequently validated with more scrutiny in a detailed and low-throughput assay (level 2). We used a recombinantly expressed and purified Kelch domain of mouse KEAP1, henceforth referred to as KEAP1. An overview of the experimental verification workflow is graphically represented in Extended Data Fig. 6, and a detailed description of the experimental procedure is provided in the Methods. Of these four biophysical methods, fluorescence polarization and BLI detect the ability of the hits to displace the NRF2 peptide from KEAP1, identifying hits that we refer to as displacers. SPR and NMR directly detect binding of hits to KEAP1, identifying hits referred to as binders. VirtualFlow identifies molecules that potentially bind to the NRF2 interaction interface on KEAP1; however, the in silico screen is performed using KEAP1 alone, in the absence of NRF2. The NRF2-binding surface on KEAP1 is part of the deep pocket/tunnel of the KEAP1 β-barrel with NRF2 binding to the entrance of this tunnel. However, some compounds could bind more tightly by inserting deep into this central tunnel of KEAP1 rather than embracing the surface like the NRF2 peptide and/or bind to parts of KEAP1 that are not engaged by NRF2. Such binders might not effectively disrupt the interaction with NRF2, while still engaging KEAP1 with high affinity (Extended Data Fig. 9). In our experimental validation, we identified both displacers and binders.

Out of the cherry-picked 590 compounds, 69 were confirmed to bind to KEAP1 by level 2 SPR. To assess the ability of the compounds to displace the NRF2 peptide, we used the fluorescence polarization assay. Ten compounds were confirmed to be displacers with a half-maximum inhibitory concentration (IC_{50}) < 60 μM by fluorescence polarization and all of the compounds were identified as a binder by level 2 SPR. Interference by autofluorescence from the compounds themselves prevented the analysis of some of the compounds by fluorescence polarization. Thus, we used BLI as an orthogonal assay to assess the ability of the compounds to displace NRF2. The binding affinity of the NRF2 peptide to KEAP1 as measured by BLI was 1.86 nM, which is similar to that measured by fluorescence polarization, which was 3.67 nM (Extended Data Fig. 4). Of the 69 SPR level 2 active compounds, 40 compounds were able to disrupt the NRF2–KEAP1 interaction as observed by BLI. Of these 40 compounds, 16 were able to displace NRF2 from KEAP1 at a compound concentration of 20 μM, while all 40 compounds could do so at 100 μM. Using BLI, we were able to identify displacers that were missed by fluorescence polarization due to autofluorescence (an example is shown in Extended Data Fig. 8). We tested all of the SPR level 2 active compounds for potential aggregation by dynamic light scattering. We identified seven compounds that aggregated in the dynamic light scattering assay and hence were not considered for further evaluation (Supplementary Table 5). On the basis of the SPR level 2 and the fluorescence polarization level 2 binding data, we selected 23 compounds for SPR level 3 experiments to determine the binding affinity. All 23 compounds had affinities in the low micromolar to nanomolar range, and 12 compounds had submicromolar K_d values. From these 23 compounds, we tested the binding of 6 compounds (iKeap1, iKeap2, iKeap7, iKeap8, iKeap9 and iKeap22) to KEAP1 by a suite of NMR-based ligand-detection experiments. Out of these six compounds, five are displacers and one (iKeap9) is a binder. These six compounds were selected on the basis of the solubility constraints of the NMR experiments, the SPR K_d value and/or their ability to displace the peptide. We used differential line broadening, saturation transfer difference, Car–Purcell–Meiboom–Gill-based transverse relaxation time experiments, and protein-observed ^{1}H–^{13}C heteronuclear multiple-quantum correlation experiments to confirm the binding of the compounds to KEAP1. The ligand-detection NMR experiments confirmed that all six of the tested SPR level 3 active compounds bind to KEAP1 (Fig. 3 and Extended Data Figs. 7, 8). Protein-detection ^{1}H–^{13}C heteronuclear multiple-quantum correlation experiments show that the compounds engage KEAP1 in a specific manner, at the targeted NRF2-binding site. In the absence of resonance assignments, we use the fact that the compounds perturb a subset of KEAP1 resonances affected by the addition of the NRF2 peptide as evidence for competitive binding. These compounds are shown in Supplementary Figs. 1–3. Details about the other active compounds are provided in Supplementary Information section B.

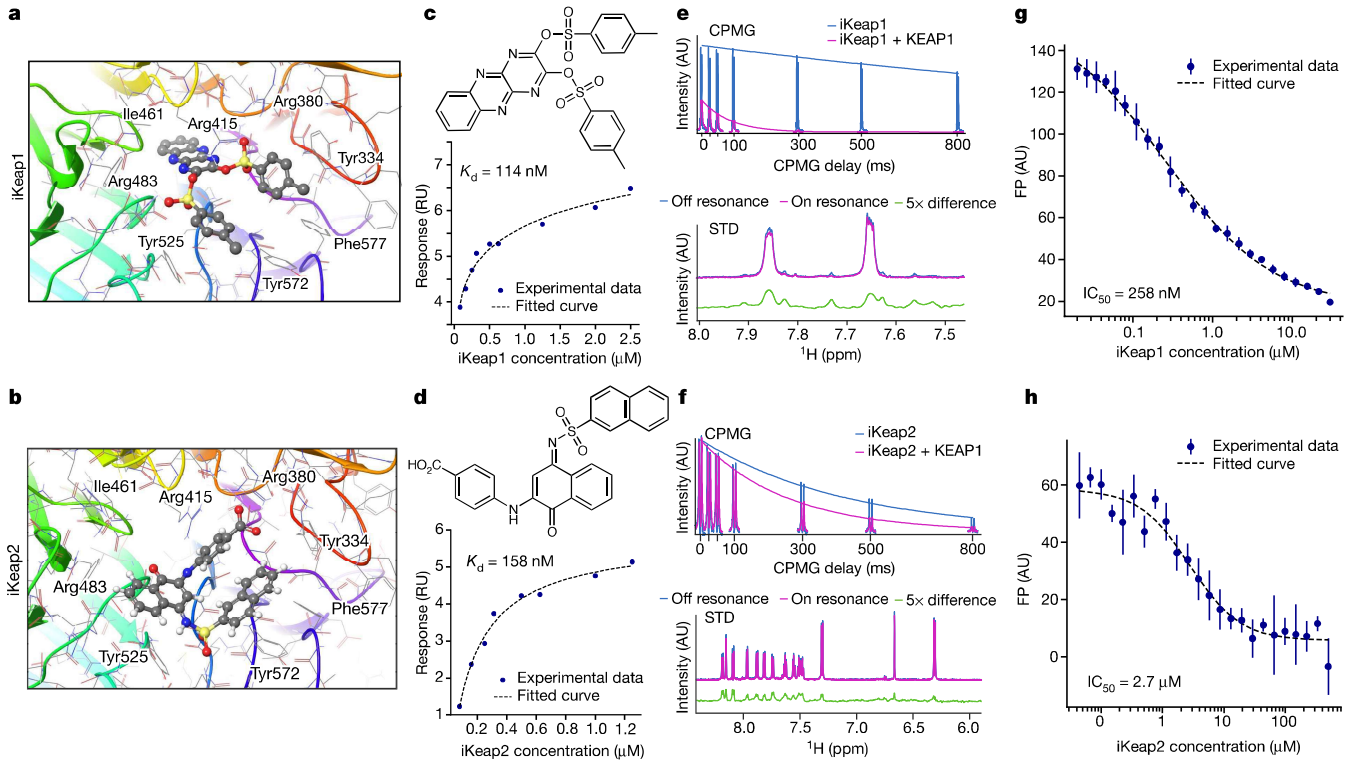


Fig. 3 | Docking poses and experimental verification of two hit compounds (iKeap1 and iKeap2). a, **b**, The docking poses of iKeap1 (**a**) and iKeap2 (**b**) were obtained from stage 2 of the virtual screening. **c**, **d**, SPR steady-state binding curves are shown for iKeap1 (**c**) and iKeap2 (**d**), showing clear binding with a K_d in the nanomolar range. We show one representative dataset from three independent experiments with similar results. RU, resonance units. **e**, **f**, Ligand-detection NMR experiments using Car–Purcell–Meiboom–Gill (CPMG)–R_2, which measures the transverse relaxation rate (R_2) of the ligand and saturation transfer difference (STD) NMR of iKeap1 (**e**) and iKeap2 (**f**) confirm the binding of the two compounds to KEAP1. AU, arbitrary units. **g**, **h**, Keap1 (**g**) and iKeap2 (**h**) were also functional in the fluorescence polarization (FP) assay, confirming that the compounds displace the peptide. The fluorescence polarization data shown here are from three technical replicates and the curve was fitted to the average value of the three technical replicates. Data are mean ± s.d. of the individual data points. The fluorescence polarization experiment was repeated independently twice with similar results and one representative result is shown here.

iKeap1 and iKeap2 are two of our top hits, and both these compounds are able to displace the NRF2 peptide from KEAP1. Both of the compounds are predicted to engage the NRF2-binding pocket of KEAP1, located at the entrance to the tunnel formed by the β-barrel (Fig. 3a, b). In comparison to iKeap2, iKeap1 descends deeper into this central tunnel of KEAP1. SPR results showed that iKeap1 and iKeap2 bind to KEAP1 with a binding affinity of 114 nM and 158 nM, respectively (Fig. 3c, d). NMR-based ligand-detection experiments confirmed that both iKeap1 and iKeap2 directly bind to KEAP1 (Fig. 3e, f). Fluorescence polarization assays showed that iKeap1 displaces NRF2 peptide with an IC_{50} of 258 nM and iKeap2 displaces the NRF2 peptide with an IC_{50} of 2.7 μM (Fig. 3g, h). BLI measurements additionally confirmed that both iKeap1 and iKeap2 are able to displace the NRF2 peptide from KEAP1. iKeap1 is similar to a previously reported naphthalene-based compound[20] with a lower IC_{50} (IC_{50} = 2.7 μM; compound C17 in Supplementary Table 1 and Extended Data Fig. 5d). C17 was identified as the best hit in a high-throughput screen of 270,000 compounds in a previous study[20].

Here, we also highlight iKeap7, which has the highest affinity as assayed by SPR (K_d = 15 nM) and displaces the NRF2 peptide with an IC_{50} of 38.2 μM (Extended Data Fig. 8). It should be noted that of the 14 hits described in the manuscript, only two hits, namely iKeap2 and iKeap7, contain pan-assay interference (PAINS) substructures. However, we performed a series of orthogonal binding assays, which confirmed that iKeap2 and iKeap7 are not experimental false positives. For details and discussion on how we verified that our experimental results were not affected by PAINS, see Supplementary Information section B.

Typically, protein–protein interactions have a larger interaction interface compared to the interface of the active site of an enzyme. Hence the in silico screen can identify binders that either partially overlap with the binding site of the interacting protein, such as iKeap9 (Extended Data Fig. 7) or those that bind in a manner that energetically favours the formation of the protein–protein complex. Examples of the latter, referred to as glues, have been previously described in the literature[21].

An open-source platform

To allow VirtualFlow to be used widely and develop dynamically, it is set up as a free and open-source project. GPU support is planned for the future and will be incorporated into VirtualFlow both natively and through external docking programs such as Gnina[22]. We encourage scientists to join the project and contribute to improving existing features, adding new features and functionality. The primary homepage of VirtualFlow, which provides additional resources, can be accessed at https://www.virtual-flow.org.

Outlook

VFVS can be used to search extremely large regions of the chemical space, which is the key to identifying promising small-molecule binders. VFVS is able to accomplish this by efficiently using high-performance computing resources, which will continue to increase in availability and power in the years to come, and novel virtual screening databases such

as the Chemical Universe Databases, which contain billions to trillions of compounds, that are still waiting to be explored[23].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2117-z.

1. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).
2. Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
3. Zhang, S., Kumar, K., Jiang, X., Wallqvist, A. & Reifman, J. DOVIS: an implementation for high-throughput virtual screening using AutoDock. *BMC Bioinformatics* **9**, 126 (2008).
4. Jiang, X., Kumar, K., Hu, X., Wallqvist, A. & Reifman, J. DOVIS 2.0: an efficient and easy to use parallel virtual screening tool based on AutoDock 4.0. *Chem. Cent. J.* **2**, 18 (2008).
5. Hassan, N. M., Alhossary, A. A., Mu, Y. & Kwoh, C.-K. Protein-ligand blind docking using QuickVina-W with inter-process spatio-temporal integration. *Sci. Rep.* **7**, 15451 (2017).
6. Bohacek, R. S., McMartin, C. & Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50 (1996).
7. Yonchuk, J. G. et al. Characterization of the potent, selective Nrf2 activator, 3-(pyridin-3-ylsulfonyl)-5-(trifluoromethyl)-2*H*-chromen-2-one, in cellular and in vivo models of pulmonary oxidative stress. *J. Pharmacol. Exp. Ther.* **363**, 114–125 (2017).
8. Pallesen, J. S., Tran, K. T. & Bach, A. Non-covalent small-molecule Kelch-like ECH-associated protein 1-nuclear factor erythroid 2-related factor 2 (Keap1–Nrf2) inhibitors and their potential for targeting central nervous system diseases. *J. Med. Chem.* **61**, 8088–8103 (2018).
9. Davies, T. G. et al. Monoacidic inhibitors of the Kelch-like ECH-associated protein 1: nuclear factor erythroid 2-related factor 2 (KEAP1:NRF2) protein–protein interaction with high cell potency identified by fragment-based discovery. *J. Med. Chem.* **59**, 3991–4006 (2016).
10. Cuadrado, A. et al. Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases. *Nat. Rev. Drug Discov.* **18**, 295–317 (2019).
11. Sterling, T. & Irwin, J. J. ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
12. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
13. Alhossary, A., Handoko, S. D., Mu, Y. & Kwoh, C.-K. Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics* **31**, 2214–2216 (2015).
14. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904 (2013).
15. Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J. & Sanner, M. F. AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLOS Comput. Biol.* **11**, e1004586 (2015).
16. Koebel, M. R., Schmadeke, G., Posner, R. G. & Sirimulla, S. AutoDock VinaXB: implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina. *J. Cheminform.* **8**, 27 (2016).
17. Nivedha, A. K., Thieker, D. F., Makeneni, S., Hu, H. & Woods, R. J. Vina-Carb: improving glycosidic angles during carbohydrate docking. *J. Chem. Theory Comput.* **12**, 892–901 (2016).
18. Amaro, R. E. et al. Ensemble docking in drug discovery. *Biophys. J.* **114**, 2271–2278 (2018).
19. Houston, D. R. & Walkinshaw, M. D. Consensus docking: improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.* **53**, 384–390 (2013).
20. Marcotte, D. et al. Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorg. Med. Chem.* **21**, 4011–4019 (2013).
21. Andrei, S. A. et al. Stabilization of protein–protein interactions in drug discovery. *Expert Opin. Drug Discov.* **12**, 925–940 (2017).
22. Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J. & Koes, D. R. Protein-ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* **57**, 942–957 (2017).
23. Reymond, J. L. The chemical space project. *Acc. Chem. Res.* **48**, 722–730 (2015).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Parallelization of the virtual screen using VirtualFlow

VirtualFlow uses four levels of parallelization in a hierarchical manner to enable it to run on batch system-managed Linux clusters of any configuration while allowing for perfect scaling behaviour. Each instance of VirtualFlow can submit multiple jobs, each job may use several job steps (currently only supported when using SLURM and Moab/TORQUE/PBS as the resource manager, whereas for SGE and LSF only single job steps per job are possible), one job step is able to execute an arbitrary number of queues, and each queue executes the external programs that are processing the ligands (Extended Data Fig. 1). These programs may be additionally parallelized internally, for instance through multithreading. Details about the workflow within a single queue are provided in Supplementary Information section G.

Workload balancing

When processing ligands in parallel, there needs to be a mechanism that makes sure that each ligand is treated only once. However, one main problem with parallelization is that most cluster file systems are too slow to work off a single simple task list. This is because, when different processes access the file at the same time, clashes can occur, as it may take up to several seconds until one job sees the changes made to a file by another job. These latency problems also mean that file locking mechanisms do not prevent these clashes. The standard solution for solving this kind of problem is to let different processes communicate directly with each other or through a central master process. However, in most cases, this results in sub-linear scaling behaviour, which normally worsens as more and more parallel running processes become involved. Moreover, many advanced parallelization methods such as MPI or OpenMP, do not allow for inter-job communication, while in many cases multiple simultaneously running jobs are needed. Therefore, to maintain perfect scaling behaviour that allows multiple jobs and a virtually unrestricted number of CPUs, we have developed an advanced task-list mechanism. The key is to minimize the number of instances that the parallel processes need to access the task list. The mechanism that we have implemented requires only a single access per batch system job, each of which can contain a large number of parallel running processes. For this purpose, we have implemented a workload balancer, which distributes the tasks from the central task list at the beginning of each job to all the queues that belong to it. The central task list contains collections of ligands as elementary components (rather than individual ligands), and the workload balancer takes into account the length of each collection when distributing them among the queues. This approach markedly reduces the number of times the central task list has to be accessed. For example, if the workflow uses 10 jobs in parallel, and each job runs on 100 nodes with a wall time (real run time) of 1 week and 24 CPU cores per node, and one ligand requires approximately 30 s to be docked, then the central task list needs to be accessed only 10 times per week to feed a total of 24,000 parallel running queues (assuming each queue runs on one CPU core). In this case, approximately 483,840,000 ligands are processed in 1 week, which means that the advanced task list approach reduced the number of accesses to the central task list by a factor of 48,384,000 in comparison to the number needed by a trivial task list approach (one access per ligand processed). This factor can be improved even further depending on the job size and the cluster wall time. In case two parallel processes want to access the central task list simultaneously, two backup mechanisms were implemented. The first mechanism is a time-dispersion mechanism, which spreads out simultaneously arriving jobs in time, and further stalls subsequent jobs until the workload balancer of the current job is finished. If this mechanism should fail to prevent a simultaneous access event, which could result in a damaged or empty task file, a second mechanism restores the task list using an automatically backed-up copy of a previous version of the central task list. More information about the input and output file structures of VirtualFlow can be found in Supplementary Information section F.

Reduction of input and output load

One of the potential bottlenecks of computer clusters is the input and output load that they can handle, even when they utilize shared cluster file systems with high bandwidth. The limit of the input and output capacities of a cluster can be easily reached if many small processes that individually handle their input and output and use the shared file system are running in parallel. This circumstance can pose a serious problem when running large-scale workflows with thousands of queues working in parallel, and can easily lead to crashing the cluster file system. To address this problem and considerably minimize the load on the shared file system, VirtualFlow is able to perform most input and output operations on the local temporary file systems of the computing nodes, which are normally fast RAM-based (virtual) drives readily available on any Linux system (usually /dev/shm). The final output files are then stored in batches at large time intervals on the permanent cluster file system.

Preparation of the ligand databases

One of the ligand databases that was screened originates from the state of the ZINC 15 database in the November of 2016. Approximately 330 million compounds were downloaded in SMILES format and converted into three-dimensional PDBQT files with VFLP because, at the time, the ZINC 15 database only provided a fraction of the compounds in a ready-to-dock format. During the conversion, the molecules were protonated with cxcalc (ChemAxon) and the three-dimensional structure of the ligand was computed by the molconvert tool of ChemAxon (https://chemaxon.com/products/jchem-engines). If protonation or the generation of the three-dimensional structure failed, Open Babel[24] was used as a fall-back option. Other preparation steps, such as desalting, were not carried out on these compounds as they had already undergone these basic preparation steps for the ZINC 15 database.

We also prepared the compounds in the REAL database provided by Enamine (https://enamine.net/library-synthesis/real-compounds). Approximately 700 million partially stereospecific SMILES were expanded into fully stereospecific SMILES, resulting in around 1.4 billion molecules. These were then prepared with VFLP into a ready-to-dock format. Specifically, the compounds were desalted and neutralized using cxcalc (ChemAxon), major tautomers were computed using cxcalc and then protonated using cxcalc (using Open Babel as fall back), the three-dimensional coordinates were computed using molconvert (ChemAxon) (using Open Babel as a fall back) and, finally, compounds were converted into the PDBQT format using Open Babel. This library has been made available through an interactive web interface (Supplementenary Information section C). The scaling behaviour of VFLP was measured on the GCP using up to 20,000 CPU cores and these data are shown in Supplementary Fig. 7.

Computation time of VFVS

The total computation time (T) is directly proportional to the number of ligands screened (N) and the processing time per ligand (P), and inversely proportional to the number of CPUs used (C):

$$T \propto (P \times N)/C$$

The processing time per ligand (P) depends mainly on the specific docking scenario (which includes the receptor and all of the possible docking options and parameters) and the speed of the CPUs used, and can be approximated by:

$$P \propto (E \times \Theta + \zeta)/\eta$$

where η is a factor that represents the CPU speed relative to a reference CPU, E is the docking exhaustiveness parameter (elaborated in 'Relationship between the exhaustiveness parameter and the docking time'), Θ is the docking time per unit exhaustiveness on the reference CPU (that is, the slope of the lines shown in Extended Data Fig. 3c) and ζ is the initial set-up time required by the docking program on the reference CPU (that is, the intersection at the y axis of the lines in Extended Data Fig. 3c). For a typical case of a large-scale first-stage virtual screen on one of the newer Intel CPUs, the average processing time per ligand (P) is roughly 5 s using the fastest docking settings. It follows that when 5,000 CPUs are used, the total screening time for 100 million compounds will be roughly 30 h. Extended Data Figure 3b illustrates the relationship between the computation time and the number of CPUs for a given number of ligands (assuming an average processing time of 5 s per compound).

Relationship between the exhaustiveness parameter and the docking time

The time to dock a single molecule depends on the number of conformations that are sampled, and this number is largely independent of the size of the docking box or surface area. The number of conformations sampled can be controlled by the exhaustiveness parameter of the docking programs. Docking time has a linear dependency on the exhaustiveness parameter, as shown in Extended Data Fig. 3c. The inset in the graph shows the slope for each of the docking programs, providing an estimate of the degree of dependency between the computational time and the exhaustiveness parameter for individual docking programs.

As most docking programs use a probabilistic search algorithm, the results of separate iterations with the same starting set-up can differ. This can be beneficial as it can be more efficient to carry out multiple less-exhaustive docking iterations than to run one highly exhaustive iteration. The exhaustiveness here is a measure of the extent to which the conformational space of the ligand, and potentially the protein side chains, is explored by the search algorithm during the docking procedure. In light of this, VFVS can be configured to carry out multiple replicas per docking scenario, thus improving the overall efficiency.

Lead optimization using VFVS

The operational flexibility enables VFVS to also be used during lead optimization (Fig. 1). In this context, a library of analogues of a chosen lead compound can be prepared with VFLP and screened by VFVS with high docking accuracy (for example, setting the exhaustive parameter to a high value, allowing specific amino acids in the binding interface to be flexible, using multiple docking programs and/or multiple receptor (backbone) conformations), which can considerably accelerate the lead optimization process.

Parameters of the virtual screen against the KEAP1 target

For the virtual screening validation test, the crystal structure of the KEAP1 Kelch domain (Protein Data Bank (PDB) ID: 5FNQ)[9] was used.

The protein was stripped of all small molecules present (including water), protonated at physiological pH and then converted into PDBQT format using AutoDockTools[25].

The NRF2-binding interface on KEAP1 was chosen as the target of the screening and the exact location was determined using previously published co-crystal structures of KEAP1 and the NRF2 peptide (PDB ID: 4IFL). The in silico screen was carried out as follows. First, VFVS used the docking program QuickVina 2 in an initial (primary) virtual screen with mouse KEAP1 as a rigid receptor structure.

In this primary virtual screening, the docking search space was a rectangular parallelepiped (that is, a cuboid) of size 15.0 × 16.5 × 14.275 Å^3. The exhaustiveness parameter was set to 1, which favours fast computational times. The quality of individual docking results, and therefore the ranking, depends largely on the external docking program chosen (which is independent of VirtualFlow).

Next, in the rescoring procedure, the following amino acid side chains at the binding interface were allowed to be flexible: Tyr334, Arg380, Asn382, Arg415, Cys434, His436, Ile461, Phe478, Arg483, Ser508, Tyr525, Tyr572 and Phe577. AutoDockTools was used to generate the rigid and flexible receptor structures in PDBQT format. The exhaustiveness was set to 1, and two replicas (iterations) were carried out of each docking scenario (with Smina Vinardo and AutoDock Vina as the docking programs). The size of the docking box was set to 27.0 × 27.0 × 24.0 Å^3.

Expression and purification of GST–KEAP1

A codon-optimized sequence of the Kelch domain (residues 322–624) of mouse KEAP1 cloned into a pGEX-6P-3 vector with BamHI and XhoI cloning sites, and an NRF2 peptide (AFFAQLQLDEETGEFL) with an N-terminal tetramethylrhodamine (TAMRA) fluorophore were purchased from GenScript USA. The pGEX-6P-3 vector contains an N-terminal glutathione *S*-transferase (GST) tag, which is expressed as a fusion with the target sequence, resulting in a gene product that will henceforth be referred to as GST–KEAP1. The vector carrying GST–KEAP1 was transformed into BL21(DE3) *Escherichia coli*. The transformed cells were grown at 37 °C to an optical density of 0.6 at a measurement wavelength of 600 nm and protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were allowed to grow for 12–16 h at 18 °C and were subsequently collected by centrifugation at 4,200 rpm for 20 min at 4 °C.

To purify GST–KEAP1, cell pellets from 2 l of culture were resuspended in 40 ml of GST-binding buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) supplemented with 3.5 mM β-mercaptoethanol and protease inhibitors (Roche). Cells were lysed by sonication, the insoluble fraction was removed by centrifugation at 16,000 rpm and the soluble fraction was applied to 10 ml of GST agarose resin (GoldBio). The suspension was nutated for 4 h at 4 °C, and the unbound fraction was removed by gravity-flow chromatography. The slurry was washed twice with GST-binding buffer supplemented with 3.5 mM β-mercaptoethanol. The bound fraction was eluted from the slurry with 20 mM reduced glutathione in GST-binding buffer. The resulting eluate was loaded on a Superdex 200 size-exclusion chromatography (SEC) column pre-equilibrated in SEC buffer (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM dithiothreitol).

Fluorescence polarization assays

Measuring the dissociation constant of the NRF2-KEAP1 interaction. We prepared 2 nM TAMRA–NRF2 peptide in fluorescence polarization buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM DTT, 2 mM 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 0.005% BSA, 1% DMSO) in 384-well plates (Corning, 3575), to establish the K_d of the interaction between TAMRA–NRF2 and GST–KEAP1. GST–KEAP1 was titrated into the TAMRA–NRF2 peptide starting at a concentration of 76 μM GST–KEAP1 followed by twofold dilutions for a total of 24 points. A K_d of 3.67 ± 0.35 nM was determined for the interaction (Extended Data Fig. 4a).

Fluorescence polarization, level 1 (high-throughput screening of VirtualFlow derived hits). All 590 compounds ordered for testing were dissolved in DMSO-d_6 to a final concentration of 10 mM. Two AB1056 (Abgene) plates were prepared as source plates for screening. The first source plate contained 11 μl of each of the 10-mM compounds. The second source plate was filled with 9 μl DMSO and 1 μl from the first source plate was transferred into the second via pin transfer with a Vprep liquid-handling pipetting station (Agilent), resulting in a final concentration of 1 mM for each compound in the second source plate. Then, 384-well (Corning, 3575) assay plates were pre-loaded with 7 nM GST–KEAP1 in fluorescence polarization buffer (30 μl per

well). Subsequently, 300 nl, 100 nl and 33 nl volumes were transferred from each source plate (the 10-mM and 1-mM plates) to pre-loaded 384-well assay plates. The assay plates were incubated for 1 h at room temperature before 2 nM TAMRA–NRF2 peptide was added to each well with an HP D300 (Hewlett-Packard). After 3 h of incubation at room temperature, fluorescence polarization (excitation, 485 nm; emission, 520 nm) was measured using an EnVision plate reader (PerkinElmer). This assay resulted in six-point titrations, which are not sufficient to calculate accurate IC_{50} values, but enable for the selection of top binders.

Fluorescence polarization, level 2 (screening of top hits). The 27 compounds that were active in the fluorescence polarization level 1 assay were subjected to a second 24-point fluorescence polarization screen (level 2), starting from 500 μM compound followed by 1.5-fold serial dilution. For the best compound, iKeap1, the starting concentration was reduced to 30 μM and the following concentrations were used in the titration: 30.00 μM, 21.60 μM, 15.50 μM, 11.10 μM, 8.00 μM, 5.76 μM, 4.14 μM, 2.98 μM, 2.15 μM, 1.54 μM, 1.11 μM, 0.80 μM, 0.576 μM, 0.414 μM, 0.298 μM, 0.215 μM, 0.154 μM, 0.111 μM, 0.080 μM, 0.0606 μM, 0.0459 μM, 0.0348 μM, 0.0264 μM and 0.02 μM. The measurements were carried out in triplicates, and the three data points for each concentration were averaged. IC_{50} values were determined by fitting the averaged data points to a four-parameter logistic curve using the nonlinear least-squares method provided by the SciPy library for Python (https://www.scipy.org/). The standard error (Supplementary Table 4) of the IC_{50} was computed by taking the square root of the diagonal of the parameter covariant matrix.

BLI assays

BLI binding and displacement assays. NRF2–KEAP1-binding BLI experiments were performed on an Octet RED384 (ForteBio) using streptavidin-coated Dip and Read Biosensors (ForteBio) and 384-well plates with 120 μl volume. The sensors were incubated for 5 min in 500 nM biotinylated NRF2 peptide in binding buffer (10 mM HEPES, pH 7.5, 50 mM NaCl, 0.1% (v/v) Tween-20 with 0.5 mM TCEP and 1% DMSO). To test for nonspecific binding of the GST–KEAP1 protein, reference tips were incubated in buffer only. The tips were washed with buffer for 2 min to obtain a baseline reading and then transferred to wells containing various concentrations of GST–KEAP1 protein (100 nM, 50 nM, 25 nM, 12.5 nM, 6.75 nM, 3.375 nM, 1.679 nM and 0.844 nM) for 10 min. After measuring association, tips were moved to wells containing buffer, and dissociation was measured for 5 min. The data were processed and analysed using the Octet data analysis software version 11.0 (ForteBio). The association–dissociation curve for each concentration was fitted using a 1:1 model given by the equations:

$$R_t^{on} = \frac{k_{on} \times C}{k_{on} \times C + k_{on}} R_{max}(1 - e^{-(k_{on} \times C + k_{off}) \times t})$$

$$R_t^{off} = R_{eq} \times e^{-k_{off} \times t}$$

where R_t^{on} and R_t^{off} are the BLI signals at time t, R_{eq} is the equilibrium response, k_{on} is the association rate constant, k_{off} is the dissociation rate constant, C is the analyte (protein) concentration and R_{eq} is the signal level at the equilibrium of association that depends on the analyte (protein) concentration and the maximal capacity (R_{max}) of the sensor surface. By computing the ratio k_{off}/k_{on}, the apparent equilibrium constant K_d is obtained. The resulting apparent K_d values were averaged.

Compound screening by BLI displacement assay. The BLI displacement assays were set up as described above. The biotinylated NRF2 peptide was used at a concentration of 500 nM and GST–KEAP1 protein was used at a concentration of 25 nM. The compounds were used at concentrations of 20 and 100 μM, and pre-incubated with GST–KEAP1 protein. The association phase was measured in the well containing compound with GST–KEAP1 protein for 10 min, and followed by a dissociation phase in buffer for 5 min. The inhibition percentage was the average BLI signal in the last 50 s of the dissociation phase, normalized against the condition of GST–KEAP1 protein in the absence of compound. The dose-dependent experiment with iKeap22 was carried out at 10 μM, 20 μM, 40 μM, 80 μM and 100 μM compound concentration and pre-incubated with 25 nM GST–KEAP1 protein.

To test for nonspecific binding of the compounds, the sensor was coupled with biotinylated NRF2 peptide and the compounds were used at 20 μM concentration without protein.

SPR binding assays

All SPR binding experiments were performed on a BiacoreT200 (GE Healthcare) instrument at 25 °C in running buffer (10 mM HEPES pH 7.5, 50 mM NaCl, 0.1% (v/v) Tween-20 with or without 0.5 mM TCEP and 1% DMSO). The running buffer was prepared fresh on each day of use, filtered and degassed before the SPR experiments. The target protein (GST–KEAP1) was anchored to a CM5 chip using a GST labelling kit (GE Healthcare)[26], in which a polyclonal goat anti-GST antibody was immobilized on a CM5 sensor chip by amine-coupling method 1.

SPR assay, level 1 (one-point high-throughput screen). The SPR level 1 screening was carried out as previously reported[27]. First, we prepared 10 mM DMSO-d_6 stock solutions of the 590 compounds that were procured in powder form. Then, 20 μM samples of the compounds were made by diluting the stock compounds in running buffer with 0.5 mM TCEP and 1% DMSO. The anti-GST immobilizing chip was saturated with GST in the reference channel and GST–KEAP1 in the target channel with resonance unit (RU) values of 750–800 for GST and 2,000–3,000 for GST–KEAP1. Binding of compounds to the immobilized protein was monitored for 60 s in both the association and dissociation phase. Additional injection of the running buffer was performed after every compound binding. All binding signals (RU_{max} = 16–29 RU, 1:1 stoichiometry) were corrected for the signals from the reference channel and buffer blank. Compounds were classified as an SPR level 1 hit if the condition RU > 4 was satisfied. This criterion was based on the positive control (iKeap1, RU = 4.65 ± 0.74).

SPR assay, level 2 (five-point high-throughput screen of the SPR level 1 hits). The hits from the SPR level 1 assay were rescreened at five different compound concentrations (0.5, 1, 5, 10 and 20 μM), in running buffer with 0.5 mM TCEP and 1% DMSO, at a rate of 30 μl min^{-1}. The hits were classified as hits if they produced a concentration-dependent SPR response and RU > 4 at a compound concentration of 20 μM.

SPR assay, level 3 (SPR experiments of selected SPR level 2 hits). We chose 23 out of the 69 SPR level 2 hits for level 3 analysis. Given the low throughput of the level 3 SPR assay, we chose a subset of the SPR level 2 hits, which included the displacers from the level 3 fluorescence polarization assay, the compounds that were tested by NMR and select SPR level 2 hits. SPR experiments were carried out in which the target protein (GST–KEAP1) was captured and regenerated in each compound cycle. All SPR data processing and analyses were performed using the BIAevaluation software (version 3.0). For steady-state binding, the R_{eq} signal was plotted against the analyte concentration and fitted to the one-site or the biphasic binding model (Supplementary Table 3) using the Levenberg–Marquardt algorithm used by the BIAevaluation software. The one-site binding model is given by the equation

$$R_{eq} = (R_{max} \times C)/(K_d + C) + b \quad (1)$$

where R_{eq} is the SPR signal at equilibrium, R_{max} is the SPR signal at saturation of the binding mode, K_d is the dissociation constant of the compound, b is the offset and C is the concentration of the compound. The biphasic binding model is given by the equation

$$R_{eq} = (R_{max,1} \times C)/(K_{d,1} + C) + R_{max,2} \times C/(K_{d,2} + C) + b \quad (2)$$

where R_{eq} is the SPR signal at equilibrium, $R_{max,1}$ and $R_{max,2}$ are the SPR signals at saturation of the two binding modes, $K_{d,1}$ and $K_{d,2}$ are the dissociation constants of the compound corresponding to the two binding modes, b is the offset and C is the concentration of the compound.

Standard errors of the estimated K_d values were computed with the BIAevaluation software, which computes them using the diagonal elements of the covariance matrix and the residual. The software operates based on the equations found on page 378 of the book *Receptor-Ligand Interactions: A Practical Approach*[28].

Ligand-detection NMR experiments

The differential line broadening experiments serve as simple one-dimensional experiments, in which the proton signal of the ligand is monitored. The ligand concentration exceeds the receptor concentration (for example, 10–20-fold) in this experiment and broadening of the resonance frequencies in the presence of the receptor is a consequence of ligand molecules shuttling between free and bound states. Differential line broadening manifests as a broadening of the ligand resonance due to binding of a protein. The ligand is in equilibrium between the free and protein-bound states dictated by the equilibrium constant. Differential line broadening is the result of the change in relaxation rate and the difference in chemical shift of the bound ligand. In the saturation transfer difference (STD) experiments, a region of the spectral space (−1 to 0.5 ppm) that has resonances from the receptor but not the ligand is selectively saturated. Resonances from methyl-bearing amino acids (Ile, Leu and Val) often populate this region of the spectral space. This saturation is transferred to the rest of the protein and eventually to the bound ligand by spin diffusion. In the implementation of STD, two one-dimensional spectra are recorded in an interleaved fashion. In the first experiment neither the receptor nor the ligand is saturated (off-resonance) and in the second the receptor is selectively saturated (on-resonance). Spectra of free ligands are observed in both experiments. However, if the ligand transiently binds to the receptor then the saturated receptor will transfer magnetization to the ligand. This transfer will be reflected as reduced intensity in the on-resonance saturated spectrum compared to the off-resonance saturation. The results are often presented as a difference spectrum between the on- and off-resonance saturation experiments. The appearance of ligand resonances in the difference spectrum is indicative of ligand binding. Measurement of the transverse relaxation rate of the ligand is another complementary strategy to detect ligand binding to a receptor. The free ligand behaves like a small molecule and experiences slow transverse relaxation; however, transient binding to the receptor enhances the transverse relaxation rate of the ligand. Thus, an increased transverse relaxation rate in the presence of a receptor directly indicates binding to the receptor. In the experimental set-up, a series of one-dimensional experiments in which the coherences of the ligand spend increasing amounts of time in the transverse plane is recorded. Ligands that engage the protein will relax faster than unbound ligands. We refer to these experiments as Car–Purcell–Meiboom–Gill (CPMG) or CPMG–R_2 experiments. Although any of these experiments are in principle sufficient to demonstrate ligand binding, false positives for either of these experiments have been reported. However, a detection of a false-positive hit is highly unlikely if all three experiments indicate binding, which is the case for all of the hits reported here.

All of the ligand-detecting experiments were performed with 50 μM compound alone or in the presence of 5 μM KEAP1 (without the GST tag) in NMR buffer (phosphate-buffered saline supplemented with 5% DMSO-d_6 and 4 mM deuterated DTT at pH 7.4) unless otherwise noted. For iKeap1 and iKeap2, the protein concentration was kept at 2.5 μM due to tight binding. ^{1}H one-dimensional spectra of the compounds were recorded in the absence and presence of KEAP1 to assess line broadening. STD spectra of the compounds in the presence of KEAP1 were recorded with 3 s saturation time on (0 ppm) and off (−20 ppm) resonance, respectively. The relaxation rate of the compounds was measured in the absence and presence of KEAP1 with a series of ^{1}H one-dimensional experiments with CPMG-based transverse relaxation time filters of various lengths: 1 ms, 25 ms, 50 ms, 100 ms, 300 ms, 500 ms and 800 ms. Data were analysed and visualized in MATLAB (MathWorks).

Protein-detection NMR experiments

The cleaved Kelch domain (residues 322–624) of mouse KEAP1 consists of 308 amino acids with close to 300 detectable amide resonances. Therefore, correlating chemical shift perturbations of small-molecule inhibitors to perturbations introduced by NRF2 would have been prohibitively difficult in ^{1}H–^{15}N HSQC spectra without full backbone assignment. Our aim was to rely on methodology that can be quickly and easily implemented even for very large proteins for which backbone assignment might not be feasible. Indeed, with a molecular mass of 33.7 kDa, the Kelch domain (322–624) of mouse KEAP1 is already on the larger side for NMR backbone assignment. To overcome the spectral crowding in a ^{1}H–^{15}N HSQC spectrum and minimize problems due to low ligand solubility, we implemented a ^{1}H–^{13}C TROSY heteronuclear multiple-quantum correlation experiment coupled with fast data acquisition. For the protein-detected ^{1}H–^{13}C heteronuclear multiple-quantum correlation experiments, a sample of KEAP1 that is selectively labelled with ^{1}H and ^{13}C at the methyl groups of isoleucine, leucine and valine (ILV) residues, in an otherwise deuterated background, was used. This labelling strategy is referred to as ILV labelling. ILV-labelled samples of KEAP1 were prepared by culturing BL21(DE3) *E. coli* cells containing a plasmid for GST–KEAP1, in perdeuterated M9 medium with 1 g ^{15}N-NH_4Cl and 2 g ^{2}H-^{12}C-glucose in ^2H_2O. Then, 1 h before induction with IPTG, 330 mg l^{-1} 2-(^{13}C)methyl-4-($^2H_3$)-acetolactate (a precursor for leucine and valine) was added. The acetolactate was activated beforehand as previously described[29]. Subsequently, 20 min before induction, 75 mg l^{-1} ^{13}C/^{1}H-methyl ketobutyrate sodium (a precursor for isoleucine), which was otherwise deuterated, was added. The use of acetolactate resulted in stereospecific ^{1}H–^{13}C labelling of only one of the leucine(λ2) and valine(γ2) methyl groups as previously described[29]. The protein was purified as described above. The GST-tag was cleaved by preScission protease cleavage and the free Kelch domain of mouse KEAP1 was eluted in NMR buffer from a SEC column. All NMR measurements for the ILV-labelled KEAP1 were performed at a protein concentration of 5 μM. The protein concentration was kept low to account for poor solubility (for NMR) of some of the compounds. The concentrations of the compounds were 50 μM, except for iKeap1 and iKeap2, for which the concentrations were 25 μM, owing to poor solubility.

Given the low concentration of the protein, we used the methyl SOFAST methyl TROSY with 46 ms and 18 ms acquisition times in the direct and indirect dimensions, respectively[30]. The spectral width was set to 14 ppm (^{1}H) and 20 ppm (^{13}C) in the direct and indirect dimensions, respectively, and the spectrum was recorded at 298 K on an 800-MHz Bruker spectrometer equipped with an AVANCE III console and a cryogenically cooled probe. A 4.5-ms Pc9_4_90.1000 pulse was used for selective excitation of the methyl ^{1}H resonances and a 1.2-ms Rsnob.1000 pulse was used to selectively refocus proton chemical shift evolution and ^{1}H–^{13}C J-coupling during ^{13}C chemical shift evolution. Proper choice and calibration of the excitation and refocusing pulses is crucial to avoid perturbing the water signal, which can significantly lower the achievable signal to noise ratio. Fast data acquisition was achieved with a 150-ms recycling delay, which enabled the recording of experiments with 512 scans in 5 h.

Detecting aggregation using dynamic light scattering

To test the potential aggregation of hits, we used dynamic light scattering experiments. The experiments were performed on a ZS90 Zetasizer instrument (Malvern Panalytical). Measurements were done in

triplicate with 10 scans per run (100 s). The compounds were used at 20 μM concentration in running buffer (10 mM HEPES pH 7.5, 50 mM NaCl, 0.1% (v/v) Tween-20 with or without 0.5 mM TCEP, 2% DMSO), which was filtered before use. The 20-μM working solution was made from a 1 mM stock of the compound in DMSO. The data were analysed by the built-in software. Compounds were classified as aggregated when the radius of the measured particles was above the minimum colloidal aggregate size (for small molecules)[31] of 50 nm.

In addition, the solubility of iKeap1, our most potent displacer, was analysed with an NMR solubility assay based on a previously described technique[32]. We made individual samples of iKeap1 at various concentrations (in PBS buffer, pH 7.4) ranging from 5 μM to 30 μM and measured the one-dimensional NMR spectrum of each sample with identical experimental conditions. The resonances of iKeap1 were then integrated and plotted as a function of the concentration. The plot shows a linear trend ($R^2 = 0.996$), indicating that iKeap1 does not aggregate in this concentration range (Supplementary Fig. 4).

Excluding interference from PAINS

PAINS comprise 480 markers initially identified as moieties postulated to cause interference in experimental high-throughput screens[33]. PAINS are often found in the databases that are commonly used for in silico screens, and the user should be cognizant of the fact that a potential hit could contain a PAINS substructure. However, it should also be noted that certain PAINS-like aspects can be mitigated by judicious use of medicinal chemistry, and some aspects of PAINS could have no effect, depending on the target of choice and/or the experimental assays used[34,35]. Attention should be paid to identifying and rigorously characterizing any PAINS among the hits identified in an in silico screen.

Two of the hit compounds (iKeap2 and iKeap7) reported in this manuscript contain PAINS substructures. We performed additional experiments to confirm that iKeap2 and iKeap7 are not false positives due to assay interference. First, we used dynamic light scattering to confirm that all of the compounds shown here did not aggregate at the concentrations used in the various experiments (Supplementary Table 5). Second, we carried out ligand-detection NMR experiments using STD NMR and CPMG performed with a tenfold excess of the compound to show that iKeap2 and iKeap7 bind KEAP1 in a reversible manner (Fig. 3 and Extended Data Fig. 8). And finally, we carried out protein-observed ^{1}H–^{13}C heteronuclear multiple-quantum correlation experiments to show that both iKeap2 and iKeap7 engaged KEAP1 in a specific manner at the NRF2-binding site and did not aggregate the protein (Supplementary Figs. 1, 3). In the event that these compounds caused KEAP1 to aggregate, all of the resonances would be broadened, which is not the case here.

Statistics and reproducibility

Characterization of the violin plots in Fig. 2b were as follows. Screening size of 100,000: minimum, −10.3 kcal mol^{-1}; maximum, −11.6 kcal mol^{-1}; median, −10.4 kcal mol^{-1}; first quartile (Q_1), −10.4 kcal mol^{-1}; third quartile (Q_3), −10.6 kcal mol^{-1}. Screening size of 1 million: minimum, −10.9 kcal mol^{-1}; maximum, −12 kcal mol^{-1}; median, −11 kcal mol^{-1}; Q_1, −11.1 kcal mol^{-1}; Q_3, −11.3 kcal mol^{-1}. Screening size of 10 million: minimum, −11.675 kcal mol^{-1}; maximum, −12.3 kcal mol^{-1}; median, −11.5 kcal mol^{-1}; Q_1, −11.4 kcal mol^{-1}; Q_3, −11.5 kcal mol^{-1}. Screening size of 100 million: minimum, −11.8 kcal mol^{-1}; maximum, −12.6 kcal mol^{-1}; median, −11.9 kcal mol^{-1}; Q_1, −11.8 kcal mol^{-1}; Q_3, −12.1 kcal mol^{-1}. Screening size of 1 billion: minimum, −12.3 kcal mol^{-1}; maximum, −13.4 kcal mol^{-1}; median, −12.4 kcal mol^{-1}; Q_1, −12.3 kcal mol^{-1}; Q_3, −12.6 kcal mol^{-1}.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The ready-to-dock library from Enamine is freely available online on the homepage of VirtualFlow at http://virtual-flow.org/real-library. Source Data for Figs. 2, 3 and Extended Data Figs. 7, 8 are available with the paper.

Code availability

VirtualFlow is mainly written in Bash (a Turing complete command language), which not only makes it simple for anyone to modify and extend the code, but also has essentially no computational overhead and is readily available in any major Linux distribution. The code for VirtualFlow is freely available on https://github.com/VirtualFlow, distributed under the GNU GPL open-source licence. The primary homepage for end users, which includes additional resources such as documentation, ligand libraries, tutorials and video demonstrations, is available at https://www.virtual-flow.org. The external docking programs discussed here are available as follows: AutoDock Vina is available at http://vina.scripps.edu, QuickVina 2 and QuickVina-W at https://qvina.github.io, Vina-Carb at http://glycam.org/docs/othertoolsservice/download-docs/publication-materials/vina-carb, Smina at https://sourceforge.net/projects/smina, AutoDockFR at http://adfr.scripps.edu and VinaXB at https://github.com/ssirimulla/vinaXB.

24. O'Boyle, N. M. et al. Open Babel: an open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
25. Morris, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
26. Hutsell, S. Q., Kimple, R. J., Siderovski, D. P., Willard, F.S. & Kimple, A. J. High-affinity immobilization of proteins using biotin- and GST-based coupling strategies. *Methods Mol. Biol.* **627**, 75–90 (2010).
27. Hämäläinen, M. D. et al. Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. *J. Biomol. Screen.* **13**, 202–209 (2008).
28. Hulme, E. C. (ed.) *Receptor–Ligand Interactions: A Practical Approach* (Oxford Univ. Press, 1992).
29. Gans, P. et al. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chem. Int. Ed.* **49**, 1958–1962 (2010).
30. Lu, M. et al. Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway. *Eur. J. Med. Chem.* **146**, 251–259 (2018).
31. Irwin, J. J. et al. An aggregation advisor for ligand discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).
32. LaPlante, S. R. et al. Compound aggregation in drug discovery: implementing a practical NMR assay for medicinal chemists. *J. Med. Chem.* **56**, 5142–5150 (2013).
33. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).
34. Baell, J. B. & Nissink, J. W. M. Seven year itch: pan-assay interference compounds (PAINS) in 2017—utility and limitations. *ACS Chem. Biol.* **13**, 36–44 (2018).
35. Capuzzi, S. J., Muratov, E. N. & Tropsha, A. Phantom PAINS: problems with the utility of alerts for pan-assay interference compounds. *J. Chem. Inf. Model.* **57**, 417–427 (2017).

Acknowledgements We thank M. Zhang for help with the binding assays; the research computing teams of the Faculty of Arts and Sciences at Harvard University (especially S. Yockel, J. Cuff, F. Pontiggia and P. Edmon), the Jülich Supercomputing Centre, the Freie Universität (especially J. Dreger), the Harvard Medical School (HMS), the HLRN and the IT support of HMS (especially K. Bayer, G. Sekmokas and D. Morgan) for their support; K. E. Leigh, N. Gray, M. Kostic, A. Dubey, B. Klein, S. Schwaninger and S. Wu for discussions and manuscript preparation; the ICCB-Longwood Screening and East Quad NMR Facilities at HMS for assistance with the ligand screen; K. Arnett and the Center for Macromolecular Interactions at the HMS for advice on the SPR and BLI experiments; A. Jaffe for his support; and the teams from the Google Cloud Platform (especially S. Fang, R. Goldenbroit and D. Payne), Amazon Web Services, and Fluid Numerics for their support. This work was partially funded by a scholarship to C.G. from the Max Planck Institute for Molecular Genetics in Berlin and a scholarship from the Einstein Center for Mathematics Berlin. C.G. and K.F. thank the ECMath and MATHEON. C.G. is grateful to C. Schütte and P. Imhof for their support and supervision during his doctoral studies. We thank Z. Alirezaeizanjani, M. Bagherpoor and Anita Nivedha for testing VirtualFlow. M.H. acknowledges funding from Deutsche Forschungsgemeinschaft (CRC 958/Project A04, CRC 1114/Project A04). A.B. was supported by an Austrian Science Fund's Schrödinger Fellowship (J3872-B21) and an American Heart Association's fellowship (19POST34380800). This research was supported in part by grant TRT 0159 from the Templeton Religion Trust and by ARO Grant W911NF1910302 to A. Jaffe. K.M.P.D. was supported by a fellowship from the Max Kade Foundation and the Austrian Academy of Sciences. H.A. acknowledges funding from the Claudia Adams Barr Program for Innovative Cancer Research. G.W. acknowledges support from NIH grant CA200913, AI037581 and GM129026.

Author contributions C.G. conceived the project, and designed and implemented the drug discovery platform (VirtualFlow). H.A. and C.G. designed the experimental workflow. A.B., H.A., Z.-F.W., P.D.F. and K.M.P.D. designed and carried out the fluorescence polarization and NMR experiments. Z.-F.W. designed and carried out the SPR and BLI experiments. K.M.P.D. and Z.-F.W. carried out the dynamic light scattering experiments. M.H. provided technical assistance regarding the code and homepage. C.G. designed the applications (screening of KEAP1 and the preparation of the REAL library). P.W.C. analysed the NMR data. C.G. carried out the computations using VFLP (preparation of the REAL database) and VFVS (screening/rescoring of KEAP1). Y.S. Malets created the web interface to the REAL database. C.G. prepared the VirtualFlow homepage. Y.S. Moroz prepared the REAL database in the initial SMILES format. C.G. and Y.S. Moroz designed the structure of the VirtualFlow version of the REAL database. Y.S. Moroz, I.I. and D.S.R. supervised and directed the synthesis and purification of the on-demand compounds from the REAL library. D.A.S. helped to evaluate the screening hits. C.G., H.A., A.B., K.F., Z.-F.W., P.W.C. and G.W. prepared the manuscript. K.F., H.A. and G.W. supervised the project.

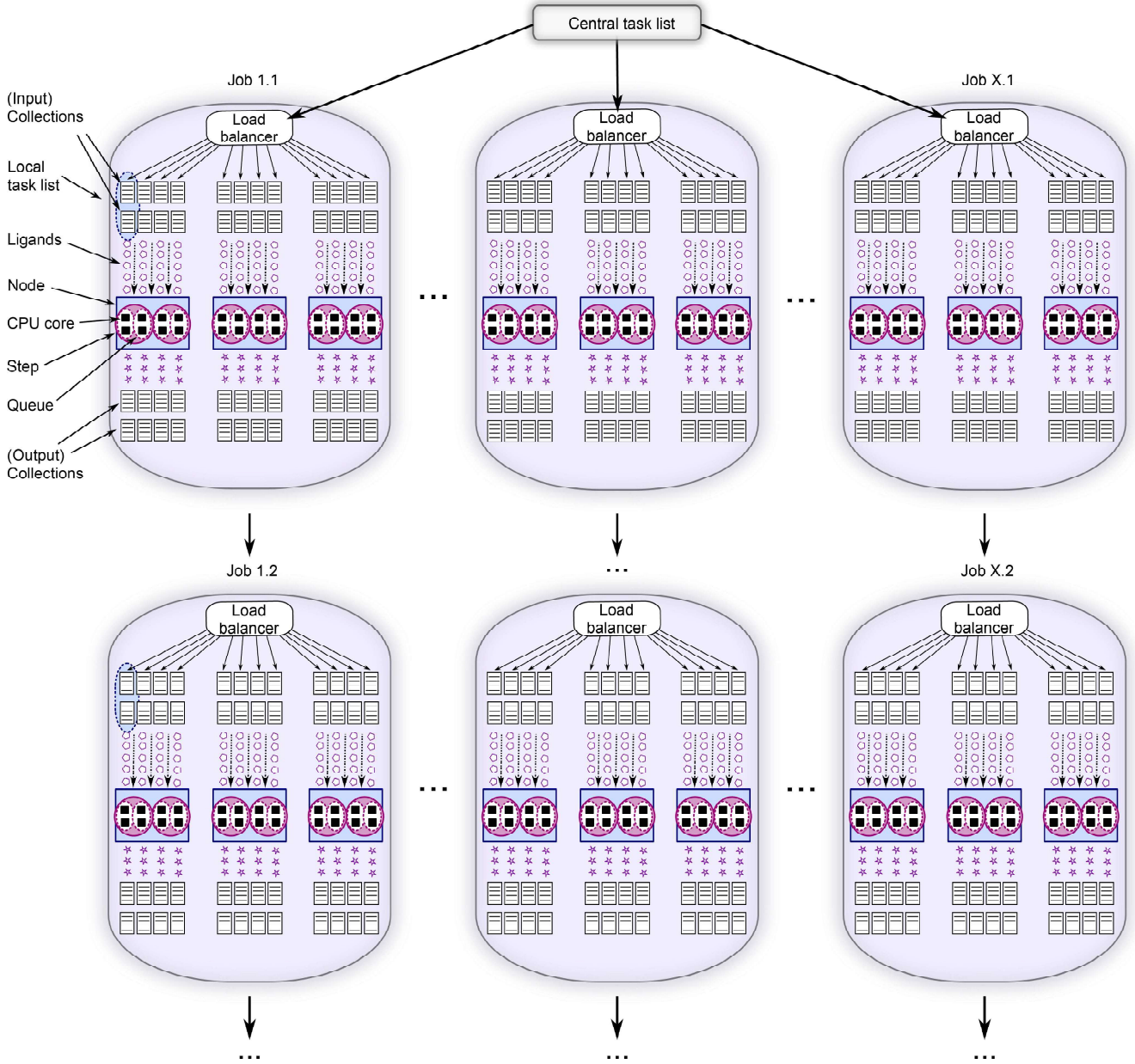
Competing interests I.I., D.S.R. and Y. S. Malets work for Enamine, a company that is involved in the synthesis and distribution of drug-like compounds. Y. S. Moroz is a scientific advisor for Enamine.

Additional information

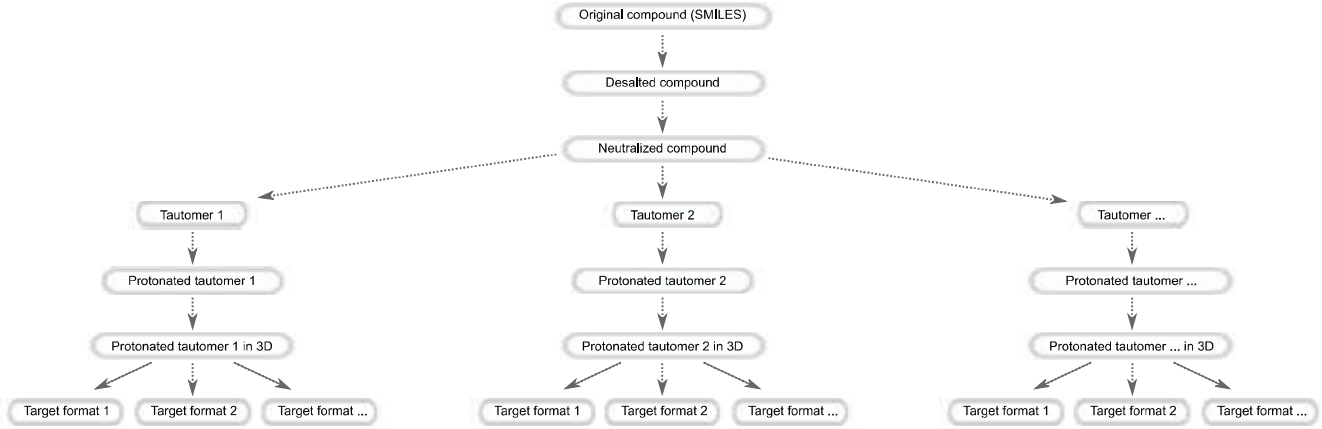
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2117-z.

Correspondence and requests for materials should be addressed to C.G. or H.A.

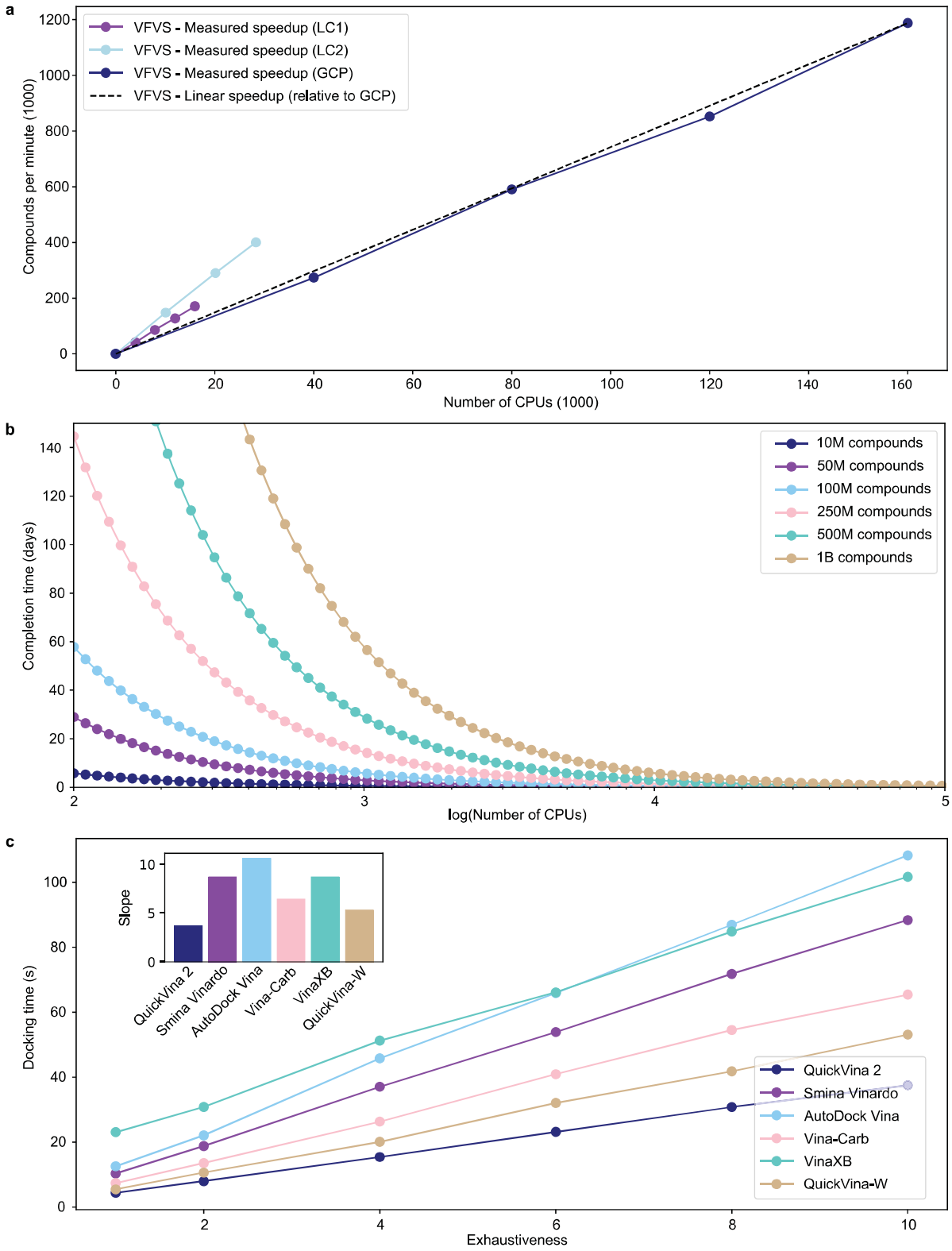
Peer review information *Nature* thanks Tara Mirzadegan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



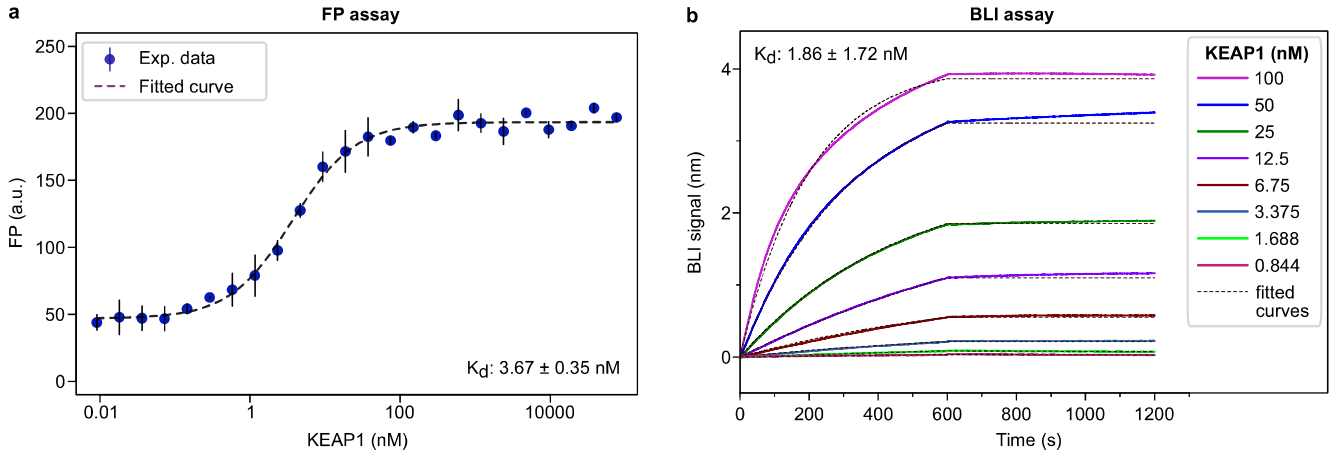
Extended Data Fig. 1 | Schematic overview of the organization of the VirtualFlow workflow on computer clusters. A computer cluster consists of compute nodes, that is, single computers (blue boxes), which contain a certain number of CPU cores (black squares inside the blue boxes). The resource manager (batch system) of the cluster generates so-called jobs (large violet ovals), each of which uses a certain number of CPU cores and nodes. In the example, each job uses three compute nodes, in which each node has eight CPU cores. Each job can contain multiple sub-jobs, referred to as job steps (purple circles). With VirtualFlow, each job step comprises multiple queues (white oval shapes within the purple circles). Often the workflow is set up such that on each CPU core one queue is running. Hierarchical multi-organization is required to allow VirtualFlow to run on any type of cluster, from the largest supercomputers (which often require that a single job has multiple nodes) to very small clusters (which often allow a job to use single CPU cores). Each queue processes ligands, which are taken from the input collections in raw form and stored in the output collection or database. The central task list contains all of the ligand collections that should be processed by the workflow, and they are distributed among the queues (into local task lists) by a workload balancer at the beginning of each job. The user can choose any number of batch system jobs (first row comprising job 1.1 to job X.1), which will automatically start successive jobs (second row comprising job 1.2 to job X.2) after their completion.



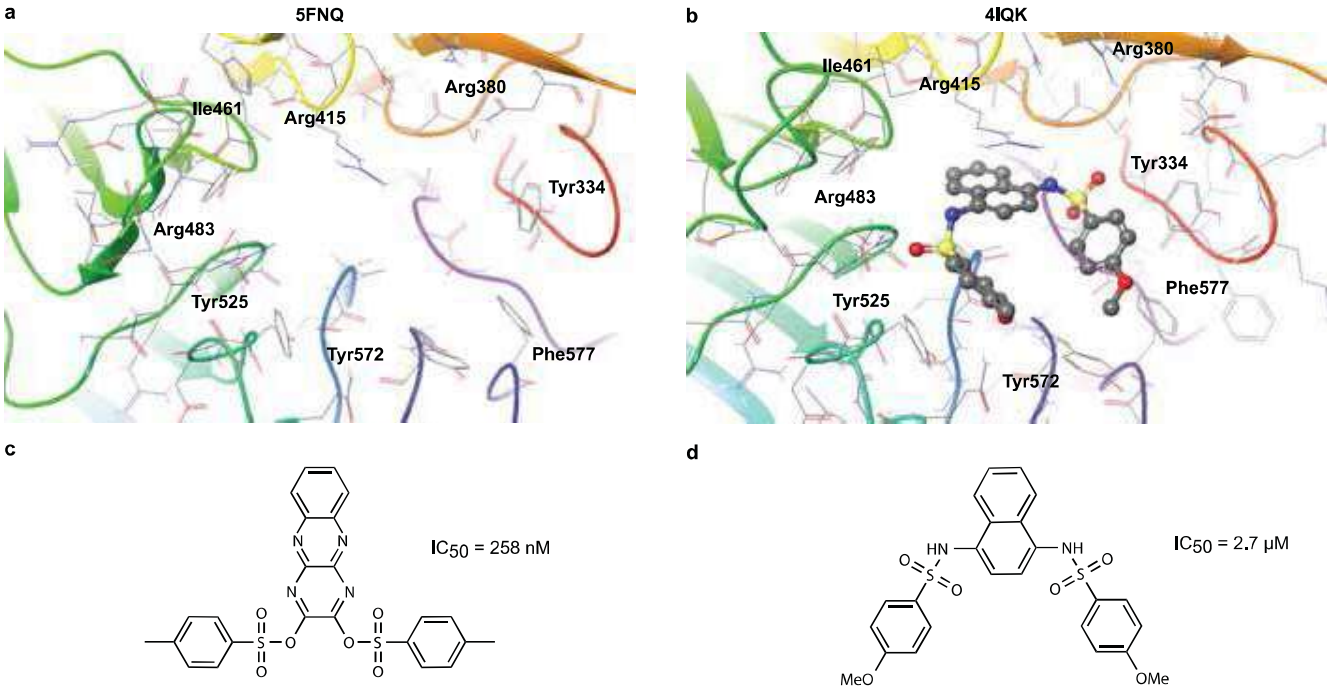
Extended Data Fig. 2 | Overview of possible processing steps during ligand preparation with VFLP. Ligands can be desalted, neutralized, and one, or possibly multiple, tautomeric state(s) as well as protonation states for each tautomer computed at specific pH values can be generated, three-dimensional coordinates can be computed and, finally, the molecules can be converted into one or potentially multiple desired target formats.



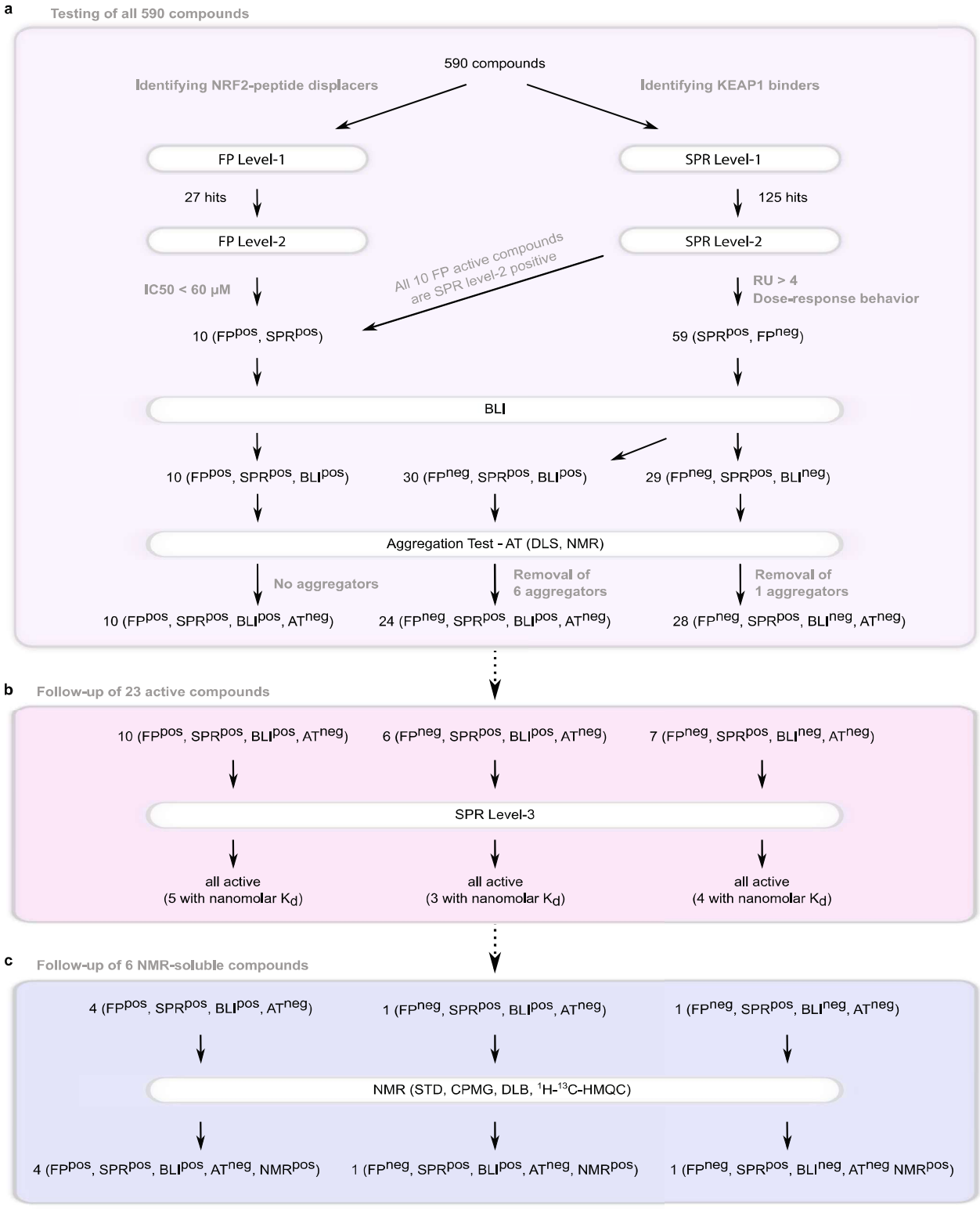
Extended Data Fig. 3 | Docking and virtual screening metrics. a, Scaling behaviour of VFVS using QuickVina 2 as the docking program. Tests with up to 30,000 cores on two local computer clusters (LC1 and LC2) and up to 160,000 CPUs on the GCP were carried out. The measured speedup is linear. DOVIS 2.0, an alternative software for virtual screenings on Linux computer clusters using AutoDock, was shown to exhibit near-linear scaling only up to 256 cores, as previously reported[4]. **b**, The computational time required (in days) for VFVS to complete virtual screens of different sizes, as a function of the number of CPUs being used in parallel. Each curve corresponds to a input ligand library with a different size, and the average computation time per ligand was assumed to be 5 s per ligand. **c**, Docking time of an average-sized ligand on a modern Intel CPU (using only a single core) as a function of the exhaustiveness parameter for different docking programs supported by VFVS. The bar plot in the inset shows the slope of the curves, which corresponds to the docking time per exhaustiveness unit. The test ligand that was used for this purpose is given by the SMILES code CN1CCN(S(=O)(=O)N2CCN(C(=O)CCCNC(=O)C3CC3)CC2)CC1. More detailed benchmarks can be found in publications related to these docking programs[5,12–17].



Extended Data Fig. 4 | Binding of the NRF2 peptide to KEAP1 as assayed by fluorescence polarization and BLI. a, a TAMRA-tagged NRF2 peptide was used for the fluorescence polarization (FP) assay. The fluorescence polarization assay was performed with three technical replicates per point. Data are mean ± s.d. for each titration point, along with the fitted curve. Two independent experiments were performed, each with similar results and one representative result is shown. **b**, A biotin-tagged NRF2 peptide was used for the BLI assay. The BLI experiment was repeated independently twicewith similar results and one representative result is shown.

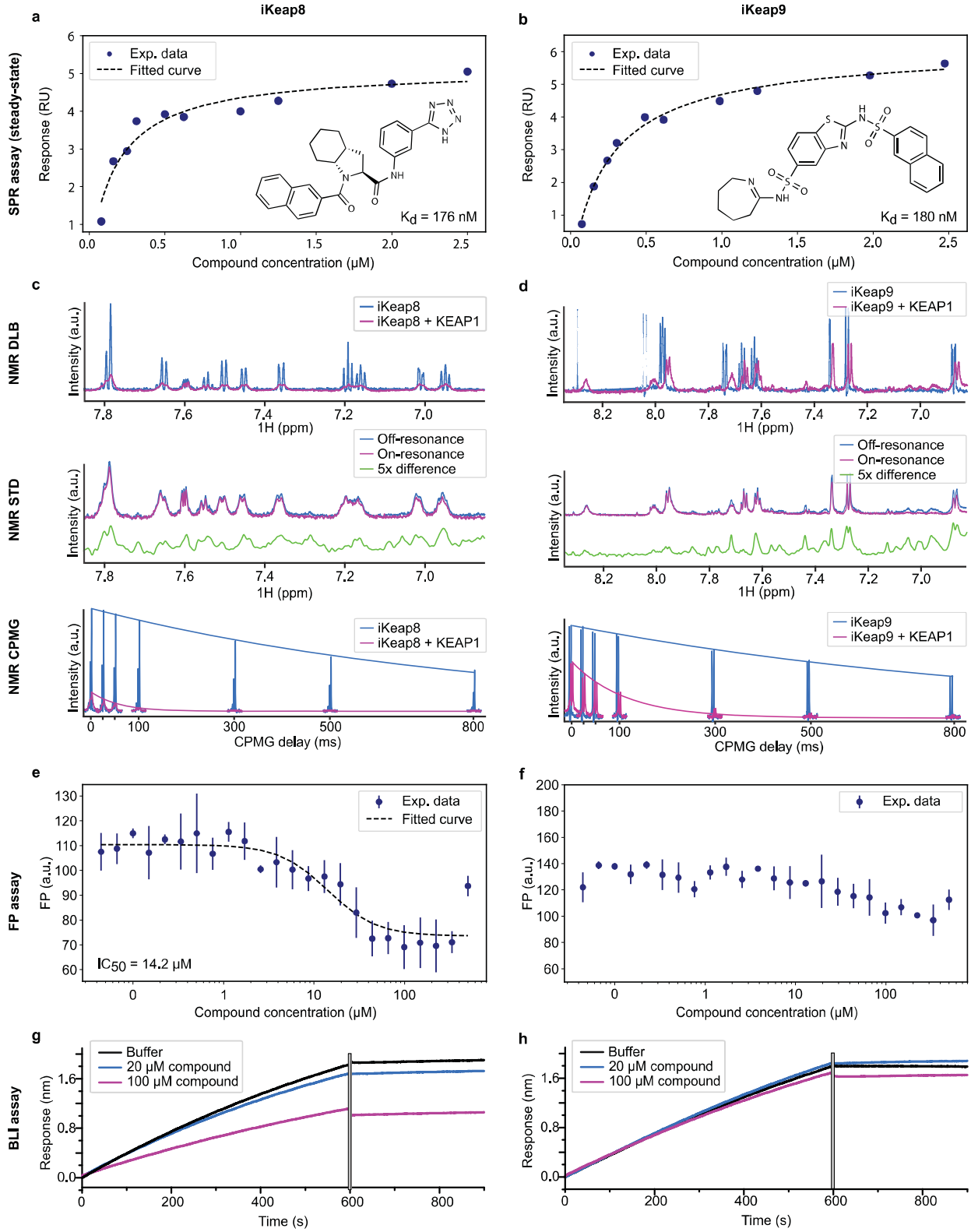


Extended Data Fig. 5 | Comparison of iKeap1 with the previously identified displacer C17. a, Crystal structure (PDB ID: 5FNQ)[9] of KEAP1 with its ligand removed, the structure used for the primary virtual screening procedure. **b**, Crystal structure of KEAP1 (PDB ID: 4IQK) with ligand C17 (Supplementary Table 1), the chemical structure of which is shown in **d**. **c**, **d**, iKeap1, the best displacer of the NFR2 peptide (**c**), is similar to compound C17, which has previously been identified by experimental methods (**d**). Although iKeap1 and C17 look similar, they differ in a number of aspects in their core scaffold (thus, analogues of the two compounds cover distinct chemical spaces, assuming that the analogues retain the core scaffold of the parent compound). This similarity, as well as the fact that the predicted docking positions (Fig. 3a) of both ligands (**b**) are nearly identical, is additional evidence that iKeap1 is binding at the predicted site.

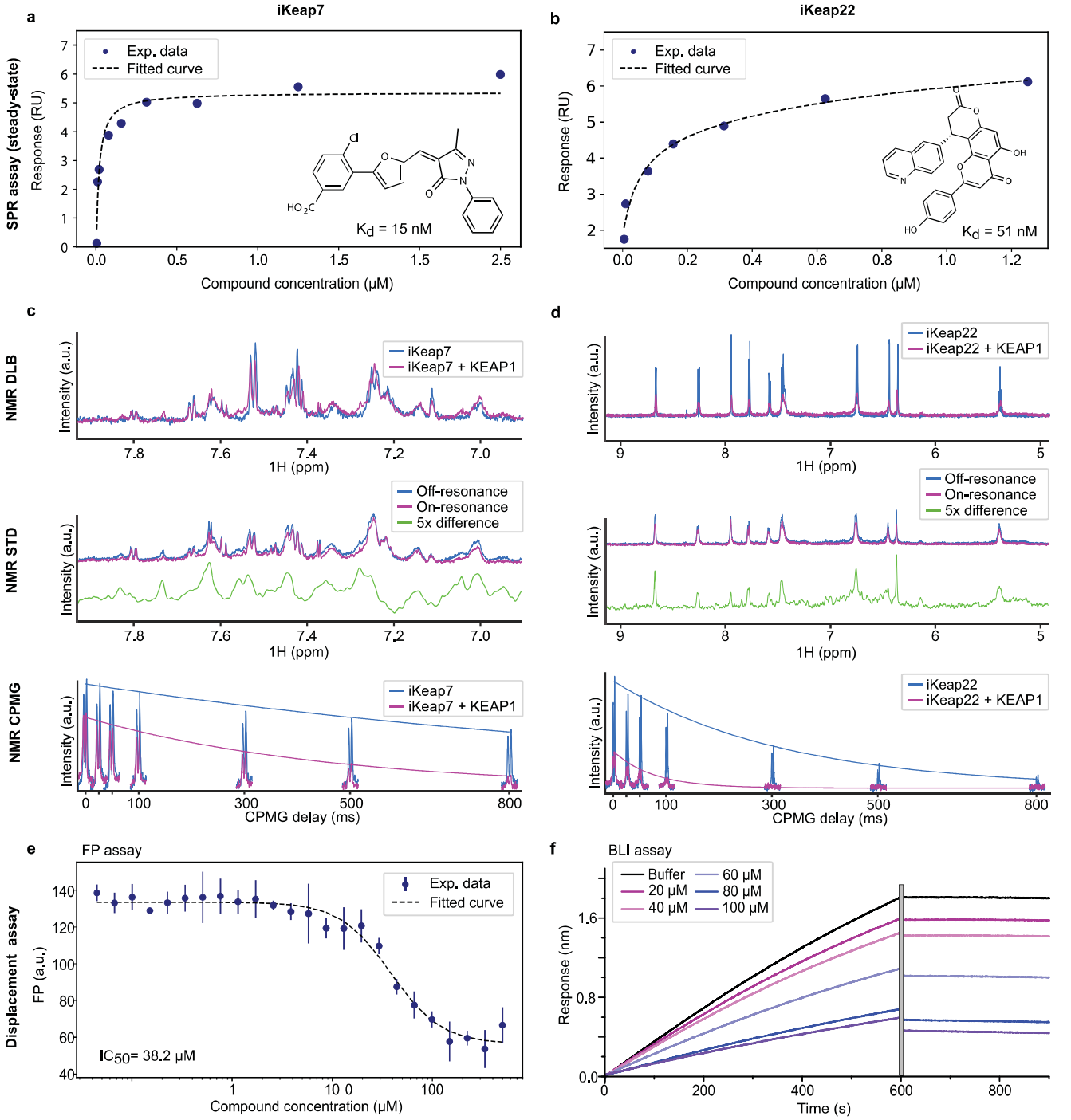


Extended Data Fig. 6 | See next page for caption.

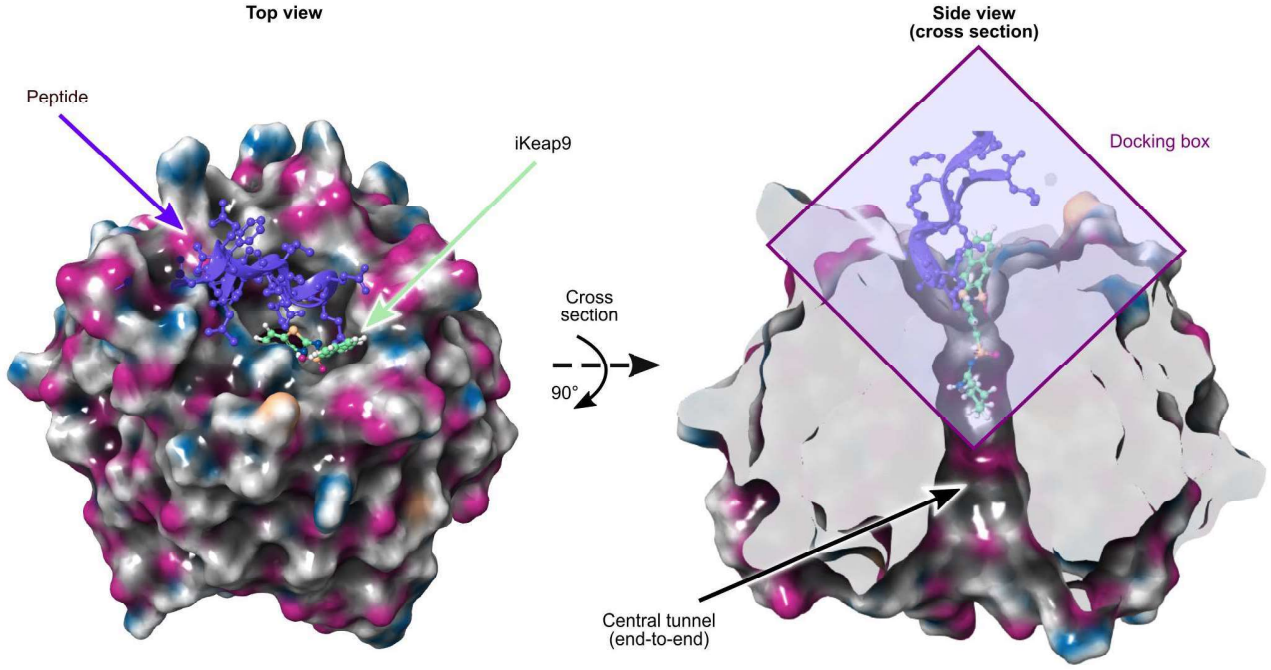
Extended Data Fig. 6 | Overview of binding assays to determine the activity of the hits identified by VirtualFlow. This schematic outlines the experimental validation workflow. The binding experiments can be broadly classified into two categories: (i) assays that directly detect the binding of the compounds to KEAP1 (SPR and NMR) and (ii) assays that detect the displacement of the NRF2 peptide from KEAP1 (fluorescence polarization and BLI). Compounds in level 2 SPR experiments were classified as active if they exhibited dose-dependent activity (measured over a range of five concentrations) and had an RU value greater than 4 at a compound concentration of 20 μM. **a**, The high-throughput workflow in which the 590 compounds identified as hits by VirtualFlow were tested using SPR and fluorescence polarization. The hits identified here were further validated by BLI and the potential of these hits to form aggregates was tested by DLS. **b**, Then, 23 of the potent hits were chosen for level 3 SPR analysis to measure accurate binding affinities. **c**, Six of the potent binders were further subjected to NMR analysis in both protein-detected and ligand-detected NMR experiments.



Extended Data Fig. 7 | Binder versus displacer. Here we highlight two scaffolds, iKeap8 and iKeap9, to illustrate the difference between binders and displacers. **a**, **b**, SPR confirms that both iKeap8 (**a**) and iKeap9 (**b**) bind to KEAP1 and with similar K_d values. Data are representative results from the SPR assay for iKeap8 and iKeap9. For each compound, three independent SPR experiments were performed, each with similar results and one representative result is shown. **c**, **d**, Ligand-detection NMR experiments shows that both iKeap8 (**c**) and iKeap9 (**d**) bind to KEAP1. **e**–**h**, However, fluorescence polarization (**e**, **f**) and BLI (**g**, **h**) assays show that iKeap8 (**e**, **g**) is able to displace the NRF2 peptide whereas iKeap9 (**f**, **h**) is not able to effectively displace the NRF2 peptide. The fluorescence polarization assay was performed with three technical replicates per concentration measured. Data are mean ± s.d. for each titration point shown together with the fitted curve.



Extended Data Fig. 8 | Displacers validated by fluorescence polarization and BLI. Here we show two more displacers, iKeap7 and iKeap22. **a**, **b**, Both iKeap7 (**a**) and iKeap22 (**b**) were confirmed as binders by SPR. **c**, **d**, Ligand-detection NMR experiments show that both iKeap7 (**c**) and iKeap22 (**d**) bind to KEAP1. **e**, iKeap7 is confirmed to be a displacer of the NRF2 peptide by both fluorescence polarization and BLI (data not shown). **f**, As the fluorescence polarization experiments of iKeap22 were affected by autofluorescence, BLI was needed to confirm that this compounds is a displacer. The fluorescence polarization assay was performed with three technical replicates per concentration measured. Data are mean ± s.d. for each titration point, shown along with the fitted curve. Two independent BLI experiments were performed with similar results and one representative result is shown here.



Extended Data Fig. 9 | NRF2 peptide- and ligand-binding sites, rationale for binder versus displacer. Here we show the docking pose of one of the hit compounds (iKeap9, green ball-and-stick representation) bound to KEAP1, together with the NRF2 peptide (PDB ID: 4IFL; peptide in violet). iKeap9 is a tight binder (180 nM by steady-state SPR) but cannot displace NRF2. Left, the top view. Right, the side view of the cross-section of KEAP1 along the central plane. The violet box indicates the docking region (where the ligands were allowed to bind), which was used in the virtual screening. The site of interest includes a part of the deep pocket/tunnel of the β-barrel-shaped KEAP1, as it enables ligands to bind more tightly by insertion into the channel than on a shallow surface. However, the deep tunnel is largely non-overlapping with the peptide-binding site (which binds to the entrance site of the tunnel). Thus, binding molecules might only partially interfere with peptide binding, which could reduce or eliminate the ability of small-molecule binders to displace the peptide. The ability of a small molecule to displace the peptide is difficult to predict, and was not attempted in this study. In some cases, small molecules can also act as molecular glues and strengthen the interaction between NRF2 and KEAP1.

nature research

Corresponding author(s): Gerhard Wagner, Haribabu Arthanari, Christoph Gorgulla

Last updated by author(s): Feb 18, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Software Autodock Vina, Smina, QuickVina and Topspin (Bruker) used in this study are all publicly available. The platform VirtualFlow which was developed for this manuscript is freely available on GitHub the following URL https://github.com/VirtualFlow Additional resources inducing documentation, ligand libraries, tutorials and video demonstration are freely available at the VirtualFlow homepage http://virtual-flow.org/
Data analysis	DataWarrior, Python, Maestro, Matlab, Matlab, R, TopSpin (Bruker); (all publicly available)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The ready-to-dock library from Enamine is available online on the homepage of VirtualFlow at http://virtual-flow.org/real-library. Other datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In experiments where we performed replicated the sample size and the nature of the replicated (technical and/or independent) is clearly mentioned. Sample size is not applicable to some of experimental measurements (NMR) described in this manuscript. The manuscript contains no animal data.
Data exclusions	No data was excluded in this study.
Replication	We had performed replicates (independent and/or technical replicates) for the experiments that were used to extract quantitative information, such as Kd (SPR) and IC50 (FP) and DLS. We have updated the figure captions and the methods to reflect this and provided explicit details of how the replicates were done. We have also provided the source data where applicable. Data replication does not apply to NMR measurements described in this manuscript.
Randomization	Randomization does not apply to this study since we are not comparing data across samples or cohorts.
Blinding	Blinding is not relevant since the study does not compare different groups of measurements

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

nature

https://doi.org/10.1038/s41586-020-2117-z

Supplementary information

An open-source drug discovery platform enables ultra-large virtual screens

In the format provided by the authors and unedited

Christoph Gorgulla✉, Andras Boeszoermenyi, Zi-Fu Wang, Patrick D. Fischer, Paul W. Coote, Krishna M. Padmanabha Das, Yehor S. Malets, Dmytro S. Radchenko, Yurii S. Moroz, David A. Scott, Konstantin Fackeldey, Moritz Hoffmann, Iryna Iavniuk, Gerhard Wagner & Haribabu Arthanari✉

Contents

A KEAP1 binders from the literature

Supplementary Table 1 lists the experimentally verified binders to the NRF2-binding domain of KEAP1 found in[9,25,44,45], which were added to the primary virtual screening for the purpose of further validation of VirtualFlow. The threshold docking score (resembling the free energy of binding ΔG) for the top 10% of the compounds is $-8.6\,\mathrm{kcal/mol}$. All 17 of the compounds have a predicted docking score above that threshold, indicating that the docking procedure in the primary virtual screening has worked well, despite the use of the lowest possible docking accuracy.

Compound	SMILES	IC_{50} (μM)	K_d (μM)	Docking Score (kcal/mol)
C1	O=C(O)[C@@H]1CCCC[C@@H]1C(=O)N3CCc2ccccc2C3Cn5c(=O)c4ccccc4c5=O	3		-8.7
C2	O=C(O)[C@H]1CCCC[C@H]1C(=O)N1CCc2ccccc2[C@@H]1CN1C(=O)c2ccccc2C1=O	2.3		-9.0
C3	O=C(O)[C@H]1CCC[C@H]1C(=O)N3CCc2ccccc2[C@H]3Cn5c(=O)c4ccccc4c5=O	2.2		-9.8
C4	O=C(O)[C@H]1CC[C@H]1C(=O)N3CCc2ccccc2[C@H]3Cn5c(=O)c4ccccc4c5=O	8.0		-9.3
C5	O=C(O)[C@H]1C[C@H]1C(=O)N3CCc2ccccc2[C@H]3Cn5c(=O)c4ccccc4c5=O	20.8		-9.6
C6	O=C(O)[C@H]1CCNC[C@H]1C(=O)N3CCc2ccccc2[C@H]3Cn5c(=O)c4ccccc4c5=O	69.7		-8.7
C7	O=C(O)[C@H]1CCCC[C@H]1C(=O)N1CCc2ccccc2[C@H]1CN1Cc2ccccc2C1=O	1.1		-9.5
C8	O=C(O)[C@H]1CCCC[C@H]1C(=O)N3CCc2ccccc2[C@H]3CN4C(=O)CCC4=O	1.2		-9.1
C9	O=C1c2ccccc2C(=O)N1C[C@@H]1c2ccccc2CCN1C(=O)[C@@H]1CCCC[C@@H]1c1nnn[nH]1	7.4		-10.8
C10	Cc1cc(C)c(C)c(S(=O)(=O)Nc2ccc(N3CC[C@@H](C(=O)O)C3)c3ccccc23)c1C	0.14		-10.5
C11	CC(C)c1ccc(S(=O)(=O)Nc2cc(Sc3ncn[nH]3)c(O)c3ccccc23)cc1		2.9	-9.4
C12	O=C(O)c4ccc(c3ccc(C=c2c(=O)[nH]n(c1cccc(Cl)c1)c2=O)o3)cc4Cl		15.2	-8.8
C13	Cc1ccccc1NC(=O)Cn4c(=O)sc(=CC2CCCN2c3cccc(C(=O)O)c3)c4=O		10.4	-10.0
C14	CN(Cc1cc([C@H](CC(=O)O)c2ccc3c(c2)nnn3C)ccc1Cl)S(C)(=O)=O	3.4		-8.7
C15	CN(Cc1cc([C@H](CC(=O)O)c2ccc3c(c2)nnn3C)ccc1Cl)S(=O)(=O)c1ccccc1	0.27		-9.7
C16	Cc3ccc([C@H](CC(=O)O)c1ccc2c(c1)nnn2C)cc3CN5C[C@@H](C)Oc4ccccc4S5(=O)=O	0.015		-10.8
C17	COc1ccc(S(=O)(=O)Nc2ccc(NS(=O)(=O)c3ccc(OC)cc3)c3ccccc23)cc1	2.7		-9.9

Supplementary Table 1: Previously identified binders to the NRF2-binding domain of KEAP1. Shown are the SMILES-formatted chemical structures, the reported IC_{50} values, and the predicted docking scores from the primary virtual screen. The goal of the primary virtual screen was to distinguish binders from non-binders for demonstration purposes, and the stringency was set to the lowest possible level.

In Extended Data Fig. 5a the crystal structure which was used for the virtual screening procedure is shown, as well as the structure of iKeap1 (Extended Data Fig. 5c). The structure of iKeap1 is similar to a previously published inhibitor of KEAP1, shown as compound C17 (N,N'-Naphthalene-1,4-Diylbis(4-Methoxybenzenesulfonamide) in Supplementary Table 1 and Extended Data Fig. 5d[25]. The predicted docking position for iKeap1 (Fig. 3a) is very similar to the co-crystal structure of compound C17 (Extended Data Fig. 5d). These similarities further substantiate that iKeap1 binds to the KEAP-1-NRF2 interface.

B Details about the experimental validation of the hits

The SPR Level-2 compounds can be filtered using several characteristics, such as their suitability for medicinal chemistry, predicted pan assay inference (PAINS), or ability to displace the NRF2 peptide. In Table 2 we are listing the hit compounds which are presented in the manuscript. Also included in the table are all the other hits which are able to displace the NRF2 peptide, which do not harbour any PAINS substructures, and which do not contain other problematic substructures (such as azo-dye compounds or compounds which are unsuitable for medicinal chemistry). It should be noted that, in addition to the compounds listed here, our experimental validation identified other potent displacers and binders which were filtered out due PAINS and/or non-drug like properties.

ID	SMILES	FP Assay	BLI Assay
iKeap1	Cc1ccc(S(=O)(=O)Oc2nc3nc4ccccc4nc3nc2OS(=O)(=O)c2ccc(C)cc2)cc1	active	active
iKeap2	O=C(O)c1ccc(NC2=C/C(=N\S(=O)(=O)c3ccc4ccccc4c3)c3ccccc3C2=O)cc1	active	active
iKeap7	CC1=NN(c2ccccc2)C(=O)/C1=C\c1ccc(-c2cc(C(=O)O)ccc2Cl)o1	active	active
iKeap8	O=C(Nc1cccc(-c2nnn[nH]2)c1)[C@@H]1C[C@H]2CCCC[C@H]2N1C(=O)c1ccc2ccccc2c1	active	active
iKeap9	O=S(=O)(N=C1CCCCCN1)C1=CC=C2SC(NS(=O)(=O)C3=CC=C4C=CC=CC4=C3)=NC2=C1	not active	not active
iKeap12	O=C(c1ccccc1)C1=C[C@@H]2[C@@H]3C(=O)N(c4cccc5ccccc54)C(=O)[C@@H]3[C@@H](C(=O)c3ccc(Cl)cc3)N2C=C1	active	active
iKeap22	O=c1cc(-c2ccc(O)cc2)oc2c1c(O)cc1c2[C@H](c2ccc3ncccc3c2)CC(=O)O1	not active	active
iKeap27	Cc1ccc(C)c(N2C(=O)c3ccc(-c4nc(-c5ccc6[nH]nnc6c5)no4)cc3C2=O)c1	not active	active
iKeap33	O=S(=O)(Nc1nc2ccccc2nc1N1CCC[C@@H](c2nc3ccccc3[nH]2)C1)c1ccccc1	not active	active
iKeap41	O=C(c1cc(-c2ccc3c(c2)OCCO3)nc2ccccc21)N1CCC(Cc2ccccc2)CC1	not active	active
iKeap48	CC(=O)N/C(=C\c1ccc(Cc2nc3c([nH]2)C(=O)c2ccccc2C3=O)cc1)c1nc2c([nH]1)C(=O)c1ccccc1C2=O	not active	active
iKeap73	C[C@H]1CCc2c(sc3ncnc(N4CCO[C@@H](CN5C(=O)c6ccccc6C5=O)C4)c23)C1	not active	active
iKeap74	C[C@@H]1Oc2ccc(C(=O)C3CCN(C(=O)[C@H]4C[C@H]4c4ccccc5ccccc45)CC3)cc2NC1=O	not active	active
iKeap75	Cc1cccn2c(=O)c3c(nc12)N1CCCCC[C@H]1[C@]1(C3)C(=O)N(Cc2ccccc2Cl)C(=O)N=C1O	not active	active

Supplementary Table 2: SPR Level-2 hit compounds, which were also tested experimentally via a fluorescence polarization (FP) assay and BLI experiments for their ability to displace the NRF2-peptide. All compounds shown in this table have also been verified by protein detected NMR experiments (STD, CPMG, DLB, ^{1}H-^{13}C HMQC), or have passed additional filters to remove compounds with problematic substructures (regarding pan assay interference or suitability for medicinal chemistry).

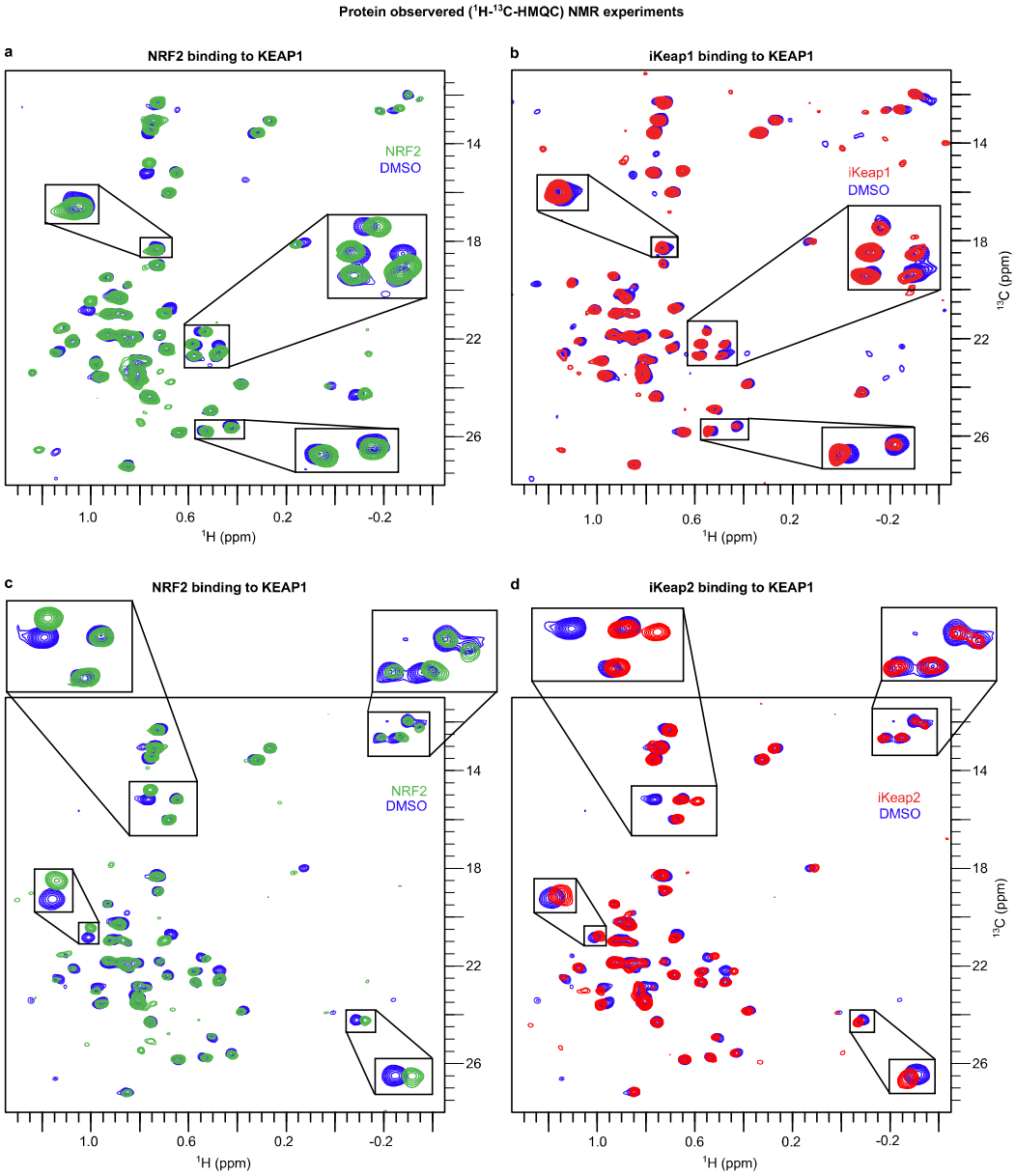
Two of the hit compounds (iKeap2 and iKeap7) contain PAINS alerts. We have carried a suite of orthogonal experiments to show that the compounds are indeed true and specific binders, and are not artifacts due to pan assay interference. DLS and 1D NMR show the compounds are not aggregating, and they are not similar to any known aggregators as assessed by the Tanimoto similarity measure[39]. STD-NMR and R2 measurements show the compounds are not covalently binding. The FP assay shows that the compounds bind at the targeted peptide-binding site by displacing it. The FP assay is done in the presence of BSA to account for non-specific binding. Protein-observed NMR experiments (^{1}H-^{13}C-HMQC) clearly show the compounds site-specifically engage KEAP1 in a manner similar to NRF2. The protein-detected NMR experiments show that both iKeap2 and iKeap7 do not aggregate the protein (KEAP1).

Compound	$K_{d,1}$ (nM)	$K_{d,2}$ (nM)	$K_{d,3}$ (nM)	χ_1^2	χ_2^2	χ_3^2	$K_d^{average}$ (nM)	SD (nM)
ikeap1*	114	178	85	0.026	0.404	0.033	125.667	47.585
iKeap2	158	263	334	0.031	0.333	0.265	251.667	88.546
iKeap7	23	15	16	0.0305	0.306	0.019	18.000	4.359
iKeap8*	176	115	187	0.118	0.130	0.397	159.333	38.786
iKeap9*	180	187	190	0.278	0.242	0.158	185.667	5.132
iKeap22*	62	51	52	0.082	0.109	0.051	55.000	6.083

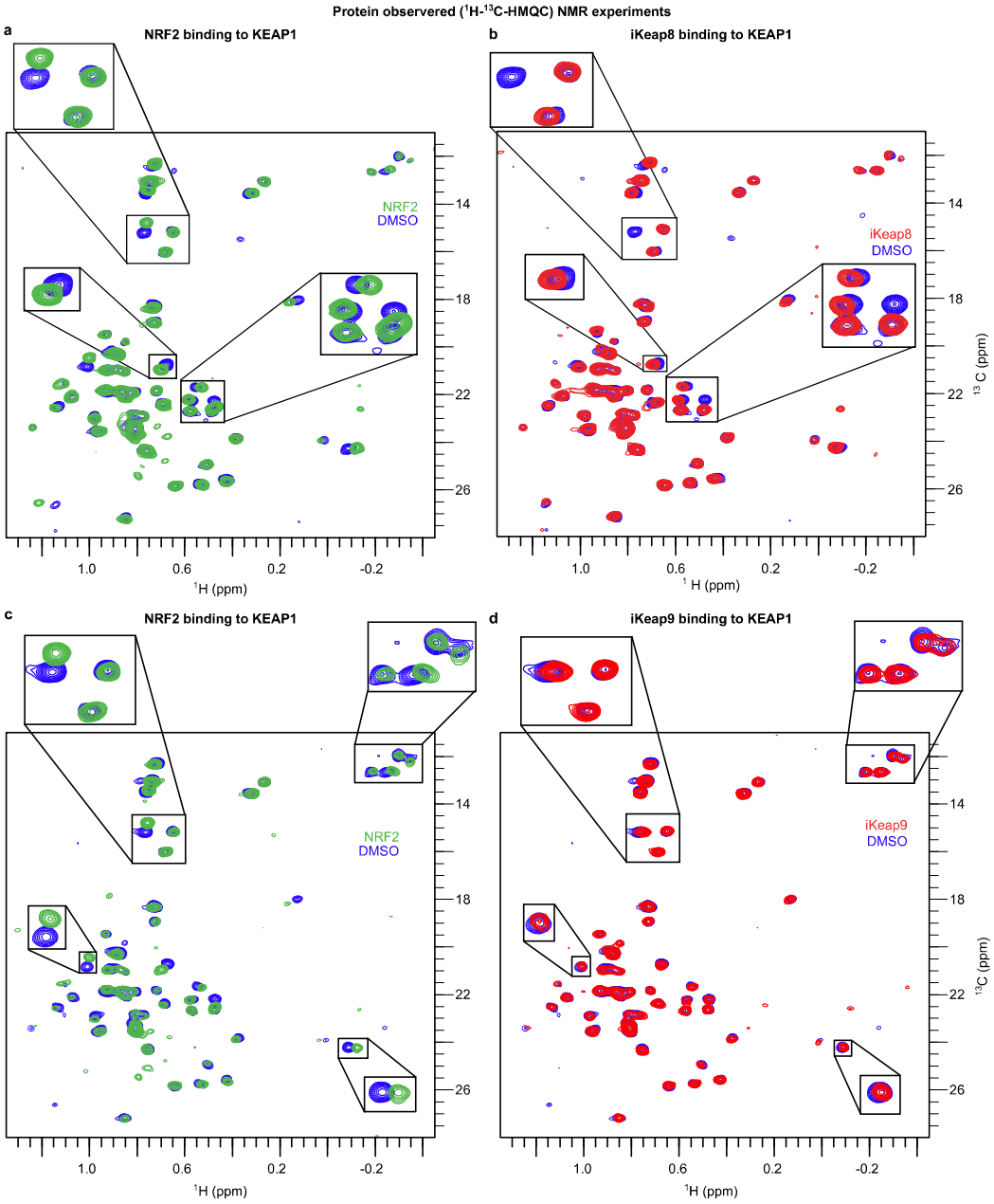
Supplementary Table 3: K_d values which were determined by SPR Level-3 experiments of the compounds shown in more detail in the manuscript (Fig. 3, Extended Data Figs. 7, 8). For each compound three independent experiments were carried out, giving rise to three $K_{d,i}$ and χ_i^2 values, where i indicates the independent experiment. Compounds marked with an asterisk were determined by fitting the biphasic binding model given by equation (6) to the experimental data, for the other K_d values the one-site binding model given by equation (5) was used. The last column (SD) contains the standard deviation of the average K_d value.

Compound	IC_{50} (μM)	Standard Error (μM)
iKeap1	0.258	0.037
iKeap2	2.7	0.6
iKeap7	38.2	3.9
iKeap8	14.2	4.0

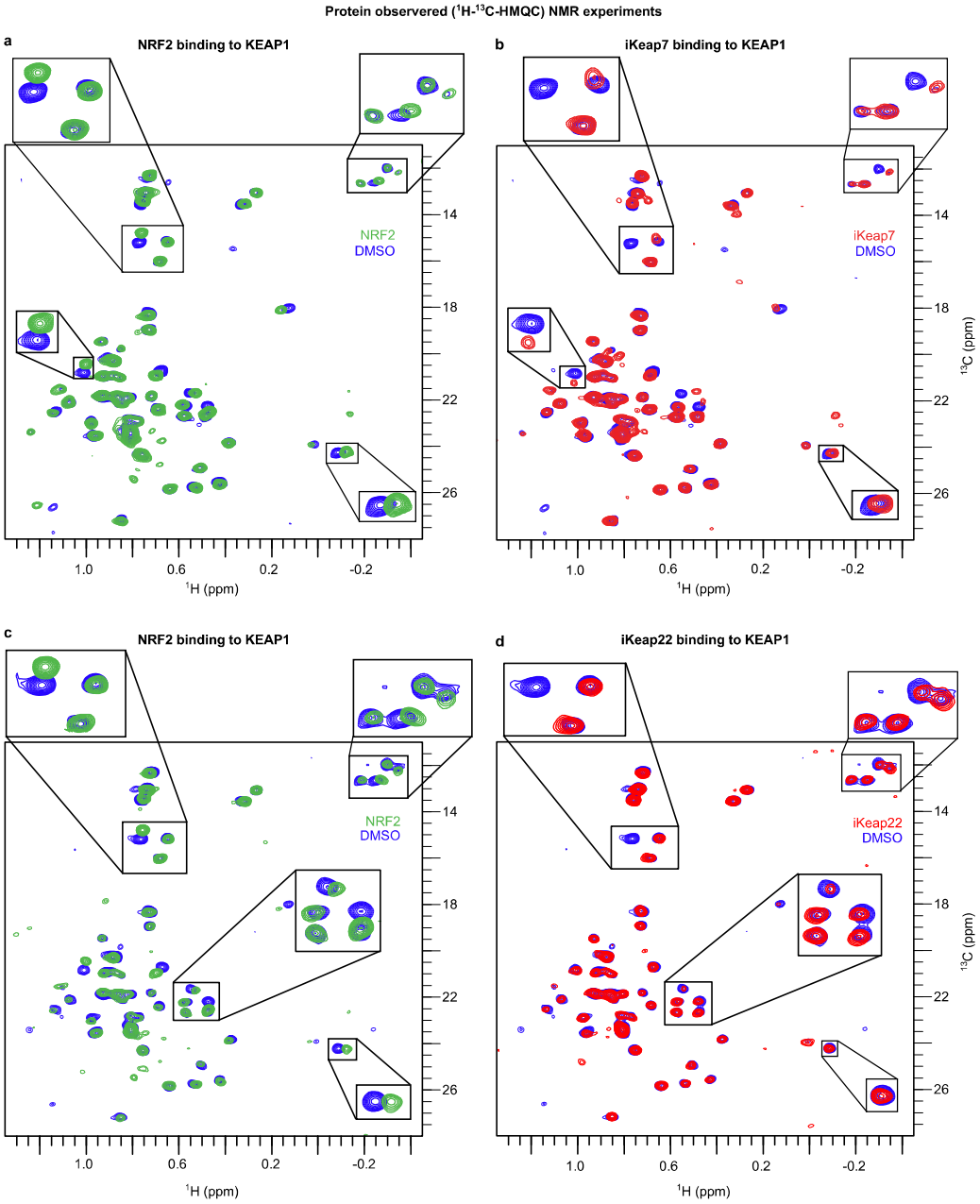
Supplementary Table 4: IC_{50} values and the associated errors which were determined by the Level-2 FP experiments are shown here for the compounds highlighted in the manuscript. Representative curves for these FP hits are shown in more detail in the manuscript (Fig. 3, Extended Data Fig. 7 and 8).



Supplementary Fig. 1: ^{1}H-^{13}C HMQC experiments showing the binding of the NRF2 peptide (**a**, **c**), iKeap1 (**b**) and iKeap2 (**d**) to KEAP1 as monitored by chemical shift perturbation to the methyl resonances of Ile, Leu and Val of KEAP1. Upon addition of iKeap1 and iKeap2 we see selective and specific changes to a subset of resonances and these correlate to the changes we observe when we add the NRF2 peptide. The rest of the resonances are largely unaffected. This indicates that the protein is folded and does not aggregate after the addition of the compounds.



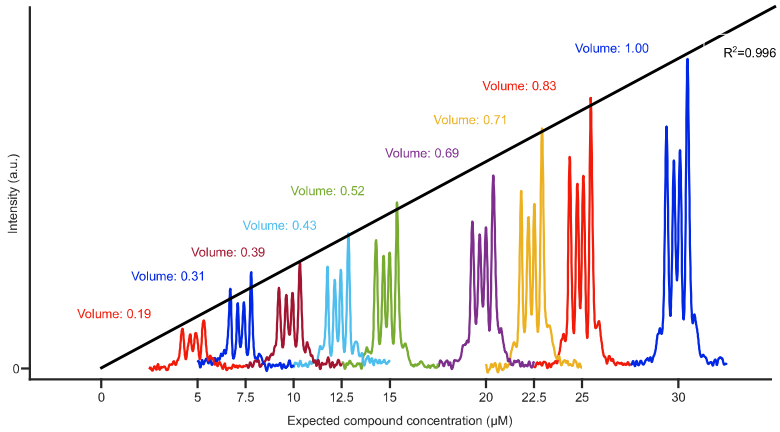
Supplementary Fig. 2: ^{1}H-^{13}C HMQC experiments showing the binding of the NRF2 peptide (**a**, **c**), iKeap8 (**b**) and iKeap9 (**d**) to KEAP1 as monitored by chemical shift perturbation to the methyl resonances of Ile, Leu and Val of KEAP1. Upon addition of iKeap8 and iKeap9 we see selective and specific changes to a subset of resonances and these correlate to the changes we observe when we add the NRF2 peptide. The rest of the resonances are largely unaffected. This indicates that the protein is folded and does not aggregate after the addition of the compounds.



Supplementary Fig. 3: ^{1}H-^{13}C HMQC experiments showing the binding of the NRF2 peptide (**a**, **c**), iKeap7 (**b**) and iKeap22 (**d**) to KEAP1 as monitored by chemical shift perturbation to the methyl resonances of Ile, Leu and Val of KEAP1. Upon addition of iKeap7 and iKeap22 we see selective and specific changes to a subset of resonances and these correlate to the changes we observe when we add the NRF2 peptide. The rest of the resonances are largely unaffected. This indicates that the protein is folded and does not aggregate after the addition of the compounds.

Compound	Radius (nm)	Standard Error (nm)
buffer	8.781	6.782
iKeap1	2.306	2.534
iKeap2	12.492	2.008
iKeap7	3.663	3.449
iKeap8	3.340	1.117
iKeap9	1.862	0.603
iKeap12	44.780	31.646
iKeap22	2.547	1.890
iKeap27	2.643	3.017
iKeap33	1.585	1.779
iKeap41	4.446	0.887
iKeap48	25.579	31.497
iKeap73	2.510	1.583
iKeap74	3.034	1.972
iKeap75	3.612	1.527

Supplementary Table 5: DLS data for the hit compounds listed in Supplementary Table 2 showing that these compounds are not aggregating. According to[39] colloidal aggregates of compounds have radii in the range of 50-800 nm. All measurements were done in triplicates and the average and standard error is shown here.



Supplementary Fig. 4: NMR solubility assay of iKeap1. Determination of the solubility of iKeap1 via an NMR solubility assay as described in[40]. As can be seen in the figure, the NMR intensity of iKeap1 remains linear over the range of concentrations measured here, indicating that iKeap1 does not aggregate at these concentrations.

C Web interface to the REAL database

The ready-to-dock REAL database that we have prepared is freely available on the VirtualFlow homepage (`https://virtual-flow.org/real-library`). In order to allow the user to select specific subsets of the database for download or screening with VFVS, the REAL database was divided into a six-dimensional lattice (table), where each dimension corresponds to one of the following properties:

- Molecular weight (MW)
- Partition coefficient (SlogP)
- Number of hydrogen bond acceptors (HBA)
- Number of hydrogen bond donors (HBD)
- Number of rotatable bonds (RotB)
- Topological polar surface area (TPSA)

For each of these properties, a set of bins (tranches) was created into which the compounds are classified (Supplementary Table 6).

Tranche	MW	SlogP	HBA	HBD	RotB	TPSA
A	0 to 200	-∞ to -1	0 to 1	0 to 1	0 to 1	0 to 20
B	200 to 250	-1 to 0	1 to 3	2	2 to 3	20 to 40
C	250 to 300	0 to 1	3 to 5	3	4 to 5	40 to 60
D	300 to 325	1 to 2	5 to 7	4	6 to 7	60 to 80
E	225 to 350	2 to 2.5	7 to 9	5	8 to 9	80 to 100
F	350 to 375	2.5 to 3	9 to 10	6 to ∞	10 to ∞	100 to 120
G	375 to 400	3 to 3.5	11 to ∞			120 to 140
H	400 to 425	3.5 to 4				140 to ∞
I	425 to 450	4 to 4.5				
J	450 to 500	4.5 to 5				
K		5 to ∞				

Supplementary Table 6: Tranche-table used for the REAL database. The compounds in the REAL database were sorted into a six-dimensional lattice. Each dimension corresponds to one of six physico-chemical properties: (1) Molecular weight (MW), (2) partition coefficient (SlogP), (3) number of hydrogen bond acceptors (HBA), (4) number of hydrogen bond donors (HBD), (5) number of rotatable bonds (RotB), and (6) topological polar surface area (TPSA). For each dimension/property a number of bins/tranches were created as specified in the above table, into which the compounds were classified.

In order to allow users to visually select subsets of the REAL database, an interactive two-dimensional tranche-table was created which is a projection of the six-dimensional lattice (Supplementary Fig. 5), which was inspired by the design of the ZINC 15 database. For each of the two dimensions of the table, one of the six physico-chemical properties from above can be selected. For each property, a slider allows the user to select a specific range of interest.

MW ▾

SLogP ▾	0 - 200	200 - 250	250 - 300	300 - 325	325 - 350	350 - 375	375 - 400	400 - 425	425 - 450	450 - 500	Row sums:
-∞ - -1	30K	300K	1.63M	2.87M	5.42M	1.08M	237K	54.6K	27.8K	16K	11.7M
-1 - 0	178K	1.58M	7.91M	11.9M	23M	5.03M	1.76M	596K	330K	227K	52.5M
0 - 1	474K	4.99M	23.3M	32.6M	64.4M	17.4M	8.73M	3.45M	2.29M	1.5M	159M
1 - 2	604K	8.16M	43.1M	**57.8M**	**116M**	**41M**	**26.4M**	**12.3M**	8.73M	5.92M	320M
2 - 2.5	221K	3.83M	23.1M	**33M**	**67.3M**	**29.1M**	**21.6M**	**11.9M**	8.77M	6.08M	205M
2.5 - 3	118K	2.83M	20.6M	**29.8M**	**60.4M**	**31.7M**	**26M**	**16.2M**	12.5M	8.59M	209M
3 - 3.5	48.7K	1.66M	14.6M	22.4M	46.9M	30.7M	27.3M	19.7M	15.7M	10.9M	190M
3.5 - 4	10.4K	678K	8.1M	11.7M	18.4M	22.9M	24.9M	21M	17.5M	12.3M	137M
4 - 4.5	707	176K	3.56M	6.56M	11.8M	16.9M	19.7M	19.1M	16.9M	12.2M	107M
4.5 - 5	0	30K	1.13M	2.78M	6.1M	10M	13.4M	14.5M	13.8M	10.4M	72.2M
5 - ∞	0	0	98	301	836	1.91K	2.93K	5.31K	5.36K	6.64K	23.4K
Column sums:	1.69M	24.2M	147M	211M	420M	206M	170M	119M	96.6M	68.2M	

Compounds selected: 581M **from** 1.46B **Tranches selected:** 9.76K **from** 46.6K

Molecular Weight (MW): 300 to 425

0 200 250 **300** 325 350 375 400 **425** 450 500

Hydrogen Bond Acceptors (HBA): 0 to 10

0 1 3 5 7 9 **10**

SLogP: 1 to 3

-∞ -1 0 **1** 2 2.5 **3** 3.5 4 4.5 5 ∞

Hydrogen Bond Donors (HBD): 0 to ∞

0 1 2 3 4 5 ∞

Topological Polar Surface Area (TPSA): 0 to 140

0 20 40 60 80 100 120 **140**

Rotatable Bonds (RotB): 0 to 10

0 1 3 5 7 9 **10**

Download format: PowerShell ▾ Download

Supplementary Fig. 5: Interactive tranche-table of the REAL database of Enamine, which was prepared in a ready-to-dock format with VFLP. The REAL database was divided into a six-dimensional lattice, where each dimension corresponds to one of the following physico-chemical properties: (1) Molecular weight (MW), (2) partition coefficient (SlogP), (3) number of hydrogen bond acceptors (HBA), (4) number of hydrogen bond donors (HBD), (5) number of rotatable bonds (RotB), and (6) topological polar surface area (TPSA). The tranche table is a two-dimensional projection of the six-dimensional lattice. The numbers within the body of the table represent the number of molecules which are contained in the corresponding cell (taking into account the selected ranges of the remaining four dimensions of the table, which are not explicitly visible). The coloring of the cells is based on the number of compounds they contain, the more molecules the darker the cell color.

D Probabilistic model of the true-positive rate in virtual screens

In this section we define the true hit rate (which is complementary to the false-positive rate) of a virtual screen and show that this rate increases with the number of compounds screened. A compound is a hit compound if its real (true) binding affinity surpasses a certain threshold (which is freely selectable). The predicted binding affinity is subject to errors resulting from inaccuracy in the docking methods. This error means that some compounds with binding affinities predicted to be above the binding affinity threshold can have true binding affinities below the threshold, resulting in false (predicted) positives. The impact of this effect reduces when increasing the size of the ligand library which is screened, because a larger (well-designed) library allows for finer sampling of the density of binding strengths in the chemical space and thus identification of higher affinity hits than might be found in a smaller library. This means, that for a fixed threshold, on average the computed affinities of the top hits are higher than the selected threshold, decreasing the probability that the true affinity falls below the hit cut-off affinity (threshold).

Let L denote the ligand and R the receptor, then the binding affinity ΔA_{bind} is defined as

$$\Delta A_{\text{bind}} = A_{\text{RL}} - A_{\text{R+L}}, \tag{7}$$

where A is the thermodynamic free energy (e.g. the Gibbs free energy if the isothermal-isobaric ensemble is considered), $\text{R}+\text{L}$ denotes the unbound state of the ligand and the receptor, and RL the bound state of the ligand-receptor system. The more negative the value of the ΔA_{bind} the stronger the binding affinity. For the purpose of simplifying the discussion, we define a more intuitive value, the *binding strength* B, to be the additive inverse of the binding affinity, *i.e.*,

$$B := -\Delta A_{\text{bind}} \tag{8}$$

The unit of free energies, and therefore also B, is normally expressed as a molar quantity, for instance kcal/mol or kJ/mol (an SI derived unit).

In virtual screening procedures, the inclusion of more compounds in the input collections has a positive effect on the true hit rates. For a given virtual screening procedure we define the true hit rate $R_{\text{THR}}(b,K)$ by

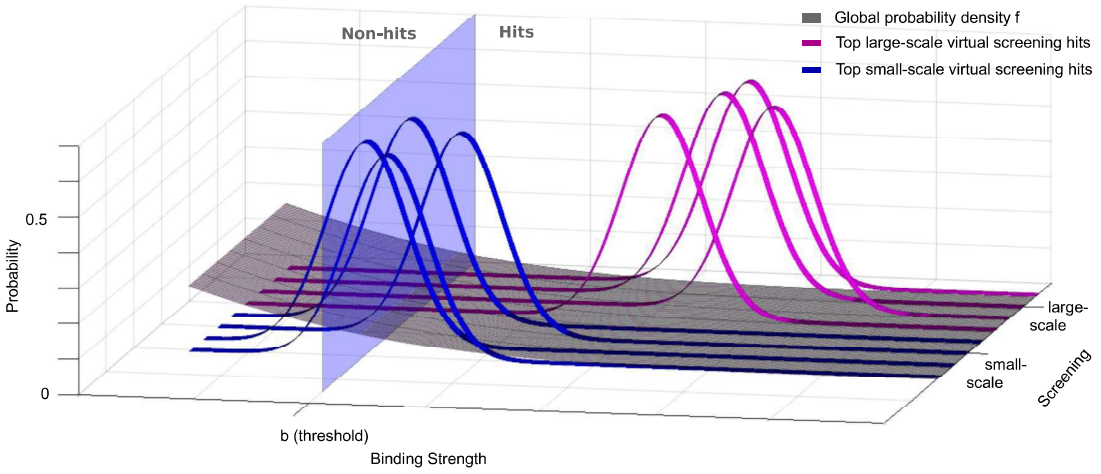
$$R_{\text{THR}}(b,K) = \frac{K_b^{\text{real}}}{K}, \tag{9}$$

where b is the binding strength threshold which needs to be surpassed for a compound being considered a true hit, K is the number of the highest-scoring virtual screening hits which are considered, and K_b^{real} is the number of these which are also hits in reality. In real drug discovery projects, b is often chosen such that it corresponds to a K_{d} value in the low (e.g. single digit) µM range. For very challenging targets, compounds with values in the 100 µM range or higher are sometimes considered to be hits as well. For a fixed binding strength threshold $b \in \mathbb{R}$ and fixed top score bracket size K, the true hit rate will improve with increasing input collection size N. Higher N will better sample rare, high-affinity compounds, resulting in higher predicted binding strengths for the K top-scoring compounds, which therefore in turn will increase the distance between the estimated binding strengths of these top compounds and the threshold value b. It is assumed here that the distribution of the predicted binding strength of these compounds was independent of the precise compound. However, we used the best estimation for the expectation value of this distribution as the predicted binding strength. Therefore, it follows that the higher the predicted binding strength the more likely it is that true binding strength surpasses the threshold b, even if the true value is lower than the predicted value, which means that the true hit rate $R_{\text{THR}}(b,K)$ rises on average with the number of compounds screened. This principle is illustrated in Supplementary Fig. 6. This effect is independent of the precise probability distributions involved.

We will briefly look at a more specific example, in which we only consider the single highest-scoring compound (i.e. $K = 1$) with predicted binding strength B_1^{virt}. If we assume that the true binding strength B_1^{real} is normally distributed with expectation value B_1^{virt} and variance σ, i.e. the probability density g is given by

$$g(x \mid B_1^{\text{virt}}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\left(x - B_1^{\text{virt}}\right)^2}{2\sigma^2}}$$

Then, the average hit rate $R_{\text{THR}}^{\text{ave}}(B_1^{\text{virt}}, b)$, which is the probability that the true binding strength B_1^{real} is higher than the threshold b, is given by



Supplementary Fig. 6: Qualitative diagram showing the probability distribution (g) of the real binding affinity of the four top hit compounds of a small-scale (blue) and a large-scale (magenta) virtual screen. Each of the compounds has a predicted binding strength B^{virt} from the docking procedure, but the real binding affinity can be different due to the error associated with molecular dockings. Here it is assumed that the true binding affinity is normally distributed with an expectation value centered at the predicted binding affinity. The standard deviation was assumed to be independent of the compound, therefore the shape of the distribution is independent of the scale of the screening. If the real binding affinity of a compound surpasses a certain threshold value (blue transparent plain), it is classified as a hit compound. Thus, the higher the predicted binding affinity of a virtual (predicted) hit compound, the more likely it is that the compound is also a true hit compound. And the larger the scale of the virtual screen, the larger the predicted binding strength of the top-scoring compounds on average. Therefore, the true-hit rate of the top-scoring compounds improves with the size of the screening. The gray surface represents the global probability density which describes how likely it is to find a compound with a certain binding affinity, which decreases with increasing binding strength.

$$R^{\text{ave}}_{\text{THR}}(B_1^{\text{virt}}, b) = \mathbb{P}(B > b) \tag{10}$$

$$= \int_b^\infty g(x \mid B_1^{\text{virt}}, \sigma^2) dx \tag{11}$$

$$= \int_b^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(x - B_1^{\text{virt}}\right)^2}{2\sigma^2}} dx \tag{12}$$

$$= \int_{-\infty}^{B_1^{\text{virt}} + \Delta B_1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(x - B_1^{\text{virt}}\right)^2}{2\sigma^2}} dx, \tag{13}$$

where

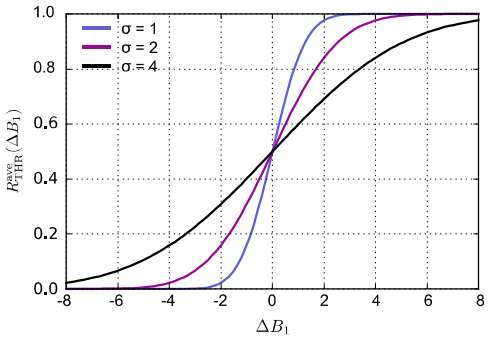
$$\Delta B_1 := B_1^{\text{virt}} - b \tag{14}$$

is the difference between the predicted binding strength and the threshold value. Equation (13) follows by the symmetry of normal distribution. Inserting (14) into (13) and evaluating the integral thus yields

$$R^{\text{ave}}_{\text{THR}}(B_1^{\text{virt}}, b) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{B_1^{\text{virt}} + \Delta B_1 - B_1^{\text{virt}}}{\sigma\sqrt{2}}\right)\right) \tag{15}$$

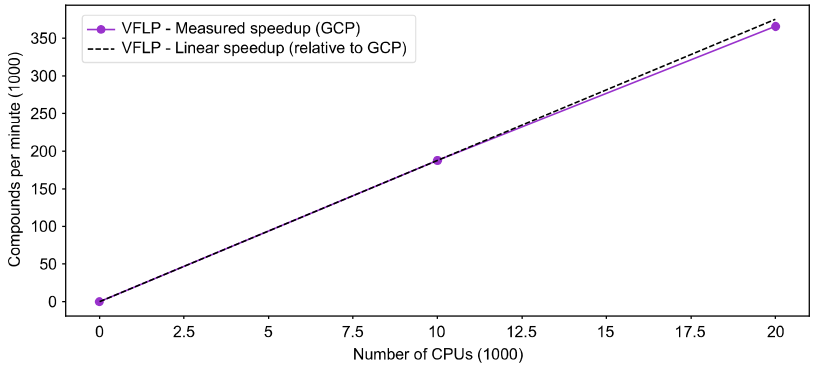
$$= \frac{1}{2}\left(1 + \text{erf}\left(\frac{\Delta B_1}{\sigma\sqrt{2}}\right)\right), \tag{16}$$

where erf is the error function. In other words, $R^{\text{ave}}_{\text{THR}}$ is the probability that the true binding strength lies between $B_1^{\text{virt}} - b$. It thus follows, that the larger the distance between the threshold and the predicted binding strength B_1^{virt}, the higher is the average hit rate. $R^{\text{ave}}_{\text{THR}}$ is plotted in Supplementary Fig. 7 for different values of the standard deviation σ, clearly illustrating the favourable dependence of the average true hit rate of the highest-scoring compound on ΔB_1.
Fast docking functions, which are normally used in virtual screening, generally have high standard deviations, making large ΔB_1 values particularly favourable. This positive effect on the true hit rate can be generalized to $K > 1$ top compounds, as the effect is the same.



Supplementary Fig. 7: The average true hit rate for the single highest-scoring compound, $R^{\mathrm{ave}}_{\mathrm{THR}}$, is shown as a function of $\Delta B_1(B_1^{\mathrm{virt}}, b)$ in kcal/mol for different values of σ. Here b is the binding strength threshold which a compound needs to surpass in order to be classified as a hit, and B_1^{virt} is the binding affinity predicted by the virtual screen with an associated standard deviation σ.

E VFLP - Scaling behaviour and file formats



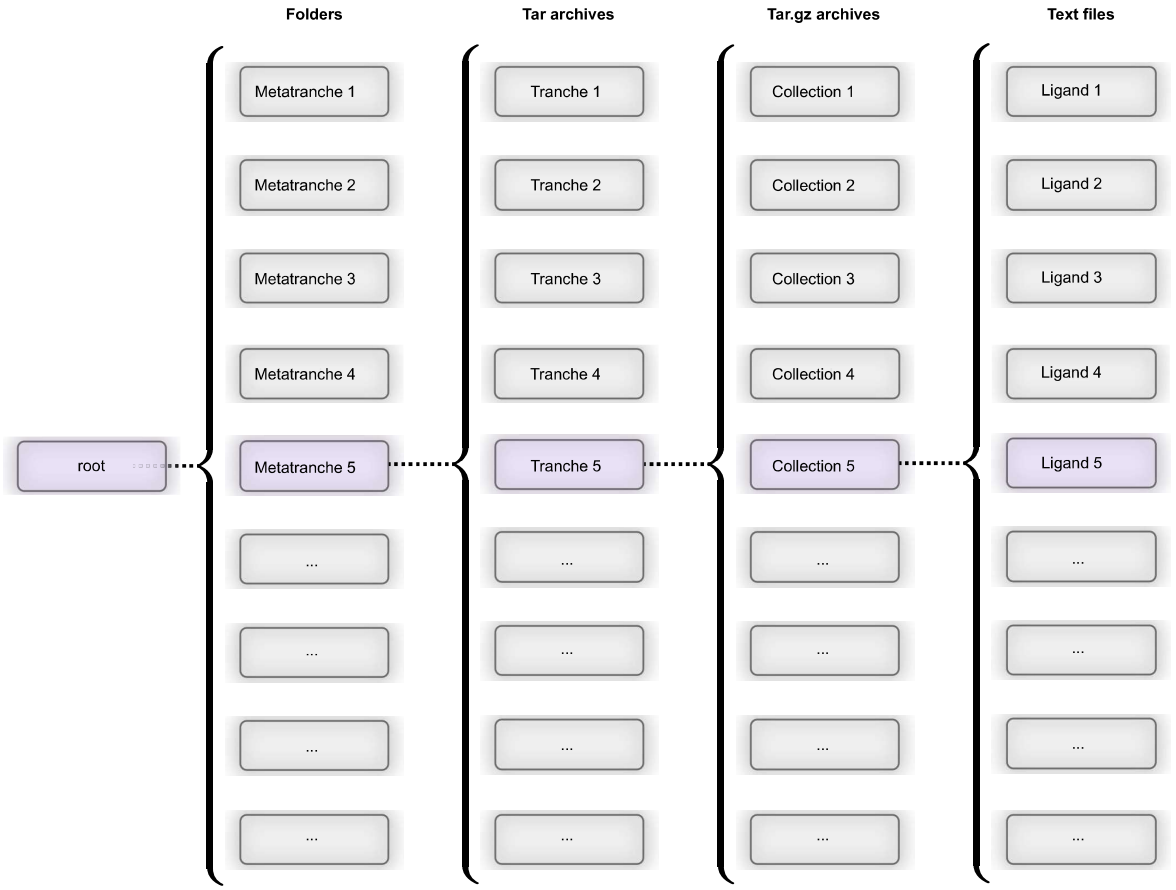
Supplementary Fig. 8: VFLP scaling behaviour on the Google Cloud Platform using up to 20,000 CPUs in parallel on n1-highcpu-16 virtual machines.

File extension	Proper Name	File extension	Proper Name
acesin	ACES input format	lmpdat	The LAMMPS data format
adf	ADF cartesian input format	mcdl	MCDL format
alc	Alchemy format	mcif	Macromolecular Crystallographic Info
ascii	ASCII format	mdl	MDL MOL format
bgf	MSI BGF format	ml2	Sybyl Mol2 format
box	Dock 3.5 Box format	mmcif	Macromolecular Crystallographic Info
bs	Ball and Stick format	mmd	MacroModel format
c3d1	Chem3D Cartesian 1 format	mmod	MacroModel format
c3d2	Chem3D Cartesian 2 format	mna	Multilevel Neighborhoods of Atoms (MNA)
cac	CAChe MolStruct format	mol	MDL MOL format
caccrt	Cacao Cartesian format	mol2	Sybyl Mol2 format
cache	CAChe MolStruct format	mold	Molden format
cacint	Cacao Internal format	molden	Molden format
can	Canonical SMILES format	molf	Molden format
cdxml	ChemDraw CDXML format	molreport	Open Babel molecule report
cht	Chemtool format	mop	MOPAC Cartesian format
cif	Crystallographic Information File	mopcrt	MOPAC Cartesian format
cml	Chemical Markup Language	mopin	MOPAC Internal
cmlr	CML Reaction format	mp	Molpro input format
com	Gaussian 98/03 Input	mpc	MOPAC Cartesian format
CONFIG	DL-POLY CONFIG	mpd	MolPrint2D format
CONTCAR	VASP format	mpqcin	MPQC simplified input format
crk2d	Chemical Resource Kit diagram (2D)	mrv	Chemical Markup Language
crk3d	Chemical Resource Kit 3D format	msms	M.F. Sanner's MSMS input format
csr	Accelrys/MSI Quanta CSR format	nw	NWChem input format
cssr	CSD CSSR format	outmol	DMol3 coordinates format
ct	ChemDraw Connection Table format	pcm	PCModel Format
dmol	DMol3 coordinates format	pdb	Protein Data Bank format
ent	Protein Data Bank format	pdbqt	AutoDock PDBQT format
fa	FASTA format	png	PNG 2D depiction
fasta	FASTA format	POSCAR	VASP format
feat	Feature format	pov	POV-Ray input format
fh	Fenske-Hall Z-Matrix format	pqr	PQR format
fhiaims	FHIaims XYZ format	pqs	Parallel Quantum Solutions format
fix	SMILES FIX format	qcin	Q-Chem input format
fps	FPS text fingerprint format (Dalke)	report	Open Babel report format
fpt	Fingerprint format	rsmi	Reaction SMILES format
fract	Free Form Fractional format	rxn	MDL RXN format
fsa	FASTA format	sd	MDL MOL format
gamin	GAMESS Input	sdf	MDL MOL format
gau	Gaussian 98/03 Input	smi	SMILES format
gjc	Gaussian 98/03 Input	svg	SVG 2D depiction
gjf	Gaussian 98/03 Input	sy2	Sybyl Mol2 format
gpr	Ghemical format	tmol	TurboMole Coordinate format
gr96	GROMOS96 format	txt	Title format
gro	GRO format	txyz	Tinker XYZ format
gukin	GAMESS-UK Input	unixyz	UniChem XYZ format
gzmat	Gaussian Z-Matrix Input	vmol	ViewMol format
hin	HyperChem HIN format	xed	XED format
inchi	InChI format	xyz	XYZ cartesian coordinates format
inchikey	InChIKey	yob	YASARA.org YOB format
inp	GAMESS Input	zin	ZINDO input format bottomrule
jin	Jaguar input format		

Supplementary Table 7: Output file formats which are currently supported by VFLP, which can be specified in the configuration file (see Supplementary Listing 2).

F VirtualFlow database structure

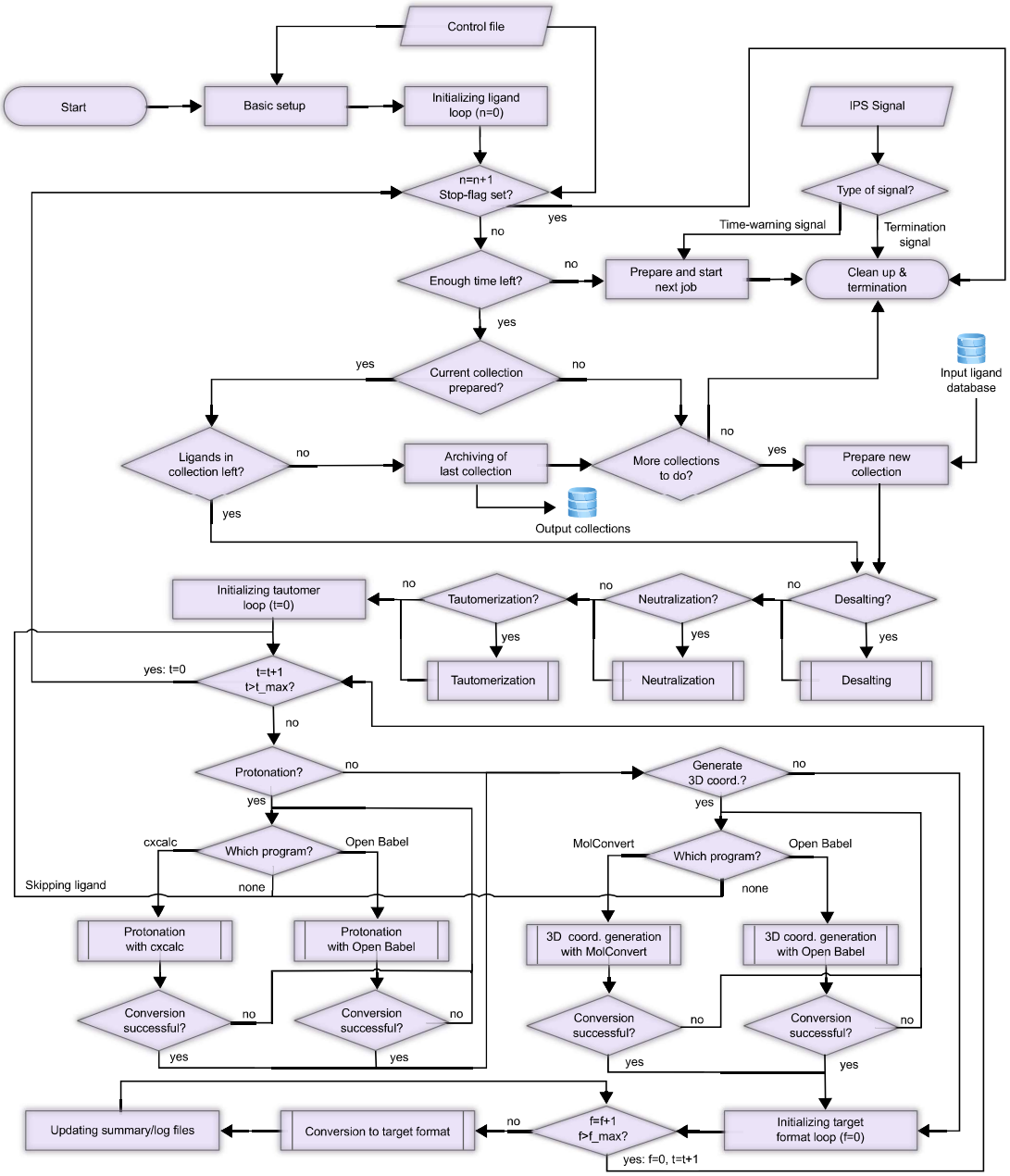
A virtual screening of 100 million ligands may involve one billion or more files (including input, output and log files), depending also on the type and number of docking scenarios. Such large numbers of files can pose major challenges for computer clusters, therefore the number of allowed files per user is normally limited. In particular, if a single folder contains too many files, the performance of the file system (and thus the entire cluster) can drop dramatically. To avoid these problems VirtualFlow uses a multidimensional file structure in the form of a hierarchical cascade of folders and archives (Supplementary Fig. 9). The input and output ligand databases consist of four levels. The first level contains so-called metatranches, the second level stores the tranches which themselves consist of collections (third level), and these collections contain the ligands (fourth level). The output file structure of VFVS can also contain a fifth level as more than one output file might be created for each ligand during the docking procedure. In addition, the collections level (third level) consists of compressed archives, which usually decreases the size of the files by more than an order of magnitude.



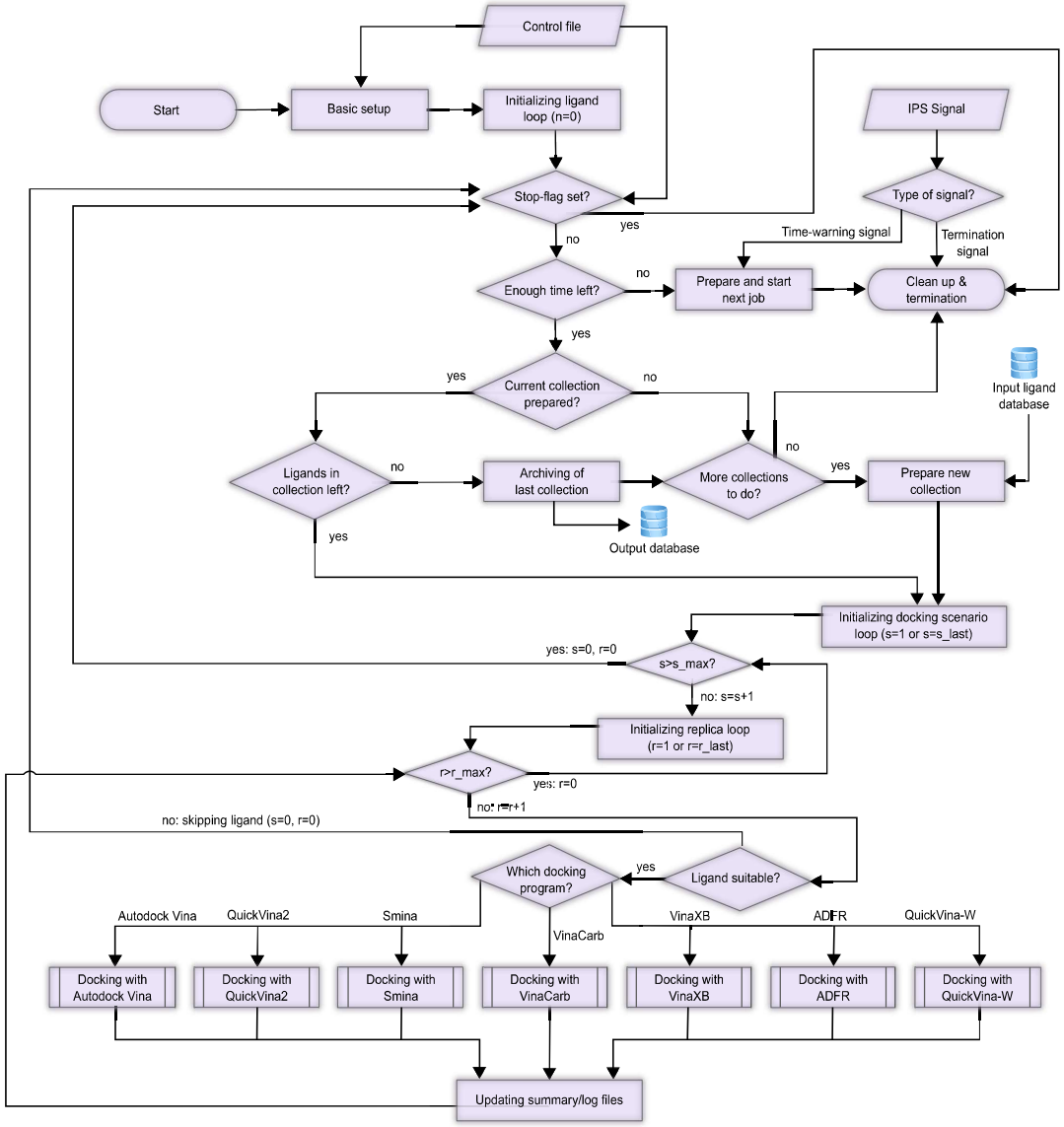
Supplementary Fig. 9: Hierarchical structure of the input ligand databases used by VirtualFlow (VFLP and VFVS). The root folder of the database contains folders called metatranches. Each metatranche folder contains tar-archives called tranches. Each tranche contains compressed (with gzip) tar-archives called collections. And each collection contains the ligands. In input ligand databases, the ligands are text files in the SMILES format (for VFLP), or in a ready-to-dock format such as PDBQT for VFLP. The hierarchical structure allows to efficiently store and retrieve between 10^{16} to 10^{24} ligands when using file systems which are currently widely used such as ext4. While the output ligand databases of VFLP are of the same structure as the input databases, the output databases of VFVS have an additional (fifth) level, in which each ligand is a tar-archive containing all the docking output files for each ligand which were generated during the virtual screening procedure.

G VirtualFlow workflow within single queues

The majority of the work done by VirtualFlow takes place on the queue level, which is unique for VFLP and VFVS. A simplified workflow diagram for VFLP on the queue level is shown in Supplementary Fig. 10, and for VFVS in Supplementary Fig.. 11. Regarding VFLP, the possible processing steps are also illustrated in Extended Data Fig. 2.



Supplementary Fig. 10: Simplified workflow diagram of a single queue of VFLP. In the beginning some basic setup and checks take place, before a loop is entered to iterate through the ligands in the input collections. After each iteration, additional checks take place, including determination if a new ligand collection has to be prepared. Processing of the individual ligands can include desalting, neutralization, and tautomerization, which may produce more than one molecule. For each tautomer generated, a loop is entered. Afterwards, each ligand may be protonated using Open Babel or ChemAxon's cxcalc tool. Then for each ligand the 3D coordinates are computed by Open Babel or by ChemAxon's molconvert tool. An additional loop is used for the generation of the target formats. After a ligand collection is completed, it is archived in the output database.



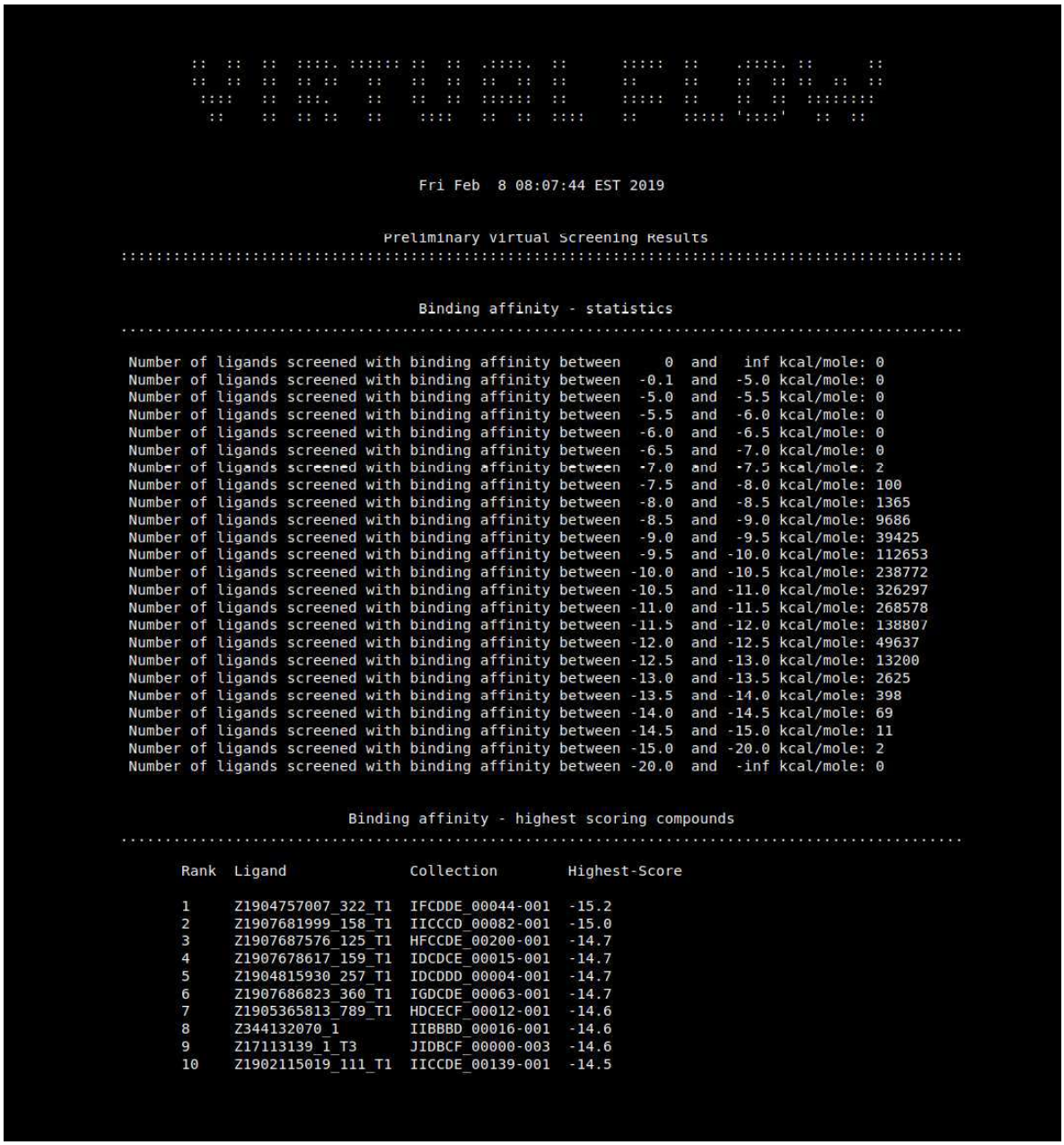
Supplementary Fig. 11: Simplified workflow diagram of a single queue of VFVS. In the beginning some basic setup and checks take place, before a loop is entered to iterate through the ligands in the screening library. After the completion of the loop, additional checks take place, including determination if a new ligand collection has to be prepared. A loop is started with one cycle for each docking scenario, followed by a second inner loop to handle multiple replicas in which the actual docking is executed. If a ligand collection is completed it is archived in the output database.

H Monitoring of VirtualFlow

The status of the workflow managed by VirtualFlow can be monitored in real-time. An example output of the monitoring command (vf_report.sh) showing information about the general workflow can be seen in Supplementary Fig. 12.

Supplementary Fig. 12: Example monitoring information of VFVS while the workflow is running. Information is given on the jobs in the batch system, the input collections and their processing status, and the status of the individual ligands and docking runs.

VFVS has additional monitoring capabilities, which can provide preliminary results of the virtual screen as it is running (Supplementary Fig. 13). The monitoring tool (vf_report.sh) can provide two types of information in this context: (1) statistical information on the docking scores (by classifying them into bins), and (2) information about the n top-scoring compounds, where n is user-defined.



Supplementary Fig. 13: Example output screen of VFVS showing the preliminary screening results of a rescoring procedure (where the top 2 million compounds of a primary virtual screening involving ~1 billion compounds were screened again with higher accuracy). The type of information to be displayed can be specified as arguments to the command script `vf_report.sh`. Two types of information can be displayed regarding the preliminary (live) data during the screening: (1) statistical information about the predicted binding affinities (upper section of the above output), and (2) information about the highest-scoring compounds of the ligands screened so far (lower section of the above output).

I Configuration file of VFVS

Many options of VFVS are specified in a central configuration file. A sample configuration file is shown in Supplementary Listing 1.

Supplementary Listing 1: VFVS Example Configuration File.

```
********************************************************************************************
*****************************   Batch System Configuration    ******************************
********************************************************************************************

job_letter=t
# One alphabetic character (i.e. a letter from a-z or A-Z)
# Should not be changed during runtime, and be the same for all joblines
# Required when running VF several times on the same cluster to distinguish the jobs in the batch system

batchsystem=SLURM
# Possible values: SLURM, TOQRUE, PBS, LSF, SGE

partition=shared
# Partitions are also called queues in some batch systems

timelimit=7-00:00:00
# Format for slurm: dd-hh:mm:ss
# Format for TORQUE and PBS: hh:mm:ss
# Format for SGE: hh:mm:ss
# Format for LSF: hh:mm
# For all batch systems: always fill up with two digits per field (used be the job scripts)

steps_per_job=1
# Not (yet) available for LSF and SGE (is always set to 1)
# Should not be changed during runtime, and be the same for all joblines

cpus_per_step=8
# Sets the slurm cpus-per-task variable (task = step) in SLURM
# In LSF this corresponds to the number of slots per node
# Should not be changed during runtime, and be the same for all joblines
# Not yet available for SGE (always set to 1)

queues_per_step=8
# Sets the number of queues/processes per step
# Should not be changed during runtime, and be the same for all joblines
# Not yet available for SGE (always set to 1)

cpus_per_queue=1
# Should be equal or higher than <cpus-per-step/queues-per-step>
# Should not be changed during runtime, and be the same for all joblines
# Not yet available for SGE (always set to 1)


********************************************************************************************
*********************************   Workflow Options   *************************************
********************************************************************************************

central_todo_list_splitting_size=10000
# When the folders are initially prepared the first time, the central todo list will be split into pieces of size <
     central_todo_list_splitting_size>. One task corresponds to one collection.
# Recommended value: < 100000, e.g. 10000
# Possible values: Positive integer
# The smaller the value, the faster the ligand collections can be distributed.
# For many types of clusters it is recommended if the total number of splitted todo lists stays below 10000.

ligands_todo_per_queue=100000
# Used as a limit of ligands for the to-do lists
# This value should be divisible by the next setting "ligands_todo_per_refilling_step"

ligands_per_refilling_step=10000
# The to-do files of the queues are filled with <ligands_per_refilling_step> ligands per refill step
# A number roughly equal to the average of number of ligands per collection is recommended

collection_folder=../../../collections/Enamine-REAL-2018q12_isomers.pdbqt
# Slash at the end is not required (optional)
# Relative pathname is required w.r.t. the folder tools/

minimum_time_remaining=10
# In minutes
# A new job if the time left until the end of the wall time is smaller than the timelimit
# This is checked before each ligand is screened
# Thus the timelimit should be larger than the maximum time which is needed to process one ligand

dispersion_time_min=3
# One positive integer, resembling the time in seconds
dispersion_time_max=10
# One positive integer, resembling the time in seconds
# The dispersion time is used when jobs try to access the central task list.
# Each job has to wait a random amount of time in the dispersion interval.
# The effect of this is that when two jobs arrive at the same time at the central task list, the random waiting time will disperse
      their access on the central task list in time
```

```
verbosity_commands=standard
# Possible values: standard, debug
# This option mainly effects the screen output and the logfiles

verbosity_logfiles=standard
# Possible values:
#   * standard
#   * debug : activates the set -x option. Increases size of log-files in average by nearly a factor of 10
# This option affects the preparation scripts for setting up the basic workflow files (before the workflow is running)

store_queue_log_files=all_compressed_error_uncompressed
# Supported values (experimental)
#   * all_uncompressed: requires most memory any and storage, but is recommending for test runs and debugging purposes
#   * all_compressed: requires less memory and storage than uncompressed, but during the last part of the log files might get lost
      (in particular during crashes) due to the on-the-fly compression
#   * only_error_uncompressed: only stderr is logged
#   * only_error_compressed: only stderr is logged and compressed. The last part of the log might get lost (in particular during
     crashes) due to the on-the-fly compression.
#   * std_compressed_error_uncompressed
#   * all_compressed_error_uncompressed
#   * none: reduces required memory and storage

keep_ligand_summary_logs=true
# Summary log files which show for each ligand the success status of conversion and the conversion time.
# If the conversion failed, a reason is stated.
# If the transformation succeeded, the conversion programs which were used are stated.
# Possible values:
#   * false
#   * true

error_sensitivity=normal
# Possible values: normal, high
# high sets the shell options "-uo pipefail". Not recommended for production runs, useful mainly for debugging. Pipefails often
     occur with tar combined with head/tail in pipes, which are not an actual problem.
# The u-option will always lead to a direct exit of the shell script when an unset variable is going to be used.

error_response=fail
# Affects most errors, but not all (e.g. not the u-option of the shell)
# Possible values:
#   * ignore    : ignore error and continue
#   * next_job  : end this job and start new job
#   * fail      : exit workflow with failure (exit code 1)

tempdir=/dev/shm
# The directory which is used for the temporary workflow files
# Should be a very fast filesystem not part of the distributed cluster filesystem, e.g. a local ramfs (as normally provided by ${
     VF_TMPDIR})
# The directory does only need to be available on the node on which the job step/queue is running
# In the tempdir, a subfolder named ${USER} will automatically be created


##############################################################################################
##############################   Virtual Screening Options    ################################
##############################################################################################

docking_type_names=smina_vinardo_flexible_receptor1:smina_vinardo_flexible_receptor2:vina_flexible_receptor1:
     vina_flexible_receptor2
# Values have to be separated by colons ":" and without spaces, e.g: docking_type_names=vina-rigid:smina-rigid
# Used for instance for the folder names in which the output files are stored
# Should not be changed during runtime, and be the same for all joblines

docking_type_programs=smina_flexible:smina_flexible:vina:vina
# Possible values: qvina02, qvina_w, vina, smina_rigid, smina_flexible, adfr
# Values have to be separated by colons ":" and without spaces, e.g: docking_type_programs=vina:smina
# The same programs can be used multiple times
# Should not be changed during runtime, and be the same for all joblines

docking_type_replicas=1:1:1:1
# Series of integers separated by colons ":"
# The number of values has to equal the number of docking programs specified in the variable "docking_programs"
# The values are in the same order as the docking programs specified in the variable "docking_type_programs
# e.g.: docking_type_replicas=1:1
# possible range: 1-99999 per field/docking program
# The docking scenario is comprised of all the docking types and their replicas
# Should not be changed during runtime, and be the same for all joblines

docking_type_inputfolders=../input-files/smina_vinardo_flexible_receptor1:../input-files/smina_vinardo_flexible_receptor2:../input
     -files/vina_flexible_receptor1:../input-files/vina_flexible_receptor2
# Relative path wrt to the tools folders
# In each input folder must be the file config.txt which is used by the docking program to specify its options
# If other input files are required by the docking type, usually specified in the config.txt file, they have to be in the same
     folder
# Should not be changed during runtime, and be the same for all joblines


##############################################################################################
################################   Terminating Variables   ###################################
##############################################################################################

stop_after_next_check_interval=false
# Determines whether the queue is stopped after the ligand batch currently in progress. The size of the ligand batches is
```

```
     determined by the <ligand_check_interval> variable.
# Possible values:
#   * false : The queue will continue to process ligands as long as there are ligands remaining for the queue
#   * true : No new ligand will be started after the current ligand is completed

ligand_check_interval=100
# Possible values
#   * Positive integer

stop_after_collection=false
# Determines whether the queue is stopped after the current collection is completed
# Possible values:
#   * false : A new collection will be started if the current collection is completed and if there are collections remaining
#   * true : No new collection will be started after the current collection is completed

stop_after_job=false
# Determines whether the queue is stopped after the current job is completed
# Possible values:
#   * false : A new job will be submitted if there are more ligands in the current collection or unprocessed collections remaining
#   * true : No new job is submitted after the current job has ended
```

J Configuration file of VFLP

Most options of VFLP are specified in a central configuration file. A sample configuration file is shown in Supplementary Listing 2.

Supplementary Listing 2: VFLP Example Configuration File.

```
********************************************************************************************
*****************************    Job Resource Configuration    *****************************
********************************************************************************************

job_letter=t
# One alphabetic character (i.e. a letter from a-z or A-Z)
# Should not be changed during runtime, and be the same for all joblines
# Required when running VF several times on the same cluster to distinguish the jobs in the batch system

batchsystem=SLURM
# Possible values: SLURM, TOQRUE, PBS, LSF, SGE

partition=shared
# Partitions are also called queues in some batch systems

timelimit=1-00:00:00
# Format for slurm: dd-hh:mm:ss
# Format for TORQUE and PBS: hh:mm:ss
# Format for SGE: hh:mm:ss
# Format for LSF: hh:mm
# For all batch systems: always fill up with two digits per field (used be the job scripts)

steps_per_job=1
# Not (yet) available for LSF and SGE (is always set to 1)
# Should not be changed during runtime, and be the same for all joblines

cpus_per_step=4
# Sets the slurm cpus-per-task variable (task = step) in SLURM
# In LSF this corresponds to the number of slots per node
# Should not be changed during runtime, and be the same for all joblines
# Not yet available for SGE (always set to 1)

queues_per_step=8
# Sets the number of queues/processes per step
# Should not be changed during runtime, and be the same for all joblines
# Not yet available for SGE (always set to 1)

cpus_per_queue=1
# Should be equal or higher than <cpus-per-step/queues-per-step>
# Should not be changed during runtime, and be the same for all joblines
# Not yet available for SGE (always set to 1)


********************************************************************************************
*********************************    Workflow Options    *********************************
********************************************************************************************

central_todo_list_splitting_size=10000
# When the folders are initially prepared the first time, the central todo list will be split into pieces of size <
     central_todo_list_splitting_size>. One task corresponds to one collection.
# Recommended value: < 100000, e.g. 10000
# Possible values: Positive integer
# The smaller the value, the faster the ligand collections can be distributed.
# For many types of clusters it is recommended if the total number of splitted todo lists stays below 10000.

ligands_todo_per_queue=500000
# Used as a limit of ligands for the to-do lists
# This value should be divisible by the next setting "ligands_todo_per_refilling_step"

ligands_per_refilling_step=10000
# The to-do files of the queues are filled with <ligands_per_refilling_step> ligands per refill step
# A number roughly equal to the average of number of ligands per collection is recommended

collection_folder=../../../collections/Enamine-REAL-2018q12_isomers.smi.splitted1000_ind_tgz_tar
# Slash at the end is not required (optional)
# Relative pathname is required w.r.t. the folder tools/

minimum_time_remaining=10
# In minutes
# A new job if the time left until the end of the wall time is smaller than the timelimit
# This is checked before each ligand is screened
# Thus the timelimit should be larger than the maximum time which is needed to process one ligand

dispersion_time_min=3
# One positive integer, resembling the time in seconds
dispersion_time_max=10
# One positive integer, resembling the time in seconds
# The dispersion time is used when jobs try to access the central task list.
# Each job has to wait a random amount of time in the dispersion interval.
# The effect of this is that when two jobs arrive at the same time at the central task list, the random waiting time will disperse
      their access on the central task list in time
```

```
verbosity_commands=standard
# Possible values: standard, debug
# This option mainly effects the screen output and the logfiles

verbosity_logfiles=standard
# Possible values:
#   * standard
#   * debug : activates the set -x option. Increases size of log-files in average by nearly a factor of 10
# This option affects the preparation scripts for setting up the basic workflow files (before the workflow is running)

store_queue_log_files=all_compressed_error_uncompressed
# Supported values (experimental)
#   * all_uncompressed: requires most memory any and storage, but is recommending for test runs and debugging purposes
#   * all_compressed: requires less memory and storage than uncompressed, but during the last part of the log files might get lost
      (in particular during crashes) due to the on-the-fly compression
#   * only_error_uncompressed: only stderr is logged
#   * only_error_compressed: only stderr is logged and compressed. The last part of the log might get lost (in particular during
     crashes) due to the on-the-fly compression.
#   * std_compressed_error_uncompressed
#   * all_compressed_error_uncompressed
#   * none: reduces required memory and storage

keep_ligand_summary_logs=true
# Summary log files which show for each ligand the success status of conversion and the conversion time.
# If the conversion failed, a reason is stated.
# If the transformation succeeded, the conversion programs which were used are stated.
# Possible values:
#   * false
#   * true

error_sensitivity=normal
# Possible values: normal, high
# high sets the shell options "-uo pipefail". Not recommended for production runs, useful mainly for debugging. Pipefails often
     occur with tar combined with head/tail in pipes, which are not an actual problem.
# The u-option will always lead to a direct exit of the shell script when an unset variable is going to be used.

error_response=fail
# Affects most errors, but not all (e.g. not the u-option of the shell)
# Possible values:
#   * ignore    : ignore error and continue
#   * next_job  : end this job and start new job
#   * fail      : exit workflow with failure (exit code 1)

tempdir=/dev/shm
# The directory which is used for the temporary workflow files
# Should be a very fast filesystem not part of the distributed cluster filesystem, e.g. a local ramfs (as normally provided by ${
     VF_TMPDIR})
# The directory does only need to be available on the node on which the job step/queue is running
# In the tempdir, a subfolder named ${USER} will automatically be created


##############################################################################################
############################    Ligand Preparation Options     ###############################
##############################################################################################

####################################    Desalting    ####################################

desalting=true
# If true, extracts the largest organic part of the molecule
# Possible values:
#   * false
#   * true

desalting_obligatory=false
# Setting only required if desalting=true
# Possible values:
#   * true: Successful desalting of the ligand is mandatory (unsuccessful desalting leads to the omission of this ligand).
#   * false: The ligand will continue to be processed even if the desalting step fails.


#################################    Neutralization    #################################

neutralization=true
# Neutralizes molecules using JChem's Standardizer of ChemAxon
# Possible values:
#   * false
#   * true

neutralization_mode=after_desalting_if_charged
# Only relevant if neutralization=true
# Possible values:
#   * always: always neutralize the molecule
#   * only_genuine_desalting: only neutralize the molecule if the input structure contained more than one component
#   * only_genuine_desalting_and_if_charged: only neutralize if the input structure contained more than one component and if the
     smallest component contained an ion

neutralization_obligatory=false
# Setting only required if neutralization=true
# Possible values:
#   * true: Successful neutralization of the ligand is mandatory (unsuccessful neutralization leads to the omission of this ligand
     ).
```

```
#   * false: The ligand will continue to be processed even if the neutralization step fails.


**********************************    Tautomerization    **********************************

tautomerization=true
# Possible values:
#   * false
#   * true

tautomerization_obligatory=false
# Setting only required if tautomerization=true
# Possible values:
#   * true: Successful tautomerization of the ligand is mandatory (unsuccessful tautomerization leads to the omission of this
     ligand).
#   * false: The ligand will continue to be processed even if the tautomerization step fails.

cxcalc_tautomerization_options=
# any options which should be passed to the tautomerization plugin of cxcalc


****************************    Protonation State Generation    ****************************

protonation_state_generation=true
# Possible values:
#   * false
#   * true

protonation_program_1=cxcalc
protonation_program_2=obabel
# Program 1 is used at first for each ligand, and if it fails program 2 is used instead. If the second program also fails, then
     the ligand is skipped.
# Setting only required if protonation_state_generation=true
# Possible values:
#   * obabel
#   * cxcalc
#   * none (only for protonation_program_2)

protonation_obligatory=false
# Setting only required if protonation_state_generation=true
# Possible values:
#   * true: Successful protonation of the ligand is mandatory (unsuccessful protonation leads to the omission of this ligand).
#   * false: The ligand will continue to be processed even if the protonation step fails. This might light to protonation states
     which are unphysiological.

protonation_pH_value=7.4
# Setting only required if protonation_state_generation=true
# Possible values: floating point number between 0.0 and 14.0


******************************    Conformation Generation    ******************************

conformation_generation=true
# Generation of 3D conformation/coordinates of the ligand
# Possible values:
#   * false
#   * true

conformation_program_1=molconvert
conformation_program_2=obabel
# Setting only required if conformation_generation=true
# Program 1 is used at first for each ligand, and if it fails program 2 is used instead. If the second program also fails, then
     the ligand is skipped.
# Possible values:
#   * obabel
#   * molconvert
#   * none (only possible for conformation_program_2)

molconvert_3D_options=-3:{fast}
# Setting only required if conformation_generation=true and one of the programs used is molconvert
# 3D conformation generation options for molconvert.
# See also the help text printed by molconvert for additional information
# Possible values:
#   * -3          Defaults to value 3{fast}
#   *  3{fine}    Find low energy conformer Leave failed fragments intact
#   *  3{fast}    Fast clean, if failed, perform fine clean, accept any generated structure (default)
#   *  3{nofaulty} Same as S{fast}, but leave failed fragments intact.

conformation_obligatory=true
# Setting only required if conformation_generation=true
# Possible values:
#   * true: Successful 3D conformation generation of the ligand is mandatory (unsuccessful conformation generation leads to the
     omission of this ligand).
#   * false: The ligand will continue to be processed even if the conformation generation step fails.


****************************    Target Format Generation    ********************************

targetformats=smi:pdb:pdbqt
# Possible values: Any format supported by the Open Babel, using the file format identifiers used by Open Babel.
# A complete list can be obtained by running the command "obabel -L formats"
```

```
# Multiple target formats can be specified by separating them with colons, e.g. pdb:sdf:pdbqt


**************************************    Open Babel    ***********************************

obabel_memory_limit=1000000
# In KB
# Recommended value: >= 500000

obabel_time_limit=50
# In seconds
# Open Babel seems to have an internal limit of 240 seconds for conformation generation


******************************    JChem-Related Packages    ******************************

jchem_package_filename=jchemsuite.tar.gz
# Required only if cxcalc or molconvert (both of ChemAxon) are used in the preparation steps.
# The filename of the JChem package located in the folder tools/packages/ in the tar.gz format (available on the ChemAxon homepage
     )
# The root folder in the archive has to have the name jchemsuite (normally distributed by ChemAxon in this way)
# Possible values:
#   * <filename>
#   * none

chemaxon_license_filename=chemaxon-license.cxl
# Required only if cxcalc or molconvert (both of ChemAxon) are used in the preparation steps.
# The filename of the license file, which has to be located in the folder tools/packages/
# Possible values:
#   * <filename>
#   * none

java_package_filename=java11_bin.tar.gz
# Required only if cxcalc or molconvert (both of ChemAxon) are used in the preparation steps.
# Any JRE binary distribution of version of at least version 8.
# If java is provided by the system (e.g. by loading a module), then no Java package needs to be provided.
# This has to be a file in the tar.gz format, which has to be located in the folder tools/packages/
# The root folder in the archive has to have the name "java" (which will be used for the JAVA_HOME variable). This normally needs
     to be manually changed in the Java package after downloading a JRE.
# Possible values:
#   * <filename>
#   * none

ng_package_filename=nailgun.tar.gz
# Required only if cxcalc or molconvert (both of ChemAxon) are used in the preparation steps.
# Nailgun package filename.
# This has to be a file in the tar.gz format, which has to be located in the folder tools/packages/
# The root folder in the archive has to have the name "nailgun"
# Possible values:
#   * <filename>
#   * none

java_max_heap_size=2
# Size in GB
# Recommended: >= 1 GB * queues_per_step
# The required memory depends mainly on how many queues are run per step (and thus per JVM/NG server), since one JVM is used per
     step


*********************************************************************************************
*********************************   Terminating Variables   *********************************
*********************************************************************************************

stop_after_next_check_interval=false
# Determines whether the queue is stopped after the ligand batch currently in progress. The size of the ligand batches is
     determined by the <ligand_check_interval> variable.
# Possible values:
#   * false : The queue will continue to process ligands as long as there are ligands remaining for the queue
#   * true : No new ligand will be started after the current ligand is completed

ligand_check_interval=100
# Possible values
#   * Positive integer

stop_after_collection=false
# Determines whether the queue is stopped after the current collection is completed
# Possible values:
#   * false : A new collection will be started if the current collection is completed and if there are collections remaining
#   * true : No new collection will be started after the current collection is completed

stop_after_job=false
# Determines whether the queue is stopped after the current job is completed
# Possible values:
#   * false : A new job will be submitted if there are more ligands in the current collection or unprocessed collections remaining
#   * true : No new job is submitted after the current job has ended
```

References

The numbering of the references is based on the numbering of the main manuscript.

9. Davies, T. G. *et al.* Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein–Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J. Med. Chem.* **59**, 3991–4006 (2016).

25. Marcotte, D. *et al.* Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorg. Med. Chem.* **21**, 4011–4019 (2013).

39. Irwin, J. J. *et al.* An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).

40. Laplante, S. R. *et al.* Compound Aggregation in Drug Discovery: Implementing a Practical NMR Assay for Medicinal Chemists. *J. Med. Chem.* **56**, 5142–5150 (2013).

44. Jnoff, E. *et al.* Binding Mode and Structure-Activity Relationships around Direct Inhibitors of the Nrf2-Keap1 Complex. *ChemMedChem* **9**, 699–705 (2014).

45. Zhuang, C., Narayanapillai, S., Zhang, W., Sham, Y. Y. & Xing, C. Rapid Identification of Keap1–Nrf2 Small-Molecule Inhibitors through Structure-Based Virtual Screening and Hit-Based Substructure Search. *J. Med. Chem.* **57**, 1121–1126 (2014).

4.4 A biphenyl inhibitor of eIF4E targeting an internal binding site enables the design of cell-permeable PROTAC-degraders

Patrick D. Fischer*, Evangelos Papadopoulos*, Jon M. Dempersmier, Zi-Fu Wang, Radosław P. Nowak, Joann Kalabathula, Christoph Gorgulla, Pierre P. M. Junghanns, Eihab Kabha, Nikolaos Dimitrakakis, Ognyan I. Petrov, Constantine Mitsiades, Christian Ducho, Vladimir Gelev, Eric S. Fischer, Gerhard Wagner and Haribabu Arthanari. (2021). A biphenyl inhibitor of eIF4E targeting an internal binding site enables the design of cell-permeable PROTAC-degraders, **EurJMedChem**, 219, 113435.

*These authors contributed equally

DOI: https://doi.org/10.1016/j.ejmech.2021.113435

URL: https://www.sciencedirect.com/science/article/abs/pii/S0223523421002841

Summary:

In this publication, a novel small molecule inhibitor of the master regulator of eukaryotic translation, the cap-binding protein eIF4E, was identified and cell-permeable PROTAC degraders were developed. The X-ray crystal structure of eIF4E in complex with the new inhibitor, i4EG-BiP, was solved with a resolution of 1.9 Å. This compound and the small molecule inhibitor 4EGI-1 are the only two compounds that inhibit eIF4E allosterically (i.e. not competitive with the mRNA cap) of which structural information is available. i4EG-BiP was compared to 4EGI-1 in terms of binding affinity to eIF4E (analyzed by fluorescence polarization), time-dependent cytotoxicity towards cancer cell lines and ability to specifically inhibit cap-dependent protein translation. It was found that, while being comparable to 4EGI-1 regarding binding affinity and cytotoxicity, i4EG-BiP displayed stronger specificity towards inhibition of cap-dependent translation. Based on the structural information at hand, PROteolysis TArgeting Chimeras (PROTACs) derived from 4EGI-1 and i4EG-BiP were designed and synthesized to induce targeted proteasomal degradation of eIF4E. To achieve this, derivatives of 4EGI-1 and i4EG-BiP were linked to the E3 ubiquitin ligase ligand thalidomide via carbon-based linkers. It was

shown by fluorescence polarization and solution NMR that PROTAC degraders derived from those two small molecules exhibited comparable binding affinities to eIF4E and bound to the same region of the protein. Furthermore, reduced cellular uptake observed for previously reported eIF4E targeting PROTACs was improved by applying a prodrug strategy (for 4EGI-1-based PROTACs) or attaching the linker to the carboxylic acid moiety to eliminate the net negative charge of the compounds (for i4EG-BiP-based PROTACs). Cellular uptake was confirmed by cell-based Cereblon engagement studies. Sadly, none of the PROTACs were able to degrade eIF4E in cells. This was contributed to the low binding affinity of the parent compounds and the observed increase in binding of eIF4E to 4EBP1, which was shown to protect eIF4E from proteasomal degradation.

Author contributions:

As the author of this thesis, I designed and synthesized the PROTAC degrader molecules studied in this publication. I carried out purification of the molecules and analyzed their purity using LC-MS-based methods and NMR. Together with Jon M. Dempersmier, he performed cytotoxicity studies for the degraders. He performed the solution NMR based comparison of binding interfaces between i4EG-BiP and PROTAC-degraders. He carried out the Western Blot analysis of eIF4E degradation. In collaboration with Evangelos Papadopoulos, he analyzed the structure of eIF4E bound to i4EG-BiP and analyzed the implications of conformational changes induced by i4EG-BiP binding on the interaction between eIF4E and 4EBP1 or eIF4G. He wrote the paper in conjunction with Evangelos Papadopoulos and Haribabu Arthanari.

European Journal of Medicinal Chemistry 219 (2021) 113435

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

A biphenyl inhibitor of eIF4E targeting an internal binding site enables the design of cell-permeable PROTAC-degraders

Patrick D. Fischer [a, b, c, 1], Evangelos Papadopoulos [b, e, 1, **], Jon M. Dempersmier [a, b], Zi-Fu Wang [a, b], Radosław P. Nowak [a, b], Katherine A. Donovan [a, b], Joann Kalabathula [a], Christoph Gorgulla [a, b], Pierre P.M. Junghanns [c], Eihab Kabha [b], Nikolaos Dimitrakakis [d], Ognyan I. Petrov [f], Constantine Mitsiades [e], Christian Ducho [c], Vladimir Gelev [f], Eric S. Fischer [a, b], Gerhard Wagner [b], Haribabu Arthanari [a, b, *]

[a] Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, 02215, USA
[b] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, 02115, USA
[c] Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbrücken, 66123, Germany
[d] Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, 02115, USA
[e] Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, 02215, USA
[f] Faculty of Chemistry and Pharmacy, Sofia University, 1 James Bourchier Blvd., 1164, Sofia, Bulgaria

ARTICLE INFO

Article history:
Received 11 November 2020
Received in revised form
23 March 2021
Accepted 1 April 2021
Available online 8 April 2021

Keywords:
Translation inhibitor
eIF4E inhibitor
eIF4E-eIF4G inhibitor
Protein-protein interaction inhibitor
eIF4E PROTAC
Prodrug

ABSTRACT

The eukaryotic translation initiation factor 4E (eIF4E) is the master regulator of cap-dependent protein synthesis. Overexpression of eIF4E is implicated in diseases such as cancer, where dysregulation of oncogenic protein translation is frequently observed. eIF4E has been an attractive target for cancer treatment. Here we report a high-resolution X-ray crystal structure of eIF4E in complex with a novel inhibitor (i4EG-BiP) that targets an internal binding site, in contrast to the previously described inhibitor, 4EGI-1, which binds to the surface. We demonstrate that i4EG-BiP is able to displace the scaffold protein eIF4G and inhibit the proliferation of cancer cells. We provide insights into how i4EG-BiP is able to inhibit cap-dependent translation by increasing the eIF4E-4E-BP1 interaction while diminishing the interaction of eIF4E with eIF4G. Leveraging structural details, we designed proteolysis targeted chimeras (PROTACs) derived from 4EGI-1 and i4EG-BiP and characterized these on biochemical and cellular levels. We were able to design PROTACs capable of binding eIF4E and successfully engaging Cereblon, which targets proteins for proteolysis. However, these initial PROTACs did not successfully stimulate degradation of eIF4E, possibly due to competitive effects from 4E-BP1 binding. Our results highlight challenges of targeted proteasomal degradation of eIF4E that must be addressed by future efforts.

1. Introduction

Cap-dependent translation in eukaryotes is initiated when eIF4E binds to the m^7GTP cap of mRNAs, a rate-limiting step that results in the formation of the eIF4F complex, which is comprised of eIF4E, the DEAD-box RNA helicase eIF4A, and the scaffold protein eIF4G [1–7]. eIF4E's interaction with eIF4G facilitates loading of the 40S small ribosomal unit onto the mRNA, triggering scanning for the start codon. eIF4E binding proteins (4E-BPs) compete with eIF4G for binding to eIF4E, and successful binding of 4E-BPs abolishes cap-dependent translation, making 4E-BPs important regulators of this process. This competition is regulated by the 4E-BP phosphorylation state. Upon phosphorylation by mTORC1, 4E-BP isoforms disassociate from eIF4E, freeing it to then engage with eIF4G and form the eIF4F complex [8] (Fig. 1A). The critical role of eIF4E in cancer was first identified when overexpression was observed to cause tumorigenic transformation of fibroblasts [9]. eIF4E has since been found to be overexpressed in a number of cancer types,

* Corresponding author. Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, 02115, USA.
** Corresponding author. Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, 02215, USA.
E-mail addresses: Epapadopoulos@partners.org (E. Papadopoulos), hari@hms.harvard.edu (H. Arthanari).
[1] Joint Authors.

https://doi.org/10.1016/j.ejmech.2021.113435
0223-5234/

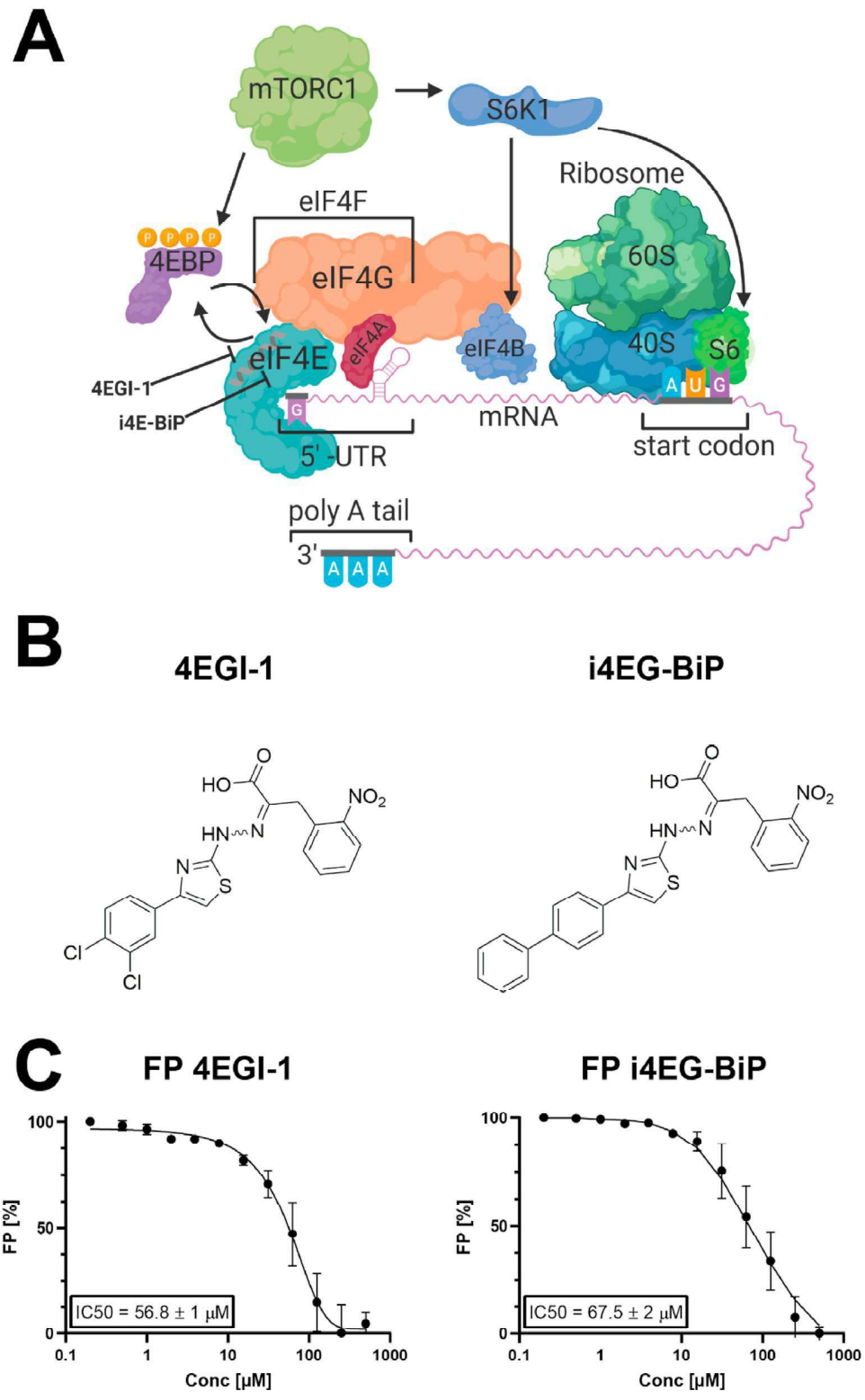


Fig. 1. A. Assembly of the translation initiation complex eIF4F, which consists of eIF4E, eIF4G and eIF4A. eIF4E is negatively regulated by hypo-phosphorylated 4E-BP, which in turn is regulated by mTORC1. B. Structures of the small molecule inhibitors 4EGI-1 and i4EG-BiP. C. Fluorescence polarization assay for displacement of an eIF4G peptide from eIF4E. Both 4EGI-1 and i4EG-BiP are capable of displacing eIF4G with similar affinities.

including breast [10], non-Hodgkin lymphoma [11] and head and neck [12].

The structure of eIF4E resembles a hand with a palm consisting of β-strands and dorsally positioned α-helices [13,14]. The m^7GTP cap binds tightly to the palm region and is stabilized by interactions with four tryptophan sidechains. Both eIF4G and 4E-BPs engage eIF4E, in part, through conserved motifs with the consensus sequence $YX_4L\Phi$, where Φ is a hydrophobic amino acid. Structural studies have revealed that the conserved 4G/4E-BP binding motif binds to the dorsal surface, opposite of the cap-binding surface [15,16]. Studies have shown crosstalk between the cap-binding and 4G/4E-BP binding events with binding at one site affecting the affinity of the other [17]. All published structures of eIF4E in complex with minimal binding epitopes from eIF4G or 4E-BPs are nearly identical and only provided information about the eIF4G and 4E-BP peptides engaging eIF4E. However, structures with larger fragments of eIF4G or 4E-BP in complex with eIF4E have been recently determined [15,16]. These structures show that eIF4G and 4E-BPs use interfaces beyond their minimal consensus sequence to engage eIF4E.

Despite this structural data, the atomic detail of how 4E-BPs outcompete eIF4G is still not fully known. Although the canonical binding helix is conserved between eIF4G and 4E-BPs, the residues beyond the consensus sequence are not, and in the case of 4E-BPs, this non-consensus sequence harbors several phosphorylation sites [15,16,18–20]. It has been reported that phosphorylation of T37 and T46 of 4E-BP2 induces folding of the protein into a four β-strand folded domain that sequesters the eIF4E binding motif [21]. More recently, it was found that hyper-phosphorylation of the C-terminal intrinsically disordered phosphor-sites of 4E-BP2 stabilizes the β-stranded folded domain, further reducing 4E-BP2 binding to eIF4E [22]. Dislodging eIF4G by binding of either 4E-BPs or a small molecule inhibitor would stop cap-dependent translation and this latter approach could be used for the therapeutic treatment of eIF4E-mediated pathogenic dysregulation of translation.

4EGI-1 was the first identified small molecule inhibitor of the eIF4E-eIF4G interaction, binding to eIF4E with low micromolar affinity (IC_{50} = 57 ± 1 μM [23,24]). 4EGI-1 has been shown to be effective in arresting proliferation in a number of cancer cell lines [25–28]. The crystal structure of 4EGI-1 in complex with eIF4E reported in 2014 (PDB: 4TPW) [24] showed that 4EGI-1 bound to eIF4E at a site that is distinctly different from the primary binding site of the eIF4G/4E-BP consensus sequence or the m^7GTP cap binding site. In the crystal structure, the binding of 4EGI-1 induces a conformational change in helix α1 of eIF4E, thus leading to the hypothesis that 4EGI-1 is an allosteric inhibitor. Another small molecule inhibitor that displaces eIF4G from eIF4E is 4E1RCat, which was discovered with a high-throughput screen using a time resolved (TR)-FRET based assay [29]. There is no high-resolution structures of 4E1RCat bound to eIF4E. Analogues of 4EGI-1 co-crystallized by our group reveal binding at the same site as 4EGI-1 [24]. This common binding site is proximal to a cavity on the surface of eIF4E.

In an effort to increase the efficacy of 4EGI-1, we previously synthesized and screened multiple analogues of this compound. Systematic variation of the di-chlorophenyl "head", the nitrophenyl "tail" or the thiazole core did not produce analogues with significantly increased binding affinity to eIF4E. Removal of the hydrazone linker consistently reduced binding affinity. In contrast, certain significant changes in the aromatic "head" could be accommodated without significant loss in binding affinity. For example, an analogue (4EGI-1A, Supplementary Fig. 2) containing an additional cyclohexyl ring showed activity similar to 4EGI-1 in an FP assay measuring displacement of an eIF4G peptide from eIF4E and was a part of the original publication [23]. We then synthesized the active head-head dimer (4EGI-dim) of 4EGI-1, and 4EGI-dim displayed a slightly better K_d than 4EGI-1 in the FP assay. Next, we synthesized i4EG-BiP, containing a biphenyl instead of cyclohexyl-phenyl "head"(here BiP denotes the biphenyl moiety in the scaffold), as a control for a series of "head-to-head" dimer series such as 4EGI-dim (Supplementary Fig. 2). We serendipitously discovered that i4EG-BiP binds to a new internal cavity on eIF4E, near the 4EGI-1 binding site.

We were also interested in leveraging information from high-resolution structures of 4EGI-1 and i4EG-BiP bound to eIF4E to engineer PROteolysis-TArgeting Chimeras (PROTACs) in an effort to trigger specific degradation of eIF4E as an alternative to traditional small molecule inhibitors that target protein-protein interfaces [30–35]. PROTACs are heterobifunctional molecules in which a small molecule ligand is conjugated to an E3 ligase ligand with a linker, frequently PEG- or carbon-based. PROTACs take advantage of the host E3 ubiquitin ligase machinery to induce targeted proteasomal degradation of the protein of interest. Studies with PROTACs based on promiscuous kinase degraders have shown that ligands with weak binding (K_D > 10 μM) can still be turned into potent degraders [30,36]. Kaur et al. previously attempted to employ PROTAC strategies against eIF4E using derivatized cap analogues coupled to lenalidomide or von Hippel-Lindau (VHL) ligands [37]. Bn^7GDP linked to lenalidomide was shown to be an effective eIF4E binder *in vitro* with an affinity of 50 μM, but all tested compounds failed to degrade eIF4E in cellular assays. This was discovered to be due to insufficient cell permeability, a known problem for highly negatively charged cap analogues. Our secondary goal in this study was therefore to generate cell-permeable degraders of eIF4E based on our inhibitors 4EGI-1 and i4EG-BiP. Here, we present the structural and biochemical characterization of i4EG-BiP and our PROTAC derivatives of 4EGI-1 and i4EG-BiP. Since there are many acronyms and abbreviations, we have included a table (Supplementary Table 2) in the supplementary section to ease readership.

2. Results

2.1. Discovery of i4EG-BiP

In an effort to increase the efficacy of 4EGI-1, we synthesized and screened multiple analogues of 4EGI-1. The analogues synthesized included variations of the di-chlorophenyl moiety, the nitrophenyl moiety or the thiazole core. However, none of these compound variations improved the binding affinity. Amongst the synthesized compounds were also dimeric versions of 4EGI-1, mirrored along the di-chlorophenyl moiety. One of the intermediates from this synthesis was the biphenyl compound i4EG-BiP (Fig. 1B). This compound caught our attention due to its similar eIF4G displacement properties when compared to 4EGI-1 (Fig. 1C); i4EG-BiP displaces an eIF4G peptide from eIF4E with an IC_{50} value of 68 ± 2 μM, which is similar to that of 4EGI-1. Since 4EGI-1 binds on the surface of eIF4E, we were intrigued as to how the extra aromatic ring would be stabilized, so to answer this question, we solved the structure of i4EG-BiP in complex with eIF4E by X-ray crystallography.

2.2. The structure of eIF4E bound to i4EG-BiP

The eIF4E structure in complex with i4EG-BiP was resolved by molecular replacement to a resolution of 1.9 Å (Fig. 2A). The structure can be accessed at the PDB databank with the PDB ID 7MEU. The refined model allowed us to unambiguously locate the position of the ligand in the electron density map (Fig. 2B). i4EG-BiP binds to a cavity on the surface of eIF4E that is near the 4EGI-1

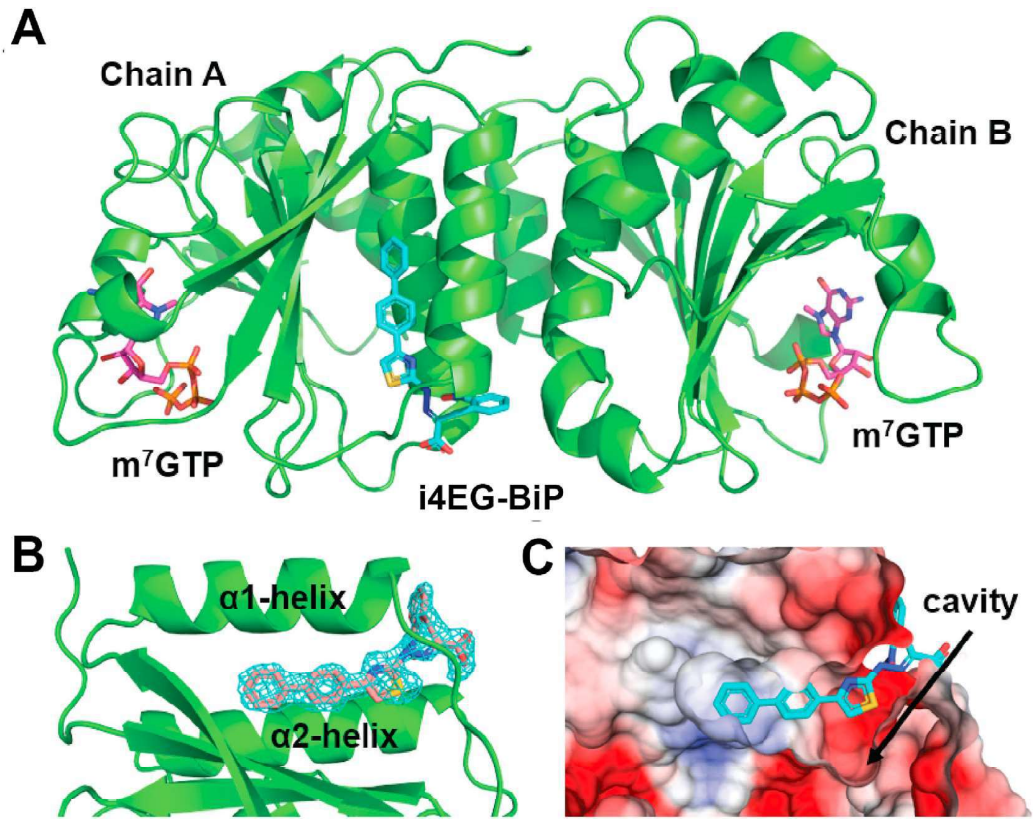


Fig. 2. A. Structure of eIF4E bound to i4EG-BiP. The asymmetric unit consists of two cap-bound eIF4E structures, of which one shows density for the small molecule. B. i4EG-BiP density is unambiguously located between helix α1 and α2. C. Surface representation of the co-crystal structure shows the cavity into which i4EG-BiP binds.

binding site, opening up possibilities for the development of a new class of eIF4E inhibitors. The structure of the protein itself has a striking similarity to that of eIF4E bound to 4EGI-1 (PDB ID 4TPW). The unit cell is comprised of a dimer with two copies of palm-like eIF4E molecules arranged at 180° orientations relative to each other around a pseudo-symmetry axis positioned between the two dorsal surfaces. m^7GTP, which constitutes the cap structure of mRNA, is bound to the palm of eIF4E. However, only one copy of the protein (chain A) is occupied by the i4EG-BiP ligand, while the binding site on the other copy (chain B) is empty. This is similar to the case of 4EGI-1, where we found the ligand bound to only one of the two proteins in the unit cell. Furthermore, as observed in the case of 4EGI-1, the 3—10 helix between residues S82 and L85 has melted into a loop while the α1 helix (residues H78-L85) is extended by one turn upon engagement of i4EG-BiP. This conformational change induced by i4EG-BiP is seen in chain A, but is obviously not present in chain B, which lacks the small molecule.

i4EG-BiP fits well within its binding cavity on the surface of eIF4E, burying 334 $Å^2$ of the eIF4E surface area, yet leaving room to expand the small molecule near the thiazole moiety, which is proximal to a deep cavity as shown in Fig. 2C. There is an estimated 200 $Å^2$ of additional surface area available that could be leveraged to increase the binding affinity of i4EG-BiP. The i4EG-BiP binding mode to eIF4E is distinct from any previously described eIF4E binder. The compound engages the pocket between helix α1 and helix α2 (Fig. 3A). The biphenyl moiety is buried deeply, while the carboxylic acid and the nitrophenyl functionalities are solvent exposed. The compound engages eIF4E using hydrophobic interactions with residues L45, L75, I79, L93 and L134. Edge-to-face π-π interactions are formed between the biphenyl moiety and Y76, Y91 and W130. Hydrogen bonds are formed between the nitro group of i4EG-BiP and the amide side chain of Q80 as well as between the hydrazone NH of i4EG-BiP and the backbone carbonyl oxygen of Q80. The carboxylic acid moiety of i4EG-BiP forms a hydrogen bond with the NH backbone of L85 (Fig. 3C), instead of the salt-bridge with K49, which represents the strongest interaction found for 4EGI-1 (Fig. 3D). Upon binding of i4EG-BiP, the α1 helix is extended by the amino acids Q80, L81, S82 and S83. These four residues constitute a loop that faces the α2 helix in other eIF4E structures without a small molecule inhibitor. This observed helix extension is needed to create room for the thiazole, hydrazone and nitrophenyl moieties of i4EG-BiP and the formation of hydrogen bonds between Q80 and the ligand. It is unclear why the same helix extension is observed with the binding of 4EGI-1, since in this case the structural rearrangement neither enables new protein interactions nor does it serve to avoid steric clashes. This helix extension does, however, appear to be a crucial factor in eIF4G displacement which is discussed in greater detail later.

2.3. i4EG-BiP is able to displace the eIF4G peptide from eIF4E

To assess the ability of inhibitors binding to eIF4E to disrupt the eIF4E-eIF4G interaction, we used a previously reported fluorescence polarization (FP) assay [23]. This assay leverages the fact that a free fluorescently labelled eIF4G peptide has a molecular correlation time that is vastly different from when this peptide is bound to eIF4E, thereby allowing displaced (free) peptide to be distinguished from bound peptide. In this displacement assay, increasing concentrations of compounds were titrated against eIF4E bound to a fluorescently labelled eIF4G peptide. Compounds that can

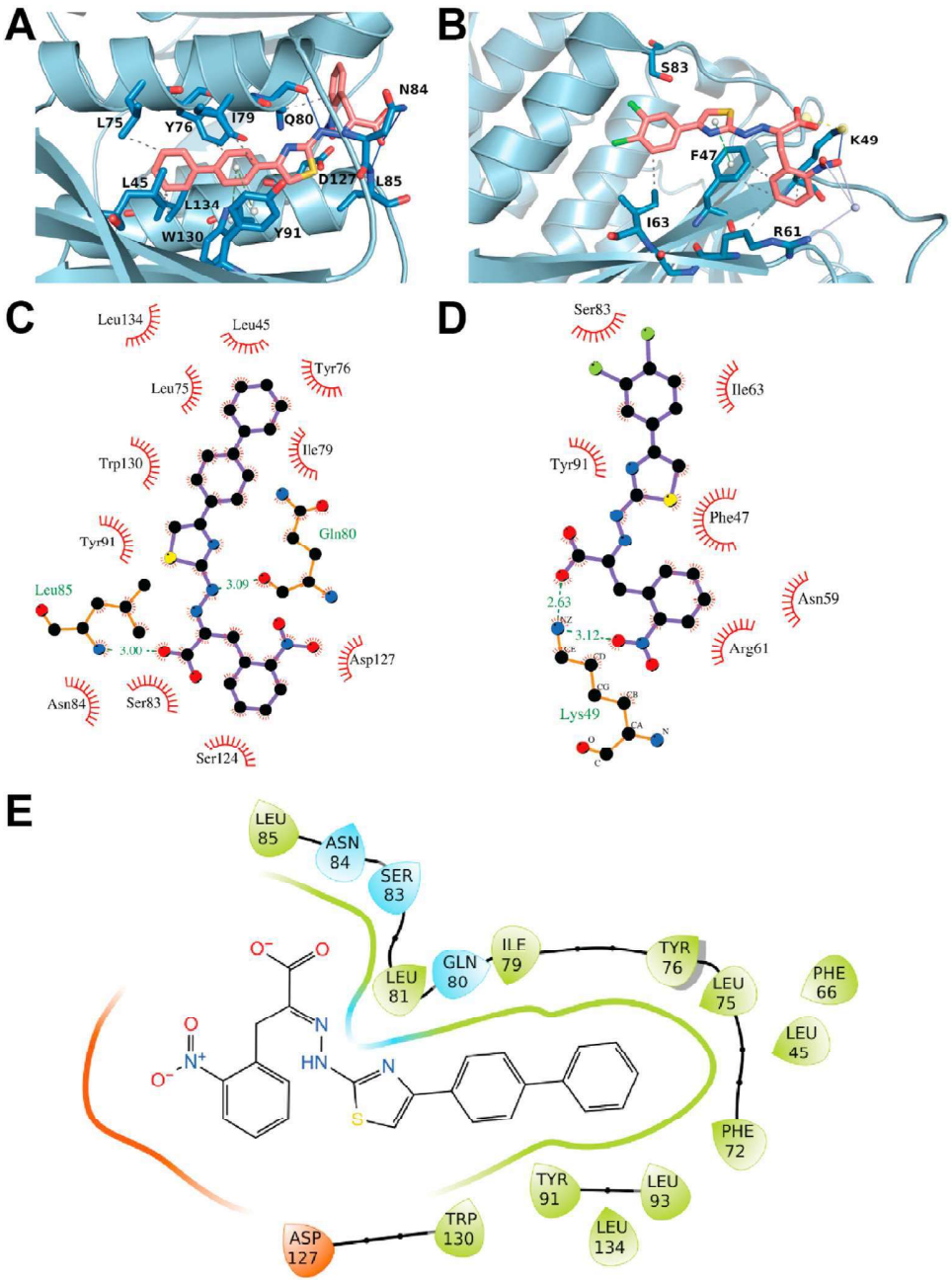


Fig. 3. A. Interaction profile of i4EG-BiP. eIF4E is depicted in light blue, i4EG-BiP is depicted in salmon and the side chains of eIF4E involved in the interaction are depicted in blue. **B.** Interaction profile of 4EGI-1. eIF4E is depicted in light blue, 4EGI-1 is depicted in salmon and the side chains of eIF4E involved in the interaction are depicted in blue. **C.** i4EG-BiP makes hydrophobic interactions with most residues of helix α1 and α2. Hydrogen bonds are observed between the small molecule and the side chains of Gln80 and Leu85. **D.** 4EGI-1 makes primarily hydrophobic interactions with different residues than those interacting with i4EG-BiP. A salt-bridge and hydrogen bond between the small molecule and Lys49 is also formed. **E.** Depiction of eIF4E residues within 4 Å of i4EG-BiP.

displace the eIF4G peptide led to a reduction in FP signals. Our results show that i4EG-BiP was able to displace the eIF4G peptide from eIF4E with an IC_{50} of 68 ± 2 μM, which is similar to that observed for 4EGI-1 (Fig. 1C).

2.4. i4EG-BiP impairs viability in MCF7 and MM1S cells

To test the activity of i4EG-BiP in cells, we performed time-dependent CS-BLI cell viability assays [38] in two different cell lines. MCF7 and MM1S cells were chosen as representative cell lines due to reported sensitivity to eIF4E ablation in the DepMap database [39]. Cells were treated once with serial dilutions of either 4EGI-1 or i4EG-BiP and assayed at 24, 48 or 72 h (Fig. 4A). 24 h of treating MCF7 cells with 4EGI-1 resulted in an IC_{50} of 71 μM, which decreased to 43 μM at 48 h and 35 μM at 72 h, thus resulting in a 2-fold decrease in viability as compared to 24 h. i4EG-BiP treatment inhibited cell viability of MCF7 cells over a treatment period of 24 h with an IC_{50} value of 59 μM, which decreased to 27 μM at 48 h and 18 μM when treatment is prolonged for 72 h. This corresponds to a decrease in viability of more than 3-fold from 72 h treatment as

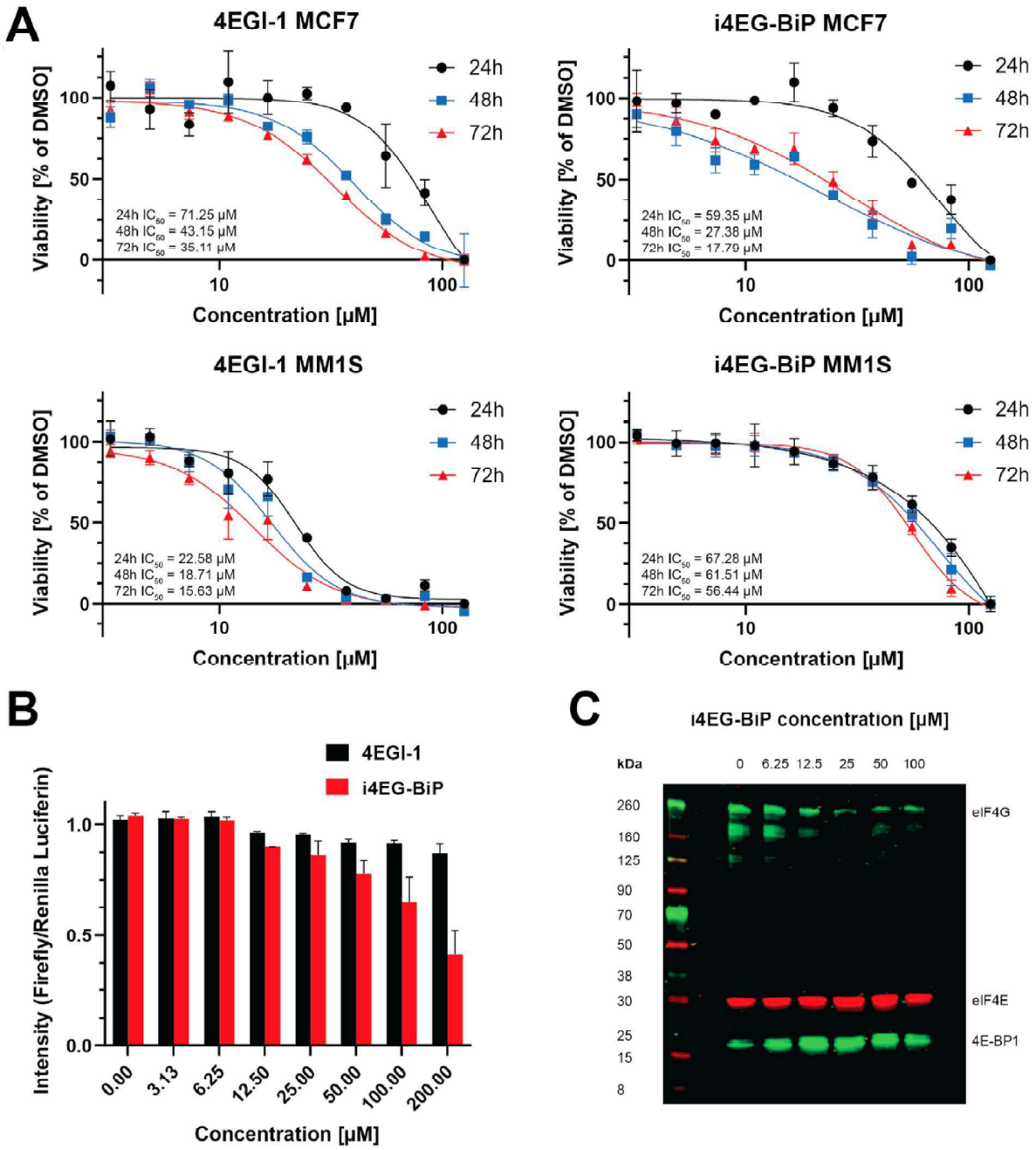


Fig. 4. A. Cell viability assay of 4EGI-1 and i4EG-BiP in two different cell lines (MCF7 and MM1S) as a function of concentration over 24 h, 48 h and 72 h. Error bars represent standard deviation and in all cases n = 3. B. A dual luciferase assay for 4EGI-1 and i4EG-BiP shows a greater decrease in cap-dependent translation with increasing compound concentration for i4EG-BiP. C. Western blot of an eIF4E pull down with m^7-GDP functionalized agarose beads in the absence and presence of increasing concentrations of i4EG-BiP. With increasing concentrations of i4EG-BiP, the eIF4E/eIF4G ratio increases, while the eIF4E/4E-BP ratio decreases.

compared to 24 h. In MM1S cells, we did not observe a significant time-dependent effect with either compound. The IC_{50} values for 4EGI-1 were 23 μM for a treatment period of 24 h, 19 μM for 48 h and 16 μM for 72 h. i4EG-BiP was less active in MM1S cells, with IC_{50} values of 67 μM after treatment for 24 h, 62 μM for 48 h and 56 μM for 72 h. This data indicates that i4EG-BiP affects the viability of MCF7 to a similar extent to 4EGI-1. In MM1S cells i4EG-BiP is less active compared to 4EGI-1, while both compounds do not exert significant time dependency in their cytotoxic effects.

2.5. i4EG-BiP inhibits cap-dependent translation better than 4EGI-1 in cellular assays

To test the inhibition of cap-dependent protein translation by i4EG-BiP in cells, we transfected a plasmid into HEK293 cells that produces a bicistronic transcript containing a 5′-Firefly luciferase followed by an IRES and a 3′-Renilla luciferase. The Firefly luciferase therefore reports on cap-dependent translation while the Renilla luciferase reports on cap-independent translation. The day after

transfection, cells were treated with increasing concentrations of compounds and following a 3 h incubation, Firefly and Renilla luciferase activity were measured by chemiluminescence. The interaction between eIF4E and eIF4G is essential for cap-dependent translation, but not IRES-dependent translation, and therefore only the Firely luciferase activity should be affected by compound treatment. Interestingly, we found a 65% reduction in the Firefly to Renilla (L/R) luciferase activity ratio with i4EG-BiP treatment while treatment with 4EGI-1 resulted in a considerably lower 15% L/R ratio reduction (Fig. 4B). This data indicates that i4EG-BiP inhibits cap-dependent translation to a greater extent than 4EGI-1.

2.6. i4EG-BiP disrupts the eIF4E-eIF4G and strengthens the eIF4E-4E-BP1 interactions

We assessed the interaction between eIF4E and its binding partners eIF4G and 4E-BP1 in pull-down experiments using a simplified cap-analogue (m^7GDP-agarose) resin [40]. HeLa cell lysates were incubated with either DMSO as a negative control or increasing concentrations of i4EG-BiP. After incubation of the mixture with m^7GDP agarose resin, beads were washed, followed by elution of bound proteins and analysis by Western Blot. As can be seen in Fig. 4C, the amount of eIF4G relative to eIF4E decreases with increasing amounts of i4EG-BiP, while the amount of 4E-BP1 increases. These results suggest that the extension of the α1-helix in eIF4E, caused by binding of i4EG-BiP, is detrimental for eIF4G binding to eIF4E, but does not adversely impact 4E-BP1 binding.

2.7. Design and synthesis of the 4EGI-1-based PROTAC d4E-1

Next, we decided to leverage the structural information we had to design eIF4E-targeting degrader molecules (PROTACs). The idea was to deal a double-blow to inhibit translation, one by blocking the eIF4E-eIF4G interaction, and the other by degrading eIF4E. We first assessed the structure of 4EGI-1 for functional groups to which a flexible linker could be coupled for the attachment of thalidomide. The carboxylic acid moiety appeared to be ideal for linkage; however, it is also involved in a crucial interaction with eIF4E (Fig. 3D). We therefore decided to attach the linker to the free carbon of the thiazole moiety, which is surface exposed in the co-crystal structure (PDB ID: 4TPW). Based on the simulated modelling of eIF4E bound to 4EGI-1 in complex with Cereblon (CRBN) bound to lenalidomide (PDB ID: 4CI2), we decided to synthesize the first test compound with a 4-carbon linker (Fig. 5A).

The synthetic challenge was to introduce the acetic acid functionality to the thiazole ring and choose a selective protection strategy for the resulting two carboxylic acids. Therefore, succinic anhydride was reacted with 1,2-dichlorobenzene in a Friedel Crafts acylation reaction to produce the phenyl-4-oxobutanoic acid **1** with 41% yield (Fig. 5B). The carboxylic acid was protected using iodomethane as a methylation agent to quantitively yield the methyl ester **2**. Bromination of **2** in the alpha-keto position produced the phenyl-3-bromo-4-oxobutanoate **3** with 96% yield. The hydrazine functionalized thiazole **4** was obtained by reacting this product with thiosemicarbazide. Due to the high reactivity of **4** it was used without further purification. To finish the synthesis is the 4EGI-1 analogue, the nitrophenyl-2-oxopropanoic acid **5** was created by hydrolysis of an oxazol-5-one derivative, which was obtained from reacting 2-Nitrobenzaldehyde with acetyl glycine. The overall yield of this 2-step reaction was 70%. Then, the acid was protected using *tert*-butyl acetate to form the *tert*-butyl ester **6** with 79% yield. The crude hydrazineylthiazol **4** was then coupled to the *tert*-butyl ester **6** to form an E/Z-isomeric mixture of the protected 4EGI-1 building block **7** with 63% yield.

As a starting point, we chose to synthesize the first PROTAC as a 4EGI-1 analogue linked to a 2-phenoxyacetamide derivative of thalidomide, as described for dBET1, a BET bromodomain degrader developed by Winter et al. [33]. To achieve this, 4-Hydroxythalidomide **8** was synthesized by reacting 3-Hydroxy phthalic anhydride and 3-Aminopiperidine-2,6-dione in glacial acetic acid with 96% yield (Fig. 6A). The linker was prepared using a benzyloxycarbonyl (CBz) protected diamine, which is more economical than the previously used Boc protected variant [33]. Benzyl (4-aminobutyl)carbamate was reacted with 2-Chloroacetyl chloride to produce the CBz-protected 2-Chloroamide **9** in 91% yield, which was subsequently transformed into the more reactive 2-Iodoamide **10** via Finkelstein reaction in a quantitative manner. **10** was then coupled to 4-Hydroxythalidomde **8** to yield the CBz-protected Thalidomide linker, which was deprotected using hydrogen gas and a Palladium black catalyst to give the free Thalidomide linker **11** in 41% yield.

To selectively deprotect the acid moiety on the thiazole ring and generate the final PROTAC d4E-1, methyl ester hydrolysis of the *tert*-butyl protected 4EGI-1 building block **7** was performed using lithium hydroxide at 4 °C to create the ready-to-couple 4EGI-1 derivative **12** (Fig. 6B). The final product d4E-1 was then generated with a total yield of 41% over two steps using HATU-mediated coupling of the 4EGI-1 derivative 12with the Thalidomide linker 11 and subsequent *tert*-butyl deprotection using 12% trifluoroacetic acid in dichloromethane.

2.8. d4E-1 interacts with eIF4E, but does not engage Cereblon in cells

After successful synthesis of the 4EGI-1 based PROTAC d4E-1, we next wanted to analyse binding of the bifunctional molecule to its targets: eIF4E and Cereblon.

2.8.1. Binding to eIF4E by displacement of the eIF4G peptide

To test whether d4E-1 can displace eIF4G in a similar manner compared to 4EGI-1, the compound was tested for its ability to displace a fluorescently tagged eIF4G peptide in an FP assay (Fig. 8A). At the three measured concentrations (33, 100 and 300 μM), both d4E-1 and 4EGI-1 displaced the 4G peptide in a similar manner. We therefore concluded that the binding affinity of our PROTAC compound is comparable to that of the parent compound, 4EGI-1.

2.8.2. Cellular assay to evaluate the ability of 4EGI-1 degraders to engage Cereblon

Next, we wanted to assess the cellular CRBN engagement of d4E-1. We used a previously described assay [41,42] in which compounds are tested for their ability to rescue dBET6-induced (a CRBN-dependent BET bromodomain degrader) degradation of $BRD4_{BD2}$ by competing for binding to CRBN. Flp293T cells stably expressing a $BRD4_{BD2}$-GFP fusion protein and an mCherry reporter were co-treated with dBET6 at 100 nM and compounds in dose response, with lenalidomide used as a positive control. The GFP/RFP signal ratio was quantified using an Acumen laser scanning cytometer (TTP Labtech). Active compounds were identified by an increased GFP/mCherry ratio resulting from inhibition of $BRD4_{BD2}$-GFP degradation by dBET6. To our surprise, d4E-1 did not prevent degradation of $BRD4_{BD2}$ in a dose dependent manner (results shown later). We hypothesized that this could be due to the net negative charge of d4E-1, which could hinder cellular uptake in a similar manner to the cap-based degraders previously reported [37]. Therefore, our next goal was to synthesize an optimized prodrug version of d4E-1.

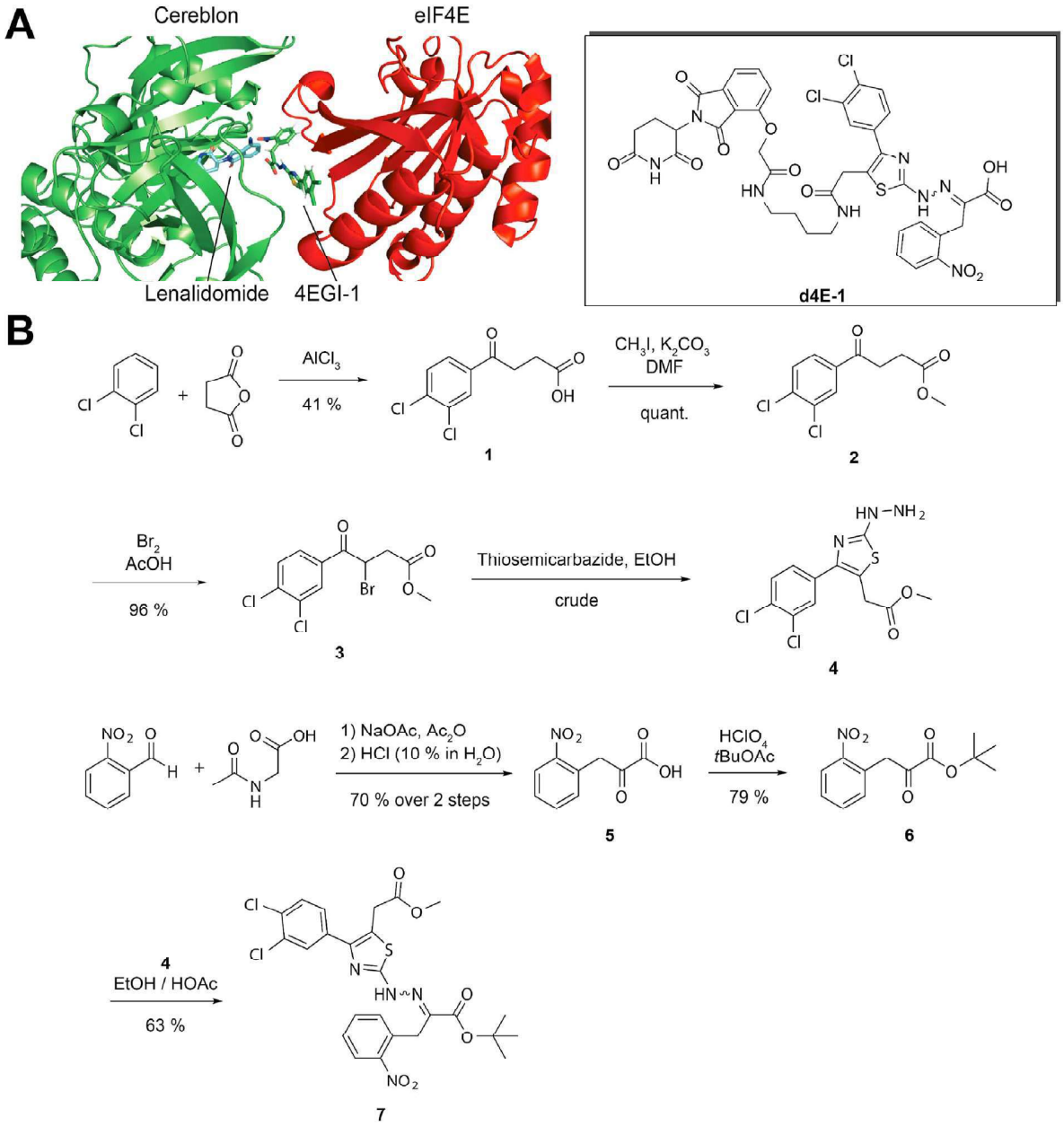


Fig. 5. A. Docking of lenalidomide bound to Cereblon (PDB ID: 4CI2) and eIF4E bound to 4EGI-1 (PDB ID: 4TPW). Surface exposed parts of the small molecule can be arranged in proximity without steric clashes. The target degrader molecule derived from this docking pose (d4E-1) is depicted on the right. B. Synthesis of the double protected 4EGI-1 building block 7.

2.9. Design, synthesis and CRBN engagement of a 4EGI-1-based prodrug PROTAC

To negate the hypothesized negative effects from the negative charge on the carboxylic acid, a commonly applied prodrug approach was employed. In this approach, the carboxylate is modified with a protective group that can be cleaved off in cells after uptake. We chose to use a pivaloyloxymethyl ester as a prodrug unit, a moiety that is also used in prodrug versions of ampicillin (pivampicillin) [43] or butyric acid (AN-9) [44], for example. In addition to the prodrug strategy, arylamine linkage to thalidomide was used as these CRBN ligands reduce synthetic efforts significantly. To further improve cellular uptake, the 2-nitrophenyl moiety was removed from the 4EGI-1 scaffold, as this functionality reduces cellular uptake (results from i4EG-BiP PROTAC analogues shown later).

The arylamine linked thalidomide derivative **15** was synthesized as previously described by others [45]. To make the free acid of the modified 4EGI-1 PROTAC, the hydrazine derivative **4** was reacted with pyruvic acid and subsequently protected using *tert*-butylacetate (Fig. 7). HATU-mediated amide coupling with **15** and acidic deprotection with 12% trifluoroacetic acid in dichloromethane yielded the free acid of the 4EGI-1-prodrug PROTAC **d4E-6**. Despite repeated efforts, the purity of the **d4E-6 free acid** could not be improved to satisfying levels, hence the crude product was reacted with pivaloyloxymethyl iodide to generate pure **d4E-6** with an overall yield of 31% over 5 steps.

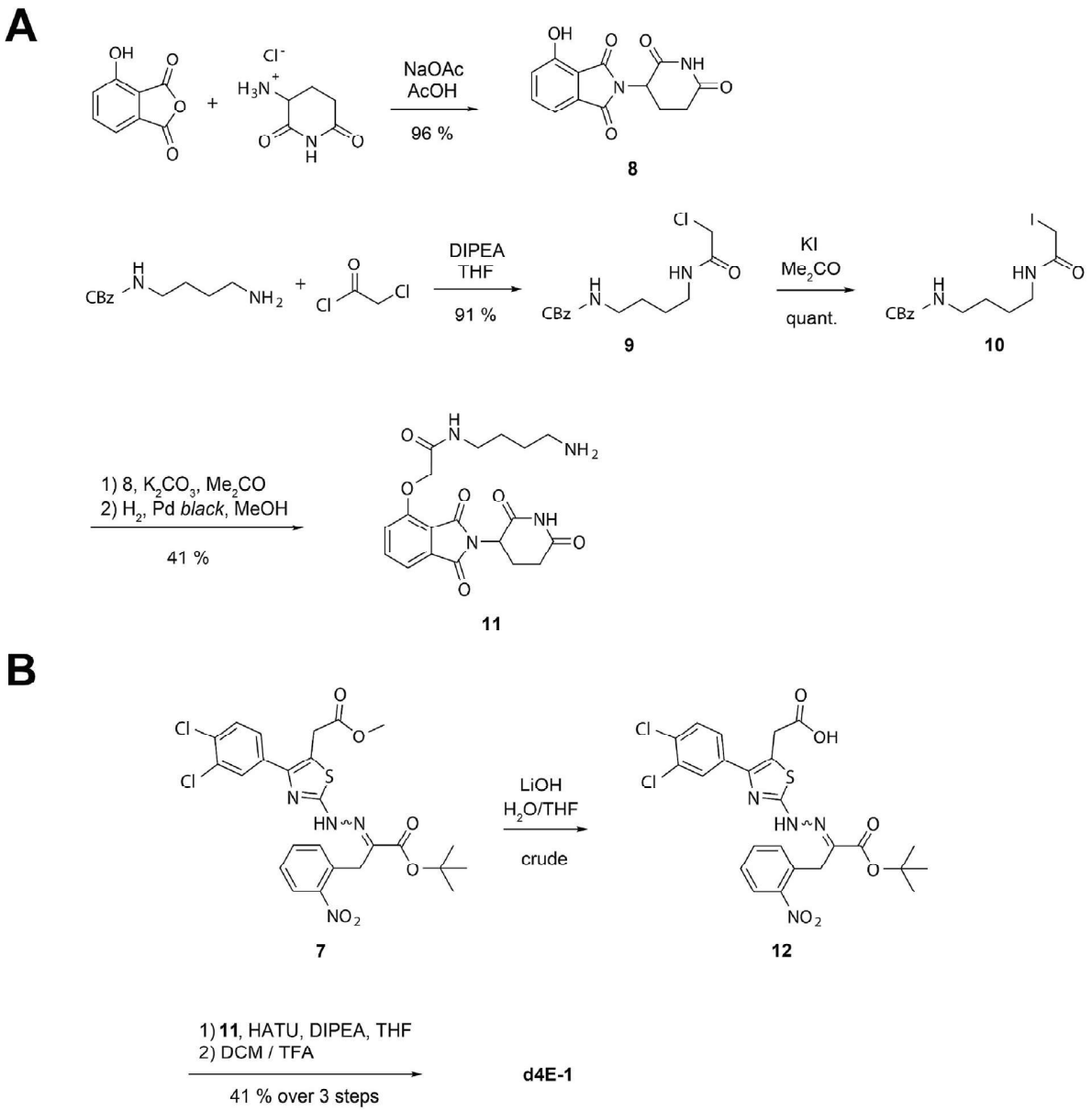


Fig. 6. A. Synthesis of the 2-phenoxyacetamide thalidomide linker 11. B. Endgame of the synthesis of target structure d4E-1.

Using the same assay for cellular CRBN engagement, we evaluated d4E-6 and its parent compound. While the free acid was not capable of rescuing dBET6-induced degradation of BRD4 (data not shown), the prodrug did with an IC_{50} value of ~20 μM (Fig. 8B). These encouraging results supported our previous hypothesis that the net negative charge in our 4EGI-1-based PROTAC d4E-1 was root cause for its lack of activity.

2.10. Design and synthesis of the i4EG-BiP-based PROTACs (d4E-2 to d4E-5)

To optimize and expand the spectrum of possible eIF4E degraders, we switched to using i4EG-BiP as a scaffold for new PROTACs. In this case, the carboxylic acid moiety is not involved in the same crucial interaction found with 4EGI-1 (Fig. 3C and D). Therefore, we hypothesized that direct linking of the acid to a thalidomide linker should not result in a significant loss of affinity, while beneficially eliminating the net negative charge of the product. Furthermore, the structure suggests that the 2-nitrophenyl moiety plays a minor role for interaction. On the basis of this observation, we decided to include structures that lack this functional group. We tried to further improve CRBN binding by switching from phthalimide ether to arylamine phthalimide. This was driven by the fact that the field developed improvements in linker design, which encouraged us to opt for arylamine phthalimides instead of the phthalimide ether used for d4E-1 [46]. Regarding chirality of the thalidomide analogues it is well established that the stereocenter interconverts *in vivo* [47].

Starting from 1-([1,1′-biphenyl]-4-yl)-2-bromoethan-1-one, the crude reactive hydrazine derivative **13** was synthesized using thiosemicarbazide (Fig. 9A). **i4EG-BiP** and the i4EG-BiP derivative lacking the 2-nitrophenyl moiety **14** were generated using 2-nitrophenyl pyruvic acid and pyruvic acid with 63% and 57% yield, respectively. The arylamine linked thalidomide derivatives

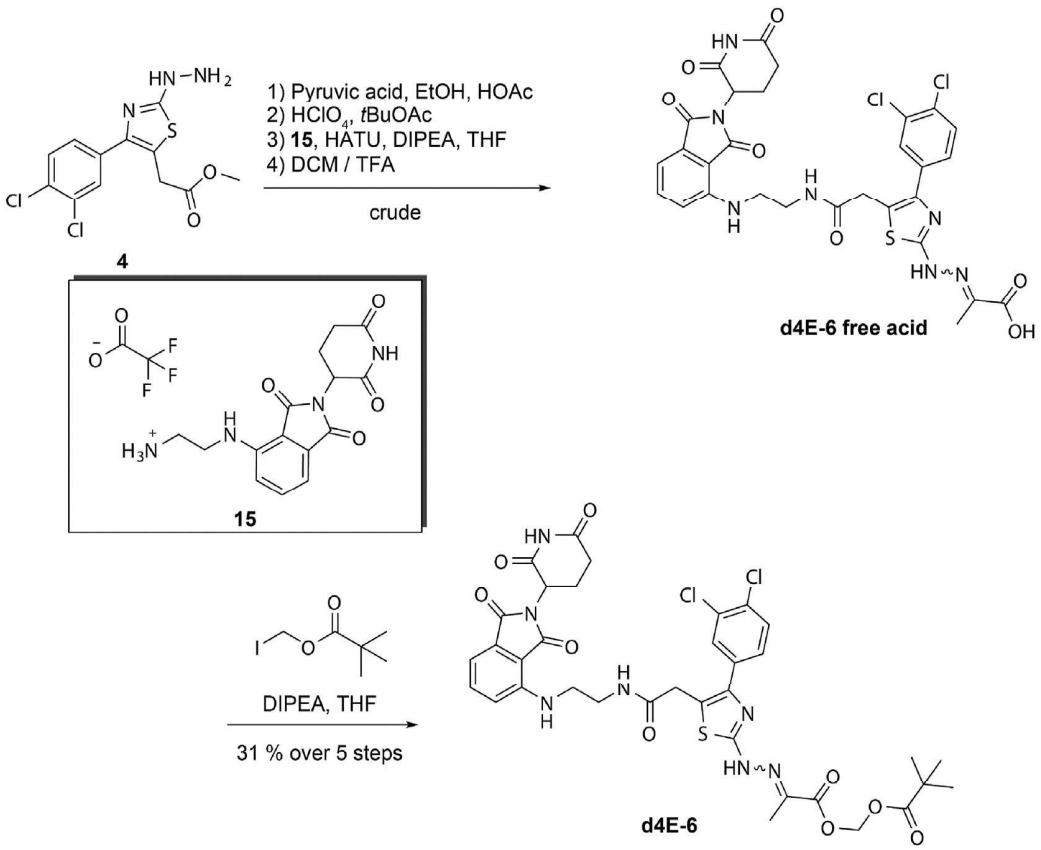


Fig. 7. Synthesis of the 4EGI-1 prodrug PROTAC d4E-6 by POMylation of the free acid 4EGI-1 analog with iodomethyl pivalate.

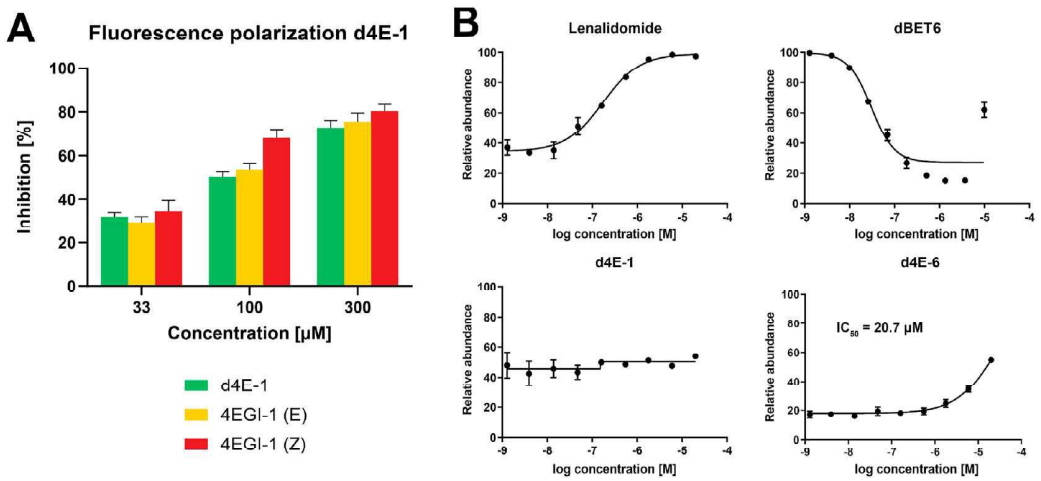


Fig. 8. A fluorescence polarization assay shows that d4E-1 displaces the eIF4G peptide in a manner comparable to 4EGI-1. B. Cellular Cereblon engagement assay used to elucidate intracellular activities of d4E-1 and d4E-6.

15, **16** and **17** were synthesized as previously described by others [45] (Fig. 9B). As before, HATU-mediated amide coupling was used to generate the PROTACs **d4E-2**, **d4E-3**, **d4E-4** and **d4E-5** with 43%, 39%, 40% and 43% yields.

2.11. i4EG-BiP-based degraders engage Cereblon in cells

To test whether the new PROTACs were able to engage CRBN in cells, we tested two model compounds to evaluate the effects of the 2-nitrophenyl moiety on cellular uptake: d4E-2 (4-carbon linker, with no nitrophenyl), and d4E-3 (4-carbon linker, nitrophenyl included). The best molecule tested was d4E-2 with an IC_{50} value of 1.6 μM. In contrast, d4E-3 was approximately 8-fold less active (with an IC_{50} value of 13 μM), suggesting the nitrophenyl moiety had a negative effect on cellular uptake (Fig. 9D). Based on these findings, we tested three nitrophenyl-free compounds with varying carbon linker lengths (d4E-2, d4E-4 and d4E-5) for cytotoxicity in HeLa cells (Fig. 9C). The efficacy of the tested compounds decreased

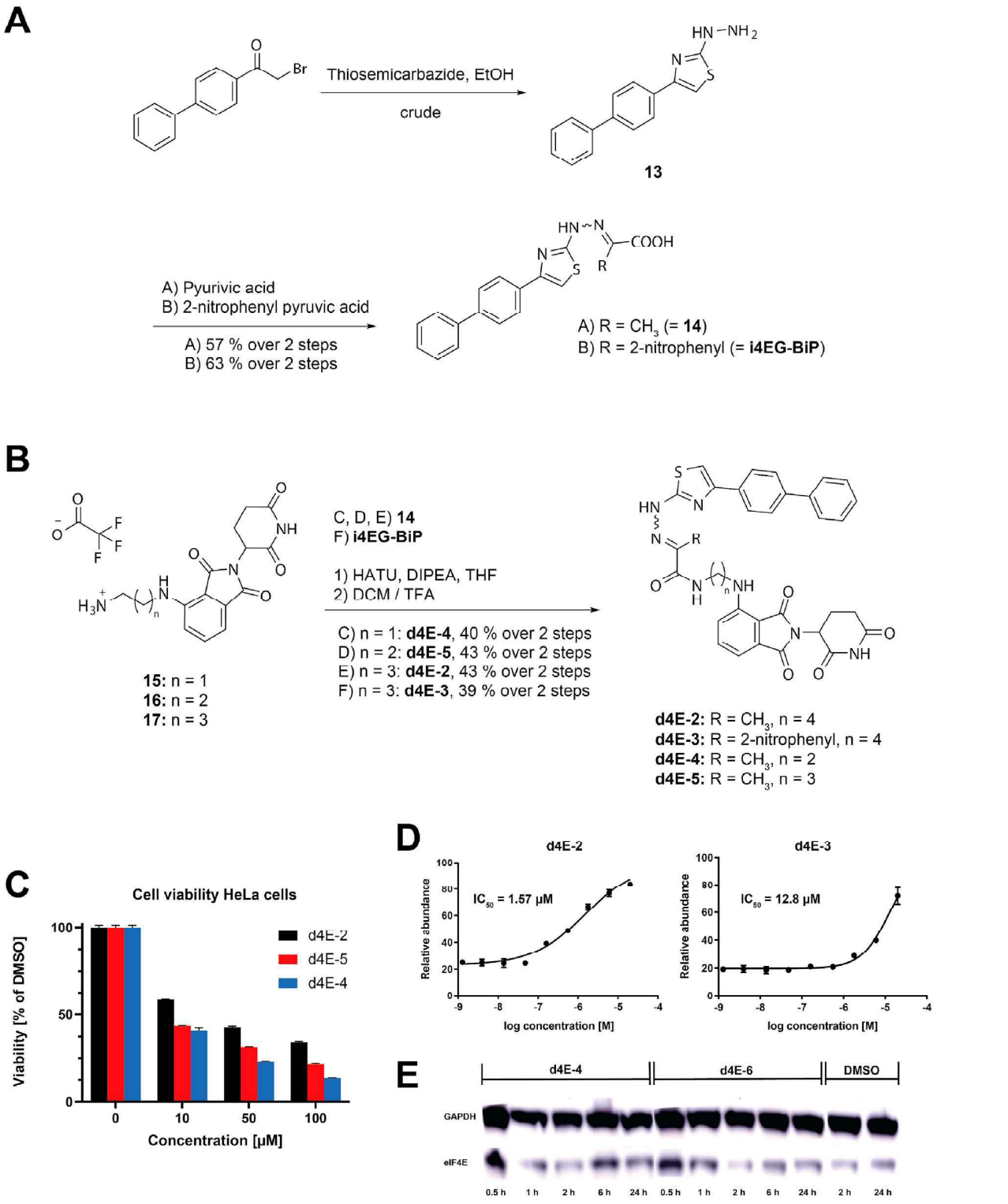


Fig. 9. A. Synthesis of i4EG-BiP and the i4EG-BiP deriative lacking the 2-nitrophenyl moiety (**14**). B. Endgame of the synthesis of target structures d4E-2, d4E-3, d4E-4 and d4E-5. C. Cell viability data for PROTACs in HeLa cells (triplicates over 24 h). D. Cellular Cereblon engagement assay used to elucidate intracellular activities of d4E-2 and d4E-3, to compare the influence of the 2-nitrophenyl moiety on cellular uptake. E. Representative Western Blot of eIF4E degradation by d4E-4 and d4E-6 at 10 µM compound concentration.

with increasing linker length, but more strikingly all PROTACs displayed strong cytotoxic effects, even at rather low concentrations, with d4E-4 and d4E-5 killing over 50% of the cells at the lowest concentration measured (10 μM).

After these promising initial biochemical and cellular results, we proceeded to test our two best compounds (d4E-4 and d4E-6) for their potential to induce eIF4E degradation in HEK293T cells. We performed two experiments: 1) treatment of HEK293T cells with PROTAC concentrations varying between 0.1 and 10 μM for a period of 24 h, and 2) treatment of the same number of HEK293T cells with 1 μM PROTAC for varying lengths of time, ranging from 0.5 h to 24 h. Unfortunately, neither experiment showed significant degradation of eIF4E, as assessed by immunoblotting (representative Western Blot shown in Fig. 9E). However, we did observe that eIF4E levels varied at different time points for both compounds as well as DMSO treated cells. This suggests that a cellular response may be counteracting eIF4E degradation. To test whether eIF4E was being ubiquitinated in the first place, HEK293T cells were incubated with 1 μM PROTACs in the presence of 100 nM Bortezomib, a known inhibitor of the 26S proteasome [48]. We compared eIF4E ubiquitination in these cells to those from cells treated with Bortezomib and a DMSO control; however, we were not able to detect any significant differences (data not shown), suggesting that our PROTACs did not induce CRBN-mediated ubiquitination of eIF4E.

We performed quantitative proteomics experiments to measure the global change in protein expression resulting from treatment with d4E-2 relative to DMSO control. Proteomics experiments have confirmed that d4E-2 is engaging CRBN in cells as we see downregulation of C2H2 zinc finger targets, SALL4, ZNF692 and ZNF827, which are commonly degraded as a result of the IMiD CRBN-binding handle of the degrader [58] (Supplementary Fig. 3). We also observed that d4E-2 did not downregulate eIF4E.

2.12. Solution NMR studies confirm binding modes of the i4EG-BiP-based degraders to eIF4E are similar to that of the parental compound

Due to these unexpected results, we wanted to confirm that the derivatization of i4EG-BiP with thalidomide did not affect the binding of the degraders to eIF4E. To do this, we used solution-state nuclear magnetic resonance (NMR) spectroscopy as an additional method to identify the binding site of i4EG-BiP on eIF4E and compare it with the i4EG-BiP-based PROTAC d4E-4 (Fig. 10A and C). Chemical shift perturbations (CSPs) were observed when ^{15}N-labelled GB1-eIF4E was titrated with i4EG-BiP or d4E-4. Significant CSPs were observed in both cases at the binding interface identified from the co-crystal structure (Fig. 10B and D). A molecular docking simulation of d4E-4 was carried out with QuickVina 2 [49], which is based on AutoDock Vina [50]. The obtained docking score was −8.2 kcal/mol. CSPs plotted onto the structure of eIF4E bound to i4EG-BiP as well as the docked structure of d4E-4 reveal significant CSPs located in similar regions for both small molecules (Fig. 10E and F). Although the overall CSP patterns look similar for i4EG-BiP and d4E-4, indicating similar binding modes, some additional CSPs were observed towards the N- and C-termini of the protein when d4E-4 was added. This could originate from non-specific binding events due to the 3-fold excess of inhibitor that was used in these experiments, in order to obtain significant CSPs. Alternatively, these additional CSPs could be the result of conformational changes induced by ligand binding. In particular, the C-terminal residues constitute the cap-binding region and there have been previous reports of crosstalk between eIF4G binding to the dorsal surface and cap binding to the palm of eIF4E. This may be the result of dynamic loops in the cap-binding region, which have been observed in multiple conformations in the various published crystal structures.

3. Discussion

The cap-dependent translation machinery represents a relatively underexplored area of vulnerability for targeted therapeutics in cancer, with many genes (eIF4A, eIF4E, eIF4G) listed as sensitive in DepMap. Inhibiting the protein-protein interaction of eIF4E with eIF4G and concomitantly disrupting the translation initiation complex has been shown to be an attractive therapeutic option. Inhibitors of the eIF4E-eIF4G interaction have been shown to inhibit translation, and exhibit antiproliferative and putative antitumor activities [51,52].

The previously described 4EGI-1 and i4EG-BiP, introduced here, bind to different allosteric sites on eIF4E, but both result in a similar conformational change (extension of the helix α1 by one turn) which in turn displaces eIF4G. i4EG-BiP binds to an internal cavity of eIF4E, rather than on the surface as done by 4EGI-1. There is a distinct possibility that an endogenous metabolite could also engage this i4EG-BiP cavity, thus modulating translation, but such a molecule has yet to be identified. We found that 4EGI-1 and i4EG-BiP can displace eIF4G but not 4E-BP1 in cells, although both eIF4G and 4E-BP1 share the same consensus binding motif. We posit that extension of the helix resulting from engagement of the small molecules hinder binding of eIF4G due to a steric clash. Specifically, the side chain of Ser-82 in eIF4E will clash with a conserved Leu-641 in eIF4G that is a part of the extended interface (Supplemnentary Fig. 1). In the case of 4E-BP1, this steric clash would not occur. Instead, we predict the complex with eIF4E would be stabilized by a putative hydrogen bond between the sidechain of Ser-82 in eIF4E and the backbone of Ser-83 from 4E-BP1 (Supplementary movie M1). This model in which the small molecule would inhibit the translation activator eIF4G and stabilize the inhibitor 4E-BP1 presents a two-fold attack on translation initiation. Since the inhibitory effect stems from a similar allosteric change, both 4EGI-1 and i4EG-BiP, although binding to different sites, have similar ability to displace the eIF4G/4E-BP consensus peptide and exhibit similar cellular activity. Exactly how the extension of the helix α1 displaces the eIF4G/4E-BP consensus peptide, which lacks the extended binding region of the full-length proteins, still remains an open question.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.ejmech.2021.113435

To further improve the potency of the inhibitors 4EGI-1 and i4EG-BiP, we designed PROTAC versions of the two molecules. Here, we employed 4EGI-1 and the novel inhibitor i4EG-BiP as scaffolds to create cell-permeable thalidomide conjugates that bind to eIF4E and engage Cereblon. Engagement of Cereblon as well as VHL with eIF4E ligands has been previously attempted by Kaur et al. using cap analogues. Even though they observed binding to eIF4E by their PROTACs, target degradation was not achieved. The authors attributed this failure to the net negative charge of the m^{7}GDP-based compounds [37]. Here we show that the design of a neutral PROTAC based on 4EGI-1 using a prodrug approach and neutral PROTACs based on i4EG-BiP using the carboxylic acid moiety as a linking point seems to alleviate this hurdle on efficacy. Using cellular Cereblon engagement assays, we were able to show a concentration-dependent cell-based effect of our compounds.

However, we were not able to detect any targeted cellular degradation of eIF4E. This could, in part, be due to the relatively low binding affinity of the parent compounds, 4EGI-1 and i4EG-BiP. Using medicinal chemistry approaches to improve the binding affinity of i4EG-BiP might help to improve the PROTAC strategy for our i4EG-BiP-based molecules. This is particularly possible as there is unoccupied space on eIF4E in the vicinity of the i4EG-BiP binding site which can be leveraged to improve molecule affinity. The failed degradation of eIF4E could also be explained by increased binding

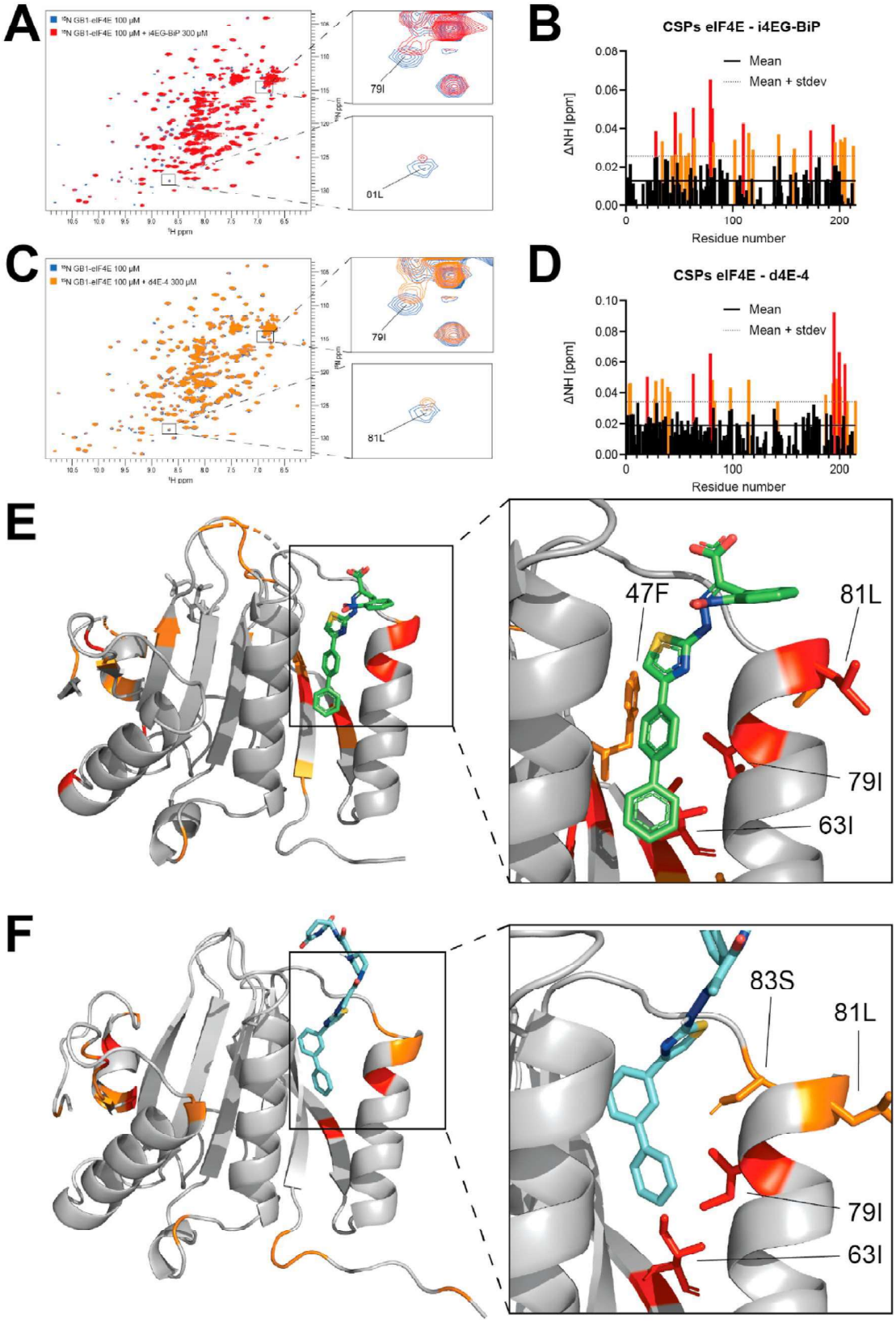


Fig. 10. A. ^{15}N–^{1}H-HSQC overlay of GB1-eIF4E in the absence (blue) and presence (red) of i4EG-BiP. B. Chemical shift perturbation (CSP) plot of changes induced by i4EG-BiP as a function of residue number. The solid line represents the mean value of all CSPs, the dashed line represents the mean value plus one standard deviation of all CSPs. C. ^{15}N–^{1}H-HSQC overlay of GB1-eIF4E in the absence (blue) and presence (orange) of d4E-4. D. CSP plot of changes induced by d4E-4 as a function of residue number. The solid line represents the mean value of all CSPs, the dashed line represents the mean value plus one standard deviation of all CSPs. E. CSPs plotted onto the eIF4E-i4EG-BiP structure: Highlighted in orange are residues with CSPs > mean + one standard deviation, highlighted in red are residues with CSPs > mean + two standard deviations. F. CSPs plotted onto the docked eIF4E-d4E-4 structure. Highlighted in orange are residues with CSPs > mean + one standard deviation, highlighted in red are residues with CSPs > mean + two standard deviations.

of 4E-BP1 to eIF4E as a result of eIF4E binding 4EGI-1 or i4EG-BiP. Increased binding of 4E-BP1 to eIF4E could result from conformational changes in eIF4E upon small molecule binding, which could increase the affinity for 4E-BP1, and/or the fact that the competing binding protein, eIF4G, is no longer able to engage eIF4E. It has been previously shown that ubiquitination of eIF4E at K159 is orchestrated by the E3 ubiquitin ligase CHIP, which leads to subsequent proteasomal degradation, and furthermore that this ubiquitination can be blocked by overexpression of 4E-BP1 [53]. Therefore, increased 4E-BP1 binding as a result of 4EGI-1 or i4EG-BiP binding could protect eIF4E from ubiquitination and stabilize the protein. It should however be noted that in our current study the ubiquitination is performed in a Cereblon-dependent manner, which might not have the same inhibition response to 4E-BP1 as CHIP. Future studies will be needed to address this possible obstructive role of 4E-BP1 on targeted degradation of eIF4E. If indeed the association of 4E-BP1 prevents the ubiquitination of eIF4E, an interesting approach would be to use 4E1RCat as a scaffold to create eIF4E targeting PROTACs, since this inhibitor has been shown to displace both eIF4G and 4E-BP1 binding to eIF4E [29]. Utilizing the synthetical method we provide in this manuscript to link the carboxylic acid moiety of 4E1RCat to lenalidomide could circumvent both cell-permeability hurdles as well as evasion of eIF4E ubiquitination by enhanced 4E-BP1 binding.

Blocking cap-dependent translation provides an attractive opportunity for future efforts to target cancer cells and eIF4E is the master regulator of this process. Disrupting the crucial interaction of eIF4E with its scaffold protein eIF4G or degrading eIF4E altogether are two independent routes to inhibit eIF4E function. Here we attempted to synergistically apply both these approaches. Although we did not achieve successful degradation of eIF4E, our successful targeting of eIF4E in this study with a newly identified small molecule inhibitor, i4EG-BiP, paves way for future efforts.

4. Material and methods

4.1. Expression and purification of eIF4E

A construct of human eIF4E was expressed in transformed *Escherichia coli* BL21(DE3). A Δ_{26}-eIF4E or GB1-eIF4E construct in a pET-28(+) backbone was used for crystallography and NMR studies, respectively. Bacteria were grown in Luria Broth (LB) at 37 °C. Protein expression was induced by the addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside at $OD_{600} = 0.6$ followed by incubation overnight at 23 °C. Cells were harvested with a yield of 3 g/L wet pellet and stored at −30 °C. Bacterial pellets were resuspended by slow pipetting in lysis buffer constituting of 50 mM Tris•HCl, pH 7.5, 100 mM NaCl, 1% Triton-X, 5 mM tris(2-carboxyethyl)phosphine (TCEP), 1 cOmplete™ Protease Inhibitor Cocktail tablet, lysozyme, RNase, and DNase.

Cells were subsequently homogenized in a cell microfluidizer and the lysates were centrifuged at 38,000×*g* for 1.5 h. After centrifugation the clarified lysate supernatant was first passed through a 0.45 μm cellulose acetate syringe-filter and then passed over a diethylaminoethylcellolose (DEAE) column previously equilibrated with the same lysis buffer. The DEAE flow-through was loaded on adipic–agarose-m^{7}GDP column and after 0.5 h of binding, the column was washed with 50 mL of wash buffer (10 mM HEPES, pH 7.5, 125 mM NaCl, and 1 mM TCEP) five times followed by elution four times with 10 mL elution buffer (10 mM HEPES, pH 7.5, 125 mM NaCl, 100 μM m^{7}GTP plus 10 mM TCEP). The eluted protein concentration was assessed by Bradford assay (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions and then concentrated to a final volume of 3 mL by ultrafiltration through a 15 mL, 10-kDa-cutoff Millipore centrifugal filter. Following concentration, up to 4 mL of eluate was subjected to size-exclusion chromatography using a Superdex75 16/10 preparative column (GE Healthcare) equilibrated with 10 mM HEPES, pH 7.5, 125 mM NaCl, and 1 mM TCEP buffer. The collected pure protein fractions were concentrated to a final concentration (by ultrafiltration) to 1 mg/mL as evaluated by NanoDrop™ at 280 nm. The total yield of the aforementioned process was approximately 3–5 mg of pure protein from 1 L of culture.

4.2. Protein crystallization

Crystallization was performed using the sitting drop method and all crystals were grown in drops containing 1 μL of protein-ligand solution and 1 μL of crystallization solution containing 10–25% (vol/vol) 3.3-kDa PEG, 100 mM MES, pH 6.0, 10% (vol/vol) isopropanol. The protein-ligand solution used for crystallization was prepared by mixing eIF4E (9 mg/mL) and the small molecule (i4EG-BiP or analogues) solution (12.5 mM in DMSO) at an approximately 1:1 stoichiometry. The resulting protein–small molecule mixture was then serially diluted from 9 to 1 mg/mL and used to set up sitting drop crystal trials. Crystals began to form in 2 days and grew to full size at day 4. They were inspected and based on their size and morphology, the optimal conditions of protein concentration and PEG were determined. Crystals were harvested and quickly transferred to a cryoprotectant solution of crystallization buffer plus 10% (v/v) glycerol before flash freezing in liquid nitrogen.

4.3. X-ray diffraction data collection

X-ray diffraction data were collected at the APS X-Ray Synchrotron Source at the Argonne National Laboratory from single protein ligand complex crystals. The data were integrated using XDS. The results indicated monoclinic space group P21 crystals with diffraction up to 1.9 Å resolution. Phases were calculated by molecular replacement using PHASER and an initial model of eIF4E (PDB ID 4TPW). Subsequently, models were manually inspected and refined in Coot followed by further rounds of phase calculations by molecular replacement in PHENIX and model structure refinement until a minimum R-free was reached. The position of the i4EG-BiP ligand was clearly seen in the Fo-Fc map before the ligand was included in the model.

4.4. Nuclear magnetic resonance experiments

^{15}N–^{1}H-TROSY-HSQC spectra were recorded on a Bruker Avance III 800 MHz spectrometer with a TXO-style cryogenically cooled probe. ^{15}N-labelled GB1-eIF4E was concentrated to a final concentration of 100 μM in a buffer composed of 50 mM sodium phosphate, pH 6.5, 50 mM potassium chloride, 2 mM DTT and 5% D_2O. A reference spectrum was recorded at 298 K by addition of DMSO to a final concentration of 1.5%. NMR titration was performed by recording ^{15}N–^{1}H-TROSY-HSQC spectra in the presence of either i4EG-BiP or d4E-4 at 50, 100 and 300 μM concentrations (diluted from 20 mM stock solutions in DMSO). NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1) [54].

4.5. Docking of d4E-4 to eIF4E

The protein structure used for docking was the structure of i4EG-BiP bound to eIF4E, with i4EG-BiP removed. The receptor structure was prepared in PDBQT format with AutoDock Tools, which is part of MGLTools [55], by assigning AutoDock atom types and merging the nonpolar hydrogen atoms. The ligand was

prepared in PDBQT format with Open Babel [56], which included computation of the 3-dimensional structure. The docking box was 18 Å × 30 Å x 18 Å, and the docking exhaustiveness was set to 4. The receptor was held rigid during the docking procedure, and 10 replicates of the docking procedure were executed.

4.6. Fluorescence polarization experiments

For the fluorescence polarization assays, a GST-eIF4E construct was used. It was expressed in *E. coli* and purified using the same procedure described above for Δ26-eIF4E without the DEAE column step. The assay was used to test the activity of various compounds against eIF4E/eIF4G—peptide complex formation. For the fluorescent probe we used a purified eIF4G-peptide conjugated with fluorescein derived by peptide synthesis with the sequence KKQYDREFLLDFQFK-FITCH. Assay mixtures consisted of 150 nM eIF4G-peptide plus 0.3 μM GST-eIF4E in 100 mM Na-phosphate, pH 7.5. A 384-well black plate was used to prepare serial dilutions of i4EG-BiP and other compounds from a 12.5 mM stock solution in DMSO at a starting ligand concentration of 500 μM. The fluorescence polarization signal was recorded using an EnVison™ plate reader.

4.7. Cell—specific bioluminescence imaging (CS-BLI) cell viability assay

CS-BLI was performed according to a previously published high-throughput cell viability assay [38]. Briefly, Luciferase-expressing MM1S cells and MCF7 were cultured in T75 flasks. RPMI or DMEM media (as appropriate), supplemented with 10% FBS, and penicillin and streptomycin were used accordingly. 10^4 cells were seeded in 50 μL of media in each well of a 96-well, tissue-culture-treated, white flat bottom plate and left to adhere overnight in a cell-culture incubator. The next day we supplemented the well volume with 50 μL of media with serially diluted compounds. After 24 h we added 5 μL of 1 mg/mL stock of beetle D-luciferin, let it equilibrate at 37 °C for 30 min, and top-read the chemiluminescence plate using a BioTek Synergy HTX plate reader. The reads were repeated at 48 and 72 h. Collected data were normalized, processed, and plotted with Scilab and GraphPad-Prism.

4.8. Cell culture

HeLa cells were grown in 4.5 g/L glucose DMEM supplemented with 10% fetal bovine serum (Gibco 16000-49) and 1% penicillin/streptomycin. Cells were seeded into opaque white 96-well assay plates at a density of $5x10^3$ cells per well. 24 h post-seeding, the medium was replaced with medium containing the drug at the specified concentrations in equal volumes of DMSO. Cells were assayed for viability 48 h later using CellTiter-Glo (Promega) and plates were read on a CLARIOstar plus (BMG Labtech) plate reader. Values were normalized to the DMSO control. A minimum of two experiments was performed with representative data shown.

4.9. Cereblon engagement assay

Cells stably expressing the $BRD4_{BD2}$-GFP with mCherry reporter [41] were seeded at 30—50% confluency in 384-well plates with 50 μL per well of FluoroBrite DMEM media (Thermo Fisher Scientific A18967) supplemented with 10% FBS a day before compound treatment. Compounds and 100 nM dBET6 were dispensed using a D300e Digital Dispenser (HP), normalized to 0.5% DMSO, and incubated with the cells for 5 h. The assay plate was imaged immediately using an Acumen High Content Imager (TTP Labtech) with 488 nm and 561 nm lasers in a 2 μm × 1 μm grid per well format. The resulting images were analyzed using CellProfiler [57]. A series of image analysis steps (an 'image analysis pipeline') was constructed. First, the red and green channels were aligned and cropped to target the middle of each well (to avoid analysis of the heavily clumped cells at the edges). A background illumination function was calculated for both red and green channels of each well individually and subtracted to correct for illumination variations across the 384-well plate from various sources of error. An additional step was then applied to the green channel to suppress the analysis of large auto fluorescent artifacts and enhance the analysis of cell specific fluorescence by way of selecting for objects under a given size (30 A U.) and with a given shape (speckles). mCherry-positive cells were then identified in the red channel by filtering for objects 8—60 pixels in diameter and by using intensity to distinguish between clumped objects. The green channel was then segmented into GFP positive and negative areas and objects were labelled as GFP positive if at least 40% of it overlapped with a GFP positive area. The fraction of GFP-positive cells/mCherry-positive cells in each well was then calculated, and the green and red images were rescaled for visualization. The values for the concentrations that led to a 50% increase in BRD4BD2-eGFP accumulation (EC50) were calculated using the nonlinear fit variable slope model (GraphPad Software).

4.10. Proteomics methods

4.10.1. Global quantitative proteomics sample preparation

Kelly cells were treated with DMSO (biological triplicate) or d4E-2 at 1 μM for 5 h and cells were harvested by centrifugation at 4 °C. Cell lysis was performed by resuspension of the cell pellet in denaturing Urea buffer (8 M Urea, 50 mM NaCl, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (EPPS) pH 8.5, Protease and Phosphatase inhibitors), followed by manual homogenization by 20 passes through a 21-gauge needle. Cell lysate was clarified by centrifugation and global protein quantified using a Bradford assay (Bio-Rad). 100 μg of protein from each treatment was reduced, alkylated, digested and TMT labelled for LC-MS analysis as previously described [58]. The TMT labelled sample was offline fractionated into 96 fractions by high pH reverse phase HPLC (Agilent LC1260) through an aeris peptide xb-c18 column (phenomenex) with mobile phase A containing 5% acetonitrile and 10 mM NH4HCO3 in LC-MS grade H2O, and mobile phase B containing 90% acetonitrile and 5 mM NH4HCO3 in LC-MS grade H2O (both pH 8.0). The resulting 96 fractions were recombined in a non-contiguous manner into 24 fractions and desalted using C18 solid phase extraction plates (SOLA, Thermo Fisher Scientific) followed by subsequent mass spectrometry analysis.

4.10.2. LC-MS data collection and analysis

Data were collected using an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) and coupled with an UltiMate 3000 RSLCnano System. Peptides were separated on an EasySpray ES803a.rev2 75 μm inner diameter microcapillary column (Thermo Fisher Scientific). Peptides were separated over a 190 min gradient of 9—32% acetonitrile in 1.0% formic acid with a flow rate of 300 nL/min. Quantification was performed using a MS3-based TMT method as described previously (McAlister et al., 2014), with the addition of Real-Time Search MS3 acquisition implemented between MS2 and MS3 scans. The data were acquired using a mass range of m/z 340—1350, 120,000 resolution, standard AGC and maximum injection time of 50 ms for the peptide measurements in the Orbitrap. Data dependent MS2 spectra were acquired in the ion trap with a normalized collision energy (NCE) set at 34%, custom AGC target and a maximum injection time of 35 ms. Real-Time Search was performed with a

Swissprot human database (December 2019), searching for tryptic peptides with maximum of 1 missed cleavage, static alkylation of cysteines (57.0215 Da), static TMT labelling of lysine residues and peptide N-termini (304.2071 Da) and variable oxidation of methionine (15.9949 Da).

MS3 scans were acquired in the Orbitrap with HCD collision energy set to 45%, custom AGC target, maximum injection time of 86 ms, resolution at 50,000 and with a maximum synchronous precursor selection (SPS) precursors set to 10.

Proteome Discoverer 2.4 (Thermo Fisher Scientific) was used for.RAW file processing and controlling peptide and protein level false discovery rates, assembling proteins from peptides, and protein quantification from peptides. The MS/MS spectra were searched against a Swissprot human database (December 2019) containing both the forward and reverse sequences. Searches were performed using a 20 ppm precursor mass tolerance, 0.6 Da fragment ion mass tolerance, tryptic peptides containing a maximum of two missed cleavages, static alkylation of cysteine (57.0215 Da), static TMT labelling of lysine residues and N-termini of peptides (304.2071 Da), and variable oxidation of methionine (15.9949 Da). TMT reporter ion intensities were measured using a 0.003 Da window around the theoretical *m/z* for each reporter ion in the MS3 scan. The peptide spectral matches with poor quality MS3 spectra were excluded from quantitation (summed signal-to-noise across channels < 100 and precursor isolation specificity < 0.5), and the resulting data was filtered to only include proteins with a minimum of 2 unique peptides quantified. Reporter ion intensities were normalized and scaled using in-house scripts in the R framework (R Development Core Team, 2014). Statistical analysis was carried out using the limma package within the R framework [59].

4.11. eIF4E pull down and Western Blot

HEK293 and HeLa cells were grown for 24 h, harvested by centrifugation and lysed by multiple freeze-thaw cycles. 300 μL (1 μg/μL) of cell lysates, prepared in freeze-thaw lysis buffer (25 mM Tris-HCl, pH 7.5,150 mM KCl, 0.1% Triton X), were treated with the indicated concentrations of 4EGI-1 and i4EG-BiP at 37 °C for 1 h. This was followed by addition of 50 μL of a 50% mixture of adipic-agarose-m^{7}GDP beads [40] followed by incubation for 1 h at 4 °C. After washing the resin three times with lysis buffer, the bound proteins were resolved by SDS-PAGE, and analyzed by immunoblotting. Briefly this consisted of transfer to an activated PVDF membrane, blocking, and staining with a polyclonal antibody against 4E-BP1 (Cell Signaling Technology) and monoclonal antibodies against eIF4E and eIF4G (Transduction Laboratories) in PBS-T buffer. Secondary antibodies labelled with IRDye® 800CW and IRDye 680RD were used to amplify the signal according to the manufacturer's protocol. The immunoblot was imaged using two-color Western blot detection with an Odyssey® Imager.

4.12. Dual luciferase assay

HEK 293T cells were cultured to 70% confluence in six-well plates. Cells were transfected with 500 ng of a bicistronic reporter construct pFL-EMCV-IRES-RL, containing firefly luciferase followed by the EMCV IRES and Renilla luciferase, using polyethyleneimine (PEI) (high molecular weight, Sigma) in a 1:6 ratio using a PEI stock prepared at 1 mg/mL in DI water for transfection. The cells were treated 16 h after transfection with the indicated concentrations of 4EGI-1 and i4EG-BiP. 3 h after treatment, the cells were lysed in 1 × passive lysis buffer (Promega), and luciferase activity was measured with a dual luciferase reporter assay system (Promega) using an EnVision™ plate reader (PerkinElmer).

Data availability

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession number 7MEU.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: E. S. F. is a founder, science advisory board member and equity holder in Civetta, Jengu (board member) and Neomorph, an equity holder in C4, and a consultant to Astellas, Novartis, Deerfield, RA Capital and EcoR1. The Fischer lab receives or has received research funding from Novartis, Astellas, Ajax and Deerfield. G.W. is co-founder of PIC Therapeutics, Cellmig biolabs, Skinap therapeutics and Virtual Discovery. H.A. is an equity holder in PIC Therapeutics. The research described here is scientifically and financially independent of the efforts in any of the above-mentioned companies.

Acknowledgement

PDF acknowledges the Chleck Foundation. HA acknowledges funding from the Claudia Adams Barr Program for Innovative Cancer Research and from NIH (GM136859). G.W. acknowledges support from NIH grant GM132079 and CA200913. We thank Dr. Kendra Leigh for her critical comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmech.2021.113435.

References

[1] Y. Jia, V. Polunovsky, P.B. Bitterman, C.R. Wagner, Cap-dependent translation initiation factor eIF4E: an emerging anticancer drug target: cap-dependent translation initiation EIF4E inhibition, Med. Res. Rev. 32 (2012) 786–814, https://doi.org/10.1002/med.21260.

[2] Y. Mamane, E. Petroulakis, L. Rong, K. Yoshida, L.W. Ler, N. Sonenberg, eIF4E — from translation to transformation, Oncogene 23 (2004) 3172–3179, https://doi.org/10.1038/sj.onc.1207549.

[3] N. Sonenberg, A.G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets, Cell 136 (2009) 731–745, https://doi.org/10.1016/j.cell.2009.01.042.

[4] B. Culjkovic, I. Topisirovic, L. Skrabanek, M. Ruiz-Gutierrez, K.L.B. Borden, eIF4E is a central node of an RNA regulon that governs cellular proliferation, JCB (J. Cell Biol.) 175 (2006) 415–426, https://doi.org/10.1083/jcb.200607020.

[5] M. Arcand, P. Roby, R. Bossé, F. Lipari, J. Padrós, L. Beaudet, A. Marcil, S. Dahan, Single-well monitoring of Protein–Protein interaction and Phosphorylation–Dephosphorylation events, Biochemistry 49 (2010) 3213–3215, https://doi.org/10.1021/bi100253p.

[6] A. Soni, A. Akcakanat, G. Singh, D. Luyimbazi, Y. Zheng, D. Kim, A. Gonzalez-Angulo, F. Meric-Bernstam, eIF4E knockdown decreases breast cancer cell growth without activating Akt signaling, Mol. Canc. Therapeut. 7 (2008) 1782–1788, https://doi.org/10.1158/1535-7163.MCT-07-2357.

[7] F. Meric, K.K. Hunt, Translation initiation in cancer: a novel target for therapy, Mol. Canc. Therapeut. 1 (2002) 971–979.

[8] M.-E. Bordeleau, F. Robert, B. Gerard, L. Lindqvist, S.M.H. Chen, H.-G. Wendel, B. Brem, H. Greger, S.W. Lowe, J.A. Porco, et al., Therapeutic suppression of translation initiation modulates chemosensitivity in a mouse lymphoma model, J. Clin. Invest. (2008), https://doi.org/10.1172/JCI34753.

[9] Lazaris-Karatzas,A., Montine,K.S. and Sonenberg,N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. Nature, 345, 544–547. https://doi.org/10.1038/345544a0.

[10] V. Kerekatte, K. Smiley, B. Hu, A. Smith, F. Gelder, A. De Benedetti, The proto-oncogene/translation factor eIF4E: a survey of its expression in breast carcinomas, Int. J. Canc. 64 (1995) 27–31, https://doi.org/10.1002/ijc.2910640107.

[11] S. Wang, I.B. Rosenwald, M.J. Hutzler, G.A. Pihan, L. Savas, J.-J. Chen, B.A. Woda, Expression of the eukaryotic translation initiation factors 4E and 2α in non-Hodgkin's lymphomas, Am. J. Pathol. 155 (1999) 247–255, https://doi.org/10.1016/S0002-9440(10)65118-8.

[12] C.-A.O. Nathan, L. Liu, B.D. Li, F.W. Abreo, I. Nandy, A. De Benedetti, Detection

of the proto-oncogene eIF4E in surgical margins may predict recurrence in head and neck cancer, Oncogene 15 (1997) 579–584, https://doi.org/10.1038/sj.onc.1201216.
[13] Matsuo,H., Li,H., McGuire,A.M., Fletcher,C.M., Gingras,A.-C., Sonenberg,N. and Wagner,G. Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. Nat. Struct. Biol., 4, 717–724. https://doi.org/10.1038/nsb0997-717.
[14] Marcotrigiano,J., Gingras,A.-C., Sonenberg,N. and Burley,S.K. Cocrystal Structure of the Messenger RNA 5 Cap-Binding Protein (eIF4E) Bound to 7-methyl-GDP. Cell, 89, 951–961. https://doi.org/10.1016/s0092-8674(00)80280-9.
[15] N. Sekiyama, H. Arthanari, E. Papadopoulos, R.A. Rodriguez-Mias, G. Wagner, M. Léger-Abraham, Molecular mechanism of the dual activity of 4EGI-1: dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) E4036–E4045, https://doi.org/10.1073/pnas.1512118112.
[16] S. Grüner, D. Peter, R. Weber, L. Wohlbold, M.-Y. Chung, O. Weichenrieder, E. Valkov, C. Igreja, E. Izaurralde, The structures of eIF4E-eIF4G complexes reveal an extended interface to regulate translation initiation, Mol. Cell 64 (2016) 467–479, https://doi.org/10.1016/j.molcel.2016.09.020.
[17] S. Grüner, R. Weber, D. Peter, M.-Y. Chung, C. Igreja, E. Valkov, E. Izaurralde, Structural motifs in eIF4G and 4E-BPs modulate their binding to eIF4E to regulate translation initiation in yeast, Nucleic Acids Res. 46 (2018) 6893–6908, https://doi.org/10.1093/nar/gky542.
[18] S. Lukhele, A. Bah, H. Lin, N. Sonenberg, J.D. Forman-Kay, Interaction of the eukaryotic initiation factor 4E with 4E-BP2 at a dynamic bipartite interface, Structure 21 (2013) 2186–2196, https://doi.org/10.1016/j.str.2013.08.030.
[19] C. Igreja, D. Peter, C. Weiler, E. Izaurralde, 4E-BPs require non-canonical 4E-binding motifs and a lateral surface of eIF4E to repress translation, Nat. Commun. 5 (2014) 4790, https://doi.org/10.1038/ncomms5790.
[20] D. Peter, C. Igreja, R. Weber, L. Wohlbold, C. Weiler, L. Ebertsch, O. Weichenrieder, E. Izaurralde, Molecular architecture of 4E-BP translational inhibitors bound to eIF4E, Mol. Cell 57 (2015) 1074–1087, https://doi.org/10.1016/j.molcel.2015.01.017.
[21] A. Bah, R.M. Vernon, Z. Siddiqui, M. Krzeminski, R. Muhandiram, C. Zhao, N. Sonenberg, L.E. Kay, J.D. Forman-Kay, Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch, Nature 519 (2015) 106–109, https://doi.org/10.1038/nature13999.
[22] J.E. Dawson, A. Bah, Z. Zhang, R.M. Vernon, H. Lin, P.A. Chong, M. Vanama, N. Sonenberg, C.C. Gradinaru, J.D. Forman-Kay, Non-cooperative 4E-BP2 folding with exchange between eIF4E-binding and binding-incompatible states tunes cap-dependent translation inhibition, Nat. Commun. 11 (2020) 3146, https://doi.org/10.1038/s41467-020-16783-8.
[23] N.J. Moerke, H. Aktas, H. Chen, S. Cantel, M.Y. Reibarkh, A. Fahmy, J.D. Gross, A. Degterev, J. Yuan, M. Chorev, et al., Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G, Cell 128 (2007) 257–267, https://doi.org/10.1016/j.cell.2006.11.046.
[24] E. Papadopoulos, S. Jenni, E. Kabha, K.J. Takrouri, T. Yi, N. Salvi, R.E. Luna, E. Gavathiotis, P. Mahalingam, H. Arthanari, et al., Structure of the eukaryotic translation initiation factor eIF4E in complex with 4EGI-1 reveals an allosteric mechanism for dissociating eIF4G, Proc. Natl. Acad. Sci. Unit. States Am. 111 (2014) E3187–E3195, https://doi.org/10.1073/pnas.1410250111.
[25] L. Chen, B.H. Aktas, Y. wang, X. He, R. Sahoo, N. Zhang, S. Denoyelle, E. Kabha, H. Yang, R.Y. Freedman, et al., Tumor suppression by small molecule inhibitors of translation initiation, Oncotarget 3 (2012), https://doi.org/10.18632/oncotarget.598.
[26] E. Santini, T.N. Huynh, A.F. MacAskill, A.G. Carter, P. Pierre, D. Ruggero, H. Kaphzan, E. Klann, Exaggerated translation causes synaptic and behavioural aberrations associated with autism, Nature 493 (2013) 411–415, https://doi.org/10.1038/nature11782.
[27] R. McMahon, I. Zaborowska, D. Walsh, Noncytotoxic inhibition of viral infection through eIF4F-independent suppression of translation by 4EGi-1, J. Virol. 85 (2011) 853–864, https://doi.org/10.1128/JVI.01873-10.
[28] C.A. Hoeffer, K.K. Cowansage, E.C. Arnold, J.L. Banko, N.J. Moerke, R. Rodriguez, E.K. Schmidt, E. Klosi, M. Chorev, R.E. Lloyd, et al., Inhibition of the interactions between eukaryotic initiation factors 4E and 4G impairs long-term associative memory consolidation but not reconsolidation, Proc. Natl. Acad. Sci. Unit. States Am. 108 (2011) 3383–3388, https://doi.org/10.1073/pnas.1013063108.
[29] R. Cencic, D.R. Hall, F. Robert, Y. Du, J. Min, L. Li, M. Qui, I. Lewis, S. Kurtkaya, R. Dingledine, et al., Reversing chemoresistance by small molecule inhibition of the translation initiation complex eIF4F, Proc. Natl. Acad. Sci. Unit. States Am. 108 (2011) 1046–1051, https://doi.org/10.1073/pnas.1011477108.
[30] D.P. Bondeson, B.E. Smith, G.M. Burslem, A.D. Buhimschi, J. Hines, S. Jaime-Figueroa, J. Wang, B.D. Hamman, A. Ishchenko, C.M. Crews, Lessons in PROTAC design from selective degradation with a promiscuous warhead, Cell Chem. Biol. 25 (2018) 78–87, https://doi.org/10.1016/j.chembiol.2017.09.010, e5.
[31] S. An, L. Fu, Small-molecule PROTACs: an emerging and promising approach for the development of targeted therapy drugs, EBioMedicine 36 (2018) 553–562, https://doi.org/10.1016/j.ebiom.2018.09.005.
[32] A.C. Lai, C.M. Crews, Induced protein degradation: an emerging drug discovery paradigm, Nat. Rev. Drug Discov. 16 (2017) 101–114, https://doi.org/10.1038/nrd.2016.211.
[33] Winter,G.E., Buckley,D.L., Paulk,J., Roberts,J.M., Souza,A., Dhe-Paganon,S. and Bradner,J.E. Phthalimide conjugation as a strategy for in vivo target protein degradation. Science, 348, 1376–1381. https://doi.org/10.1126/science.aab1433.
[34] J. Lu, Y. Qian, M. Altieri, H. Dong, J. Wang, K. Raina, J. Hines, J.D. Winkler, A.P. Crew, K. Coleman, et al., Hijacking the E3 ubiquitin ligase cereblon to efficiently target BRD4, Chem. Biol. 22 (2015) 755–763, https://doi.org/10.1016/j.chembiol.2015.05.009.
[35] P. Ottis, C.M. Crews, Proteolysis-targeting chimeras: induced protein degradation as a therapeutic strategy, ACS Chem. Biol. 12 (2017) 892–898, https://doi.org/10.1021/acschembio.6b01068.
[36] H.-T. Huang, D. Dobrovolsky, J. Paulk, G. Yang, E.L. Weisberg, Z.M. Doctor, D.L. Buckley, J.-H. Cho, E. Ko, J. Jang, et al., A chemoproteomic approach to query the degradable kinome using a multi-kinase degrader, Cell Chem. Biol. 25 (2018) 88–99, https://doi.org/10.1016/j.chembiol.2017.10.005, e6.
[37] T. Kaur, A. Menon, A.L. Garner, Synthesis of 7-benzylguanosine cap-analogue conjugates for eIF4E targeted degradation, Eur. J. Med. Chem. 166 (2019) 339–350, https://doi.org/10.1016/j.ejmech.2019.01.080.
[38] D.W. McMillin, J. Delmore, E. Weisberg, J.M. Negri, D.C. Geer, S. Klippel, N. Mitsiades, R.L. Schlossman, N.C. Munshi, A.L. Kung, et al., Tumor cell-specific bioluminescence platform to identify stroma-induced changes to anticancer drug activity, Nat. Med. 16 (2010) 483–489, https://doi.org/10.1038/nm.2112.
[39] M. Ghandi, F.W. Huang, J. Jané-Valbuena, G.V. Kryukov, C.C. Lo, E.R. McDonald, J. Barretina, E.T. Gelfand, C.M. Bielski, H. Li, et al., Next-generation characterization of the cancer cell line encyclopedia, Nature 569 (2019) 503–508, https://doi.org/10.1038/s41586-019-1186-3.
[40] I. Edery, M. Altmann, N. Sonenberg, High-level synthesis in Escherichia cofi of functional cap-binding eukaryotic initiation factor eIF-4E and affhity purification using a simplified cap-analog resin, Gene 74 (1988) 517–525.
[41] R.P. Nowak, S.L. DeAngelo, D. Buckley, Z. He, K.A. Donovan, J. An, N. Safaee, M.P. Jedrychowski, C.M. Ponthier, M. Ishoey, et al., Plasticity in binding confers selectivity in ligand-induced protein degradation, Nat. Chem. Biol. 14 (2018) 706–714, https://doi.org/10.1038/s41589-018-0055-y.
[42] M. de Wispelaere, G. Du, K.A. Donovan, T. Zhang, N.A. Eleuteri, J.C. Yuan, J. Kalabathula, R.P. Nowak, E.S. Fischer, N.S. Gray, et al., Small molecule degraders of the hepatitis C virus protease reduce susceptibility to resistance mutations, Nat. Commun. 10 (2019) 3468, https://doi.org/10.1038/s41467-019-11429-w.
[43] F. Jeppesen, P. Ilium, Concentration of ampicillin in antral mucosa following administration of ampicillin sodium and pivampicillin, Acta Otolaryngol. 73 (1972) 428–432, https://doi.org/10.3109/00016487209138962.
[44] A. Patnaik, E.K. Rowinsky, M.A. Villalona, L.A. Hammond, C.D. Britten, L.L. Siu, A. Goetz, S.A. Felton, S. Burton, F.H. Valone, et al., A phase I study of pivaloyloxymethyl butyrate, a prodrug of the differentiating agent butyric acid, in patients with advanced solid malignancies, Clin. Canc. Res. 8 (2002) 2142–2148.
[45] Nowak,R.P., Fischer,E.S., Gray,N.S., Zhang,T. and He,Z. Heterobifunctional compounds with improved specificity for the bromodomain of brd4. (U.S. Patent No. WO2019079701A1), WIPO (PCT), https://www.wipo.int/pct/en/.
[46] K.A. Donovan, F.M. Ferguson, J.W. Bushman, N.A. Eleuteri, D. Bhunia, S. Ryu, L. Tan, K. Shi, H. Yue, X. Liu, et al., Mapping the degradable kinome provides a resource for expedited degrader development, Cell 183 (2020) 1714–1731, https://doi.org/10.1016/j.cell.2020.10.038, e10.
[47] T. Mori, T. Ito, S. Liu, H. Ando, S. Sakamoto, Y. Yamaguchi, E. Tokunaga, N. Shibata, H. Handa, T. Hakoshima, Structural basis of thalidomide enantiomer binding to cereblon, Sci. Rep. 8 (2018) 1294, https://doi.org/10.1038/s41598-018-19202-7.
[48] L. Borissenko, M. Groll, 20S proteasome and its inhibitors: crystallographic knowledge for drug development, Chem. Rev. 107 (2007) 687–717, https://doi.org/10.1021/cr0502504.
[49] A. Alhossary, S.D. Handoko, Y. Mu, C.-K. Kwoh, Fast, accurate, and reliable molecular docking with QuickVina 2, Bioinformatics 31 (2015) 2214–2216, https://doi.org/10.1093/bioinformatics/btv082.
[50] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. (2009), https://doi.org/10.1002/jcc.21334.
[51] H. Wang, F. Huang, J. Wang, P. Wang, W. Lv, L. Hong, S. Li, J. Zhou, The synergistic inhibition of breast cancer proliferation by combined treatment with 4EGI-1 and MK2206, Cell Cycle 14 (2015) 232–242, https://doi.org/10.4161/15384101.2014.977096.
[52] M. Wu, C. Zhang, X.-J. Li, Q. Liu, S. Wanggou, Anti-cancer effect of cap-translation inhibitor 4EGI-1 in human glioma U87 cells: involvement of mitochondrial dysfunction and ER stress, Cell. Physiol. Biochem. 40 (2016) 1013–1028, https://doi.org/10.1159/000453158.
[53] T. Murata, K. Shimotohno, Ubiquitination and proteasome-dependent degradation of human eukaryotic translation initiation factor 4E, J. Biol. Chem. 281 (2006) 20788–20800, https://doi.org/10.1074/jbc.M600563200.
[54] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 59 (2005) 687–696, https://doi.org/10.1002/prot.20449.
[55] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, https://doi.org/10.1002/jcc.21256.
[56] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminf. 3 (2011) 33, https://doi.org/10.1186/1758-2946-3-33.

[57] A.E. Carpenter, T.R. Jones, M.R. Lamprecht, C. Clarke, I.H. Kang, O. Friman, D.A. Guertin, J.H. Chang, R.A. Lindquist, J. Moffat, et al., CellProfiler: image analysis software for identifying and quantifying cell phenotypes, Genome Biol. 7 (2006) R100, https://doi.org/10.1186/gb-2006-7-10-r100.

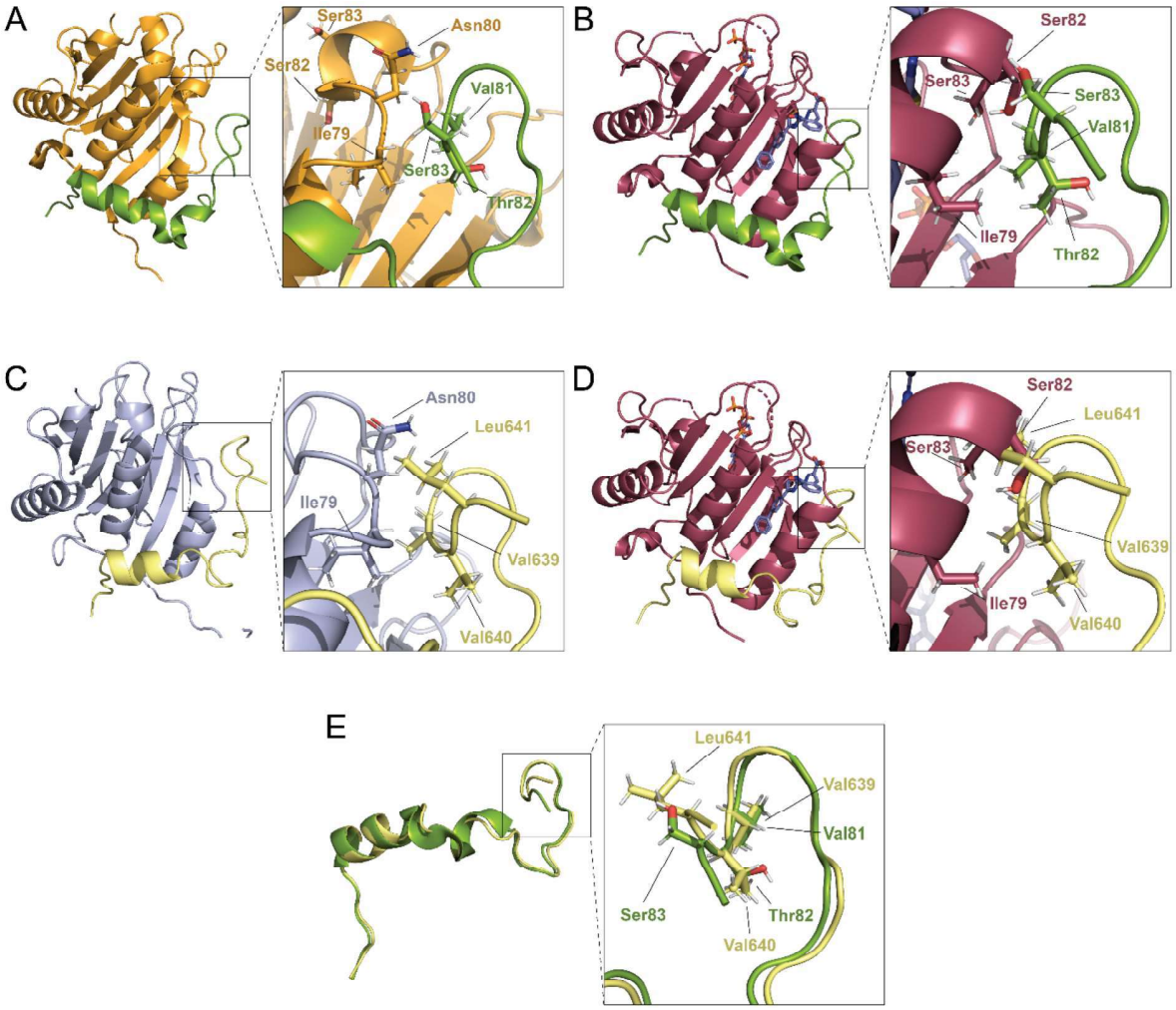
[58] K.A. Donovan, J. An, R.P. Nowak, J.C. Yuan, E.C. Fink, B.C. Berry, B.L. Ebert, E.S. Fischer, Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome, eLife 7 (2018), e38430, https://doi.org/10.7554/eLife.38430.

[59] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015), https://doi.org/10.1093/nar/gkv007 e47–e47.

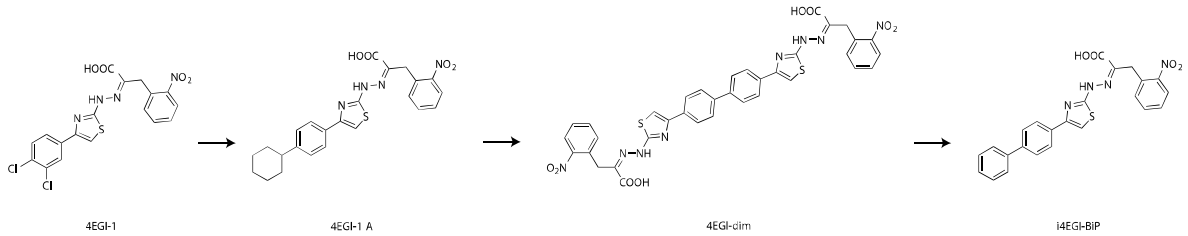
Supplementary Table 1 Data collection and refinement statistics

Source	APS Argonne National Laboratory
Wavelength	0.979180 Å
Resolution range	46.75 - 1.914 (1.983 - 1.914) Å
Space group	P 1 21 1
Unit cell	38.812 73.364 61.866 90 101.36 90
Total reflections	86728 (2621)
Unique reflections	25424 (1853)
Multiplicity	3.4 (1.4)
Completeness (%)	96.64 (70.35)
Mean I/sigma(I)	18.94 (2.91)
Wilson B-factor	17.85
R-merge	0.05771 (0.3349)
R-meas	0.06825 (0.4711)
R-pim	0.03603 (0.3309)
CC1/2	0.997 (0.788)
CC*	0.999 (0.939)
Reflections used in refinement	25416 (1853)
Reflections used for R-free	2000 (146)
R-work	0.1611 (0.2389)
R-free	0.2188 (0.3217)
CC(work)	0.962 (0.891)
CC(free)	0.927 (0.764)
Number of non-hydrogen atoms	3492
macromolecules	3117
ligands	99
solvent	276
Protein residues	370
RMS(bonds)	0.021
RMS(angles)	1.99
Ramachandran favored (%)	95.33
Ramachandran allowed (%)	3.85
Ramachandran outliers (%)	0.82
Rotamer outliers (%)	0.88
Clashscore	12.37
Average B-factor	25.39
macromolecules	24.77
ligands	29.47
solvent	31.00

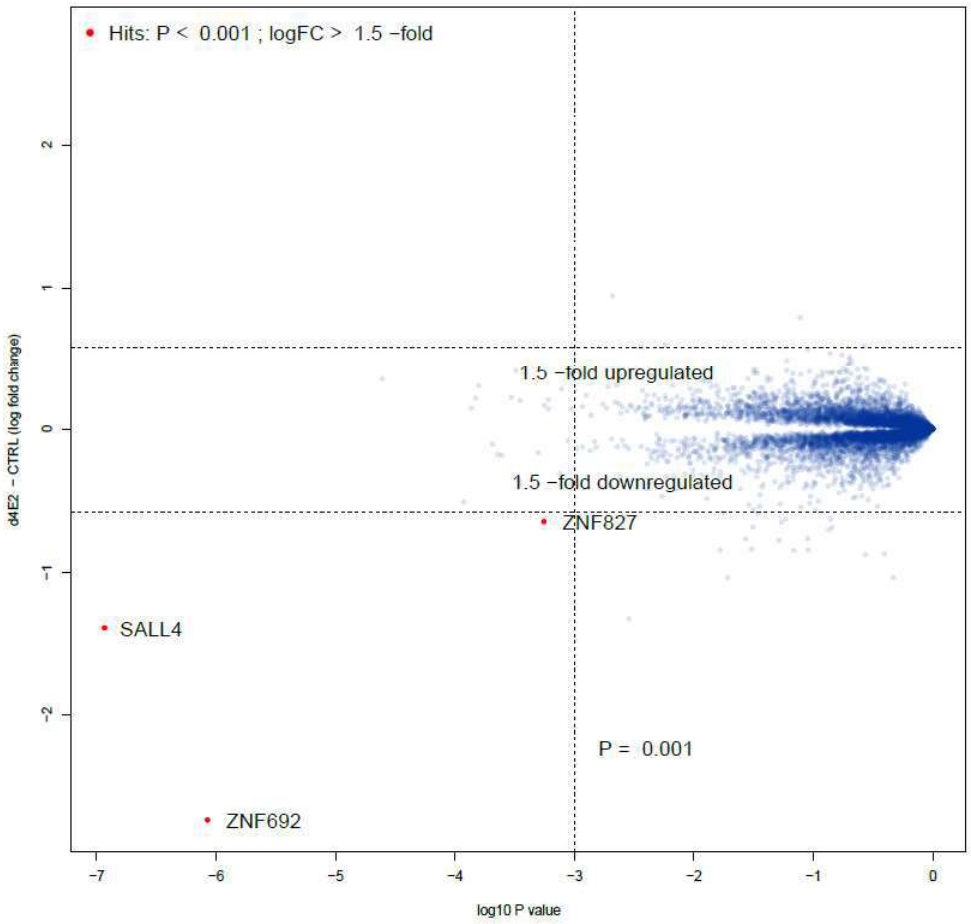
Supplementary Figure 1



Supplementary Figure 1 A. Structure of eIF4E bound to 4E-BP (pdb ID: 5BXV). eIF4E in orange and 4E-BP in green. **B.** Overlay of 4E-BP from 5BXV with the cocrystal structure of eIF4E bound to i4EG-BP (red). The extended helix with Ser82 and Ser83 of eIF4E is in proximity of the C-terminus of 4E-BP. **C.** Structure of eIF4E bound to eIF4G (pdb ID: 5T46). eIF4E in blue and eIF4G in yellow. **D.** Overlay of eIF4G from 5T46 with the cocrystal structure of eIF4E bound to i4EG-BP (red). Leu641 of eIF4G sterically clashes with the extended helix of eIF4E. **E.** Overlay of the binding modes of 4E-BP from 5BXV (green) and eIF4G from 5T46 (yellow). The structures superimpose almost perfectly identical, with the difference being the longer side chain of Leu641 from eIF4G compared to its counterpart Ser83 of 4E-BP.



Supplementary Figure 2: Evolution of the i4EG-BiP scaffold from 4EGI-1



Supplementary Figure 3: Scatterplot depicting the relative fold change in abundance of proteins following a 5-hour treatment of Kelly cells with 1 µM of d4E2 compared to DMSO vehicle control treatment. Significant changes were assessed using limma's moderated t-test (Ritchie et al., 2015), with log2 fold change displayed on the y-axis and negative log10 p-value on the x-axis (n=1 independent replicate for d4E2 and n=3 independent replicate for DMSO control).

Protein-related abbreviations and acronyms	
4E- BP	4E-binding protein (Inhibits translation by binding to eIF4E and preventing eIF4G from binding)
BET	Bromo- and Extra-Terminal domain
CRBN	Cereblon (Substrate recognition component of a DCX E3 protein ligase)
eIF4A	eukaryotic translation initiation factor 4A (helicase that unwinds secondary structure in the 5'UTR)
eIF4B	eukaryotic translation initiation factor 4B
eIF4E	eukaryotic translation initiation factor 4E (cap binding protein)
eIF4F	eukaryotic translation initiation factor 4F (complex of eIF4E, eIF4A and eIF4G)
eIF4G	eukaryotic translation initiation factor 4G (scaffold protein)
GFP	green fluorescent protein
mTORC	mammalian target of rapamycin complex
RFP	red fluorescent protein

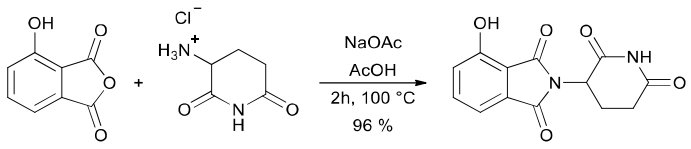
Chemistry related abbreviations	
Ac	Acetyl
Ac_2O	acetic anhydride
Bn	Benzyl
Boc	tert-butoxycarbonyl
CBz	benzyloxycarbonyl
EtOH	ethanol
HOAc	acetic acid
m^7GDP	7-methyl guanosine diphosphate
m^7GTP	7-methyl guanosine triphosphate
Me_2CO	acetone
NaOAc	sodium acetate
PEG	polyethylene glycol
PROTAC	proteolysis targeted chimera
*t*BuOAc	*tert*-butyl acetate

Other abbreviations	
5'-UTR	5'-untranslated region
CS-BLI	Compartment-Specific Bioluminescence Imaging
CSP	chemical shift perturbation
FP	fluorescence polarization
IRES	internal ribosomal entry site

Supplemental Table 2: List of acronyms and abbreviations used in the manuscript.

Chemistry

2-(2,6-dioxopiperidin-3-yl)-4-hydroxyisoindoline-1,3-dione



3-Hydroxyphthalic anhydride (1 g, 6.15 mmol, 1 Eq.) was dissolved in glacial acetic acid (3.5 mL). Sodium acetate (0.66 g, 7.99 mmol, 1.3 Eq.) was added, followed by 3-Aminopiperidine-2,6-dione (1.32 g, 7.99 mmol, 1.3 Eq.). The mixture was heated to 100 °C for 2 h. The color of the reaction mixture turned from yellow to purple. After cooling the reaction to room temperature, saturated sodium bicarbonate solution is added slowly until the mixture turns yellow. The resulting product is freeze dried and purified by chromatography (isco Combiflash, 12 g column, silica gel, ethylacetate).

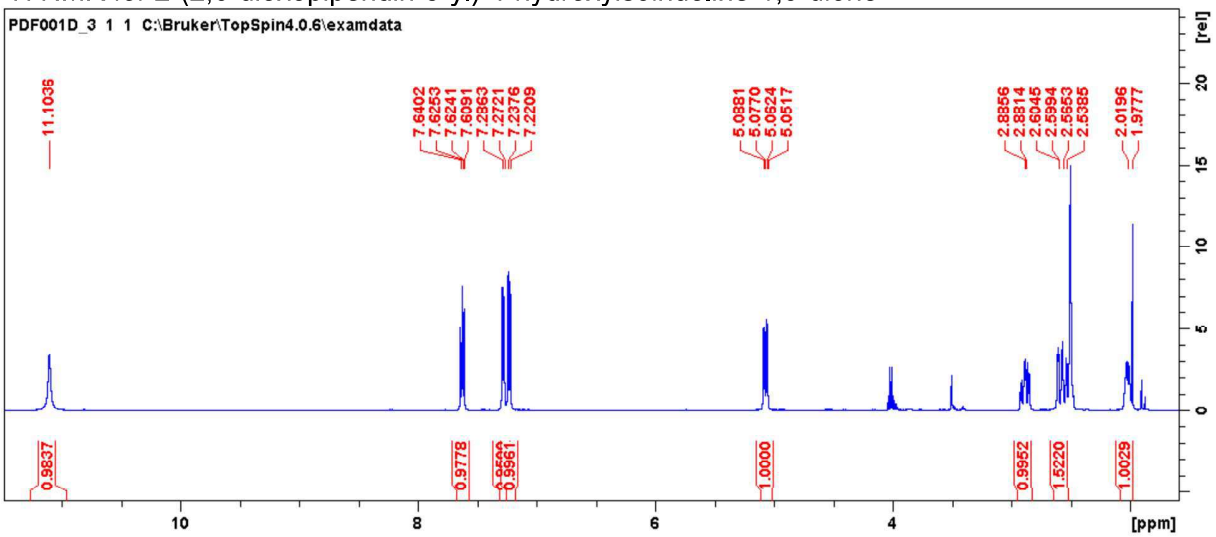
Yield: 1.62 g (5.90 mmol, 96 %) as a yellow solid.

^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 2.00 (m, 1H), 2.54 (m, 1H), 2.56 (m, 1H), 2.88 (m, 1H), 5.06 (dd, 1H, $^{3}J_{HH}$ = 12.09 Hz, 5.6 Hz), 7.21 (d, 1H, $^{3}J_{HH}$ = 8.4 Hz), 7.25 (d, 1H, $^{3}J_{HH}$ = 6.7 Hz), 7.61 (dd, 1H, $^{3}J_{HH}$ = 8.1 Hz, 7.3 Hz), 11.10 (s, 1H)

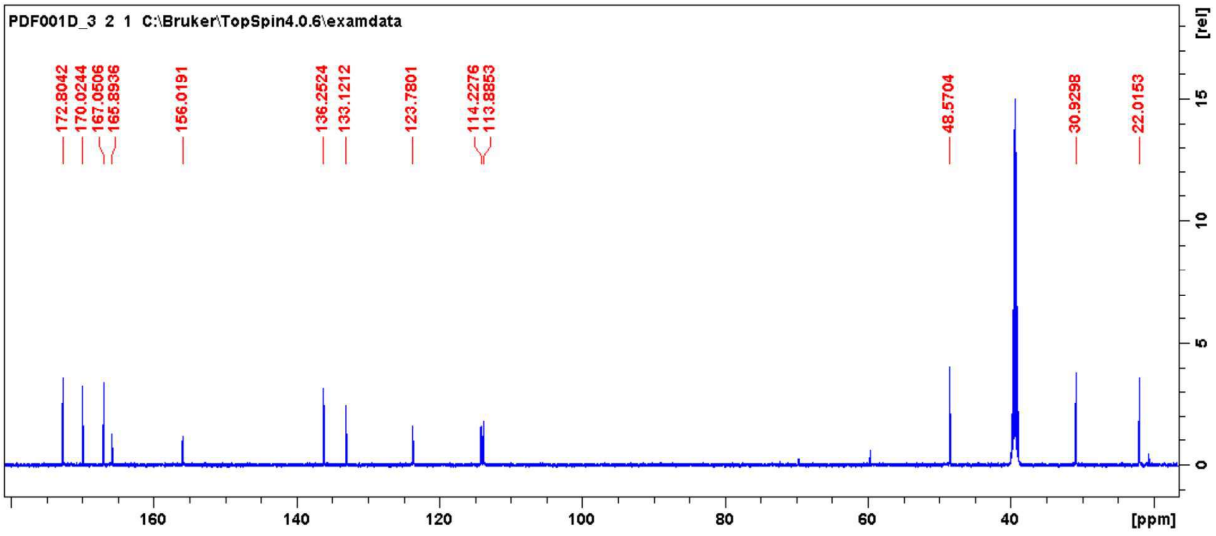
^{13}C NMR (126 MHz, DMSO-d6): δ [ppm] = 22.21, 31.00, 48.60, 113.53, 114.19, 124.15, 133.22, 136.21, 156.73, 166.06, 167.17, 170.10, 172.86

UPLC-MS: calc. [M + H]: 275.07, found [M + H]: 275.06

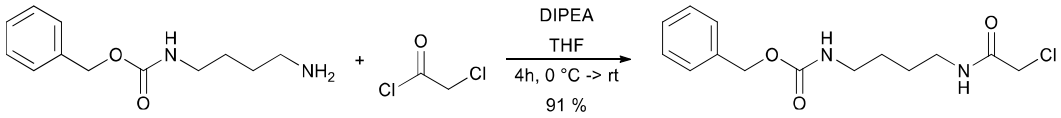
^{1}H NMR for 2-(2,6-dioxopiperidin-3-yl)-4-hydroxyisoindoline-1,3-dione



^{13}C NMR for 2-(2,6-dioxopiperidin-3-yl)-4-hydroxyisoindoline-1,3-dione



Benzyl-(4-(2-chloroacetamido)butyl)carbamate



Benzyl (4-aminobutyl)carbamate (58 mg, 0.26 mmol, 1 Eq.) was dissolved in THF (2 mL) and cooled to 0 °C. N,N-Diisopropylethylamine (DIPEA, 45.5 µL, 0.26 mmol, 1 Eq.) was added, followed by 2-chloroacetyl chloride (22.9 µL, 0.29 mmol, 1.1 Eq.) After 1.5 h, the ice bath is removed to allow heating to room temperature for another 1.5 h. The reaction mixture was extracted with ethylacetate to 10 mL and washed with water (3 x). The product was dried *in vacuo* and purified by chromatography (isco combiflash, 4 g column, silica gel, hexanes/ethylacetate 1-100% gradient).

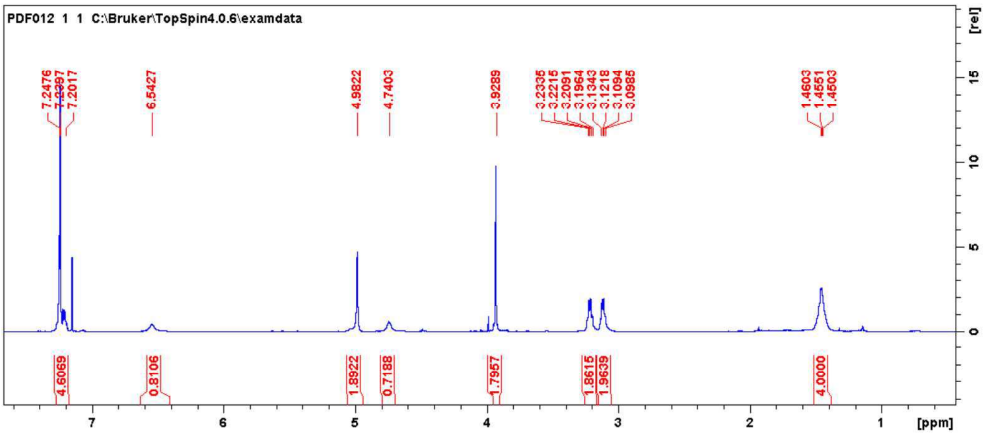
Yield: 71 mg (0.24 mmol, 91 %) of a white solid.

^{1}H-NMR (500 MHz, $CDCl_3$): δ [ppm] = 1.46 (m, 4H), 3.12 (m, 2H), 3.22 (m, 2H), 3.93 (s, 2H), 4.74 (s, 1H), 4.98 (s, 2H), 6.54 (s, 1H), 7.20-7.25 (m, 5H)

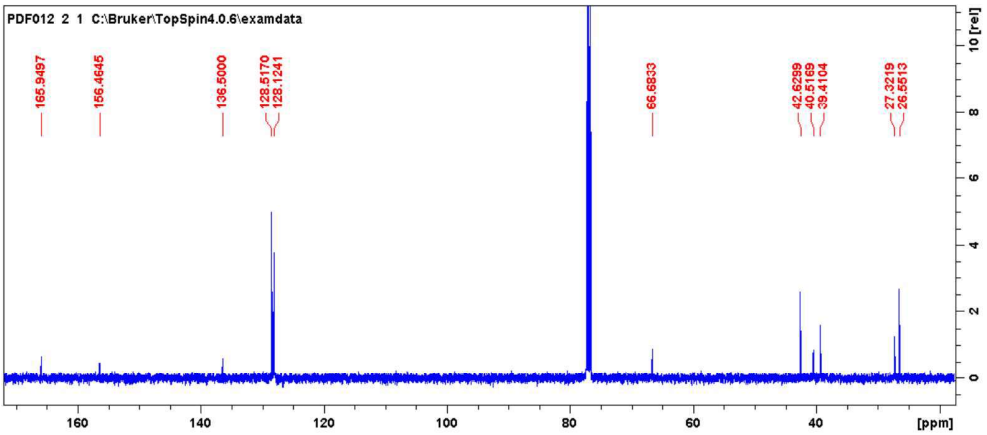
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 26.55, 27.32, 39.41, 40.52, 42.63, 66.68, 128.12, 128.52, 136.50, 156.46, 165.95

UPLC-MS: calc; [M +H]: 299.12, found [M + H]: 299.12

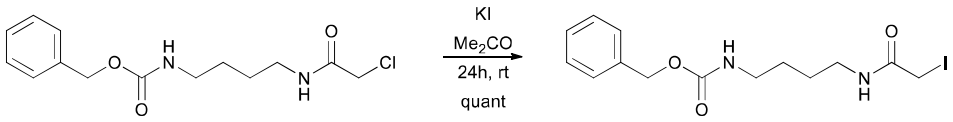
^{1}H NMR for Benzyl-(4-(2-chloroacetamido)butyl)carbamate



^{13}C NMR for Benzyl-(4-(2-chloroacetamido)butyl)carbamate

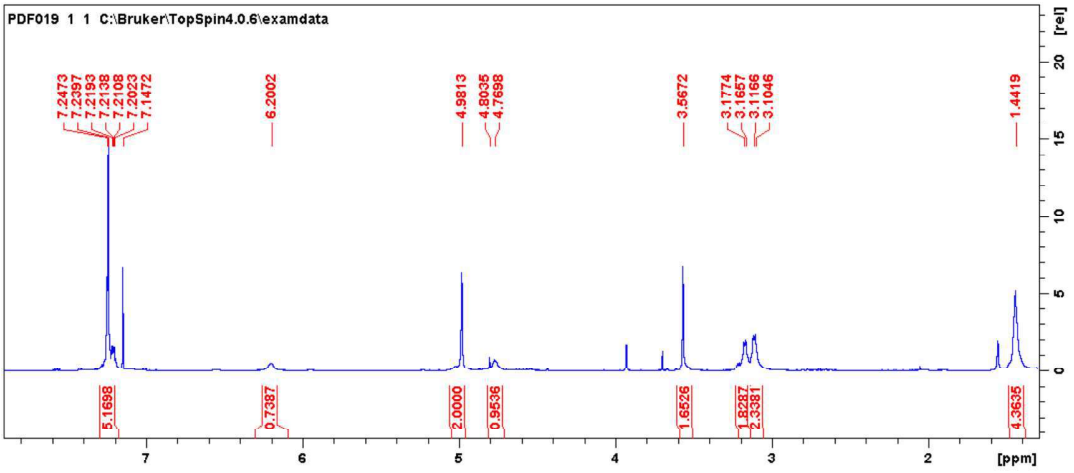


Benzyl (4-(2-iodoacetamido)butyl)carbamate

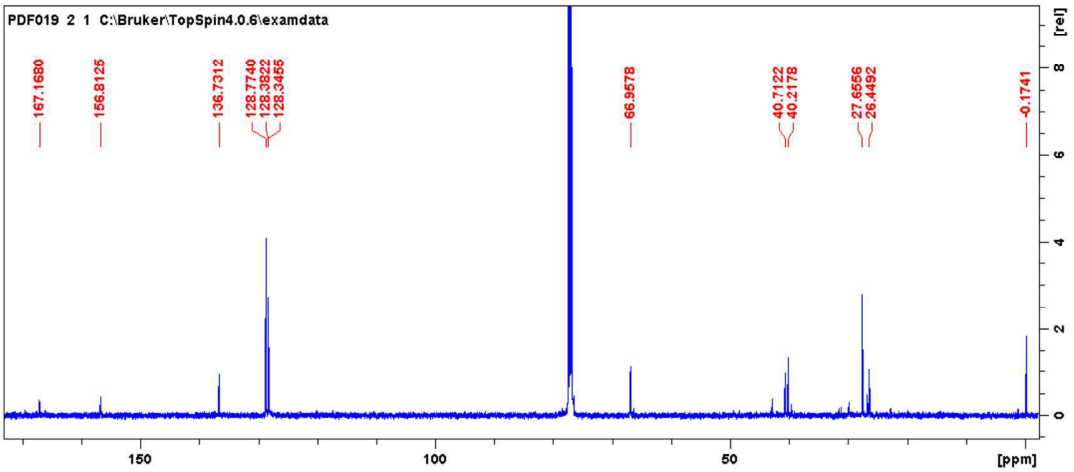


Benzyl-(4-(2-chloroacetamido)butyl)carbamate (50 mg, 0.17 mmol, 1 Eq.) was dissolved in acetone (2 mL). Potassium iodide (83.3 mg, 0.50 mmol, 3Eq.) was added and the reaction mixture was stirred at room temperature overnight (18 h). Then, the solvent was removed *in vacuo* and the remaining solid was resuspended in water (10 mL). The product was extracted with dichloromethane (3 x). No further purification was required.
Yield: 65.2 mg (0.17 mmol, 100 %) of a light yellow solid.
^{1}H-NMR (500 MHz, $CDCl_3$): δ [ppm] = 1.44 (m, 4H), 3.11 (m, 2H), 3.17 (m, 2H), 3.57 (s, 2H), 4.77 (s, 1H), 4.98 (s, 2H), 6.20 (s, 1H), 7.20-7.25 (m, 5H)
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = -0.17, 26.45, 27.66, 40.22, 40.71, 66.96, 128.35, 128.38, 128.77, 136.73, 156.81, 167.17
UPLC-MS: calc; [M +H]: 391.05, found [M + H]: 391.06

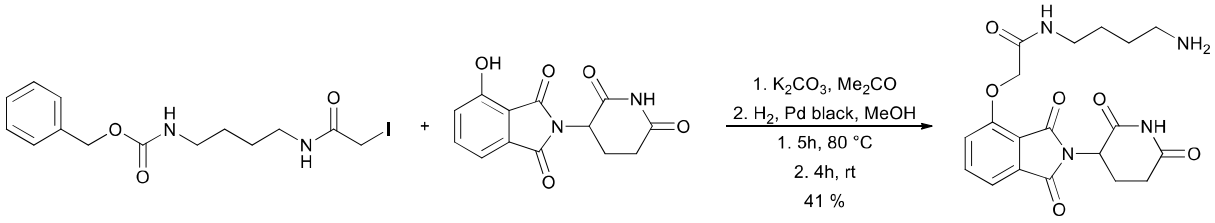
^{1}H NMR for Benzyl (4-(2-iodoacetamido)butyl)carbamate



^{13}C NMR for Benzyl (4-(2-iodoacetamido)butyl)carbamate

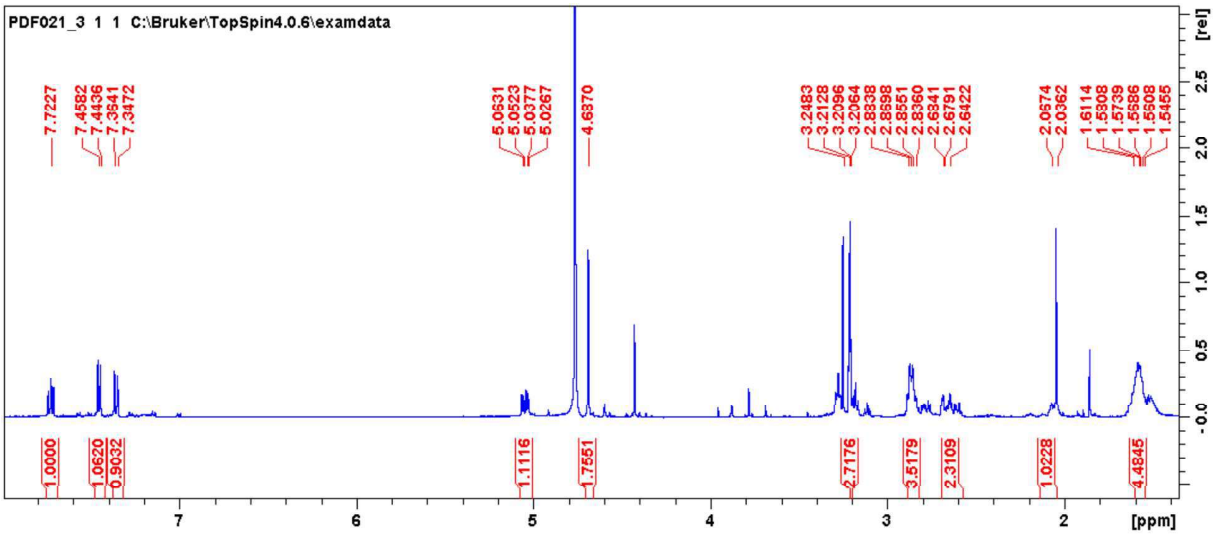


N-(4-aminobutyl)-2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)oxy)acetamide

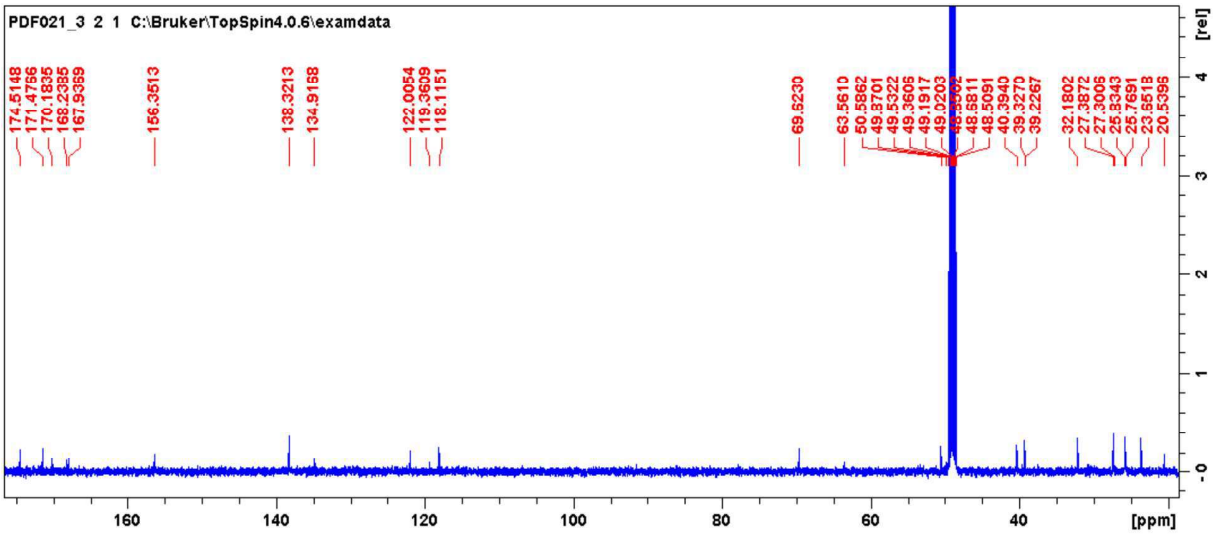


4-Hydroxythalidomide (45.8 mg, 0.17 mmol, 1 Eq.) and potassium carbonate (23.1 mg, 0.17 mmol, 2 Eq.) were dissolved in dimethylformamide (1 mL). Benzyl (4-(2-iodoacetamido)butyl)carbamate (65.2 mg, 0.17 mmol, 1 Eq.) was dissolved in dimethylformamide (2 mL) and added dropwise to the reaction mixture. The reaction was heated to 100 °C for 5 h. The product was extracted with dichloromethane and used without further purification after removal of the organic phase *in vacuo*. The crude product (30 mg, 55.9 µmol, 1 Eq.) was dissolved in dry methanol (1 mL). Palladium black (2 mg, 18.8 µmol, 0.3 Eq.) was added while stirring. A balloon filled with hydrogen gas was attached to the flask and the reaction mixture was stirred at room temperature for 1 day. The product was purified by removal of palladium black using a syringe filter and drying *in vacuo*. No further purification was required.
Yield: 22.5 mg (55.9 µmol, 41 %) of a white solid.
^{1}H-NMR (500 MHz, MeOD): δ [ppm] = 1.58 (m, 4H), 2.07 (m, 1H), 2.63 (m, 1H), 2.68 (m, 1H), 2.79 (m, 1H), 2.87 (m, 2H), 3.18-3.29 (m, 2H), 4.69 (s, 2H), 5.05 (dd, 1H, $^3J_{HH}$ = 12.7 Hz, 5.4 Hz), 7.36 (d, 1H, $^3J_{HH}$ = 8.2 Hz), 7.45 (d, 1H, $^3J_{HH}$ = 7.3 Hz), 7.72 (dd, 1H, $^3J_{HH}$ = 8.6 Hz, 7.6 Hz)
^{13}C-NMR (126 MHz, MeOD): δ [ppm] = 23.65, 25.83, 27.30, 32.18, 39.33, 40.39, 50.60, 69.62, 118.12, 119.36, 122.01, 134.92, 138.32, 156.35, 167.94, 170.18, 171.48, 174.51
UPLC-MS: calc; [M +H]: 403.16, found [M + H]: 403.15

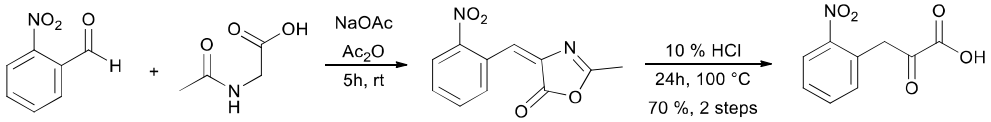
^{1}H NMR for N-(4-aminobutyl)-2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)oxy)acetamide



^{13}C NMR for N-(4-aminobutyl)-2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)oxy)acetamide



3-(2-nitrophenyl)-2-oxopropanoic acid



2-Nitrobenzaldehyde (500 mg, 3.3 mmol, 1 Eq.), N-acetylglycine (503 mg, 4.3 mmol, 1.3 Eq.) and sodium acetate (352 mg, 4.3 mmol, 1.3 Eq.) were dissolved in acetic anhydride (20 mL). The mixture was stirred for 3 h at 140 °C and extracted with dichloromethane and dried *in vacuo*. The crude product (2-methyl-5-(2-nitrobenzylidene)oxazol-4(5H)-one) was dissolved in 10 % hydrochloric acid in water (10 mL), heated to 100 °C and stirred overnight (24 h). The product was extracted with dichloromethane, dried *in vacuo* and purified by chromatography (isco Combiflash, 4 g column, silica gel, hexanes/ethylacetate 1-100% gradient).

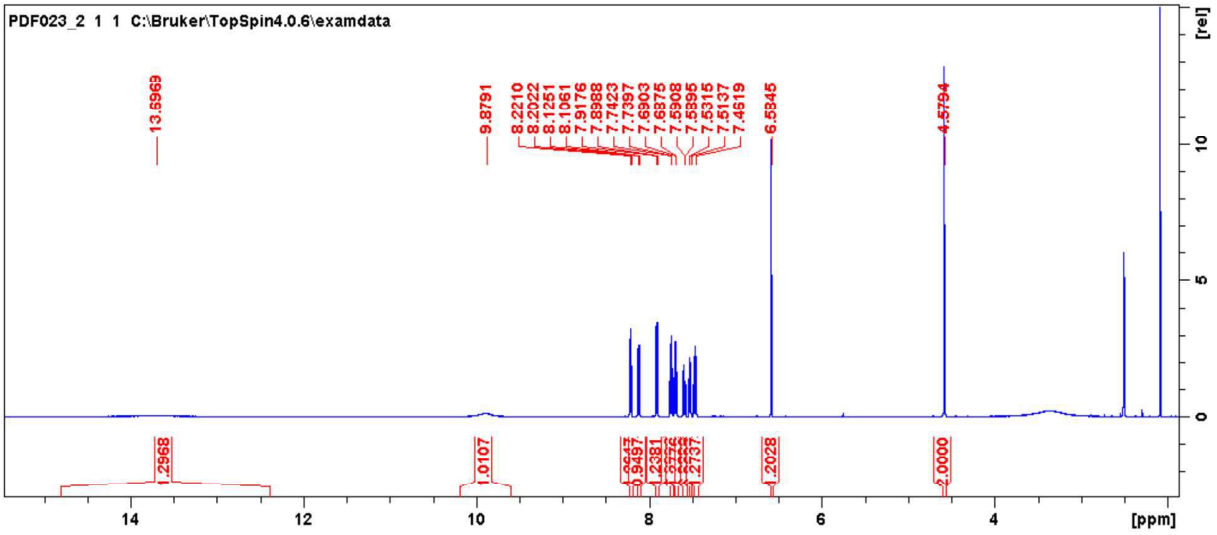
Yield: 483 mg (2.3 mmol, 70 %) of a yellow solid.

^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 4.58 (s, 2H), 6.58 (s, 1H), 7.46 (dd, 1H, $^{3}J_{HH}$ = 9.3 Hz, 7.3 Hz, $^{4}J_{HH}$ = 1.4 Hz), 7.52 (d, 1H, $^{3}J_{HH}$ = 7.6 Hz, $^{4}J_{HH}$ = 1.3 Hz), 7.59 (dd, 1H, $^{3}J_{HH}$ = 9.4 Hz, 8.1 Hz, $^{4}J_{HH}$ = 1.6 Hz), 7.69 (dd, 1H, $^{3}J_{HH}$ = 8.8 Hz, 7.65 Hz, $^{4}J_{HH}$ = 1.2 Hz), 7.74 (dd, 1H, $^{3}J_{HH}$ = 8.8 Hz, 7.6 Hz, $^{4}J_{HH}$ = 1.3 Hz), 7.91 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz, $^{4}J_{HH}$ = 1.3 Hz), 8.12 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz, $^{4}J_{HH}$ = 1.3 Hz), 8.21 (d, 1H, $^{3}J_{HH}$ = 8.0 Hz, $^{4}J_{HH}$ = 1.3 Hz), 9.88 (s, 1H), 13.70 (s, 2H)

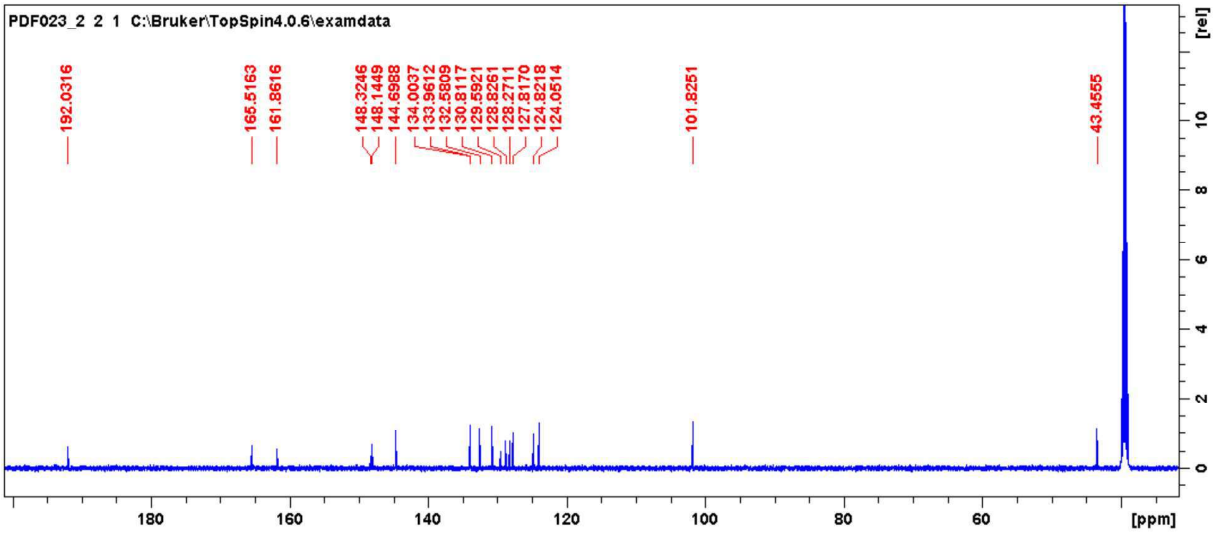
^{13}C-NMR (126 MHz, DMSO-d6): δ [ppm] = 43.46, 101.83, 124.05, 124.82, 127.82, 128.27, 128.83, 129.59, 130.81, 132.58, 133.96, 134.00, 144.70, 148.14, 148.32, 161.86, 165.52, 192.03

UPLC-MS: calc; [M +H]: 210.04, found [M + H]: 210.03

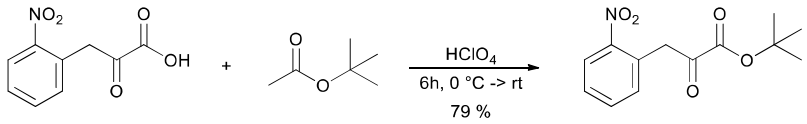
^{1}H NMR for 3-(2-nitrophenyl)-2-oxopropanoic acid



^{13}C NMR for 3-(2-nitrophenyl)-2-oxopropanoic acid

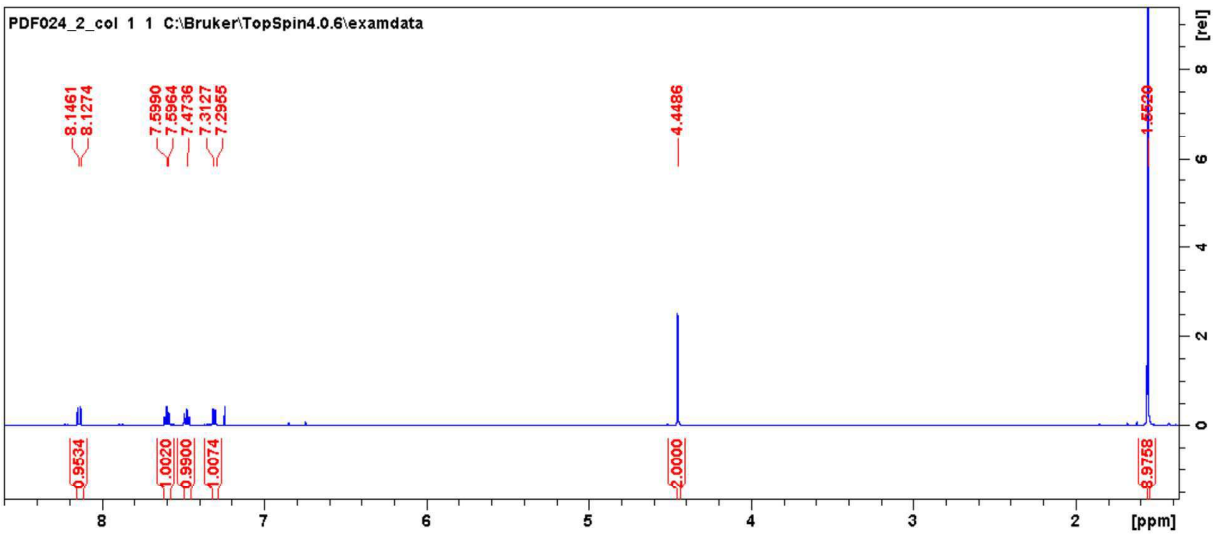


tert-butyl 3-(2-nitrophenyl)-2-oxopropanoate

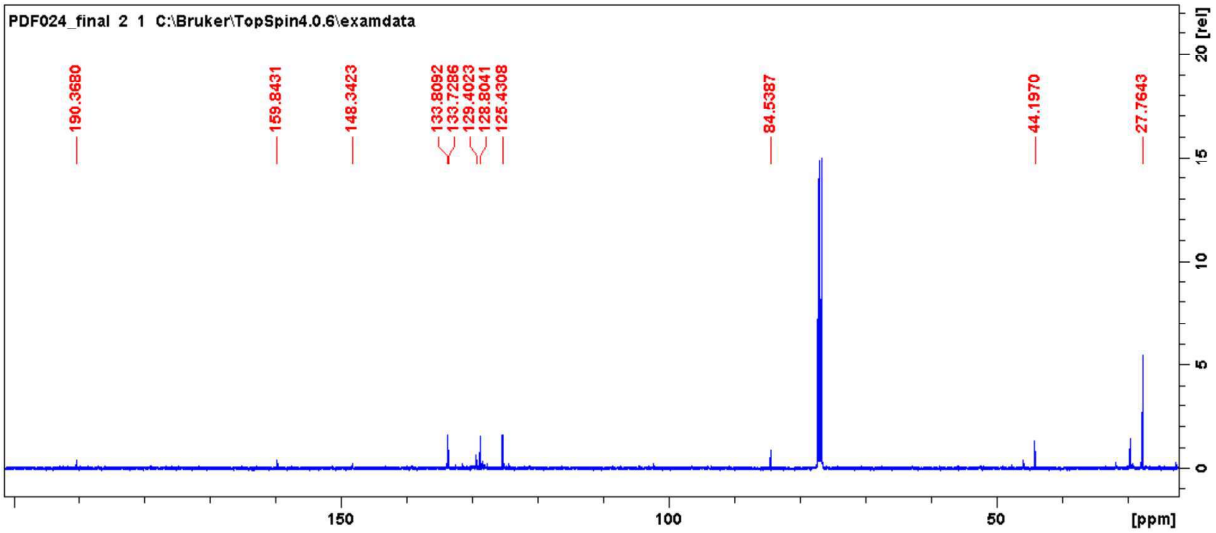


3-(2-nitrophenyl)-2-oxopropanoic acid (50 mg, 0.24 mmol, 1 Eq.) was dissolved in *tert*-butyl acetate (2 mL) and the mixture was cooled to 0 °C. Perchloric acid, 70 % in water, (29 µL, 0.39 mmol, 1.5 Eq.) were added and the ice bath was removed. The mixture was stirred at room temperature overnight (18 h). The solvent was removed *in vacuo* and the remaining solid was resuspended in water. The product was extracted with dichloromethane (3 x), dried *in vacuo* and purified by chromatography (isco Combiflash, 4 g column, silica gel, ethylacetate).
Yield: 50 mg (0.19 mmol, 79 %) of a yellow solid.
^{1}H-NMR (500 MHz, $CDCl_3$): δ [ppm] = 1.55 (s, 9H), 4.45 (s, 2H), 7.30 (d, 1H $^{3}J_{HH}$ = 7.5 Hz, $^{4}J_{HH}$ = 1.3 Hz), 7.47 (dd, 1H, $^{3}J_{HH}$ = 9.5 Hz, 8.2), 7.60 (dd, 1H, $^{3}J_{HH}$ = 8.8 Hz, 7.6 Hz, $^{4}J_{HH}$ = 1.5 Hz), 8.14 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz)
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 27.76, 44.20, 84.54, 125.43, 128.80, 129.40, 133.73, 133.81, 148.34, 159.84, 190.37
UPLC-MS: calc; [M +H]: 266.10, found [M + H]: 266.10

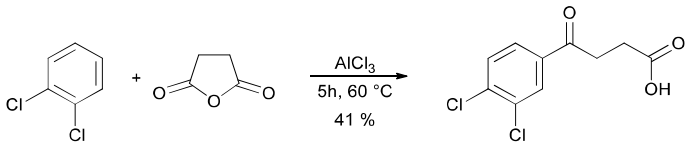
^{1}H NMR for *tert*-butyl 3-(2-nitrophenyl)-2-oxopropanoate



^{13}C NMR for *tert*-butyl 3-(2-nitrophenyl)-2-oxopropanoate

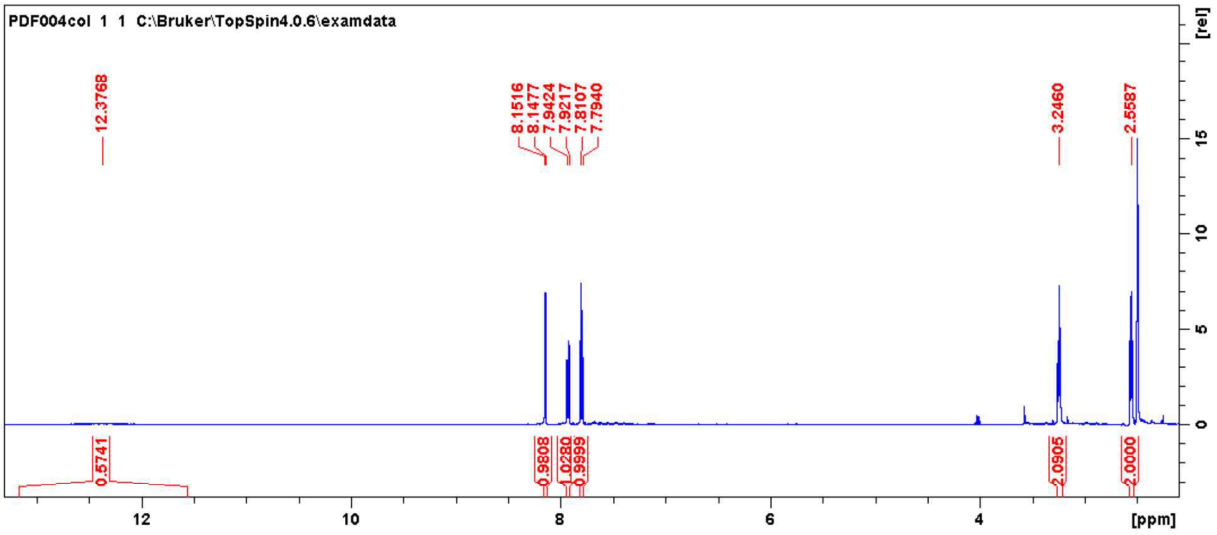


4-(3,4-dichlorophenyl)-4-oxobutanoic acid

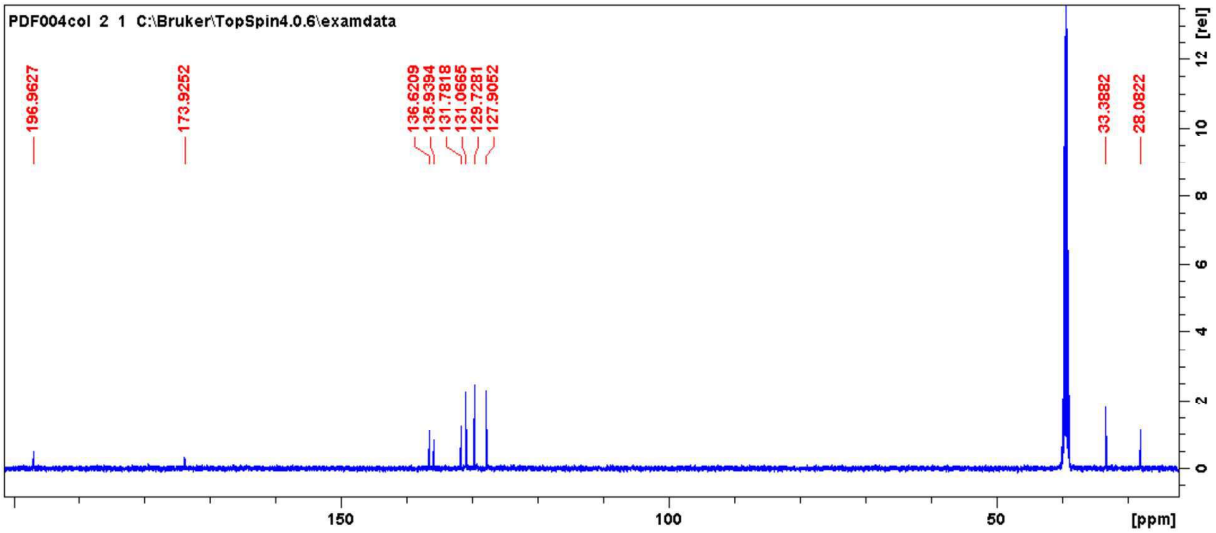


Succinic anhydride (5 g, 0.05 mol, 1 Eq.) was dissolved in 1,2-dichlorobenzene (44.1 g, 0.3 mol, 6 Eq.). Aluminum chloride (19.9 g, 0.15 mol, 3 Eq.) were added while stirring. The mixture was heated to 60 °C for 5 h. After cooling to 45 °C, the mixture was slowly poured into ice cold water (120 mL) and stirred for 0.5 h. Then, hexane (60 mL) was added and the mixture was stirred for 2 h. The resulting solid was filtered and dried *in vacuo*. No further purification was required.
Yield: 5.04 g (0.021 mol, 41 %) of a yellow solid.
^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 2.56 (t, 2H, $^{3}J_{HH}$ = 6.3 Hz), 3.25 (t, 2H, $^{3}J_{HH}$ = 6.3 Hz), 7.80 (d, 1H, $^{3}J_{HH}$ = 8.4 Hz), 7.93 (d, 1H, $^{3}J_{HH}$ = 8.4 Hz), 8.15 (s, 1H), 12.37 (s, 1H)
^{13}C-NMR (126 MHz, DMSO-d6): δ [ppm] = 28.08, 33.39, 127.91, 129.73, 131.01, 131.78, 135.94, 136.62, 173.93, 196.96
UPLC-MS: calc; [M +H]: 246.99, found [M + H]: 246.99

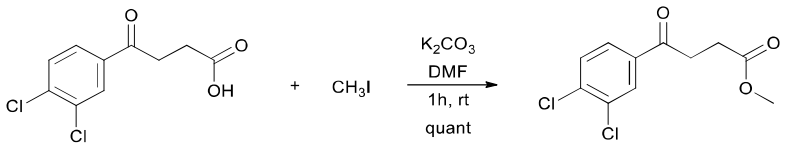
^{1}H NMR for 4-(3,4-dichlorophenyl)-4-oxobutanoic acid



^{13}C NMR for 4-(3,4-dichlorophenyl)-4-oxobutanoic acid

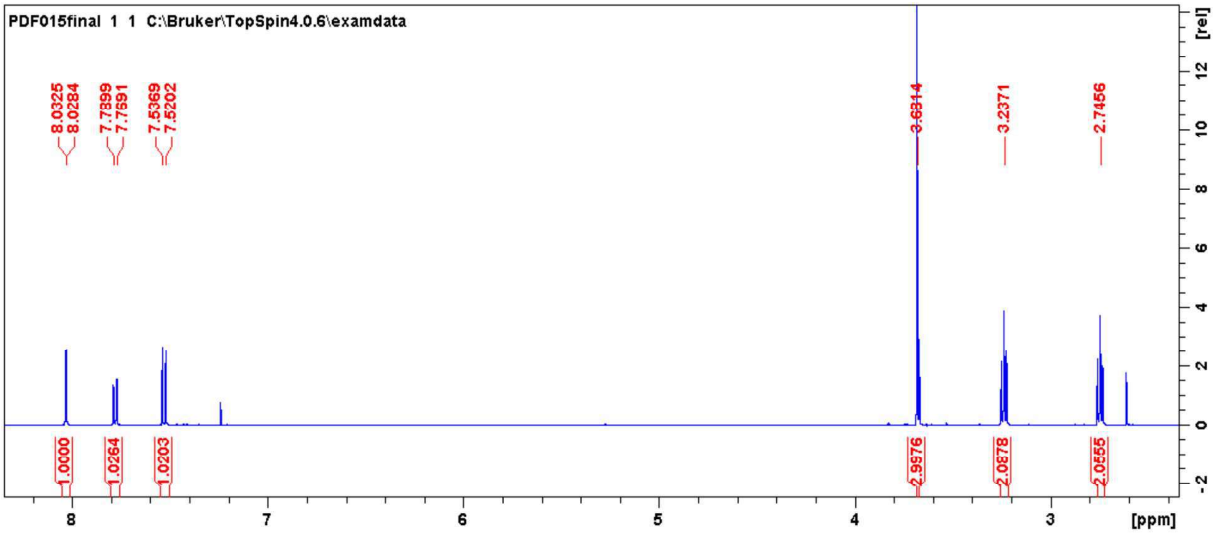


methyl 4-(3,4-dichlorophenyl)-4-oxobutanoate

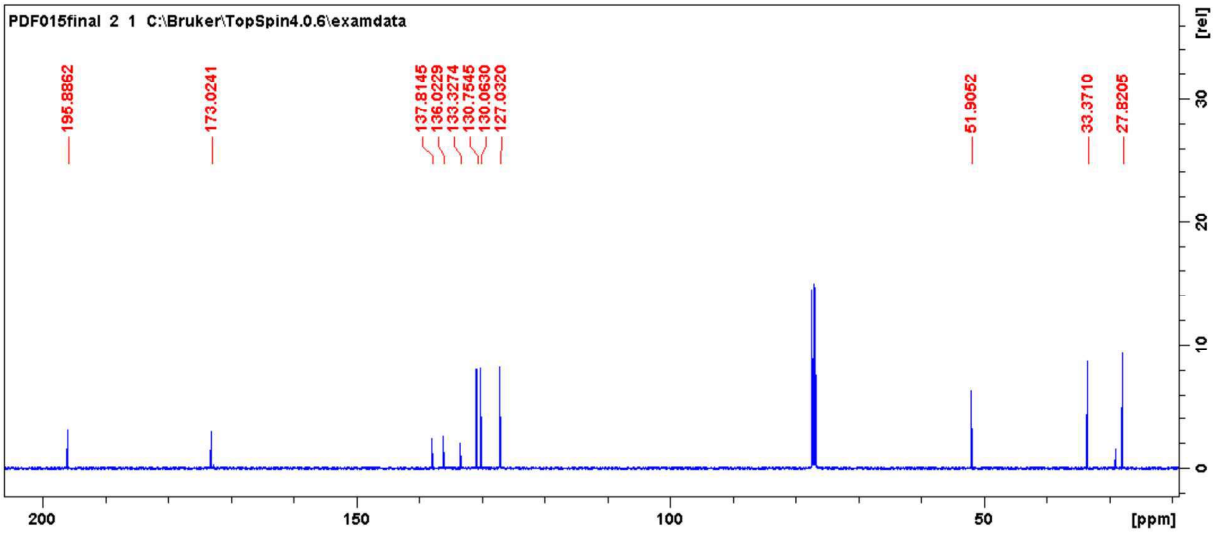


4-(3,4-dichlorophenyl)-4-oxobutanoic acid (3.19 g, 13 mmol, 1 Eq.) was dissolved in dimethyl formamide (10 mL). Potassium carbonate (2.34 g, 17 mmol, 1.3 Eq.) was added and the mixtures was stirred until it turns turbid. Then, iodomethane (1.37 mL, 22 mmol, 1.7 Eq.) was added and the mixture was stirred for 1 h at room temperature. The product was extracted with ethylacetate (100 mL) and washed with a mixture of saturated sodium carbonate solution and saturated sodium chloride solution (300 mL, 1:1). The organic solvent was removed in vacuo and the product was purified by chromatography (isco Combiflash, 24 g column, silica gel, dichloromethane/methanol 1-10 % gradient).
Yield: 3.39 g (13 mmol, 100 %) of an orange solid.
^{1}H-NMR (500 MHz, $CDCl_3$): δ [ppm] = 2.75 (t, 2H, $^{3}J_{HH}$ = 6.5 Hz), 3.24 (t, 2H, $^{3}J_{HH}$ = 6.4 Hz), 3.68 (s, 3H), 7.53 (d, 1H, $^{3}J_{HH}$ = 8.3 Hz), 7.78 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz, , $^{4}J_{HH}$ = 2.2 Hz), 8.03 (d, 1H, $^{4}J_{HH}$ = 2.1 Hz)
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 27.82, 33.37, 51.91, 127.03, 130.06, 130,73, 133.33, 136.03, 137.81, 173.02, 195.89
UPLC-MS: calc; [M +H]: 261.01, found [M + H]: 261.00

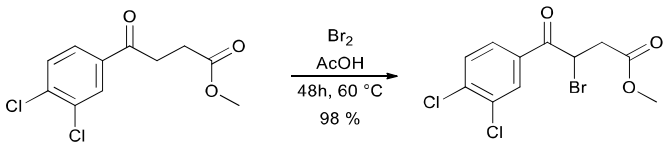
^{1}H NMR for methyl 4-(3,4-dichlorophenyl)-4-oxobutanoate



^{13}C NMR for methyl 4-(3,4-dichlorophenyl)-4-oxobutanoate

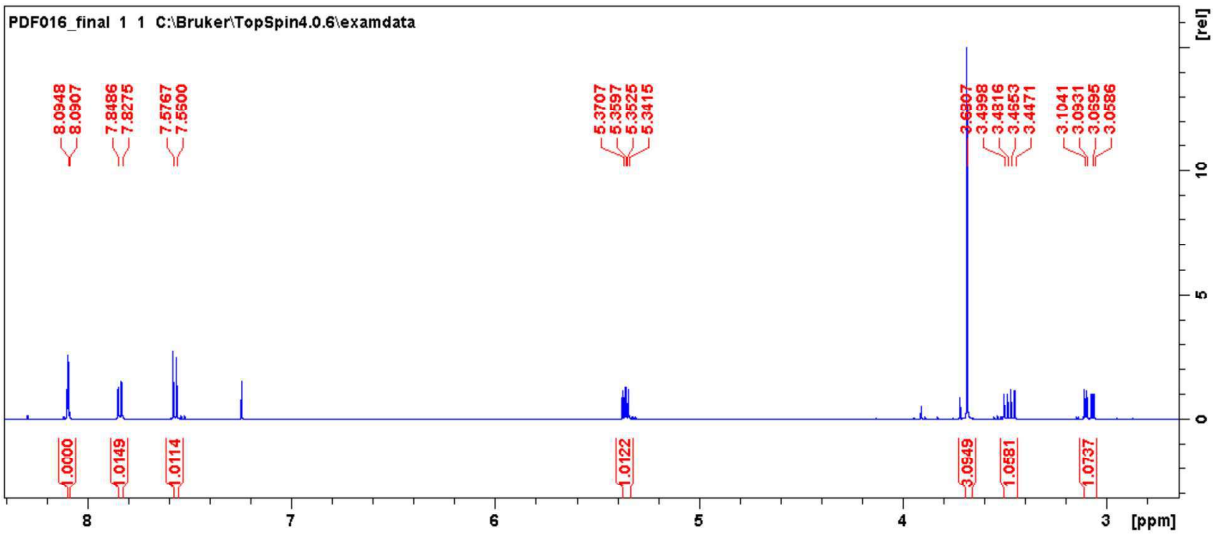


methyl 3-bromo-4-(3,4-dichlorophenyl)-4-oxobutanoate

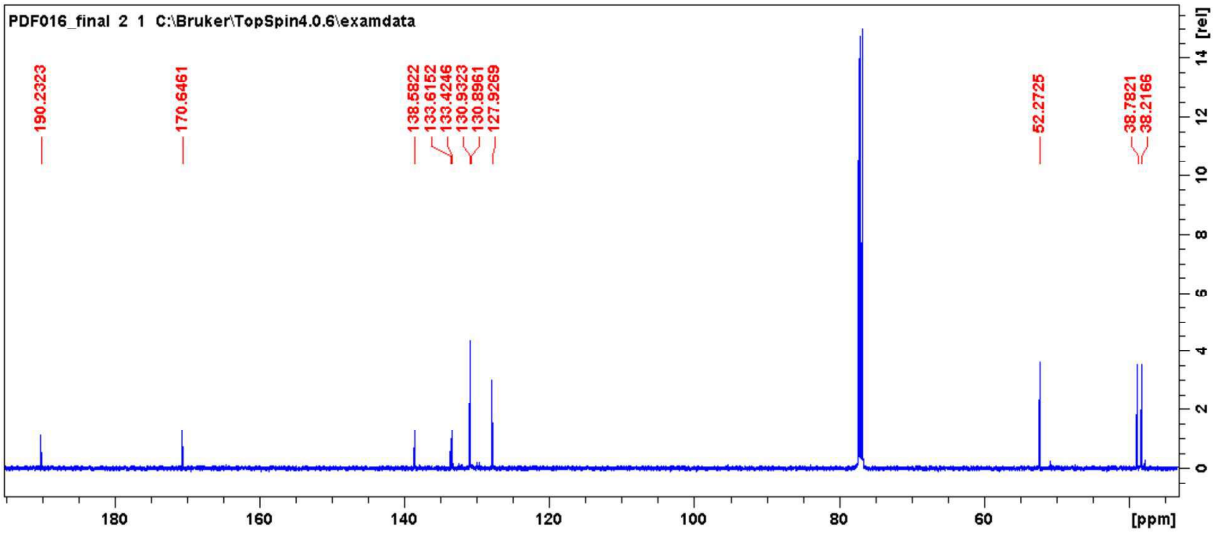


Methyl 4-(3,4-dichlorophenyl)-4-oxobutanoate (2.08 g, 8 mmol, 1 Eq.) was dissolved in glacial acetic acid (20 mL). The mixture was stirred at room temperature until it became clear. Then, bromine (410 µL, 8 mmol, 1 Eq.) was added slowly and dropwise. The reaction mixture was heated to 60 °C and stirred for 48 h. The resulting product was diluted with water (20 mL) and extracted with dichloromethane until the organic phase is colorless (4 x 60 mL). The organic solvent was removed in vacuo. No further purification was required.
Yield: 2.58 g (7.64 mmol, 96 %) of an orange oil.
^{1}H-NMR (500 MHz, $CDCl_3$): δ [ppm] = 3.08 (dd, 1H, $^{2}J_{HH}$ = 17.3 Hz, $^{3}J_{HH}$ = 5.5 Hz), 3.47 (dd, 1H, $^{2}J_{HH}$ = 17.3 Hz , $^{3}J_{HH}$ = 9.1 Hz), 3.68 (s, 3H), 5.36 (dd, 1H, $^{3}J_{HH}$ = 9.1 Hz, 5.5 Hz), 7.57 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz), 7.84 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz), 8.09 (d, 1H, $^{4}J_{HH}$ = 1.9 Hz)
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 38.22, 38.78, 52.27, 127.93, 130.90, 130.93, 133.42, 133.62, 138.58, 170.65, 190.23
UPLC-MS: calc; [M +H]: 338.92, found [M + H]: 338.91

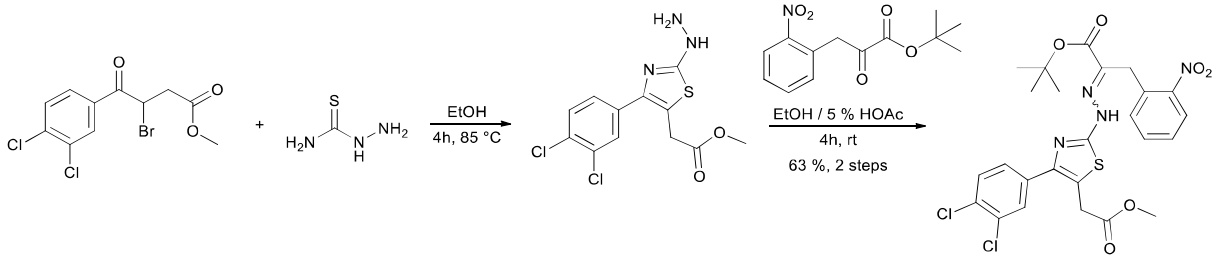
^{1}H NMR for methyl 3-bromo-4-(3,4-dichlorophenyl)-4-oxobutanoate



^{13}C NMR for methyl 3-bromo-4-(3,4-dichlorophenyl)-4-oxobutanoate



tert-butyl 2-(2-(4-(3,4-dichlorophenyl)-5-(2-methoxy-2-oxoethyl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoate



Methyl 3-bromo-4-(3,4-dichlorophenyl)-4-oxobutanoate (150 mg, 0.44 mmol, 1 Eq.) was dissolved in ethanol (10 mL). Thiosemicarbazide (40 mg, 0.44 mmol, 1 Eq.) was added and the mixture was heated to 85 °C for 4 h. The organic solvent was removed *in vacuo*. An aliquot of the crude product (16 mg, 60.4 µmol, 1 Eq.) and *tert*-butyl 3-(2-nitrophenyl)-2-oxopropanoate (20 mg, 60.4 µmol, 1 Eq.) were dissolved in acetic acid (2 mL, 5 % in ethanol). The mixture was stirred at room temperature for 4 h. The solvent was removed *in vacuo* and the product (a mixture of E/Z-isomers) was purified by chromatography (isco Combiflash, 4 g column, silica gel, methanol/dichloromethane gradient).

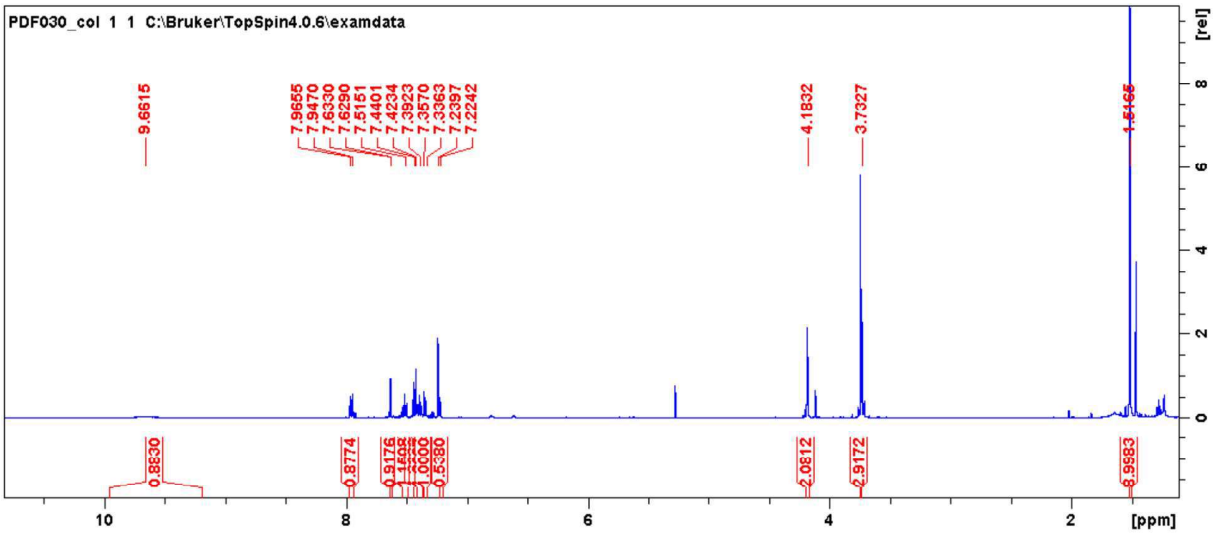
Yield: 22 mg (38.1 µmol, 63%) of a yellow solid.

^{1}H-NMR (500 MHz, $CDCl_3$): δ [ppm] = 1.52 (s, 9H), 3.73 (s, 2H), 3.74 (s, 3H), 4.28 (s, 2H), 7.23 (d, 1H, $^{3}J_{HH}$ = 7.9 Hz), 7.35 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz), 7.39 (dd, 1H, $^{3}J_{HH}$ = 8.6 Hz, 7.4 Hz), 7.43 (d, 1H, $^{3}J_{HH}$ = 8.5 Hz), 7.51 (dd, 1H, $^{3}J_{HH}$ = 8.6 Hz, 7.4 Hz), 7.63 (d, 1H, $^{4}J_{HH}$ = 1.9 Hz) 7.96 (d, 1H, $^{3}J_{HH}$ = 8.2 Hz)

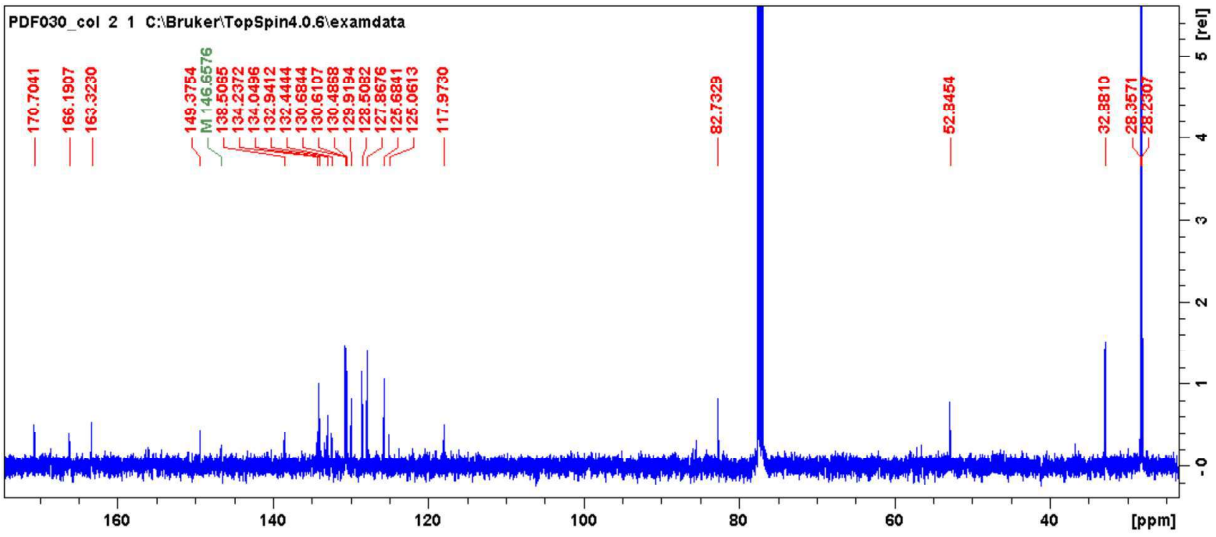
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 28.23, 28.36, 32.88, 52.85, 82.73, 117.97, 125.06, 125.68, 127.87, 128.51, 129.92, 130.49, 130.61, 130.68, 132.44, 132.94, 134.05. 134.24. 138.51, 146.66, 149.38, 163.32, 166.19, 170.70

UPLC-MS: calc; [M +H]: 579.09 found [M + H]: 579.08

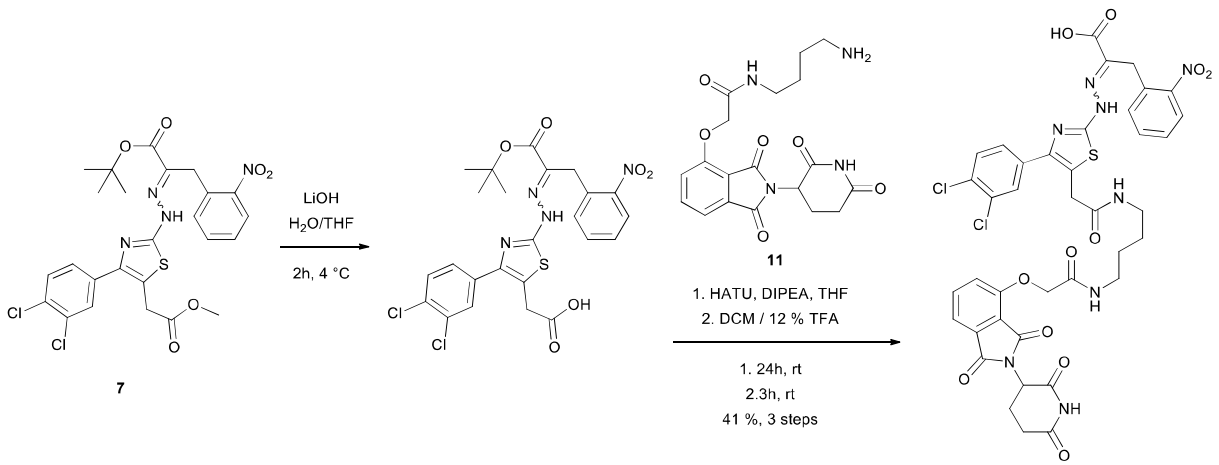
^{1}H NMR for tert-butyl 2-(2-(4-(3,4-dichlorophenyl)-5-(2-methoxy-2-oxoethyl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoate



^{13}C NMR for tert-butyl 2-(2-(4-(3,4-dichlorophenyl)-5-(2-methoxy-2-oxoethyl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoate

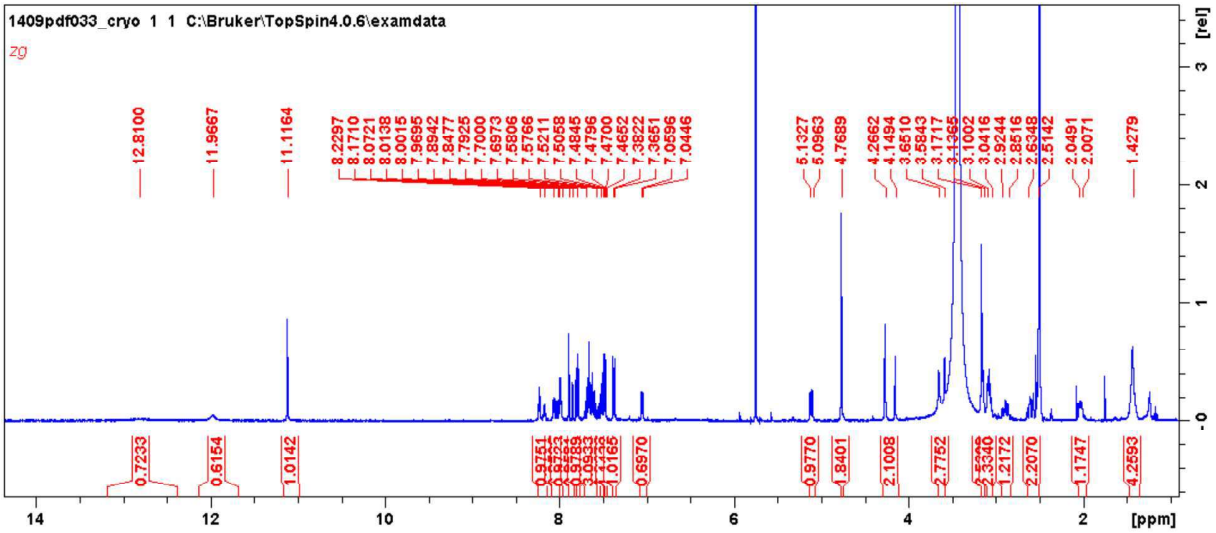


2-(2-(4-(3,4-dichlorophenyl)-5-(2-((4-(2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)oxy)acetamido)butyl)amino)-2-oxoethyl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoic acid (d4E-1)

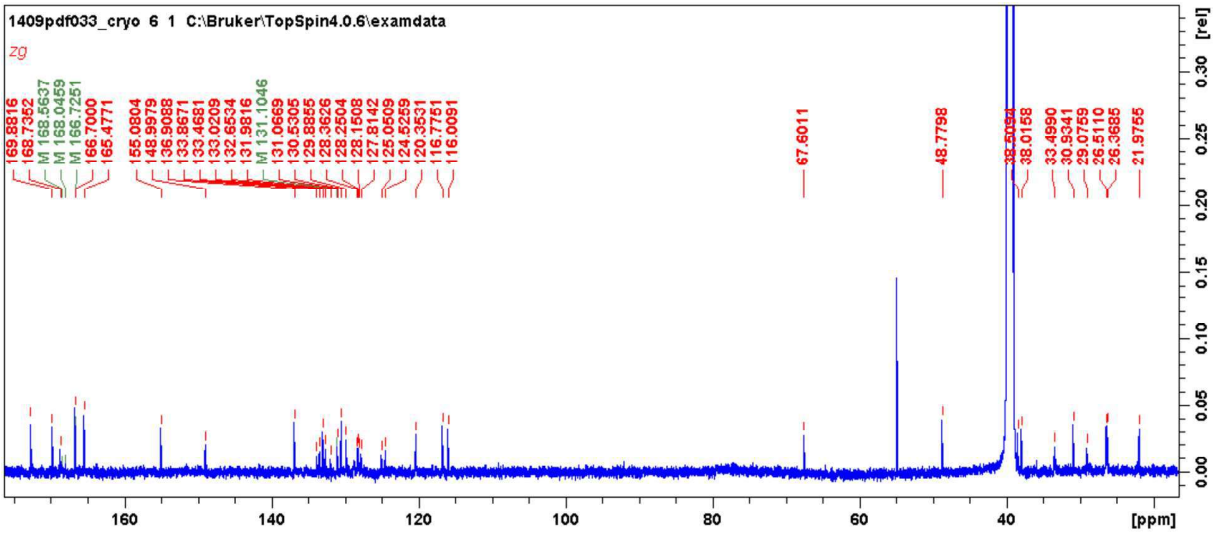


Tert-butyl 2-(2-(4-(3,4-dichlorophenyl)-5-(2-methoxy-2-oxoethyl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoate (11 mg, 19 µmol, 1 Eq.) was dissolved in tetrahydrofurane (1 mL) at 4 °C. Lithium hydroxide (1.6 mg, 38 µmol, 2 Eq.) was dissolved in water (1 mL) and the resulting solution was added dropwise to the reaction mixture. After stirring for 2 h at 4 °C, the mixture was diluted with 1 M hydrochloric acid (2 mL) and extracted with dichloromethane. The organic solvent was removed *in vacuo*. An aliquot of the crude product (5 mg, 8.86 µmol, 1 Eq.), N-(4-aminobutyl)-2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)oxy)acetamide (3.6 mg, 8.86 µmol, 1 Eq.), and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU, 3.4 mg, 8.86 µmol, 1 Eq.) were dissolved in tetrahydrofurane (2 mL). N,N-Diisopropylethylamine (DIPEA, 1.6 µL, 8.86 µmol, 1 Eq.) was added to the reaction mixture and stirred at room temperature for 24 h. The solvent was removed *in vacuo* and the coupling product was used without further purification. Trifluoroacetic acetic (12 % in dichloromethane, 2 mL) was added and the mixture was stirred at room temperature for 3 h. The solvent was removed in vacuo and the product was purified using high performance liquid chromatography (HPLC, C18 reversed phase preparative column, water/acetonitrile 5/95).
Yield: 3.9 mg (4.34 µmol, 41 %) of a yellow solid.
^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 1.43 (m, 4H), 2.03 (m, 1H), 2.52 (m, 1H), 2.59 (m, 1H), 2.89 (m, 1H), 3.08 (m, 2H), 3.15 (m, 2H), 3.65 (s 2H), 4.15/4.27 (s, 2H), 4.77 (s, 2H), 5.11 (dd, 1H, $^{3}J_{HH}$ = 6.1 Hz), 7.05 (d, 1H, $^{3}J_{HH}$ = 7.6 Hz), 7.38 (d, 1H, $^{3}J_{HH}$ = 8.5 Hz), 7.48 (d, 1H, $^{3}J_{HH}$ = 7.2 Hz), 7.51 (d, 1H, $^{3}J_{HH}$ = 7.8 Hz), 7.60 (m, 1H), 7.64 (m, 1H), 7.66 (m, 1H), 7.79 (dd, 1H, $^{3}J_{HH}$ = 8.0 Hz), 7.87 (d, 1H, $^{4}J_{HH}$ = 2.0 Hz), 7.99 (t, 1H, $^{3}J_{HH}$ = 5.4 Hz), 8.02/8.07 (d, 1H, $^{3}J_{HH}$ = 8.4 Hz), 8.17/8.23 (t, 1H, $^{3}J_{HH}$ = 5.5 Hz), 11.12 (s, 1H), 11.97 (s, 1H), 12.80 (s, 1H)
^{13}C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 21.98, 26.37, 26.52, 29.08, 30.94, 33.37/33.50, 35.90, 38.02, 38.52, 48.79, 67.61, 116.02, 116.78, 120.36, 124.53, 125.05, 127.82, 128.25, 128.80, 129.89, 130.54, 131.07, 131.10, 132.00, 132.66, 133.03, 133.47, 133.87, 135.29, 136.91, 149.00/149.06, 155.08, 164.13/165.48, 166.70, 168.54/168.74, 169.89, 172.78, 172.91, 174.32
UPLC-MS: calc; [M +H]: 893.15 found [M + H]: 893.15

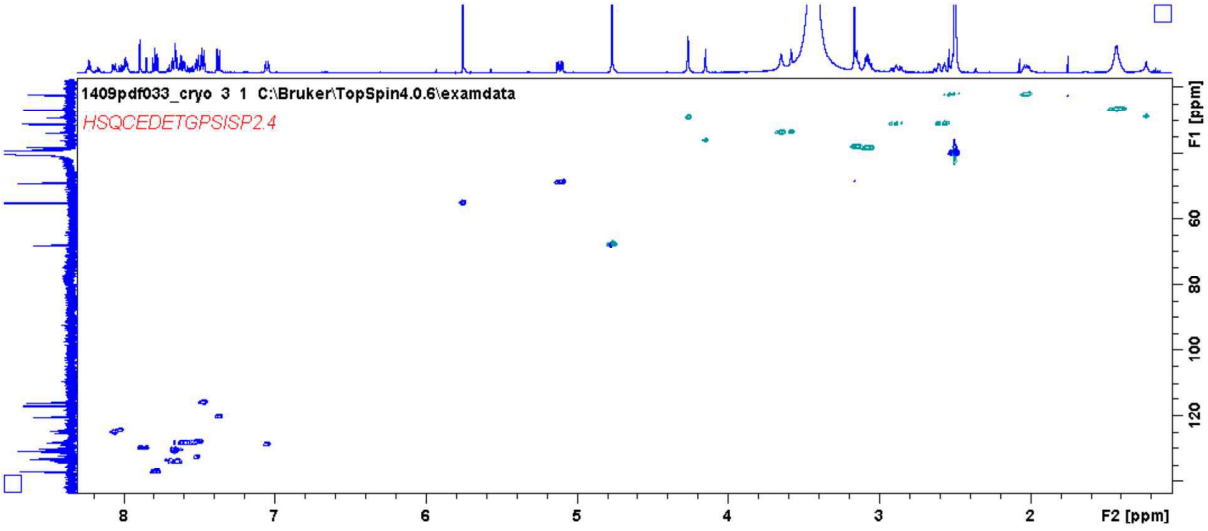
^{1}H NMR for d4E-1



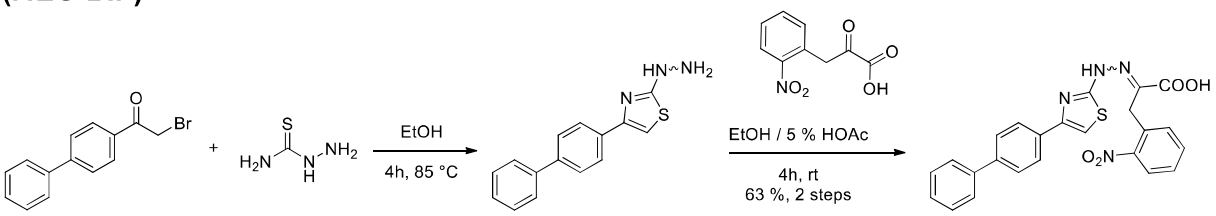
^{13}C NMR for d4E-1



^{1}H-^{13}C-HSQC for d4E-1

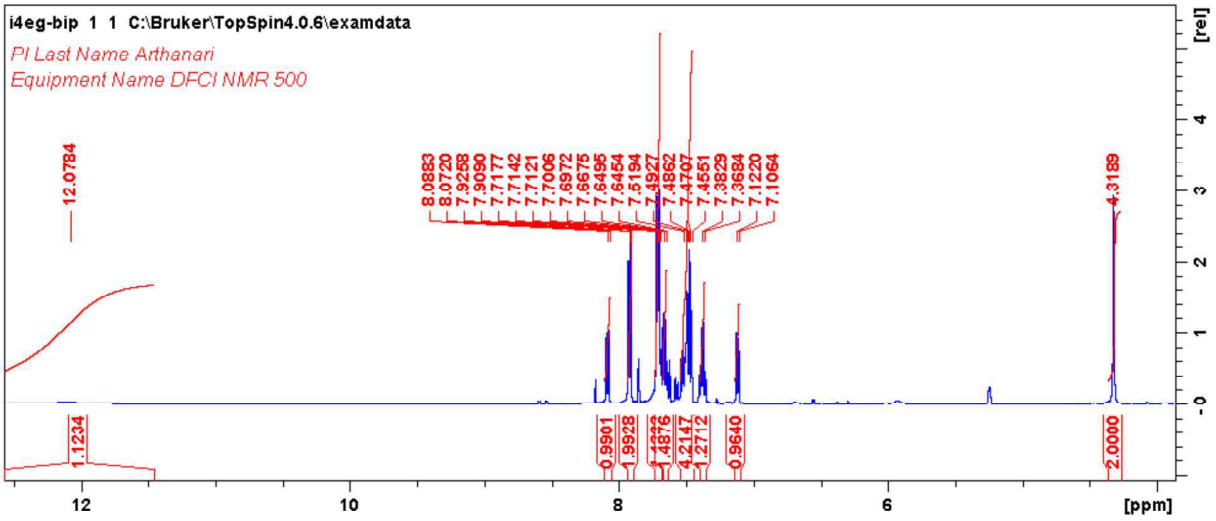


2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoic acid (i4EG-BiP)

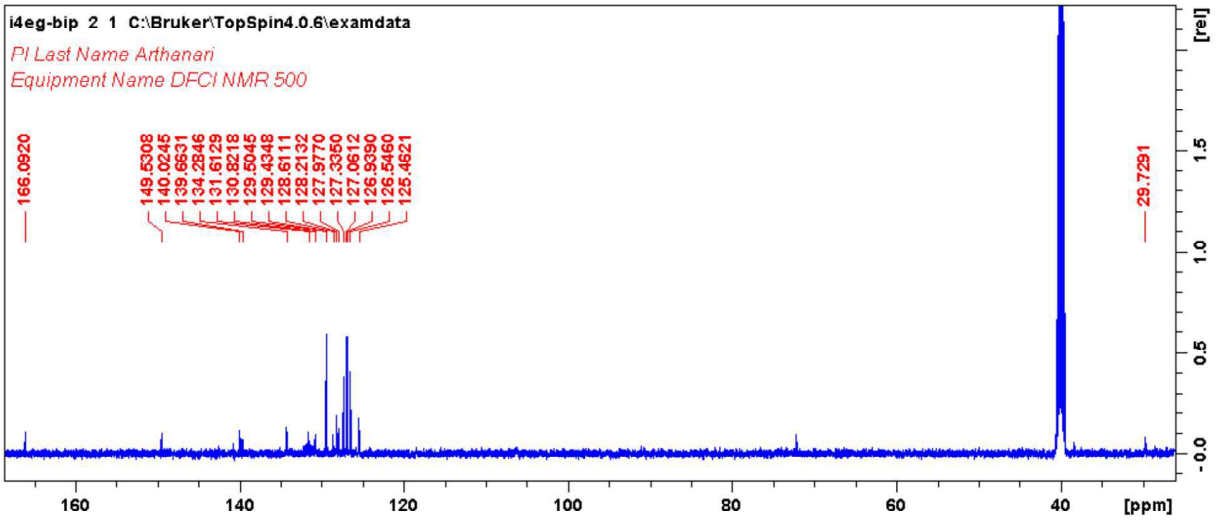


1-([1,1'-biphenyl]-4-yl)-2-bromoethan-1-one (500 mg, 1.82 mmol, 1 Eq.) was dissolved in ethanol (10 mL). Thiosemicarbazide (166 mg, 1.82 mmol, 1 Eq.) was added and the mixture was heated to 85 °C for 4 h. The organic solvent was removed *in vacuo*. An aliquot of the crude product (50 mg, 187 µmol, 1 Eq.) and 3-(2-nitrophenyl)-2-oxopropanoic acid (39 mg, 187 µmol, 1 Eq.) were dissolved in acetic acid (2 mL, 5 % in ethanol). The mixture was stirred at room temperature for 4 h. The solvent was removed *in vacuo* and the product (a mixture of E/Z-isomers) was purified by chromatography (isco Combiflash, 4 g column, silica gel, methanol/dichloromethane gradient).
Yield: 54 mg (117.8 µmol, 63%) of a yellow solid.
^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 4.32 (s, 2H), 7.12 (d, 1H, $^{3}J_{HH}$ = 7.8 Hz), 7.38 (dd, 1H, $^{3}J_{HH}$ = 13.1 Hz, 5.5 Hz), 7.46-7.52 (m, 4H), 7.65-7.72 (m, 6H), 7.92 (d, 2H, $^{3}J_{HH}$ = 8.4 Hz). 8.08 (d, 1H, $^{3}J_{HH}$ = 8.1 Hz), 12.08 (s, 1H)
^{13}C-NMR (126 MHz, DMSO-d6): δ [ppm] = 29.73, 125.46, 126.55, 126.94, 127.06, 127.34, 127.98, 128.21, 128.61, 129.43, 129.50, 130.82, 131.61, 134.28, 139.66, 140.02, 149.53, 166.09
UPLC-MS: calc; [M +H]: 459.11 found [M + H]: 459.11

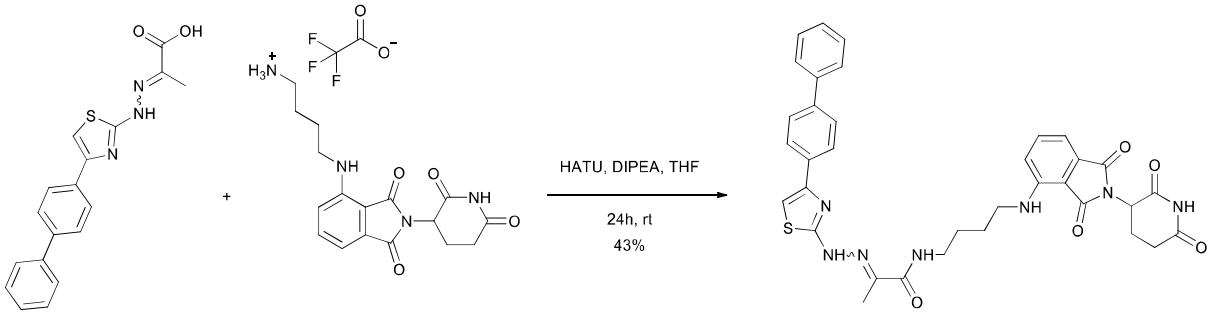
^{1}H NMR for i4EG-BiP



^{13}C NMR for i4EG-BiP

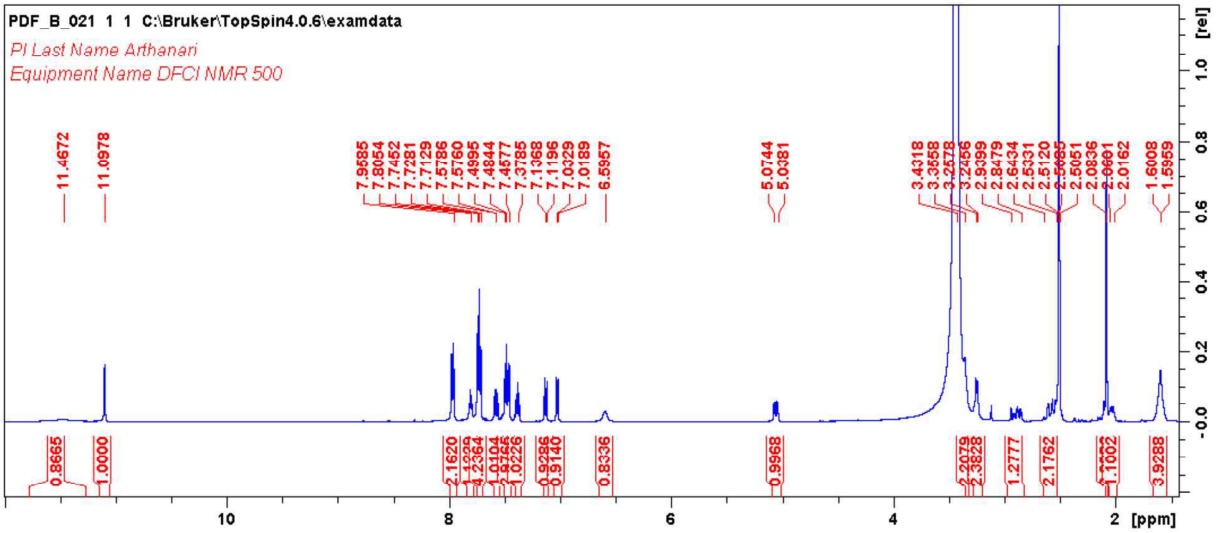


2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)-N-(4-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)butyl)propenamide (d4E-2)

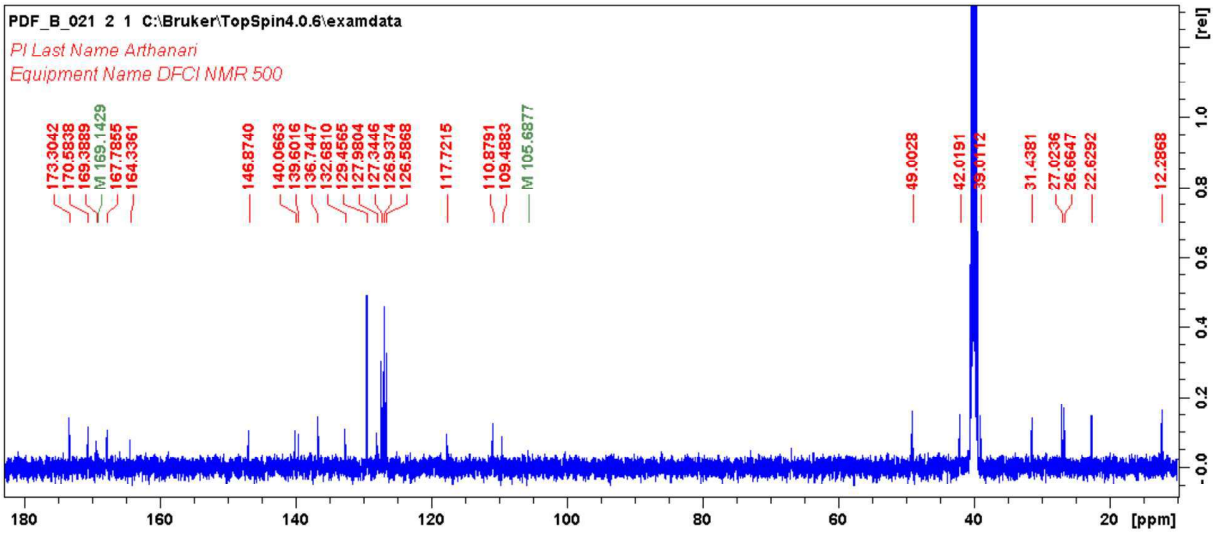


2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)propanoic acid (8.9 mg, 26.4 µmol, 1 Eq.), 4-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)butan-1-aminium trifluoroacetate (12.1 mg, 26.4 µmol, 1 Eq.), and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU, 10 mg, 26.4 µmol, 1 Eq.) were dissolved in tetrahydrofurane (2 mL). N,N-Diisopropylethylamine (DIPEA, 9.3 µL, 52.8 µmol, 2 Eq.) was added to the reaction mixture and stirred at room temperature for 24 h. The solvent was removed in vacuo and the product was purified using high performance liquid chromatography (HPLC, C18 reversed phase preparative column, water/acetonitrile 5/95).
Yield: 7.5 mg (11.4 µmol, 43 %) of a yellow solid.
^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 1.60 (m, 4H), 2.02-2.06 (m, 1H), 2.08 (s, 3H), 2.53-2.64 (m, 2H), 2.85-2.94 (m, 1H), 3.25 (m, 2H), 3.36 (m, 2H), 5.06 (dd, 1H, $^{3}J_{HH}$ = 13.1 Hz, 5.5 Hz), 6.60 (s, 1H), 7.03 (d, 1H, $^{3}J_{HH}$ = 7.0 Hz), 7.13 (d, 1H, $^{3}J_{HH}$ = 8.6 Hz), 7.38 (t, 1H, 7.3 Hz), 7.46-7.50 (m, 3H), 7.58 (dd, 1H, $^{3}J_{HH}$ = 8.6 Hz, 7.0 Hz), 7.73 (t, 4H, $^{3}J_{HH}$ = 8.1 Hz), 7.81 (t, 1H, $^{3}J_{HH}$ = 6.1 Hz), 7.97 (d, 1H, $^{3}J_{HH}$ = 8.5 Hz), 11.10 (s, 1H), 11.47 (s, 1H)
^{13}C-NMR (126 MHz, DMSO-d6): δ [ppm] = 12.29, 22.63, 26.66, 27.02, 31.44, 39.01, 42.02, 49.00, 105.69, 109.49, 110.88, 117.72, 126.59, 126.94, 127.34, 127.98, 129.46, 132.68, 136.74, 139.60, 140.07, 146.87, 164.34, 167.79, 169.14, 169.39, 170.58, 173.30
UPLC-MS: calc; [M +H]: 664.23 found [M + H]: 664.24

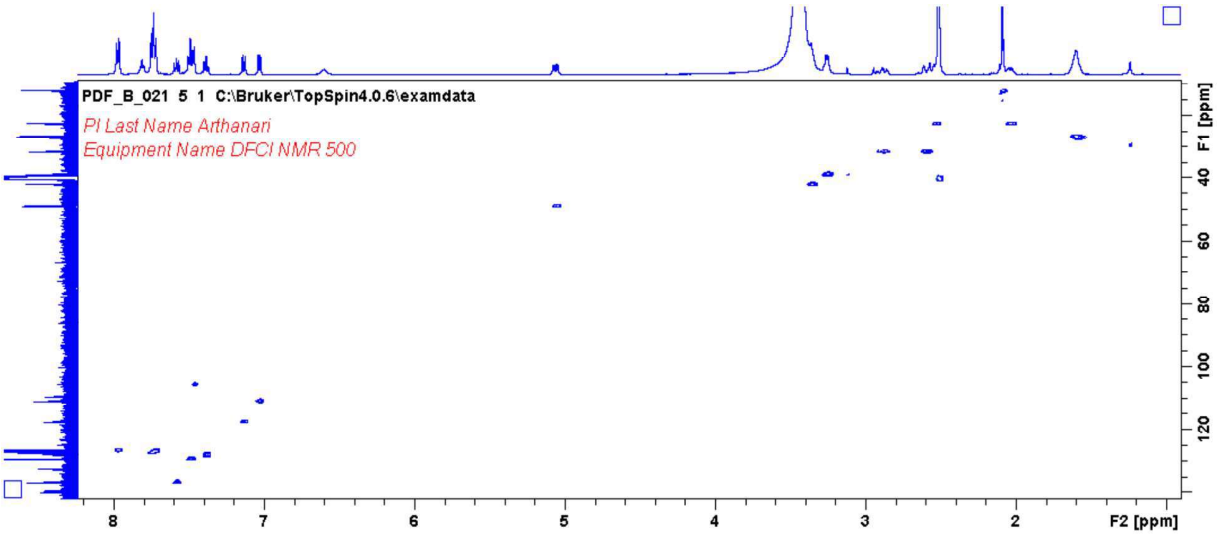
^{1}H NMR for d4E-2



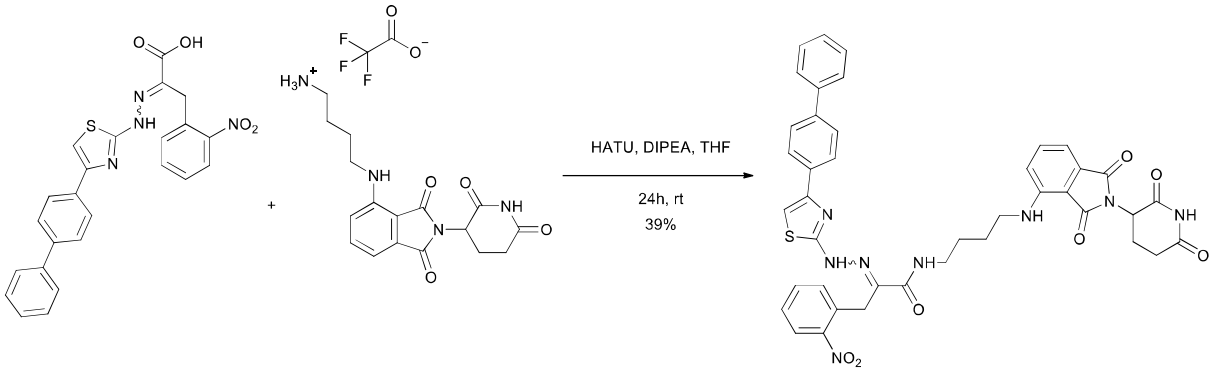
^{13}C NMR for d4E-2



^{1}H-^{13}C-HSQC for d4E-2



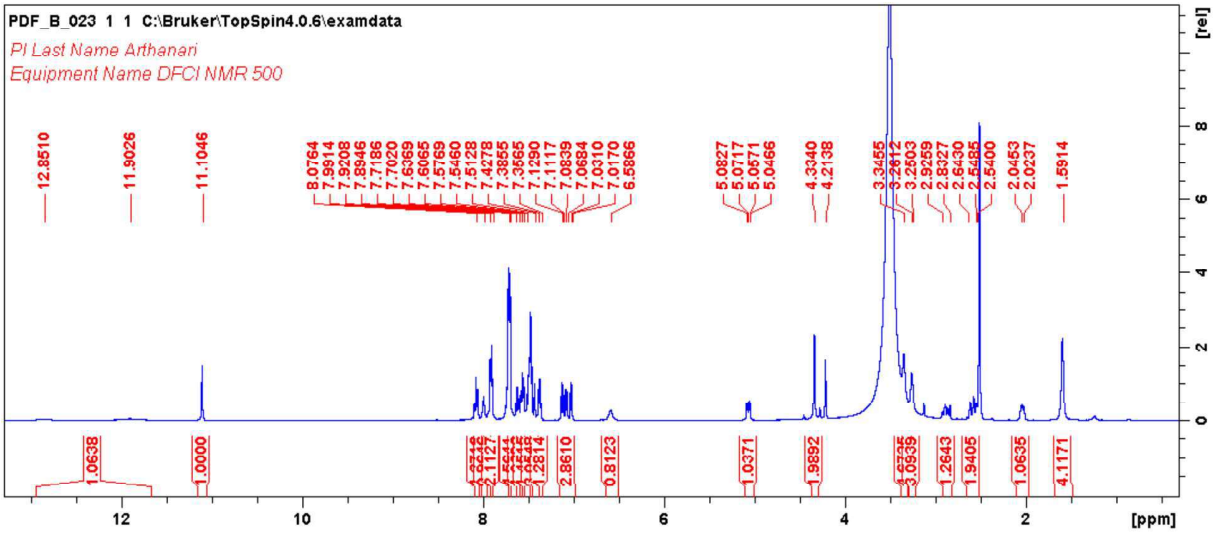
2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)-N-(4-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)butyl)-3-(2-nitrophenyl)propenamide (d4E-3)



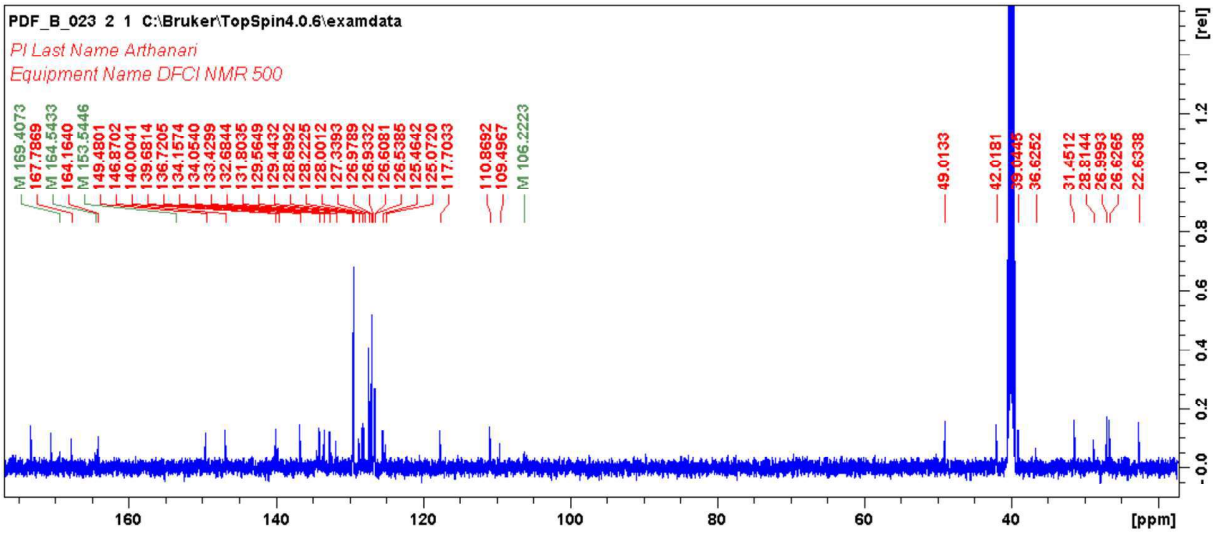
2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)-3-(2-nitrophenyl)propanoic acid (10.9 mg, 23.8 µmol, 1 Eq.), 4-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)butan-1-aminium trifluoroacetate (10.9 mg, 23.8 µmol, 1 Eq.), and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU, 9 mg, 23.8 µmol, 1 Eq.) were dissolved in tetrahydrofurane (2 mL). N,N-Diisopropylethylamine (DIPEA, 8.3 µL, 47.6 µmol, 2 Eq.) was added to the reaction mixture and stirred at room temperature for 24 h. The solvent was removed *in vacuo* and the product was purified using high performance liquid chromatography (HPLC, C18 reversed phase preparative column, water/acetonitrile 5/95).
Yield: 7.3 mg (9.28 µmol, 39 %) of a yellow solid.
^{1}H-NMR (500 MHz, DMSO-d6): δ [ppm] = 1.59 (m, 4H), 2.02-2.05 (m, 1H), 2.54-2.64 (m, 2H), 2.83-2.93 (m, 1H), 3.25 (m, 2H), 3.35 (m, 2H), 4.21/4.33 (s, 2H), 5.06 (dd, 1H, $^{3}J_{HH}$ = 12.8 Hz, 5.5 Hz), 6.59 (s, 1H), 7.02 (d, 1H, $^{3}J_{HH}$ = 7.0 Hz), 7.08 (d, 1H, $^{3}J_{HH}$ = 7.7 Hz) 7.12 (d, 1H, $^{3}J_{HH}$ = 8.8 Hz), 7.37 (t, 1H, 7.3 Hz), 7.43-7.51 (m, 4H), 7.56 (dd, 1H, $^{3}J_{HH}$ = 8.6 Hz, 7.0 Hz), 7.62 (t, 1H, $^{3}J_{HH}$ = 7.6 Hz) 7.71 (t, 4H, $^{3}J_{HH}$ = 8.1 Hz), 7.89-7.92 (m, 1H), 7.99 (t, 1H, $^{3}J_{HH}$ = 6.1 Hz), 8.08 (t, 1H, $^{3}J_{HH}$ = 8.4 Hz) 11.10 (s, 1H), 11.90/12.85 (s, 1H)

^{13}C-NMR (126 MHz, DMSO-d6): δ [ppm] = 22.63, 26.63, 27.00, 28.81/36.63, 31.45, 39.04, 42.02, 49.01, 106.22, 109.49, 110.88, 117.72, 125.08, 125.47, 126.54, 126.61, 126.94, 126.98, 127.34, 128.01, 128.23, 128.70, 129.45/129.57, 131.81, 132.68, 133.43, 134.05, 134.15, 136.72, 139.68, 140.00, 146.87, 149.48, 149.60, 164.16, 164.53, 167.79, 169.14, 169.39, 170.58, 173.30
UPLC-MS: calc; [M +H]: 785.25 found [M + H]: 785.25

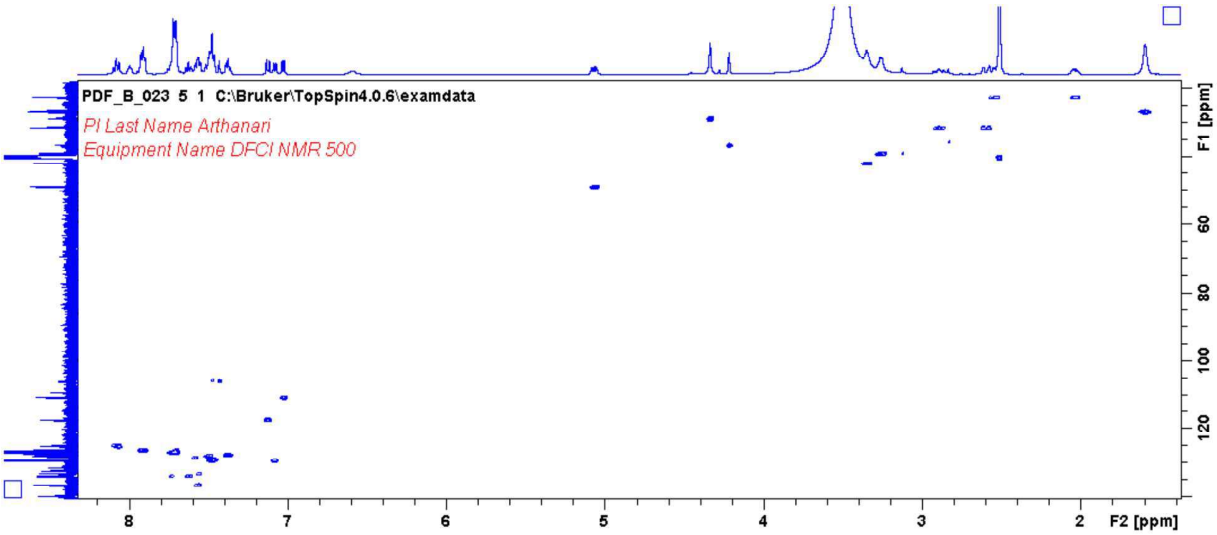
^{1}H NMR for d4E-3



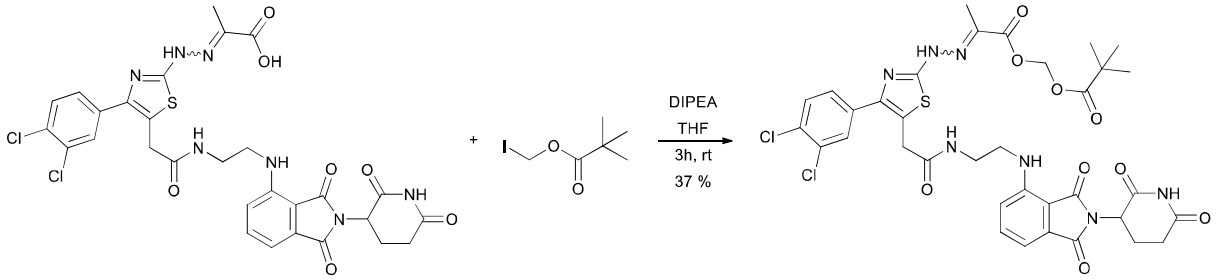
^{13}C NMR for d4E-3



^{1}H-^{13}C-HSQC for d4E-3

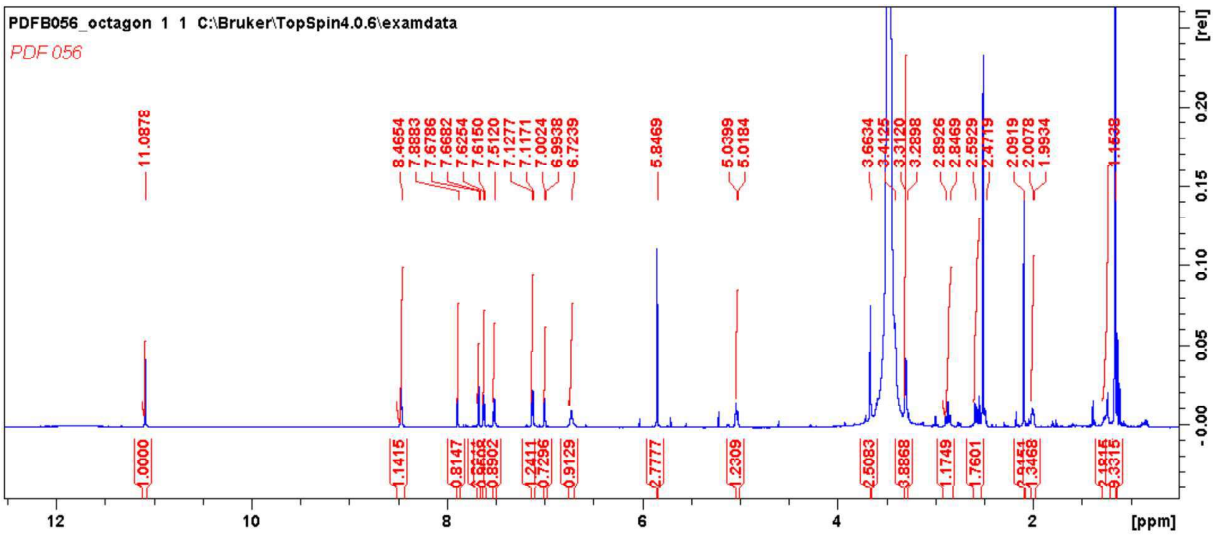


((2-(2-(4-(3,4-dichlorophenyl)-5-(2-((2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)ethyl)amino)-2-oxoethyl)thiazol-2-yl)hydrazineylidene)propanoyl)oxy)methyl pivalate (d4E-6)

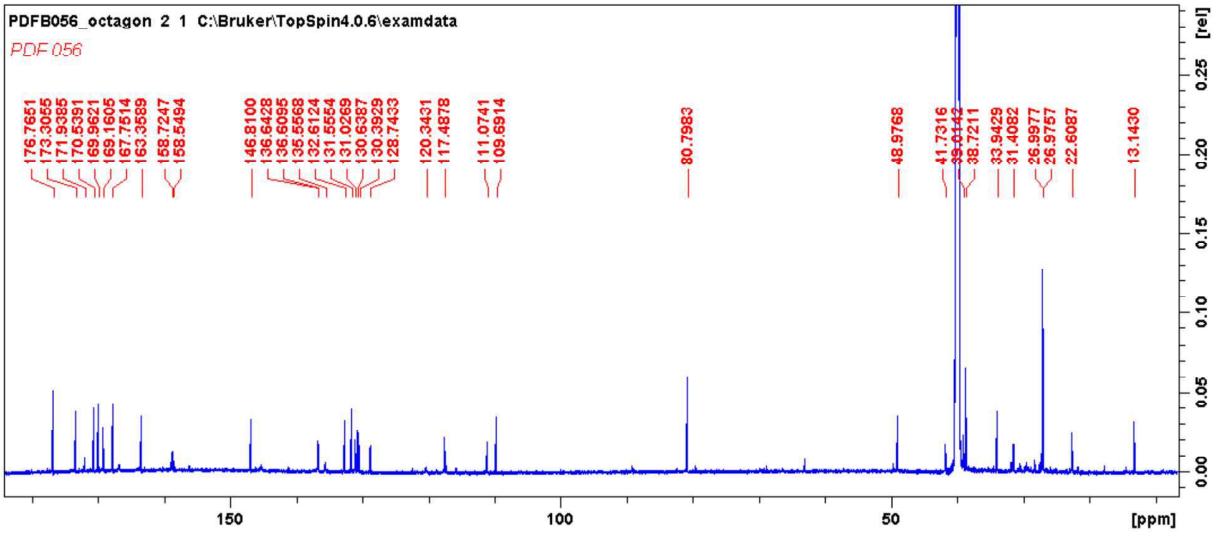


2-(2-(4-(3,4-dichlorophenyl)-5-(2-((2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)ethyl)amino)-2-oxoethyl)thiazol-2-yl)hydrazineylidene)propanoic acid (11 mg, 16 µmol, 1 Eq.) was dissolved in tetrahydrofurane (2 mL). Pivaloyloxymethyl iodide (4 µL, 16 µmol, 1 Eq.) and N,N-Diisopropylethylamine (DIPEA, 5.7 µL, 32 µmol, 2 Eq.) were added. The mixture was stirred at room temperature for 3 h. The solvent was removed *in vacuo* and the product was purified using high performance liquid chromatography (HPLC, C18 reversed phase preparative column, water/acetonitrile 5/95).
Yield: 4.7 mg (5.92 µmol, 37 %) of a yellow solid.
^{1}H-NMR (800 MHz, DMSO-d6): δ [ppm] = 1.15 (s, 9H), 1.99-2.01 (m, 1H), 2.09 (s, 3H), 2.47-2.59 (m, 2H), 2.85-2.89 (m, 1H), 3.29-3.31 (m, 2H), 3.41 (m, 2H), 3.66 (s, 2H), 5.02-5.04 (dd, 1H, $^{3}J_{HH}$ = 12.9 Hz, 5.3 Hz), 5.85 (s, 2H), 6.72 (s, 1H), 7.00 (d, 1H, $^{3}J_{HH}$ = 6.9 Hz), 7.12 (d, 1H, $^{3}J_{HH}$ = 7.9 Hz), 7.51 (t, 1H, $^{3}J_{HH}$ = 7.6 Hz), 7.62 (d, 1H, $^{3}J_{HH}$ = 8.4 Hz), 7.67 (d, 1H, $^{3}J_{HH}$ = 8.3 Hz), 7.89 (s, 1H), 8.47 (t, 1H, $^{3}J_{HH}$ = 5.3 Hz), 11.09 (s, 1H)
^{13}C-NMR (201 MHz, DMSO-d6): δ [ppm] = 12.86, 22.44, 31.29, 32.66, 33.87, 38.95, 48.91, 109.68, 111.08, 117.49, 120.04, 128.72, 128.76, 130.36, 130.42, 131.02, 131.53, 132.62, 135.86, 136.65, 136.71, 138.84, 146.83, 166.11, 167.75, 169.19, 170.03, 170.55, 173.23
UPLC-MS: calc; [M +H]: 800.17 found [M + H]: 800.15

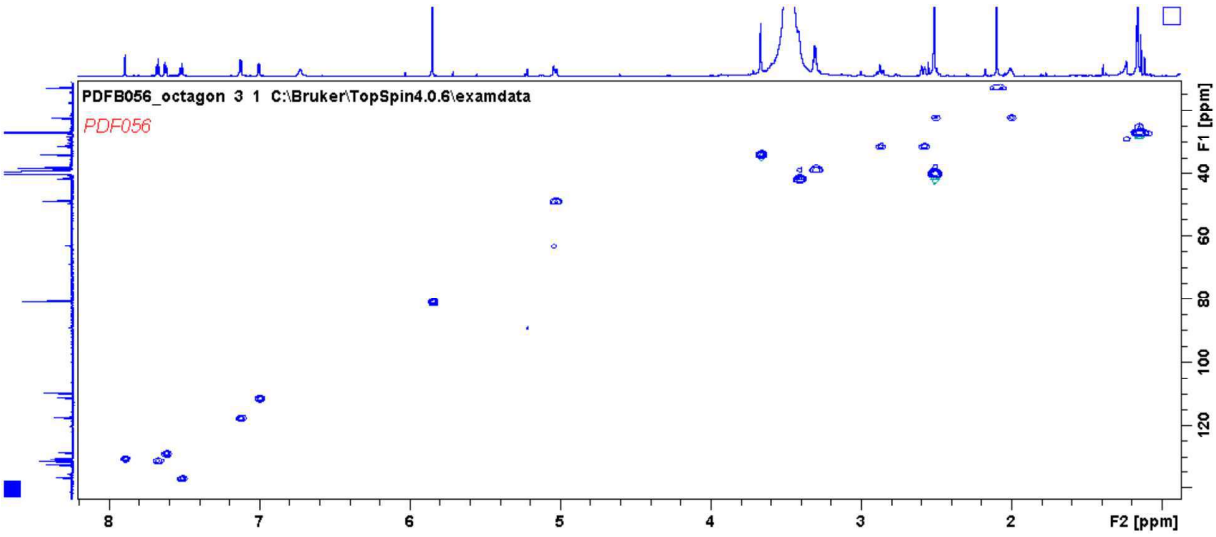
^{1}H NMR for d4E-6



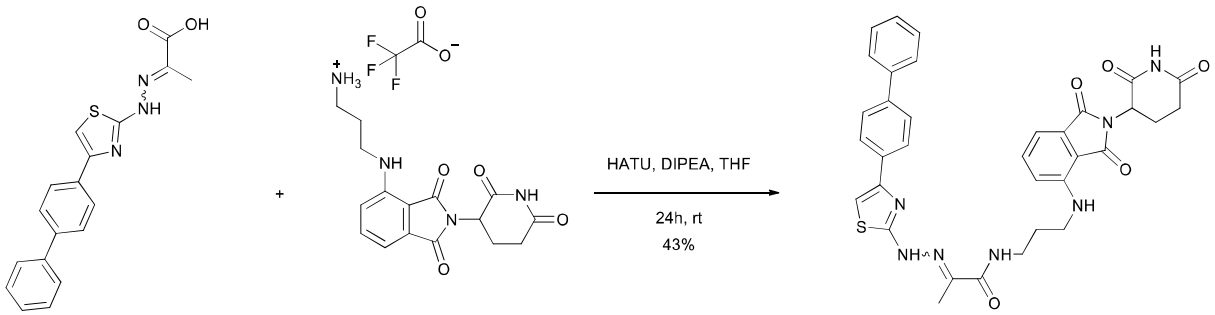
^{13}C NMR for d4E-6



^{1}H-^{13}C HSQC for d4E-6



2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)-N-(3-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)propyl)propenamide (d4E-5)



2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)propanoic acid (7.5 mg, 22.3 µmol, 1 Eq.), 3-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)propan-1-aminium trifluoroacetate (9.9 mg, 22.3 µmol, 1 Eq.), and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU, 8.5 mg, 22.3 µmol, 1 Eq.) were dissolved in tetrahydrofurane (2 mL). N,N-Diisopropylethylamine (DIPEA, 8 µL, 44.6 µmol, 2 Eq.) was added to the reaction mixture and stirred at room temperature for 24 h. The solvent was removed *in vacuo* and the coupling product was used without further purification. Trifluoroacetic acetic (12 % in dichloromethane, 2 mL) was added and the mixture was stirred at room temperature for 3 h. The solvent was removed in vacuo and the product was purified using high performance liquid chromatography (HPLC, C18 reversed phase preparative column, water/acetonitrile 5/95).

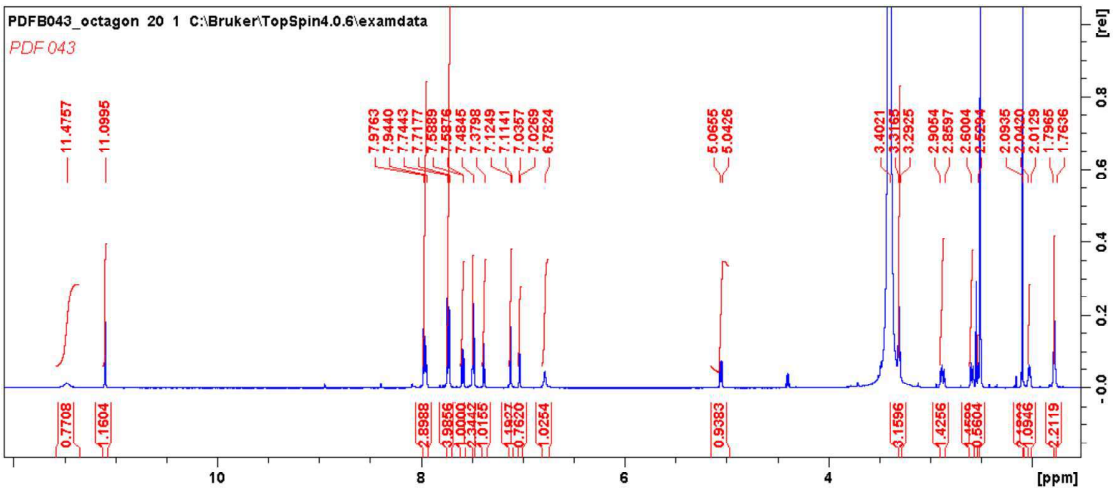
Yield: 6.2 mg (9.6 µmol, 43 %) of a yellow solid.

^{1}H-NMR (800 MHz, DMSO-d6): δ [ppm] = 1.76-1.80 (dt, 2H, $^{3}J_{HH}$ = 6.4 Hz, 6.8 Hz), 2.01-2.04 (m, 1H), 2.09 (s, 3H), 2.53-2.60 (m, 2H), 2.86-2.91 (m, 1H), 3.29-3.32 (m, 4H), 5.04-5.07 (dd, 1H, $^{3}J_{HH}$ = 12.6 Hz, 4.9 Hz), 6.78 (s, 1H), 7.03 (d, 1H, $^{3}J_{HH}$ = 7.0 Hz), 7.12 (d, 1H, $^{3}J_{HH}$ = 8.6 Hz), 7.38 (t, 1H, $^{3}J_{HH}$ = 7.3 Hz), 7.48 (t, 2H, $^{3}J_{HH}$ = 7.6 Hz), 7.59 (dd, 1H, $^{3}J_{HH}$ = 8.4 Hz, 7.3 Hz), 7.72-7.74 (m, 4H), 7.94-7.98 (m, 3H), 11.10 (s, 1H), 11.47 (s, 1H)

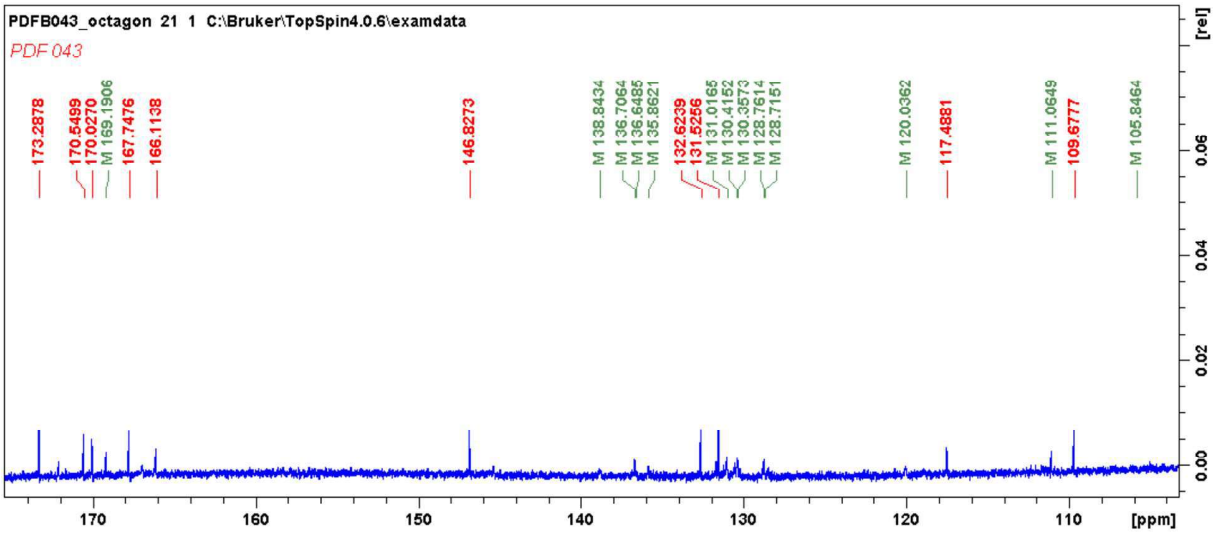
^{13}C-NMR (201 MHz, DMSO-d6): δ [ppm] = 13.14, 22.61, 26.98, 27.00, 31.41, 33.94, 38.72, 39.01, 41.73, 80.80, 105.85, 109.69, 111.07, 117.49, 120.34, 128.74, 130.39, 130.64, 131.03, 131.56, 132.61, 135.56, 136.61, 136.64, 146.81, 158.55, 158.72, 163.36. 167.75, 169.16, 170.54, 171.94, 173.31, 176.77

UPLC-MS: calc; [M +H]: 650.22 found [M + H]: 650.21

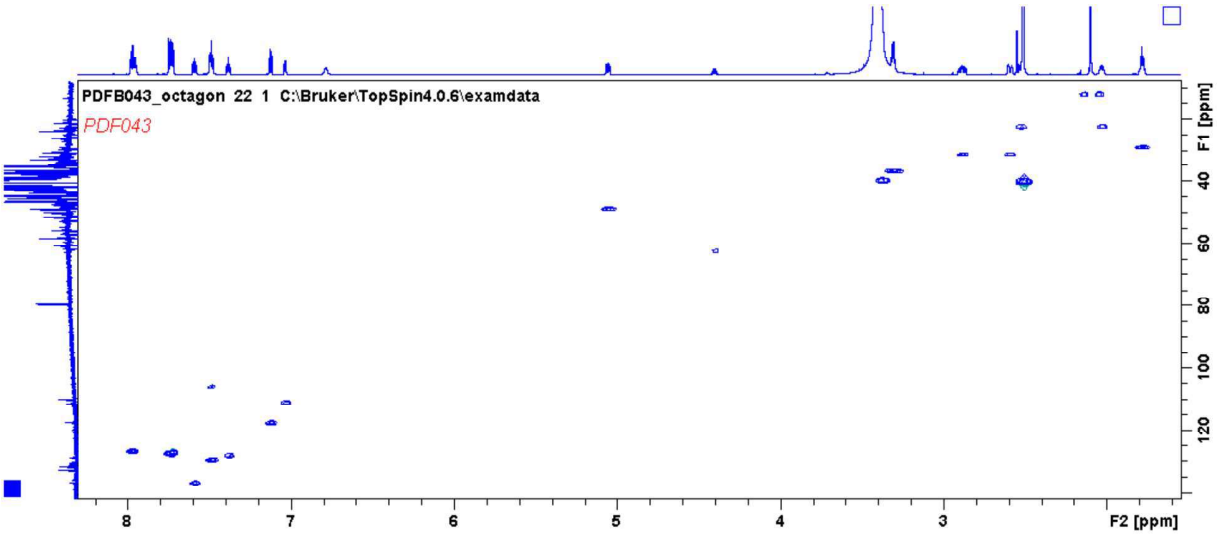
^{1}H NMR for d4E-5



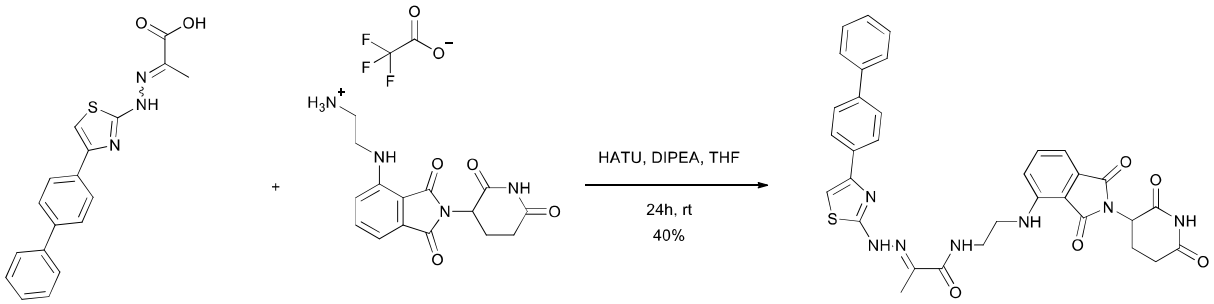
^{13}C NMR for d4E-5 (aromatics and carbonyl, aliphatic carbon peaks were assigned from HSQC)



^{1}H-^{13}C HSQC for d4E-5

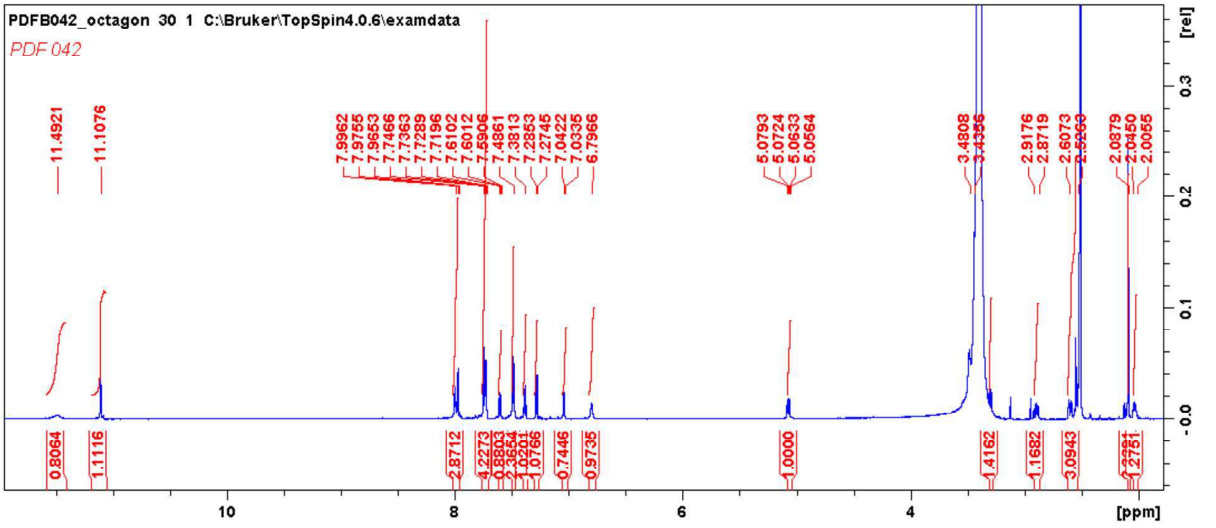


2-(2-(4-([1,1'-biphenyl]-4-yl)thiazol-2-yl)hydrazineylidene)-N-(2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino)ethyl)propenamide (d4E-4)

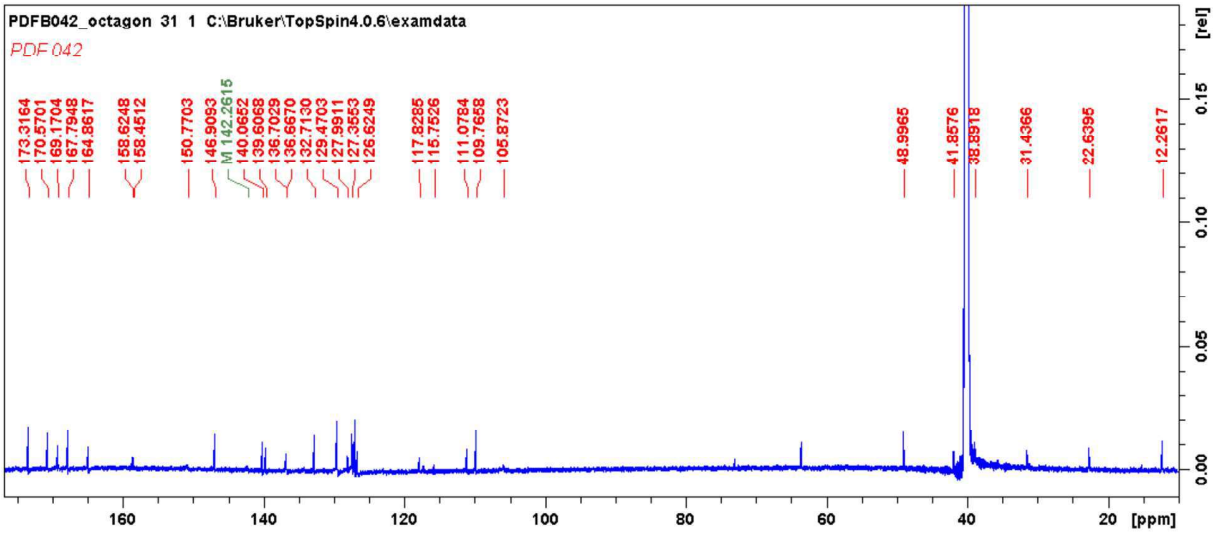


^{1}H-NMR (800 MHz, DMSO-d6): δ [ppm] = 2.01-2.05 (m, 1H), 2.09 (s, 3H), 2.53-2.61 (m, 2H), 2.87-2.92 (m, 1H), 3.43 (m, 2H), 3.48 (m, 2H), 5.06-5.05 (dd, 1H, $^{3}J_{HH}$ = 13.2 Hz, 5.9 Hz), 6.80 (s, 1H), 7.04 (d, 1H, $^{3}J_{HH}$ = 7.0 Hz), 7.28 (d, 1H, $^{3}J_{HH}$ = 8.7 Hz), 7.38 (t, 1H, $^{3}J_{HH}$ = 7.3 Hz), 7.48-7.50 (t, 2H, $^{3}J_{HH}$ = 7.9 Hz), 7.59-7.61 (dd, 1H, , $^{3}J_{HH}$ = 7.4 Hz, 6.7 Hz), 7.72-7.75 (m, 4H), 7.97 (d, 2H, $^{3}J_{HH}$ = 8.2 Hz), 8.00 (t, 1H, $^{3}J_{HH}$ = 5.9 Hz), 11.11 (s, 1H), 11.49 (s, 1H)
^{13}C-NMR (201 MHz, DMSO-d6): δ [ppm] = 12.26, 22.64, 31.44, 38.89, 41.86, 49.00, 105.87, 109.76, 111.08, 115.75, 117.83, 126.62, 127.36, 127.99, 129.47, 132.71, 136.67, 136.70, 139.61, 140.07, 142.26, 146.91, 150.77, 158.45, 158.62, 164.86, 167.96, 169.17, 170.57, 173.32
UPLC-MS: calc; [M +H]: 650.22 found [M + H]: 650.21

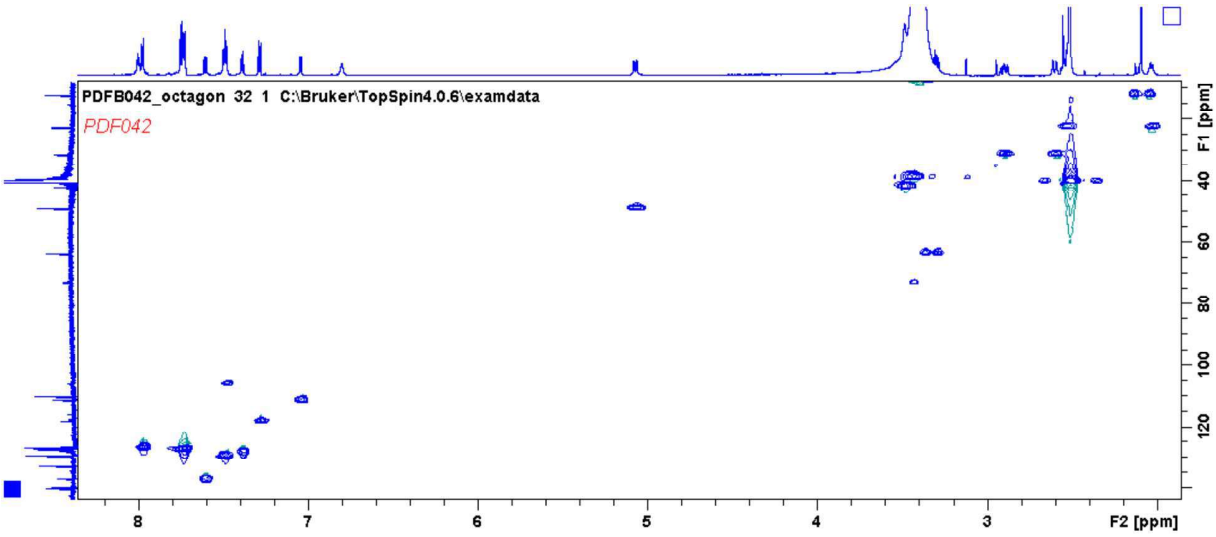
^{1}H NMR for d4E-4



^{13}C NMR for d4E-4



^{1}H-^{13}C HSQC for d4E-4



4.5 VirtualFlow Ants – Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization

Christoph Gorgulla, Süleyman Selim Çınaroğlu, Patrick D. Fischer, Konstantin Fackeldey, Gerhard Wagner, Haribabu Arthanari. (2021). VirtualFlow Ants – Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization, **IntJMolSci**, 22, 5807.

DOI: https://doi.org/10.3390/ijms22115807

URL: https://www.mdpi.com/1422-0067/22/11/5807

Summary:

This work presents the addition of the docking program PLANTS, which is a software based on ant colony optimization (ACO) algorithm, to VirtualFlow. In doing so, primary virtual screenings and rescoring policies are added to the open source platform presented in **4.3 An open-source drug discovery platform enables ultra-large virtual screens.** Additionally, ligand libraries will receive support in MOL2 format as well as on-the-fly conversion of libraries in PDBQT formation to MOL2 format to be used in VirtualFlow. Virtual Flow Ants was used on Keap1 as a test system, leveraging up to 128,000 CPUs on the Google Cloud, to reveal approximately linear scaling behavior. 10 million compounds for each of the 10 generated docking scenarios were used to fine tune docking parameters of PLANTS. Docking scores and average docking times were subsequently analyzed. Using Keap1 as a test system, a new compound (VANTS-1) was identified that displayed similarities to the previously published Keap1 ligands Cmp16 and iKeap1. Despite the structural commonalities, VANTS-1 possesses distinct differences from the other two compounds, making it an interesting candidate for potential new drug development efforts.

Author contributions:

As the author of this thesis, I evaluated and compared the ligands used for the test system Keap1. Together with Cristoph Gorgulla and Haribabu Arthanari, I evaluated the docked

model of ligand and protein and assessed the chemical functionalities of the aforementioned ligands. I assisted Christoph Gorgulla and Haribabu Arthanari in writing, reviewing and editing the paper.

International Journal of
Molecular Sciences

MDPI

Article

VirtualFlow Ants—Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization

Christoph Gorgulla [1,2,3,*], Süleyman Selim Çınaroğlu [4], Patrick D. Fischer [2,3,5], Konstantin Fackeldey [6,7], Gerhard Wagner [2] and Haribabu Arthanari [2,3,*]

1 Department of Physics, Harvard University, Cambridge, MA 02138, USA
2 Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA; PatrickD_Fischer@dfci.harvard.edu (P.D.F.); gerhard_wagner@hms.harvard.edu (G.W.)
3 Department of Cancer Biology, Dana Farber Cancer Institute, Boston, MA 02115, USA
4 Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK; suleyman.cinaroglu@bioch.ox.ac.uk
5 Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, 66123 Saarbrücken, Germany
6 Zuse Institute Berlin, 14195 Berlin, Germany; fackeldey@zib.de
7 Institute of Mathematics, Technical University Berlin, 10623 Berlin, Germany
* Correspondence: cgorgulla@g.harvard.edu (C.G.); hari@hms.harvard.edu (H.A.)

Citation: Gorgulla, C.; Çınaroğlu, S; Fischer P.D.; Fackeldey, K.; Wagner, G.; Arthanari, H. VirtualFlow Ants—Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization. *Int. J. Mol. Sci.* **2021**, 22, 5807. https://doi.org/10.3390/ijms22115807

Academic Editor: Alexandre G. de Brevern and Jean-Christophe Gelly

Received: 20 April 2021
Accepted: 14 May 2021
Published: 28 May 2021

Abstract: The docking program PLANTS, which is based on ant colony optimization (ACO) algorithm, has many advanced features for molecular docking. Among them are multiple scoring functions, the possibility to model explicit displaceable water molecules, and the inclusion of experimental constraints. Here, we add support of PLANTS to VirtualFlow (VirtualFlow Ants), which adds a valuable method for primary virtual screenings and rescoring procedures. Furthermore, we have added support of ligand libraries in the MOL2 format, as well as on the fly conversion of ligand libraries which are in the PDBQT format to the MOL2 format to endow VirtualFlow Ants with an increased flexibility regarding the ligand libraries. The on the fly conversion is carried out with Open Babel and the program SPORES. We applied VirtualFlow Ants to a test system involving KEAP1 on the Google Cloud up to 128,000 CPUs, and the observed scaling behavior is approximately linear. Furthermore, we have adjusted several central docking parameters of PLANTS (such as the speed parameter or the number of ants) and screened 10 million compounds for each of the 10 resulting docking scenarios. We analyzed their docking scores and average docking times, which are key factors in virtual screenings. The possibility of carrying out ultra-large virtual screening with PLANTS via VirtualFlow Ants opens new avenues in computational drug discovery.

Keywords: structure based virtual screening; molecular docking; swarm intelligence; artificial intelligence; computer aided drug design; CADD; KEAP1; drug discovery

1. Introduction

When a drug (ligand) enters the body, it typically engages a specific binding site on a chosen target protein and triggers the desired therapeutic effect. The tighter the ligand binds to the target protein, the higher its efficacy at much lower concentrations. This would require much lower concentrations of the drug, avoiding toxicity due to off target effects. Since the ligand as well as the target protein are not static in nature, but flexible and dynamic in solution, different conformations and orientations of the ligand–target protein complex are possible, and each of these conformations would result in different binding affinities (binding strengths). Hence, the process of computationally finding the optimal ligand–protein complex can roughly be divided into two steps: In the first step,

all possible spatial arrangements of the ligand and the target protein are enumerated. To this end, the two structures must be brought together and arranged in various different spatial arrangements; this process is called docking. However, not all orientations and conformations of the ligand–target complex lead to the same binding affinity or binding strength. Therefore, in a second step, a score is assigned to each spatial arrangement of the ligand and the protein. The spatial constellation with the lowest score (i.e., the free energy of binding) then represents the strongest binding of the ligand and the binding site of the target, in other words: the "best" fit. Assuming that one assigns a score for each possible arrangement of the ligand with respect to the protein, one obtains a scoring function or binding affinity function, which assigns a free energy of binding to each conformation. This binding affinity function can be considered as the ground truth whose absolute minimum constitutes the best fit of the protein–ligand complex. However, unfortunately, this function is hardly computable, even for small systems. Therefore scoring functions are used which try to approximate the "true" energy binding functions.

Analogous to the binding energy function, a scoring function assigns a value to each computed orientation/conformation scenario, a score which typically correlates with the binding energy. In the same way, the absolute minimum represents the constellation with the highest binding affinity. For most docking programs, this means that the more negative the docking score, the stronger the predicted binding affinity. Analogous to the energy binding function, one is interested in the absolute minimum of the scoring function. In the last decades, different methods have been developed, which mainly differ in the way of posing (ligand flexible/protein fixed, ligand flexible/protein flexible, ...), in the choice of degrees of freedom of the scoring function and in the search method for the minimum, among them are DOCK [1], AutoDock [2], GOLD [3] or FlexX [4]), to mention a few.

In mathematical terms, the scoring function of a given target–ligand complex described by the n degrees of freedom $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$, is an objective function $f : \mathbb{R}^n \to \mathbb{R}$. Hence, the docking problem can be expressed as a (global) optimization problem, i.e.,

$$\min_{x \in \mathbb{R}^n} f(x).$$

Summing up, finding the best arrangement of the protein and the ligand is equivalent to finding the (absolute) minimum of f. In the following section, we will treat this problem by taking advantage of agent based methods developed in the context of swarm intelligence.

2. Ant Intelligence in Molecular Docking with PLANTS

In agent-based models, the global behavior of a system is employed by local interactions between agents on a micro scale. The PLANTS (Protein-Ligand ANT System) [5–7] is a stochastic optimization method for non-covalent protein–ligand docking algorithms, which is based on ant colony optimization (ACO) [8,9].

ACO itself is inspired by the swarm intelligence of ants: Each ant deposits a substance (pheromone) that other ants sense. If an ant has several choices, it follows with a higher probability the way with the highest pheromone concentration. If a certain trail is shorter than other trails, more ants will use it in the same time interval, which leads to a higher pheromone concentration. Contrariwise, if a trail is not used for a longer time span, the concentration of pheromone decreases (evaporation). At the end, the whole ant colony has optimized its way. Setting this into the context of the protein–ligand docking, the shortest trail can be related to the structure with the lowest binding energy.

The ACO algorithm optimizes the trail between the ant nest (=initial position of the molecule) and the food source (=target position). ACO belongs to the class of agent-based swarm models, where the behavior of a complex system is investigated by interaction between agents (ants). Originally, the ant algorithm was used for combinatorial problems like finding the shortest path in a given network. Here, we have no network, but a function

and seek for its absolute minimum. Thus, in order to apply the ACO algorithm to our problem, we need a discretization for the degree of freedom and an iteration rule of the ants.

2.1. Discretization

Let us restate that the spatial orientation of the target–ligand complex is given by a n dimensional vector $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$ where each degree of freedom $x_i \in [a^i, b^i]$ is a real number. However, for applying the ACO, we can only permit finitely many values for each degree of freedom.

We therefore partition each interval $[a^i, b^i]$ by a finite series

$$a^i = z_i^0 < z_i^1 < .... < z_i^{n_i} = b^i$$

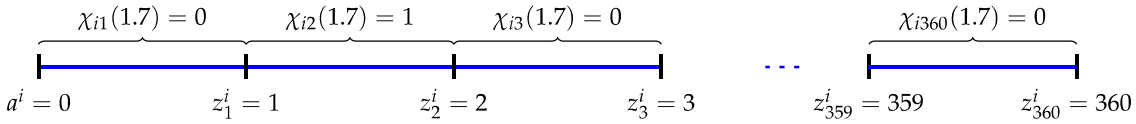
into n_i sub intervals, i.e.,

$$[a^i, b^i] = \bigcup_{\ell=1}^{n_i} [z_i^{\ell-1}, z_i^\ell].$$

By introducing the characteristic function

$$\chi_{i,\ell}(x_i) = \begin{cases} z_i^\ell & \text{if } x_i \in [z_i^{\ell-1}, z_i^\ell] \\ 0 & \text{otherwise.} \end{cases} \qquad \ell = 1, ..., n_i$$

we can assign to each real value x_i an interval value z_i^ℓ. For instance, let x_i be the rotational degree of freedom, then the partition $[0, 360] = [0, 1] \cup [1, 2] \cup ... \cup [359, 360]$ can be used. A value of 1.7 would be assigned to the second interval and $\chi_{i2}(1.7) = 1$



To date, we have just employed a discretization scheme for the degrees of freedom, such that the real valued degrees of freedom can be assigned to integers.

2.2. Iteration Rule of the Ants

The movement of the ants is represented as an iterative process. At time t, the ant a is in a certain position $x^a(t)$. Then, for each degree on freedom i, a local minimization by the simplex algorithm is employed, i.e.,

$$s_i^a = \text{argmin}_{x_i^a \in \mathbb{R}^n} f(x_i^a), \ i = 1, ..., n$$

such that s^a is a local minimizer of f in the iteration step t. In the original ant algorithm, the ants walk on a graph. An ant that is at time step t on node i then chooses the edge ℓ with probability

$$p_{i\ell}(t) = \frac{\tau_{i\ell}(t)}{\sum_{k=1}^{n_i} \tau_{ik}(t)}, \tag{1}$$

where $\tau_i^\ell(t)$ is the so called pheromone trail, which assigns a desirability of edge ℓ at time step t. In the context of PLANTS (1) is the probability to assign x_i to z_i^ℓ. With this, the pheromone trails τ_i^ℓ can be updated via

$$\tau_{i\ell}(t+1) = (1-\rho)\tau_{i\ell}(t) + \mathbf{1}_{i\ell}\Delta\tau_i,$$

where

$$\mathbf{1}_{i\ell} = \begin{cases} 1 & \text{if } s_i^a \in [z_i^{\ell-3}, z_i^{\ell+2}] \text{ and } x_i \text{ is rotational dof} \\ 1 & \text{if } s_i^a \in [z_i^{\ell-2}, z_i^{\ell+1}] \text{ and } x_i \text{ is not rotational dof} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\Delta\tau(s_i^{\mathsf{a}}) = \begin{cases} |f(x_i^{\mathsf{a}})| & \text{if } f(x_i^{\mathsf{a}}) < 0 \\ 1 & \text{if } s_i^{\mathsf{a}} \in [z_i^{\ell-2}, z_i^{\ell+1}] \text{ and } x_i \text{ is not rotational dof} \\ 0 & \text{otherwise.} \end{cases}$$

This principle is illustrated in Figure 1.

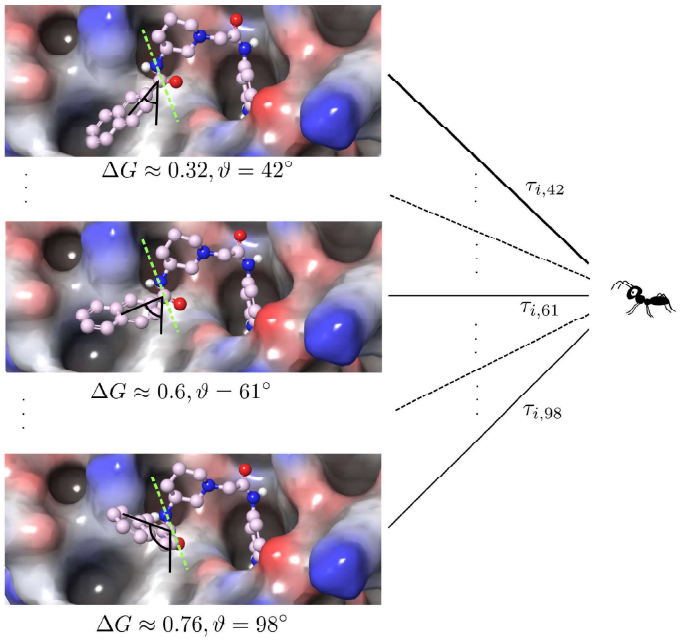


Figure 1. Sketch of a one the i th degree of freedom with $x_i = \vartheta \in [0, 360]$. The spatial arrangement with the angle $\vartheta = 42°$ has the lowest energy ($\Delta G \approx 0.32$), such that $\tau_{i,42}$ has the highest value and consequently the ant chooses the trail "42" with the highest probability.

For an upper and lower bound of the pheromone trails for each degree of freedom, we refer to [7]. In PLANTS, the candidate solutions are optimized by the simplex local search algorithm [10].

2.3. PLANTS Features for Molecular Docking

PLANTS provides many special features for molecular docking studies, as well as nature-inspired molecular docking algorithm. PLANTS contains three empirical scoring functions: PLP (piece-wise linear potential), PLP95, and ChemPLP [6,11,12]. Piece-wise linear potentials are used to model the steric complementarity between the protein and ligand in all of the three scoring functions. All scoring functions use potentials for van der Waals interactions and repulsive terms. An internal score of the ligand contains an empirical heavy-atom potential to avoid internal ligand clashes. Furthermore, the torsional potential from the Tripos force field [13] is calculated for rotatable bonds in the ligand. ChemPLP, in addition, contains angle-dependent hydrogen bond terms from GOLD's ChemScore [14,15]. ChemPLP is the default scoring function in PLANTS and also in GOLD. PLANTS allows one to modify and customize many intermolecular and intramolecular terms which are used by these three scoring functions. PLANTS cannot only be used in classical dockings and virtual screenings, it also allows to rescore compounds for given protein–ligand complexes. PLANTS uses the MOL2 file format for both the receptor and the ligands, having the advantage of being one of the most commonly used chemical file formats, which is able to store atom positions, connectivity, and arbitrary meta information in a single file. This information is extremely important for the identification of rotatable

bonds and functional groups in molecular docking. MOL2 is a standardized file format that can be read by other modeling programs.

Many features of the PLANTS docking program make it a powerful tool for molecular docking. With the fastest setting PLANTS can be used to economically sample vast chemical spaces. Additionally, PLANTS allows flexibility on the receptor side, i.e., on side chains of the amino acid residues in the protein. Even if no flexibility is specified, PLANTS partially applies the flexibility to the protein by optimizing the positions of hydrogen atoms. On the ligand side, PLANTS can also perform rigid-body docking, allowing for externally generated ligand conformations. It is also possible to perform rigid ligand docking with flexible protein side chains.

PLANTS has also been compared to other docking tools, among them are [16–20]. In [16], for instance PLANTS has been tested on the human cluster of differentiation 38 (CD38), in the ranking based on scoring power with 42 compounds PLANTS was in the upper class.

One of the challenges of high-throughput docking routines is the treatment of the effect of water and its contribution to the binding free energy. Before the ligand engages the protein, both the protein and the ligand are solvated. Upon binding, there will be changes in the solvation. It is easy to conceive that water molecules will be freed from the ligand upon binding the protein and this would result in an entropic gain. On the side of the protein, solvation is more complex. It has been shown that in the case of protein–small molecule interactions, one of the major entropic contributions to the binding free energy stems from desolvation of the small molecule upon binding, especially in case of small molecule harboring hydrophobic groups [21,22]. High-throughput docking routines traditionally use a version of an implicit water model which enables screening large libraries of compounds in reasonable time. However if the docking interface on the protein has bound water molecules with long resident times, and they could contribute significantly to the binding free. This contribution will not be appropriately captured by the implicit water mode. Sometimes, the water molecules bound to the protein might help for stabilizing the ligand–protein complex via mediating hydrogen-bonding interactions. In high-resolution crystal structures, typically with resolution better than 2.8 Å, we potentially observe highly ordered water molecules. Occasionally, these water molecules would aid connecting the protein to the ligand in the binding pocket [23]. PLANTS has features to include explicit water molecules during the molecular docking, which allows the water molecule to move by translation or rotation. It is also possible to fix all of the water molecule's degrees of freedom. This capability also allows to optimize hydrogen bond parameters of the water.

PLANTS can help lead optimization by scaffold hopping via restraining the position of a ring system or a non-ring atom in molecular docking. All degrees of freedom in the scaffold can be completely neglected using this approach. So, many different substituents on the scaffold can be screened. This is a useful feature for lead optimization since all manipulations in the use of scaffold hopping in medicinal chemistry generally involve cyclic systems [24].

While it is desired to have a high-resolution crystal structure of the ligand bound to the protein, there are several instances where we are unable to co-crystallize the protein–ligand complex or instances where we do not see the density for the ligand. In some of these instances, nuclear magnetic resonance (NMR) spectroscopy can provide information that can help docking routines. In an ideal case nuclear Overhauser effect (NOE), restraints between the ligand and protein would provide distance information to obtain the co-structure. However, in many cases, we might not be able obtain these constraints due to a number of reasons, including the broadening of the ligand signals due to increase in size or exchange kinetics or the absence of hydrogen atoms on the ligand at the binding interface. In the absence of well defined NOE restraints, there are a series of ligand-detected NMR experiments like trNOE [25], saturation transfer difference (STD) [26], DIRECTION [27] and INPHARMA [28] experiments that can provide qualitative information on pharmacophore

mapping. PLANTS also allows molecular docking with NMR constraints derived from ligand-detected NMR experiments [29,30]. PLANTS can use these experimental data to evaluate the poses and to identify the pose fitting the experimental constraints. This approach can be helpful in case of having highly flexible ligands since it may be very difficult to predict all degrees of freedom for these ligands. Thus, the experimental data can be a part of the docking algorithm to help finding correct binding poses of the ligands with more reliable predictions.

Using molecular interaction fingerprints (IFP) can even increase the accuracy in molecular docking using PLANTS. PLANTS has an ability to generate IFP by running PLANTS in rescore-mode with one or multiple ligands for a given protein structure. After IFP identification, more robust structure-based virtual screening can be performed using these IFP, which can be set as a constraint. Multiple IFP files can be specified during the docking. In addition to these constraints, many other constraints can be specified in molecular docking with PLANTS [31]. The docking predictions will be influenced by the constraints, including hydrogen bond constraint, shape constraint and distance constraints.

3. VirtualFlow Ants—Virtual Screenings Using Ant Intelligence via PLANTS

We have added support to the docking program PLANTS in our recently developed drug discovery platform VirtualFlow [32,33]. This new feature is called VirtualFlow Ants; it also adds support for a new ligand preparation routine involving the MOL2 format, to enable screening using PLANTS. We have tested the scaling behavior of VirtualFlow Ants using a protein Kelch-like ECH-associated protein 1 (KEAP1) as a model system, and studied how varying some of the key parameters in the docking program PLANTS affects the results of the virtual screenings with our test system.

3.1. Ligand Preparation and Chemical File Formats

The virtual screening module of Virtual Flow, VirtualFlow for Virtual Screening (VFVS), previously only supported ligands in the PDBQT format [34,35] because all previously supported docking programs were to use this chemical file format and some of them only supported this single file format like AutoDock Vina or QuickVina 2. PLANTS, however, requires the ligands to be in the MOL2 format. To meet this requirement, we have added two different implementations. First, we added the support of ligands in the MOL2 format to VirtualFlow Ants. The overall structure of the input ligand libraries has to be in the canonical VirtualFlow format with multiple hierarchical levels as described previously [32], with the only difference in the chemical file format (MOL2 instead of PDBQT). A new parameter *ligand_library_format* was added to the VirtualFlow Ants version of the VFVS control file which specifies the parameters of the virtual screening. The new parameter can have the values "PDBQT" or "MOL2". Alternatively, the ligands can also be in the PDBQT format, in which case VirtualFlow Ants converts the ligands into the MOL2 format on the fly during the virtual screening. This mode allows to use previously prepared ligand libraries which are in the VirtualFlow PDBQT format, such as the Enamine REAL library containing 1.4 billion compounds which we had previously prepared [32]. The on the fly conversion of the ligands to MOL2 format happens in two steps. In the first step, Open Babel is used to convert the ligand from the PDBQT format to the PDB format, retaining the tautomerization states of the ligands [36,37]. In the second step, SPORES is used with the *completepdb* mode to prepare the MOL2 version of the molecule [38,39]. SPORES prepares ligands and proteins separately, which makes straightforward the input file preparation for virtual screening. During the preparation, it assigns atom and bond types as well as the protonation states using atom and bond information. Additionally, it can generate multiple protonation and tautomeric states, as well as multiple stereoisomers. The complete workflow of the VirtualFlow Ants module can be seen in Figure 2.

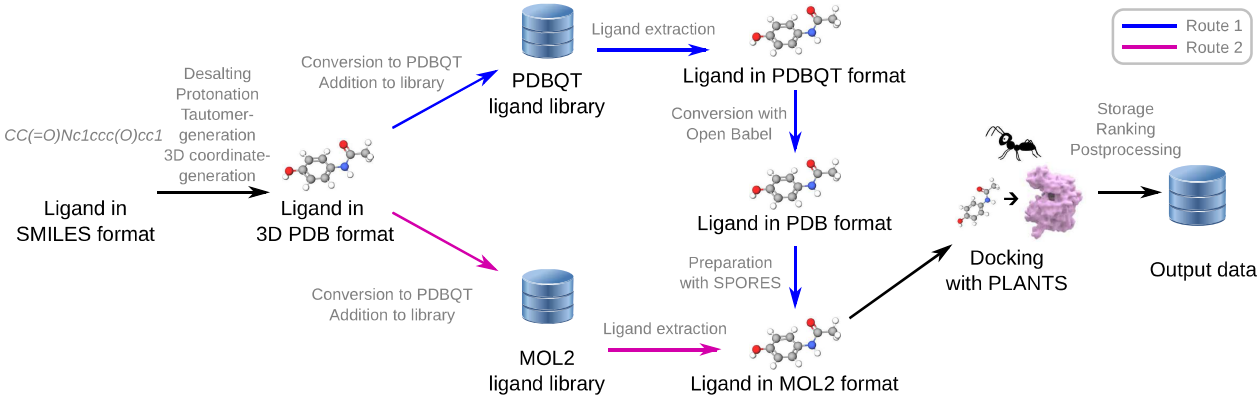


Figure 2. Overall ligand workflow in VirtualFlow Ants. Ligand libraries can be provided in two formats, firstly the PDBQT format and secondly the MOL2 format. New libraries in either format can be prepared with VirtualFlow for Ligand Preparation (VFLP). If the ligand library is provided in the PDBQT format (Route 1 in blue), then VirtualFlow Ants will convert and prepare the molecules on the fly into the MOL2 format (with Open Babel and SPORES) which is required by PLANTS. If the ligand library is provided in the MOL2 format, then no on the fly conversion is required (Route 2 in magenta), and the ligand can simply be extracted from the ligand library during the virtual screens.

The on the fly conversion is very fast (typically a fraction of a second), and therefore does not increase the virtual screening time significantly.

3.2. I/O and File Management

Docking using PLANTS generates a relatively large number of output files for each individual molecule, which is virtual screens multiplied by the number of ligands screened. Many of these files can be classified as log files, while others contain the results. As with earlier versions of VirtualFlow, the results files and the log files in VirtualFlow Ants are stored in separate final output folders. All files which contain any of the following regular expressions: `.*log`, `.*constraints.*`, `.*correspondingNames.*`, `.*descent.*`, `.*optimizer.*`, `.*plantsconfig.*`, and `.*skippedligands.*`. All other files are treated as results files, which includes the docking pose file and the ranking (docking score) files. It is recommended to disable unnecessary output files of PLANTS, such as the writing of protein conformation, in case the protein is rigid and thus identical for all the ligands.

All the processes regarding the on the fly ligand preparation with SPORES and Open Babel, as well as docking with PLANTS, are executed on the local compute nodes and use their memory and local disks if specified in the control file to reduce the I/O for the shared cluster file system.

3.3. Configuration and Set Up of VirtualFlow Ants

VirtualFlow Ants is mostly setup in a way similar to how VirtualFlow is setup in general. The main configuration file of VirtualFlow Ants is therefore also the *control file*. Regarding the ligand library format, a new parameter *ligand_library_format* was added to the control file. The new parameter can have the values "PDBQT" or "MOL2". If the value is "PDBQT", then the ligands will be converted on the fly from the PDBQT to the MOL2 format as described above.

For the docking scenario which is to be deployed during the virtual screening, one folder has to be created in the `input-files` folder, and contain a file with the filename `config.txt` as the main configuration file for the PLANTS docking program. All path names which are specified in this file have to be relative to the `tools` folder. Protein receptor structures are stored in the folder `input-files/receptors`.

PLANTS can also be used in a consensus docking procedure within VirtualFlow Ants. This can be done by setting up multiple docking scenarios, which are all carried out for each ligand during the virtual screening. Any number of the supported docking programs and scoring functions of VirtualFlow can take part in the consensus scoring besides PLANTS (e.g., AutoDock Vina [34], QuickVina 2 [35], ADRF [40], QuickVina-W [41], Smina Vinardo [42,43], VinaXB [44], or Vina-Carb [45]), which can be specified in the control file in the section defining the docking scenarios. During the virtual screening, the docking score of each ligand is stored for each docking scenario. After the virtual screening is completed, the consensus score can be obtained by computing the average value of the individual docking scores.

The dedicated web page for VirtualFlow Ants was created (https://virtual-flow.org/virtualflow-ants, accessed 14 May 2021), which describes this new feature to potential users and lists all available additional resources, including documentation.

3.4. Test System

As a biomolecular test system for the benchmark studies of VirtualFlow Ants the protein KEAP1 was selected, and as the target site the nuclear factor erythroid 2-related factor 2 (NRF2) binding site. NRF2 orchestrates cellular response to oxidative stress by upregulating detoxifying and antioxidant defense genes [46]. In normal physiological conditions, NRF2 is degraded with the help of KEAP1, an E3 ubiquitin ligase substrate adaptor protein that directly engages NRF2 [47]. The half-life of NRF2 is around 25 min. In response to oxidative, inflammatory, and metabolic stress, key cysteine residues in KEAP1 are covalently modified, preventing KEAP1 from engaging NRF2. This free NRF2 translocates to the nucleus and activates its transcriptional program of approximately 250 genes [48]. Dysregulation of the NRF2 pathway is implicated in a number of diseases, including cancer, diabetes, autoimmune disorders, neurodegenerative disease, metabolic syndrome and diseases of the gastrointestinal tract [49]. Activating NRF2 by inhibiting the protein–protein interaction between KEAP1 and NRF2 is a promising therapeutic strategy pursued by several pharmaceutical companies.

For KEAP1, multiple structures of high quality are available. The structure that we used has PDB ID 5fnq with the water molecules removed. This structure was selected because it has a relatively high resolution of less than two Angstroms, and no heavy atoms were missing. The structure was prepared with SPORES into the MOL2 file format. The docking sphere used by PLANTS was specified to have a radius of 13 Å, and its center is positioned at coordinates $(x, y, z) = (17\,\text{Å}, 66\,\text{Å}, 31\,\text{Å})$ (see Figure 3).

3.5. Scaling Behavior

To verify the ability of VirtualFlow Ants to scale to a large number of CPUs, we have tested the scalability in the Google Cloud (https://cloud.google.com/, accessed 14 May 2021). VirtualFlow Ants was run with the KEAP1 as the test system (see above). The docking parameters of PLANTS were set to the default values, except for the speed which was set to *speed4*.

Regarding the ligand library used for this benchmark, we have created a test library of a total of 1.44 billion compounds. This library contains 10 meta-tranches, and each meta-tranche contains 1000 tranches. Each tranche contains 1000 collections, and each collection consists of an identical set of 144 different compounds. Each of the 144 compounds has a molecular weight between 400 and 425 daltons and between three and five rotatable bonds. The ligands were in the PBDQT format, and therefore the newly implemented on the fly conversion from the PDBQT format into the MOL2 format.

VirtualFlow Ants was run on the Google Cloud using an autoscaling Slurm cluster. The shared cluster filesystem which was deployed was an Elastifile filesystem, which is a high-performance shared network filesystem. The compute nodes of the Slurm cluster were of the type n2d-highcpu-64 (AMD Epyc Rome processors), with external IPs disabled. VirtualFlow was run on 100 compute nodes (using 6400 vCPUs), 500 compute nodes

(32,000 vCPUs), 1000 nodes (64,000 vCPUs), and 2000 nodes (128,000 vCPUs). As can be seen in Figure 4, the scaling behavior is virtually linear with respect to the number of vCPUs used.

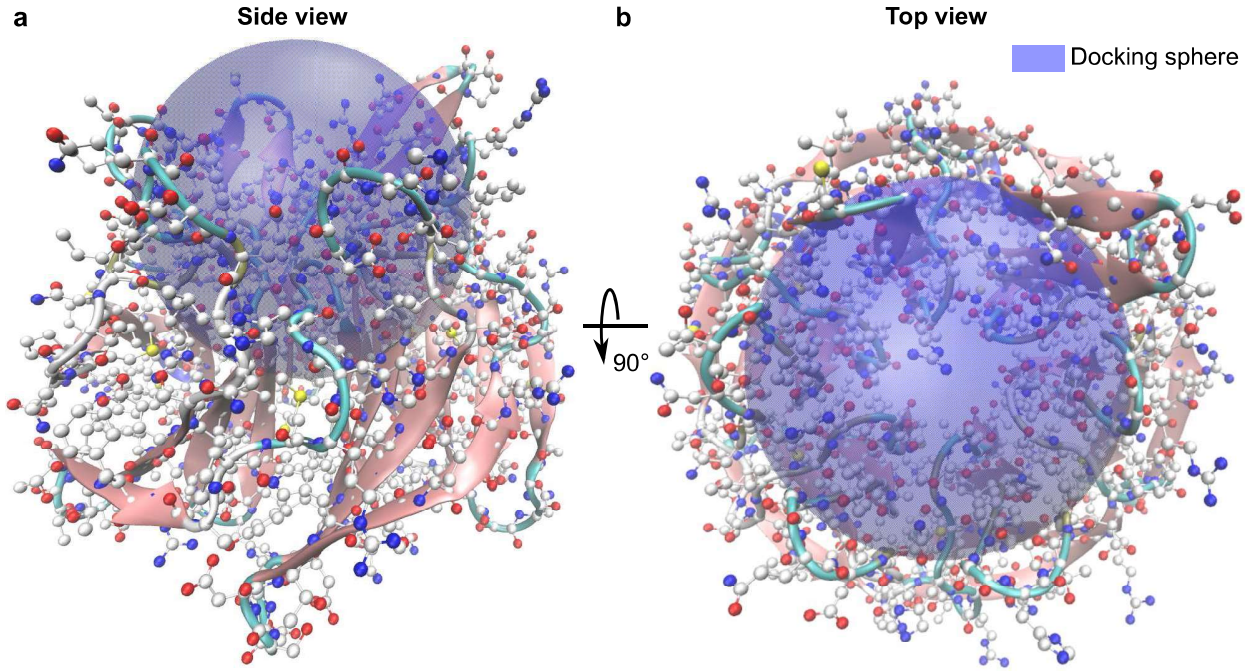


Figure 3. The target protein in the test system, which was chsen to be KEAP1. The target site is the NRF2 protein–protein interaction interface on KEAP1 (PDB ID 5fnq). The side view (**a**) and the top view (**b**) of KEAP1 are shown. The docking sphere of PLANTS which was used in all test runs (blue sphere) was centered at the NRF2 binding interface on KEAP1 and has a diameter of 13 Å.

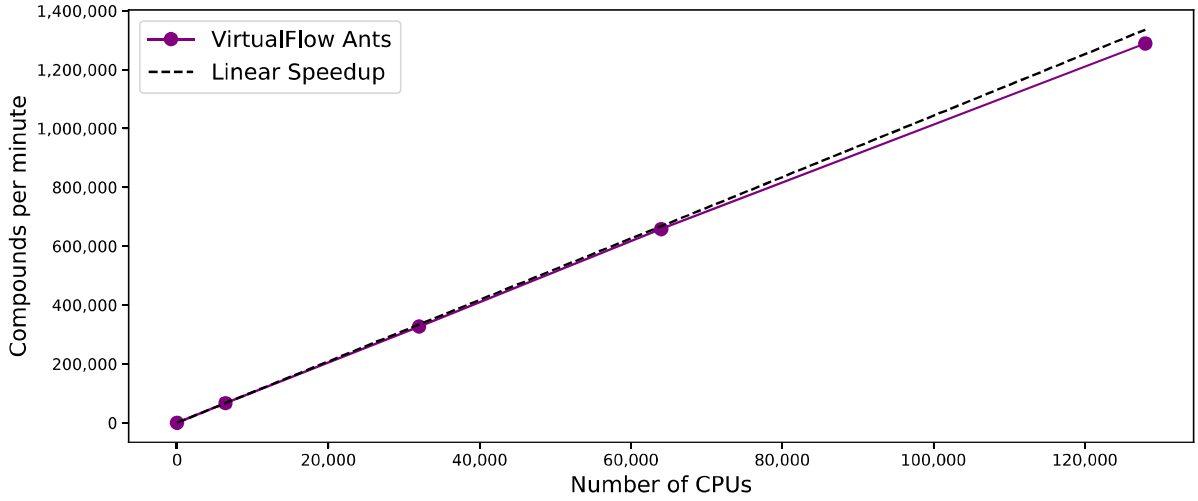


Figure 4. Scaling behavior of VirtualFlow Ants with respect to the number of vCPUs. VirtualFlow Ants was run on the Google Cloud using 6400, 32,000, 64,000 and 128,000 vCPUs. Ligands were in the PDBQT format, and were converted into the MOL2 format required by PLANTS on the fly with Open Babel and SPORES. The number of ligands docked (y-axis) scales approximately linearly with the number vCPUs (purple line). The dashed black line indicates a perfectly linear scaling behavior.

3.6. Parameter Variation

To study the effects of the docking parameters on the virtual screening, we have varied their values and created eight different docking scenarios, which all use the ChemPLP scoring function so that the values can be compared with each other. These eight docking scenarios are listed in Table 1.

Table 1. In this table, the docking scenarios are listed which were included in the parameter variation benchmark studies. The default values which PLANTS uses for the parameters `search_speed`, `aco_ants`, and `aco_sigma` depend on the values of the parameters `scoring_function` and `search_speed`. the values highlighted in light blue are the values which were actively set in the configuration file of PLANTS. The remaining values (white background) are the default values which were set by PLANTS. For each of these docking scenarios, approximately 10 million compounds were screened.

Docking Scenario	scoring_Function	Search_Speed	aco_Ants	aco_Eevap	aco_Sigma
1	chemplp	4	default (20)	default (0.15)	default (0.25)
2	chemplp	2	default (20)	default (0.20)	default (0.5)
3	chemplp	1	default (20)	default (0.20)	default (1.25)
4	chemplp	4	10	default (0.15)	default (0.25)
5	chemplp	4	50	default (0.15)	default (0.25)
6	chemplp	4	default (20)	0.10	default (0.25)
7	chemplp	4	default (20)	0.25	default (0.25)
8	chemplp	4	default (20)	default (0.15)	1
9	plp	4	default (20)	default (0.2)	default (0.5)
10	plp95	4	default (20)	default (0.2)	default (1.25)

For each of these docking scenarios, we have screened approximately 10 million compounds from the Enamine REAL library with a molecular weight between 450 and 500 daltons. The input library was in the PDBQT format, and therefore the ligands were converted on the fly into the MOL2 format. The distribution of the docking scores of the top 100 compounds can be seen in Figure 5.

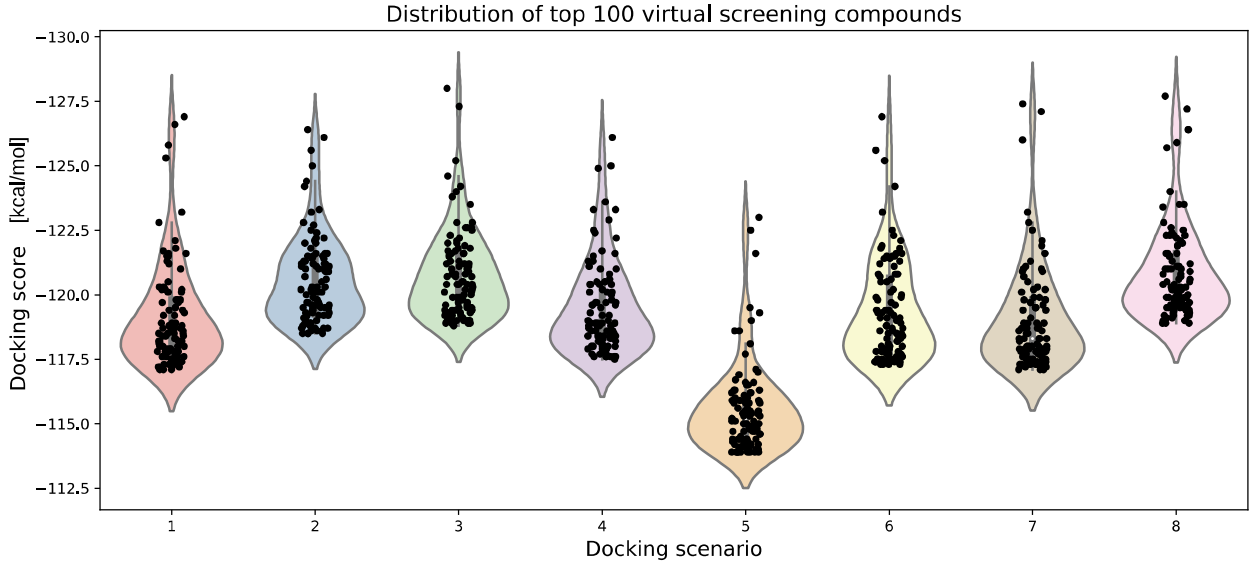


Figure 5. Violin plots of the docking scores of the top 100 ranking compounds for the eight different docking scenarios which use the chemplp scoring function. For each of the eight docking scenarios, approximately 10 million compounds with a molecular weight between 450 and 500 daltons from the Enamine REAL library were screened.

As could be expected, with slower speed, the average docking time per ligand increases, and the increase in time is approximately linear relative to the speed parameter

(see Figure 6). The accuracy of the docking only marginally decreases with the docking speed, as the average docking scores of the top 100 compounds (AVE100) of docking scenario 1 (speed 4), docking scenario 2 (speed 2) and docking scenario 3 (speed 1) are −118.5, −120.1, and −120.4, respectively, (see Figure 5). The more negative the docking score, the tighter the predicted binding. Reducing the number of ants to 10 seems to slightly improve the results (AVE100 = −118.9). Increasing the number of ants to 50 has reduced the AVE100 value to −115.5. Alternation of the `aco_evap` parameter to 0.10 and 0.25 (docking scenarios 6 and 7) had minimal effects on the docking scores of the top 100 ranking compounds (AVE100 of −118.75 and −118.3). In docking scenario 8, the `aco_sigma` value was increased to 1, which slightly increased the docking scores of the top 100 compounds (−120.2) when compared with the default settings (−120.2).

In virtual screenings, an important aspect is the speed of the docking programs, and this depends on the various docking parameters used. We have therefore obtained the average docking times per ligand for each of the ten docking scenarios of Table 1 (see Figure 6). Only docking scenarios 1 to 8 are included, because docking scenarios 9 and 10 use a different scoring function, and thus the docking scores cannot be directly compared with each other.

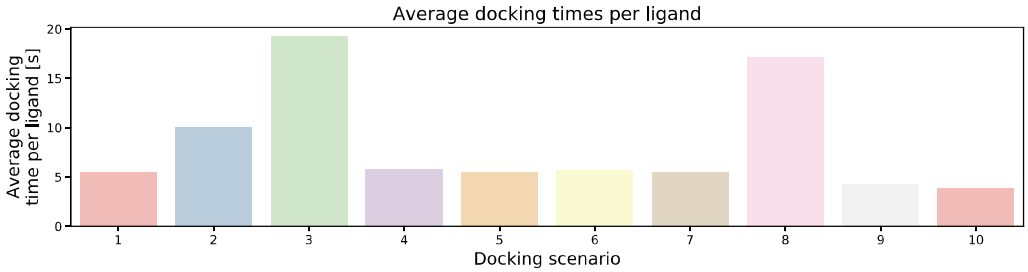


Figure 6. Bar graph showing the average docking time per ligand of the test set, which was used in the above parameter variation benchmark for each of the ten docking scenarios.

The fastest two docking scenarios were docking scenarios 9 and 10, in which the plp and plp95 scoring functions were used, respectively. Regarding the docking scenarios 1 to 8 in which the ChemPLP scoring function were used. The docking parameter `search_speed`, which can have the values 1, 2, and 4, indeed sped the docking times up by a factor approximately represented by the corresponding value.

4. Discussion

In this study, we docked 10 million compounds to the NRF2 binding interface of KEAP1 and compared several docking scenarios where we varied the docking parameters. The top scoring compounds (rank 1) of the benchmark virtual screen involving docking scenario 1 (named VANTS-1) is shown in Figure 7a. This compound has striking similarities with two previously published compounds, Cmp16, experimentally identified by the pharmaceutical company Biogen Idec [50] (Figure 7b), and iKeap1, the inhibitor from our previously published paper [32] (Figure 7c). The predicted binding mode of compound VANTS-1 (Figure 7d) is very similar to the experimentally confirmed binding modes of Cmp16 (Figure 7f) and the docking pose iKeap1 (Figure 7e).

The three molecules, VANTS-1, Cmp16 and iKEAP1, can be compared using their structural motifs, or pharmacophores: All three molecules share an aromatic core structure (depicted in blue in Figure 7), which serves as a mirror plane, along which all molecules are symmetrical. In the case of of iKEAP1, this pharmacophore is comprised of a tricyclic, heteroaromatic pyrazino[2,3-b]quinoxaline, which is simplified to a bicyclic naphthalene in Cmp16. The new molecule VANTS-1 has a further simplification by replacing one aromatic ring with *tert*-butyl as a lipophilic bioisostere. In a symmetrical way, all compounds branch out from the aromatic system using a linker functionality (depicted in red, Figure 7). In the case of iKEAP1, this linker is represented by sulfonates. In the case of Cmp16, sulfonamides

are serving the same function, while additionally gaining the ability to provide hydrogen donor functionalities. In VANTS-1, the linker is comprised of 2-oxo-ethyl benzoate linkers, which provide greater flexibility while retaining the hydrogen bond acceptor functionalities found in the original linkers. The last pharmacophore is represented by a cyclic system (depicted in green), Figure 7. Both iKeap1 and Cmp16 harbor aromatic features here in the form of toluene and anisole, respectively. In VANTS-1, this part of the molecule is where the most dramatic change in pharmacophores is found, with a non-aromatic 2-oxopyrrolidine functionality. While having lost its aromaticity compared to the other two molecules, the new moiety in VANTS-1 is still lipophilic. Additionally, the hydrogen acceptor functionality of the anisole moiety in Cmp16 is conserved in the new 2-oxopyrrolidine moiety. The 10 million compounds we screened here are a part of the Enamine REAL library and iKEAP1 belongs to the NCI collection and Cmp16 was from an experimental effort. Hence, we docked iKEAP1 and Cmp16 using PLANTS and we obtained a similar docking score for all the three compounds (VANTS-1 −126.1; Cmp16 −106.4 ; iKeap1 −107.6). Even though the compound VANTS-1 is similar to Cmp16 and iKeap1, the scaffold is different, and can be highly valuable in drug discovery efforts.

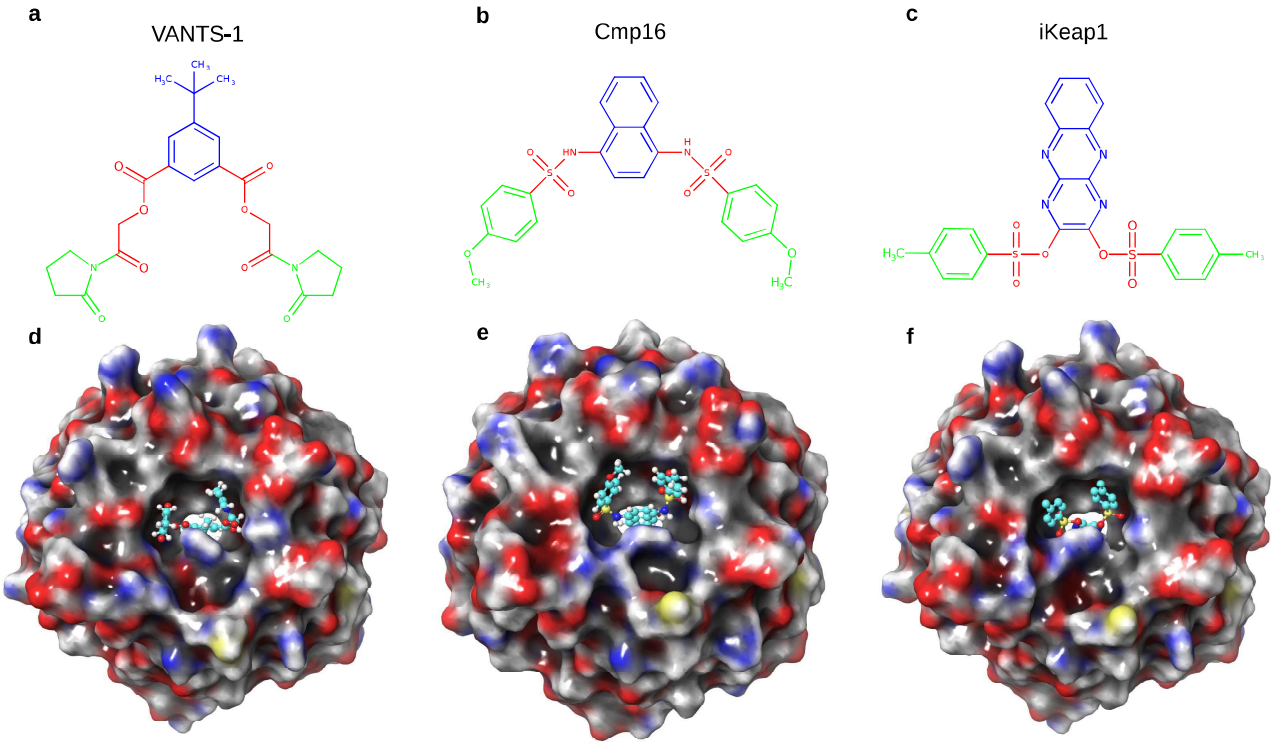


Figure 7. (**a**) Compound VANTS-1 (**b**) the top scoring compound of the 10 million compound screen with docking scenario 1. (**b**,**c**) Compounds Cmp16 and iKeap1, which are similar to compound VANTS-1, and which are experimentally confirmed inhibitors of KEAP1 published previously [32,50]. (**d**) the docking pose of VANTS-1 obtained via PLANTS in docking scenario 1. (**e**,**f**) crystal structure of Cmp16 (PDB ID 4iqk) [50], and docking pose of iKeap1 [32].

PLANTS will be a great addition to the arsenal of docking programs that VirtualFlow supports. The default setting in PLANTS works well and the speed is conducive for screening ultra-large virtual screening. With the speed setting of 4, one can potentially screen one billion compounds in approximately 15 h day when using 100,000 CPUs. It should be noted that the precise docking time depends on the size of the docking box, the target protein, and other docking parameters. We have noted that the on the fly conversion from PDBQT to MOL2 does not increase the screening time significantly. Since we have already prepared the entire Enamine REAL library in a ready to dock PDBQT format, this

on the fly conversion to MOL2 format can be adapted for other docking programs such as GOLD [11,51,52], GLIDE [53–55] and LeDOCK [56]. In addition to ultra-large scale screenings PLANTS will be an useful tool for second-stage re-scoring utilizing features like explicit water molecules, protein flexibility and IFP. In challenging cases where a crystal structure is not feasible, PLANTS can be used to include experimental constraints to obtain a docked model of the ligand to the protein, which can be further used to perform virtual medicinal chemistry.

Author Contributions: Conceptualization, C.G., K.F. and H.A.; methodology, K.F., S.S.Ç. and C.G.; software, C.G. and S.S.Ç.; validation, C.G., H.A. and P.D.F.; resources, C.G., H.A. and G.W.; data curation, C.G.; writing—original draft preparation, K.F., S.S.Ç. P.D.F. and C.G.; writing—review and editing, H.A. and G.W.; visualization K.F. and C.G.; supervision, H.A. and G.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NIH grants CA200913, AI037581 and GM129026. This research was supported in part by the ARO Grant W911NF1910302 to Arthur Jaffe. H.A. acknowledges funding from the Claudia Adams Barr Program for Innovative Cancer Research. P.D.F. would like to acknowledge the Chleck Foundation for providing a doctoral fellowship.

Acknowledgments: We thank Google for computing time on the Google Cloud, ChemAxon for a free academic license for the JChem and Marvin Sketch packages. K.F. would like to thank MATH+. We would like to thank Julia and Nina Fackeldey for designing the ant in Figures 1 and 2. We would like to thank Brigitte Klein for proofreading the manuscript, and Arthur Jaffe for their support. We thank Joachim Ansorg for a free license for the BashSupport Pro plugin for JetBrains' IDEs.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. G.W. and C.G. are co-founders of the company Virtual Discovery, Inc., which provides virtual screening services. G.W. serves as the director of this company.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism
AVE100	Average docking scores of the top 100 ranking compounds
IFP	Interaction fingerprints
STD	Saturation transfer difference
ACO	Ant colony optimization
KEAP1	Kelch-like ECH-associated protein 1
NRF2	Nuclear factor erythroid 2-related factor 2

References

1. Ewing, T.J.A.; Makino, S.; Skillman, A.G.; Kuntz, I.D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411–428. [CrossRef]
2. Goodsell, D.S.; Lauble, H.; Stout, C.D.; Olson, A.J. Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins Struct. Funct. Bioinform.* **1993**, *17*, 1–10. [CrossRef]
3. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230–252. [CrossRef] [PubMed]
4. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489. [CrossRef] [PubMed]
5. Korb, O.; Stützle, T.; Exner, T.E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Ant Colony Optimization and Swarm Intelligence*; Dorigo, M., Gambardella, L.M., Birattari, M., Martinoli, A., Poli, R., Stützle, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 247–258.
6. Korb, O.; Stützle, T.; Exner, T.E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96. [CrossRef]

7. Korb, O.; Stützle, T.; Exner, T.E. An Ant Colony Optimization Approach to Flexible Protein-Ligand Docking. *Swarm Intell.* **2007**, *1*, 115–134. [CrossRef]
8. Dorigo, M.; Birattari, M.; Stützle, T. Ant Colony Optimization. *Comput. Intell. Mag. IEEE* **2006**, *1*, 28–39.. [CrossRef]
9. Dorigo, M.; Stützle, T. *Ant Colony Optimization*; MIT Press: Cambridge, MA, USA, 2004.
10. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [CrossRef]
11. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein–ligand docking using GOLD. *Proteins Struct. Funct. Bioinform.* **2003**, *52*, 609–623. [CrossRef]
12. Verkhivker, G.M. Computational analysis of ligand binding dynamics at the intermolecular hot spots with the aid of simulated tempering and binding free energy calculations. *J. Mol. Graph. Model.* **2004**, *22*, 335–348. [CrossRef]
13. Clark, M.; Cramer, R.D., III; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012. [CrossRef]
14. Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425–445. [CrossRef] [PubMed]
15. Murray, C.W.; Auton, T.R.; Eldridge, M.D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput. Aided Mol. Des.* **1998**, *12*, 503–519. [CrossRef]
16. Boittier, E.D.; Tang, Y.Y.; Buckley, M.E.; Schuurs, Z.P.; Richard, D.J.; Gandhi, N.S. Assessing Molecular Docking Tools to Guide Targeted Drug Discovery of CD38 Inhibitors. *Int. J. Mol. Sci.* **2020**, *21*, 5183. [CrossRef]
17. Ballante, F.; Marshall, G.R. An Automated Strategy for Binding-Pose Selection and Docking Assessment in Structure-Based Drug Design. *J. Chem. Inf. Model.* **2016**, *56*, 54–72. [CrossRef] [PubMed]
18. Çınaroğlu, S.S.; Timuçin, E. Comparative Assessment of Seven Docking Programs on a Nonredundant Metalloprotein Subset of the PDBbind Refined. *J. Chem. Inf. Model.* **2019**, *59*, 3846–3859. [CrossRef]
19. Ren, X.; Shi, Y.S.; Zhang, Y.; Liu, B.; Zhang, L.H.; Peng, Y.B.; Zeng, R. Novel Consensus Docking Strategy to Improve Ligand Pose Prediction. *J. Chem. Inf. Model.* **2018**, *58*, 1662–1668. [CrossRef]
20. Poli, G.; Martinelli, A.; Tuccinardi, T. Reliability analysis and optimization of the consensus docking approach for the development of virtual screening studies. *J. Enzym. Inhib. Med. Chem.* **2016**, *31*, 167–173. [CrossRef]
21. Kawasaki, Y.; Freire, E. Finding a better path to drug selectivity. *Drug Discov. Today* **2011**, *16*, 985–990. [CrossRef]
22. Freire, E. Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today* **2008**, *13*, 869–874. [CrossRef]
23. Huang, N.; Shoichet, B.K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51*, 4862–4865. [CrossRef]
24. Hu, Y.; Stumpfe, D.; Bajorath, J. Recent advances in scaffold hopping: Miniperspective. *J. Med. Chem.* **2017**, *60*, 1238–1246. [CrossRef]
25. Fejzo, J.; Lepre, C.A.; Peng, J.W.; Bemis, G.W.; Murcko, M.A.; Moore, J.M. The SHAPES strategy: An NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **1999**, *6*, 755–769. [CrossRef]
26. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.:12<1784::AID-ANIE1784>3.0.CO;2-Q. [CrossRef]
27. Mizukoshi, Y.; Abe, A.; Takizawa, T.; Hanzawa, H.; Fukunishi, Y.; Shimada, I.; Takahashi, H. An Accurate Pharmacophore Mapping Method by NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2012**, *51*, 1362–1365. [CrossRef] [PubMed]
28. Sánchez-Pedregal, V.M.; Reese, M.; Meiler, J.; Blommers, M.J.J.; Griesinger, C.; Carlomagno, T. The INPHARMA Method: Protein-Mediated Interligand NOEs for Pharmacophore Mapping. *Angew. Chem. Int. Ed.* **2005**, *44*, 4172–4175. [CrossRef]
29. Korb, O.; Möller, H.M.; Exner, T.E. NMR-Guided Molecular Docking of a Protein–Peptide Complex Based on Ant Colony Optimization. *ChemMedChem* **2010**, *5*, 1001–1006. [CrossRef] [PubMed]
30. Onila, I.; ten Brink, T.; Fredriksson, K.; Codutti, L.; Mazur, A.; Griesinger, C.; Carlomagno, T.; Exner, T.E. On-the-Fly Integration of Data from a Spin-Diffusion-Based NMR Experiment into Protein–Ligand Docking. *J. Chem. Inf. Model.* **2015**, *55*, 1962–1972. [CrossRef] [PubMed]
31. Korb, O.; Exner, T. *Protein-Ligand ANT System User Manual for Version 1.2*; Universität Konstanz.
32. Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.F.; Fischer, P.D.; Coote, P.W.; Padmanabha Das, K.M.; Malets, Y.S.; Radchenko, D.S.; Moroz, Y.S.; Scott, D.A.; et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **2020**, *580*, 663–668. [CrossRef] [PubMed]
33. Gorgulla, C.; Fackeldey, K.; Wagner, G.; Arthanari, H. Accounting of Receptor Flexibility in Ultra-Large Virtual Screens with VirtualFlow Using a Grey Wolf Optimization Method. *Supercomput. Front. Innov.* **2020**, *7*, 4–12. [CrossRef]
34. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef] [PubMed]
35. Alhossary, A.; Handoko, S.D.; Mu, Y.; Kwoh, C.K. Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics* **2015**, *31*, 2214–2216. [CrossRef] [PubMed]
36. The Open Babel Package. 2020. Available online: http://openbabel.org (accessed on 15 March 2021).
37. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]

38. ten Brink, T.; Exner, T.E. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein-Ligand Docking Results. *J. Chem. Inf. Model.* **2009**, *49*, 1535–1546. [CrossRef]
39. ten Brink, T.; Exner, T.E. pK_a based protonation states and microspecies for protein–ligand docking. *J. Comput. Aided Mol. Des.* **2010**, *24*, 935–942. [CrossRef] [PubMed]
40. Ravindranath, P.A.; Forli, S.; Goodsell, D.S.; Olson, A.J.; Sanner, M.F. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput. Biol.* **2015**, *11*, e1004586. [CrossRef]
41. Hassan, N.M.; Alhossary, A.A.; Mu, Y.; Kwoh, C.K. Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. *Sci. Rep.* **2017**, *7*, 15451. [CrossRef]
42. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904. [CrossRef] [PubMed]
43. Quiroga, R.; Villarreal, M.A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS ONE* **2016**, *11*, e0155183. [CrossRef]
44. Koebel, M.R.; Schmadeke, G.; Posner, R.G.; Sirimulla, S. AutoDock VinaXB: Implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina. *J. Cheminform.* **2016**, *8*, 27. [CrossRef]
45. Nivedha, A.K.; Thieker, D.F.; Makeneni, S.; Hu, H.; Woods, R.J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.* **2016**, *12*, 892–901. [CrossRef]
46. Yonchuk, J.G.; Foley, J.P.; Bolognese, B.J.; Logan, G.; Wixted, W.E.; Kou, J.P.; Chalupowicz, D.G.; Feldser, H.G.; Sanchez, Y.; Nie, H.; et al. Characterization of the Potent, Selective Nrf2 Activator, 3-(Pyridin-3-Ylsulfonyl)-5-(Trifluoromethyl)-2 H -Chromen-2-One, in Cellular and In Vivo Models of Pulmonary Oxidative Stress. *J. Pharmacol. Exp. Ther.* **2017**, *363*, 114–125. [CrossRef]
47. Pallesen, J.S.; Tran, K.T.; Bach, A. Non-covalent Small-Molecule Kelch-like ECH-Associated Protein 1–Nuclear Factor Erythroid 2-Related Factor 2 (Keap1–Nrf2) Inhibitors and Their Potential for Targeting Central Nervous System Diseases. *J. Med. Chem.* **2018**, *61*, 8088–8103. [CrossRef]
48. Davies, T.G.; Wixted, W.E.; Coyle, J.E.; Griffiths-Jones, C.; Hearn, K.; McMenamin, R.; Norton, D.; Rich, S.J.; Richardson, C.; Saxty, G.; et al. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein–Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J. Med. Chem.* **2016**, *59*, 3991–4006. [CrossRef] [PubMed]
49. Cuadrado, A.; Rojo, A.I.; Wells, G.; Hayes, J.D.; Cousin, S.P.; Rumsey, W.L.; Attucks, O.C.; Franklin, S.; Levonen, A.L.; Kensler, T.W.; et al. Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases. *Nat. Rev. Drug Discov.* **2019**, *18*, 295–317. [CrossRef] [PubMed]
50. Marcotte, D.; Zeng, W.; Hus, J.C.; McKenzie, A.; Hession, C.; Jin, P.; Bergeron, C.; Lugovskoy, A.; Enyedy, I.; Cuervo, H.; et al. Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorg. Med. Chem.* **2013**, *21*, 4011–4019. [CrossRef]
51. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748. [CrossRef] [PubMed]
52. Verdonk, M.L.; Chessari, G.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Nissink, J.W.M.; Taylor, R.D.; Taylor, R. Modeling Water Molecules in Protein-Ligand Docking Using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515. [CrossRef]
53. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [CrossRef]
54. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [CrossRef]
55. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [CrossRef] [PubMed]
56. Zhang, N.; Zhao, H. Enriching screening libraries with bioactive fragment space. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 3594–3597. [CrossRef] [PubMed]

5 Final discussion

5.1 Publication A: ^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) are difficult targets due to their inherent dynamic behavior and evasion of traditional structural biology methods such as X-ray crystallography and cryo-EM. This work expanded the contingent of solution NMR experiments to study IDPs. By leveraging the enhanced resolution of nitrogen-detected NMR experiments, a suite of triple resonance backbone experiments comprised of the hCBCACON, hcaCOCAN, hCACON, HBHACBCAN, HBHACBCAcoN and hCcaCON-TOCSY experiments was developed that significantly improve backbone assignment techniques of IDPs. Using the regulatory domain of the transcription factor nuclear factor of activated T-cells (NFAT), the principal workflow of these experiments and their application to solve biological problems was presented. It was shown how nitrogen-detected experiments help overcome difficulties in the assignment of overlapping peaks in NMR spectra of IDPs, due to the characteristic shallow chemical shift distribution of ~1 ppm in the proton dimension of these proteins. Furthermore, the development of nitrogen-detected experiments enabled the assignment of proline residues, which are not accessible by traditional proton-detected NMR experiments due to the lack of NH-resonances in their backbone. It was additionally shown that, despite of the reduced intrinsic sensitivity of ^{15}N nuclei compared with ^{13}C nuclei, the slower relaxation rates of nitrogen nuclei compensate for this fact. Therefore, the nitrogen-detected experiments can provide the same information that carbon-detected experiments for IDPs developed by others[362–365] can. The ^{15}N-detected experiments presented in this publication allowed for the assignment of 100 % of resonances in NFAT$_{128\text{-}394}$, including prolines. This enabled the investigation of a fundamental question regarding NFAT activation: How is the nuclear localization signal (NLS) within the regulatory domain of NFAT masked while NFAT is phosphorylated, and how does it become accessible once

dephosphorylation has occurred? By using the newly developed assignment techniques, two phosphorylation sites addressed by protein kinase A (PKA) were assigned to S225 and S255. Analysis of ^{15}N-labeled phosphorylated and unphosphorylated NFAT in the presence of unlabeled 14-3-3 allowed the localization of a 14-3-3 binding site on NFAT that overlaps with the NLS. Importantly, this binding was only observed when NFAT was phosphorylated by PKA. Therefore, sequestering of 14-3-3 by dephosphorylation was established to be the mechanism of making the NLS accessible for nuclear translocalization of NFAT.

5.2 Publication B: Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics

Muraymycins are one of many classes of nucleoside antibiotics that inhibit the bacterial membrane-bound enzyme MraY (translocase I). Along with other uridine-derived natural products (caprazamycins, capuramycins, mureidomycins, and liposidomycins), muraymycins are potent MraY inhibitors that therefore represent interesting candidates for the development of novel antibiotics. So far, structure activity relationship (SAR) studies on muraymycins have focused on their antibacterial effects in cells, by comparing minimal inhibitory concentrations (MIC) for example. However, this methodology does not allow comparison of isolated effects directly related to MraY inhibition, as the antibacterial activity of a compound is a function of its cellular uptake, target interaction, intracellular stability, potential off-target interactions and efflux effects. Therefore, the work presented in this publication established an SAR study on representatives of the muraymycin family and their inhibitory effects measured in MraY activity assays. Direct comparison of IC_{50} values obtained from these activity measurements was performed using a variety of different MraY preparations (i.e. purified *Aquifex aeolicus* MraY and *Clostridium bolteae* MraY in DM micelles as well as *Staphylococcus aureus* MraY and *Clostridium bolteae* MraY in crude membranes of overexpressing cells). Muraymycins can be subdivided into four families A-D, each of which was represented by at least one compound in this publication. All muraymycins are comprised of a typically 5'-*O*-aminoribosyl-functionalized (5'*S*,6'*S*)-glycyluridine (GlyU) core structure, connected by

an alkyl linker to a peptide moiety. The presented publication enabled the comparison of individual functional elements of the four muraymycin subgroups and pointed out differences in inhibitory activities depending on the purification method or isoform of MraY that was used. The work revealed strong MraY inhibition of muraymycins with IC_{50} values in the low pm range. By comparing naturally occurring muraymycins with a synthetic 5'-defunctionalized analogue, it was concluded that the 5'-aminoribosyl moiety significantly contributes to inhibitory activities. On the other hand, 2'-*O*-methylation of this unit was determined to have no effect on MraY inhibition. The fatty acid moiety that is part of muraymycins from the A and B subgroups, but not the C and D subgroups, was found to have an insignificant effect on enzyme inhibition. Therefore, it was concluded that this functionalization likely plays a role in cellular uptake of these nucleoside antibiotics. While significant differences between different MraY preparation methods or isoforms were noted, the general SAR trend between individually tested compounds remained similar. Lastly, a comparison of direct inhibitory data of the tested muraymycins with their antibacterial effects (i.e. MIC values) revealed no direct correlation, thereby indicating the involvement of different uptake and/or efflux mechanisms for individual compounds.

5.3 Publication C: An open-source drug discovery platform enables ultra-large virtual screens

The development of new drugs to combat diseases is a time- and money-consuming endeavor. This is, in part, because expensive laboratory experiments are conducted, often over a period of many years, to find an effective drug candidate. This publication presented an open-source platform that is capable of screening billions of compounds for targets of which structural information is available, thus improving on traditional virtual high-throughput screenings by an order of magnitude. With a higher number of molecules screened in *in silico* screenings, the number of true hits increases significantly.[355] Using the protein-protein interaction between the nuclear factor erythroid-derived 2-related factor 2 (Nrf2) and Kelch-like ECH-associated protein 1 (Keap1), over 1 billion compounds were screened and their effects were evaluated in

orthogonal experiments. The Nrf2-Keap1 pathway has been shown to play a critical role in a variety of diseases including cancer, autoimmune diseases and neurodegenerative diseases, for example. Of over 1 billion screened compounds evaluated in rigid docking methods, ~ 3 million compounds were selected for *in silico* rescoring by flexible docking, from which ~ 10,000 hits were determined. Using an average of approximately 8,000 cores on a heterogeneous Linux cluster, these screening efforts were accomplished in approximately 4 weeks. From the ~10,000 hits, the top 590 compounds were ordered and experimentally validated. Molecules were subdivided based on their ability to displace a fluorescently labeled Nrf2 peptide from Keap1 (displacers) and those that exhibited binding affinity without displacing the peptide (binders). Hit molecules were further refined by SPR, BLI and NMR methods to quantify their binding affinities. Ultimately, multiple molecules were identified that inhibited the Nrf2-Keap1 interaction in the low nM range.

5.4 Publication D: A biphenyl inhibitor of eIF4E targeting an internal binding site enables the design of cell-permeable PROTAC-degraders

The master regulator of eukaryotic translation initiation eIF4E has long been considered to be involved in the development and negative assessments of cancers. Elevated levels of the mRNA cap-binding protein eIF4E in tumors emphasize the crucial role that protein translation plays in oncogenic cell growth. In this publication, the high-resolution X-ray crystallography structure of eIF4E with a novel inhibitor bound to a new cavity was presented. Based on a previously described small molecule inhibitor, 4EGI-1, the new molecule i4EG-BiP was designed and found to bind to a new binding fold. The work presents the comparison of binding affinities of i4EG-BiP and 4EGI-1 as well as a suite of cellular data analyzing the cytotoxic effect of the new compound in different cancer cell lines. Furthermore, a dual-luciferase assay was used to distinguish cap-dependent translation from cap-independent translation. It is shown how i4EG-BiP inhibits cap-dependent translation initiated by eIF4E in a concentration-dependent manner.

Leveraging the structural information of the two molecules 4EGI-1 and i4EG-BiP, so-called PROteolysis TArgeting Chimeras (PROTACs) were developed to induce targeted degradation of eIF4E. PROTACs are heterobifunctional molecules in which a small molecule ligand is conjugated to an E3 ligase ligand using a linker. The recruitment of the E3 ligase into proximity of the target facilitates target ubiquitination and subsequent proteasomal degradation. Improving on previously reported eIF4E PROTACs[250], the compounds developed in this publication were shown to be cell-permeable. By either generating a prodrug PROTAC of 4EGI-1 in which the carboxyl moiety is POM-protected or using the carboxyl moiety of i4EG-BiP to connect the linker and the thalidomide unit, the disadvantage of negative charges on cellular uptake was abolished. Even though no cellular degradation of eIF4E was observed in the presence of these PROTACs, the results presented in this work enabled a better understanding of inhibitory mechanisms of eIF4E: It was shown that binding of either 4EGI-1 or i4EG-BiP induces a conformational change in a helix of eIF4E that is involved in the binding of interaction partners. The observed extension of helix α1 induced by binding of the small molecules likely interferes with the bulky amino acids of eIF4G that are involved in binding this part of eIF4E. Without binding of the scaffold protein eIF4G, the translation initiation complex eIF4F cannot assemble and cap-dependent protein translation is stalled. 4EBP, an eIF4E-regulating protein, was shown to bind to the same interface. However, our results show that binding to 4EBP is enhanced by the extension of helix α1, thus contributing to the negative regulation of eIF4E. Furthermore, it was shown by others that binding of eIF4E to 4EBP blocks Chip-dependent ubiquitination of eIF4E[199], thus likely contributing to the negative results found regarding proteasomal degradation of eIF4E by PROTACs.

5.5 Publication E: VirtualFlow Ants – Ultra-Large Virtual Screenings with Artificial Intelligence Driven Docking Algorithm Based on Ant Colony Optimization

As a direct addition to VirtualFlow, presented in **4.3 An open-source drug discovery platform enables ultra-large virtual screens**, VirtualFlow Ants will be of great value in the structure-

assisted discovery of ligands for challenging drug targets. Using Keap1 as a test system, the publication directly connects to the important Nrf2-Keap1 interaction explained in **2.6 Keap1**. The identification of novel hits will open up new avenues in the intelligent design of drug candidates for targets involved in a variety of diseases.

5.6 Outlook

The results published on MraY inhibition by multiple members of the muraymycin family, in addition with published structural information by others[77–79], helped to create an SAR model for the inhibition of MraY. Interestingly, it was also found that the preparation method and isoform of the proteins significantly influences the results of activity assays. Future experiments will be conducted to compare different purification methods of MraY (i.e. micelles and nanodiscs) with overexpressed MraY in crude membranes. Also, MraY isoforms in identical preparations will be compared to elucidate the effects of species-dependent enzyme inhibition.

In addition, preliminary data presented in the appendix of this work further highlights the potential issues with different protein preparations. While it was shown that ligand binding evaluated by ITC with MraY stabilized in DM micelles can suffer from unspecific detergent interactions and that micellar MraY is prone to aggregation in MST experiments at elevated temperatures, it was also concluded that enzymatic activity of MraY in nanodiscs rapidly declines as a function of time. The production of overexpressed MraY in crude membranes on the other hand results in varying enzymatic activities, which can be hard to reproduce. Therefore, methods will be developed to quantify MraY in all preparation methods to calculate specific activities which will enable an unambiguous comparison between different methods.

Initial experiments presented in the appendix of this work have shown the importance of specific residues in the MraY sequence and the possibility of resistance development by point mutation at the target. Studies will be conducted that focus on the replacement of the first histidine of the well-conserved ‘HHH’-motif in MraY with glutamine, which is found in muraymycin-producing *Streptomyces sp.* MraY. Inhibition analysis of muraymycins, muraymycin analogues and muraymycin precursors with MraY from

Streptomyces platensis ($MraY_{SP}$), $MraY_{SP}$ in which the glutamine is back-mutated to histidine, MraY from *Staphylococcus aureus* ($MraY_{SA}$) and $MraY_{SA}$ in which the histidine is mutated to glutamine will provide insights into the possibility of *Streptomyces sp.* self-resistance by point mutation at the target. MraY isoforms from *Streptomyces platensis* will also be purified and subjected to structural analysis to further understand the effects of this specific point mutation.

In additional work, an interaction between MraY and the soluble enzyme MurF was found and confirmed using biochemical methods. The impact of this interaction on MraY inhibition was shown when MurF was able to restore the activity of muraymycin A1-inhibited MraY. Preliminary results using negative stain electron microscopy pave the way to solve the structure of the MraY-MurF complex using cryo-EM, which will help to understand the influence that MurF has on MraY inhibition. This protein-protein interaction can then be used to screen for a new class of molecules that inhibit MraY-MurF complex formation and thereby possibly interfere with the biosynthetic pathway for peptidoglycan synthesis.

The methods developed in this work led to new NMR experiments to improve the study of IDPs. The hCBCACON, hcaCOCAN, hCACON, HBHACBCAN, HBHACBCAcoN and hCcaCON-TOCSY triple resonance backbone experiments helped with the assignment of 100 % of the regulatory domain of NFAT, thus enabling the study of its interactions with the chaperone protein 14-3-3. The presented experiments can further be improved to allow their application for other proteins, not restricted to IDPs. For example, the constant-time periods used are replaceable by selective pulses on ^{13}C alpha carbons, thus minimizing signal loss due to relaxation. These pulse sequences would significantly enhance sensitivity for most amino acids whose C-α chemical shifts are well separated from their C-β chemical shifts (all amino acids other than glycine, serine and threonine). Furthermore, the presented 3D experiments can readily be transformed into 4D experiments, in which the fourth dimension could be protons attached to C-α or aliphatic carbons. Also, the conversion of the 3D experiments into 2D experiments without loss of resolution could present an addition to classical ^{15}N-^{1}H-HSQC chemical shift perturbation experiments to make use of the sharp line widths or study effects on prolines. The ability to leverage information gained from ^{15}N-detected experiments to assign phosphorylation

sites of PKA was shown in the publication. Elucidation of a broader set of phosphoproteomes of IDPs and other proteins by ^{15}N-detection is a future application of this method that will significantly improve our understanding of this posttranslational modification.

The preliminary experiments conducted on the transactivation domain of NFAT showed a phosphorylation-dependent change in protein dynamics. Future experiments will aim to collect data that will allow the calculation of a structural ensemble of NFAT conformations and compare the unphosphorylated transactivation domain to the phosphorylated version at Ser23. Combination of this data with interaction studies of the TAD with mediator and scaffold proteins will allow a detailed understanding of the relevance of this posttranslational modification. Furthermore, efforts made within this work have generated a construct to study the next phosphorylation of NFAT induced by p38 MAPK: Ser110. In similar fashion, an NFAT$_{1\text{-}130}$ S23A construct will be used to study this phosphorylation site independently. This phosphorylation site will likely influence NFAT activation, since it is directly preceding the first calcineurin binding motif. Relaxation experiments probing the protein's dynamics and studies involving purified calcineurin will help understand the importance of the second p38 MAPK target.

Lastly, full-length GFP-tagged NFAT was expressed in mammalian cells to study the effects of introducing PP2A binding motifs into NFAT's intrinsically disordered N-terminal domain. It was shown that subcellular localization of NFAT can be influenced depending on which binding motif of PP2A is incorporated into the protein. More experiments will be conducted that will investigate the effects on transcriptional activity caused by these incorporations by using IL-2 luciferase reporter gene assays. Also, additional binding motifs that allow the protein's activation by other triggers such as pH, oxidative species or light will be employed to further expand the possibility of creating a biomolecular tool for protein manipulation.

A new inhibitor of eIF4E was identified and structurally characterized in the context of this work. Also, synthetic efforts were made to create PROTACs that target eIF4E for proteasomal degradation. Sadly, this effort did not result in the desired degradation of eIF4E in a cellular context. However, insights gained from these studies will be used for future developments of eIF4E-targeting molecules. The shortcomings of the PROTACs

developed against eIF4E are likely a result of the increased 4EBP binding connected with binding of the small molecule. Therefore, eIF4E-targeting molecules that do not cause this conformational change will likely show improved results, assuming they are cell-permeable. In addition, the presented inhibitor molecule i4EG-BiP represents a lead structure to improve inhibition of eIF4E using this binding site. As the co-crystal structure shows, the addressed binding pocket of i4EG-BiP offers additional space to accommodate more functionalities, and thus expanding the small molecule's structure might improve eIF4E inhibition.

Next, preliminary results presented in the appendix show the effects of eiF4E phosphorylation on its structure. eIF4E phosphorylated at Ser209 shows minimal changes in protein structure, retaining the fold of wild-type eIF4E. Thus, the resulting phenotype observed for eIF4E phosphorylation is likely not based on its structure. However, the structure of uncapped eIF4E pS209 will be solved to investigate effects of phosphorylation prior to cap binding events. Furthermore, differences in binding affinities of phosphorylated eIF4E and wild-type eIF4E were analyzed for a suite of different cap analogues. Interestingly, it was found that the identity of the first transcribed nucleotide has a significant effect on the discrepancy between binding affinities of the phosphorylated or unphosphorylated protein. Therefore, future experiments will focus on understanding how the identity of the first nucleotide in mRNA transcripts might affect the phenotype of eIF4E phosphorylation.

In the appendix of this thesis, experiments are presented that aim to decipher the substrate specificity of AHL-synthases. Solution NMR experiments leveraging chemical shift perturbations and cross-saturation transfer have enabled the mapping of the binding interface between 3-oxo-C6i-ACP and EsaI as well as C8i-ACP and BmaI1. Further NMR experiments will be conducted that will probe the dynamics of loaded and unloaded ACP to identify changes in flexibility of amino acids upon substrate loading. In addition, substrate specificity will further be investigated using C6i-CoA as a non-native substrate for both AHL-synthases to create a clear picture of acyl-ACP/AHL-synthase interactions. Also, the influence of SAM and the order of assembly of the four subunits (ACP, acyl-CoA, SAM and AHL-synthase) will be investigated. Together, these experiments will shed light

on the dynamic interplays of quorum sensing mechanisms employed by Gram-negative bacteria.

Preliminary data on another protein, GILZ, is presented in another part of the appendix. Here, it was shown that GILZ forms large homooligomeric complexes with alpha-helical secondary structures. Circular dichroism and negative stain electron microscopy were used to eliminate the possibility of aggregation as a reason for the observed apparent high molecular weight of the purified protein and foundations were laid to solve its structure. Additionally, an interaction between GILZ and HDAC7 was shown using immunoprecipitation assays and the binding interface on HDAC7 was reduced to the first 219 amino acids of the protein. Further experiments will be carried out to improve the conditions of the immunoprecipitation and to narrow down the binding interface with respect to GILZ. Therefore, truncated GILZ constructs will be used and immunoprecipitated with truncated HDAC7 constructs to find the minimal binding interface. Additional insights into the interaction can be gained by employing X-ray crystallography or NMR spectroscopy to study the two interacting domains individually. Lastly, the Keap1-Nrf2 interaction was used to develop an ultra-large virtual high-throughput screening software called VirtualFlow. It was shown that screenings of more than 1 billion compounds were carried out in 4 weeks to yield several nanomolar inhibitors of the Keap1-Nrf2 interaction. Future experiments will be conducted to solve the structure of Keap1 in complex with these inhibitors based on the preliminary experiments presented in the appendix.

Moreover, VirtualFlow has been used to tackle one of the greatest health crises in modern humanity: the COVID-19 pandemic. Using the software, a total of more than 40 billion compounds were screened against 17 different proteins of either SARS-CoV2 or host proteins involved in viral infection. To date, several of the hits found in these screenings have been experimentally confirmed to target viral proteins and be effective in cellular infectivity assays. In addition, multiple approved FDA drugs that are used in clinical trials for COVID-19 without knowledge about their mechanisms of action have been confirmed to target viral proteins by our preliminary experiments. Experimental validation of found hits will be carried out in the hope to find effective therapeutic candidates to treat COVID-19. The work on SARS-CoV-2 screens has been published in Publication **F: A multi-**

pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening.

Overall, the work presented in this dissertation led to publications on multiple challenging drug targets such as NFAT, eIF4E, MraY and Keap1-Nrf2. The publications on these provided insights into the biology of these targets, as well as generated knowledge about inhibition of them using small molecules. Furthermore, additional work on the targets NFAT, eIF4E, MraY, Keap1, ACP and GILZ was performed, leading to the accumulation of data providing new information about these targets, which will ultimately lead to more publications in these fields.

6 Appendix

6.1 MraY

6.1.1 Interaction of nucleoside antibiotics with MraY in different environments

6.1.1.1 Solubilized MraY in detergent micelles

An effort was made to express and purify MraY homologues according to the protocols provided by CHUNG et al. and HAKULINEN et al.[76,78] Eventually, $MraY_{AA}$ as well as $MraY_{CB}$ were prepared in pure solubilized form, stabilized in decyl-*β*-D-maltopyranoside (DM) micelles. Compared to the crude membranes, solubilized proteins have the advantage of being preparations of enzyme without impurities that may influence the ligand-target interaction. However, detergents do not represent a native membrane protein environment and effects caused by ligand interaction between detergent molecules and empty micelles cannot be excluded.

Notably, inhibitors were generally less potent against solubilized protein compared to the same homologue overexpressed in crude membranes (i.e. $MraY_{CB}$, see Publication B, **4.2 Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**). Muraymycins from the A, B and D series displayed a ca. 10-fold decrease in inhibitory potency, while the loss of efficacy for muraymycin C1 was ca. 5-fold.

The difference in inhibitory activities might be explained by the presence of the non-denaturing detergent DM, which could hinder the necessary structural rearrangements of the protein needed for effective enzymatic inhibition by nucleoside antibiotics. Also, the dynamic interplay between other parts of crude membrane fractions with MraY and inhibitors might play a significant role in the differentiation between potency towards differently prepared proteins.

A less elaborated explanation might be the amount of enzyme present in the assay. Purification and solubilization of the enzyme in micelles can produce misfolded protein

incapable of converting the substrates into dansylated lipid I whilst at least partially retaining the ability to bind the inhibitors. Consequently, higher concentrations of

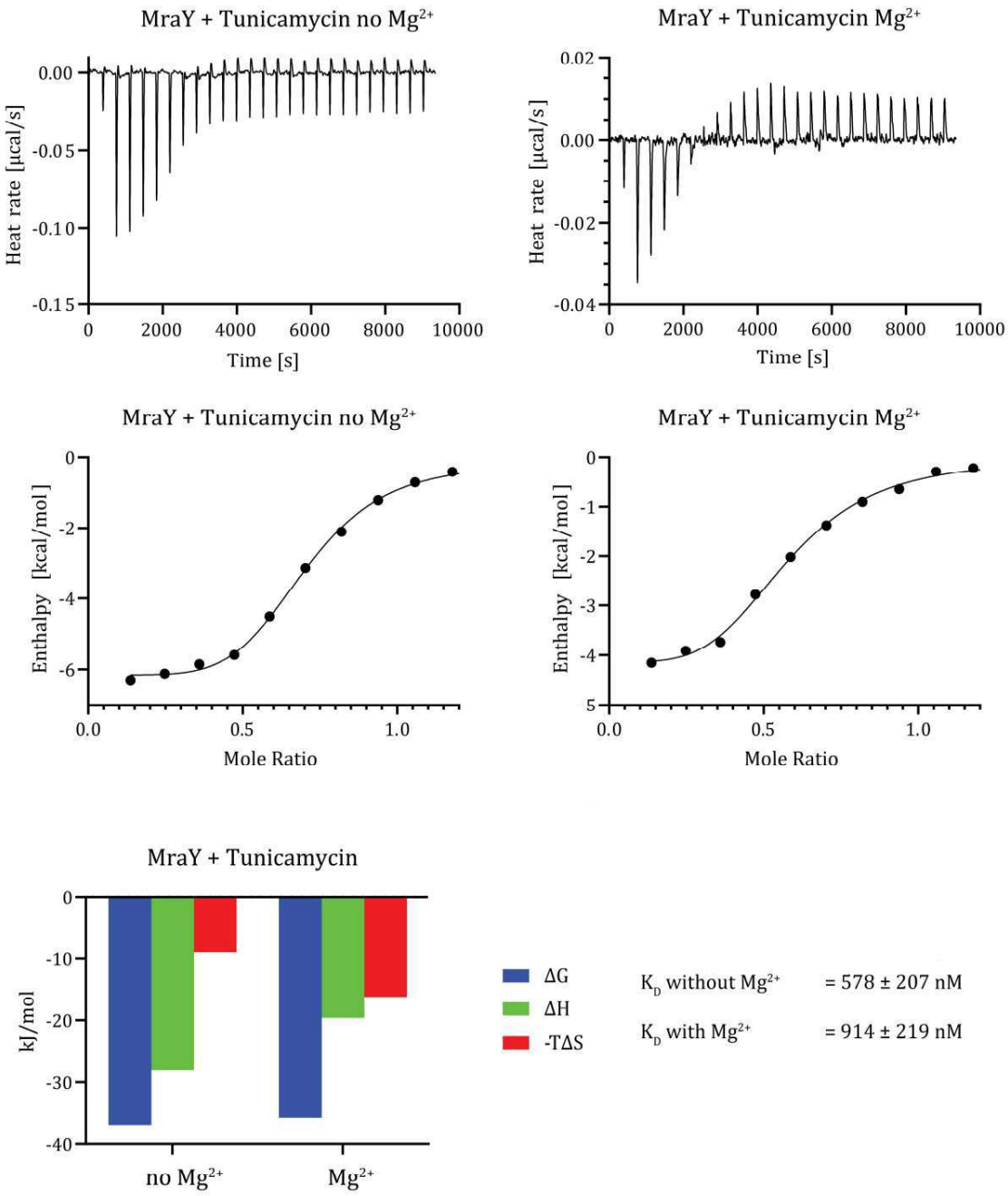


Fig. 6-1 *ITC titration of $MraY_{AA}$ with tunicamycin in the presence and absence of magnesium.*

inhibitors would be needed to achieve the same inhibitory effects towards the active enzymes.

With purified protein, it was also possible to study protein-ligand interactions in the absence of substrates using methods such as isothermal titration calorimetry (ITC). An optimized protocol was established using protein concentrations between 10-20 µM, ligand concentrations between 100-200 µM, 37 °C and optimized buffer conditions. K_D values were determined for tunicamycin, muraymycin D2 and muraymycin B6.

Comparison with buffer containing 10 mM $MgCl_2$ and buffer without $MgCl_2$ revealed no significant impact of magnesium on the binding constant of MraY and tunicamycin (see **Fig. 6-1**). The interactions' enthalpic share, however, is lower in the presence of

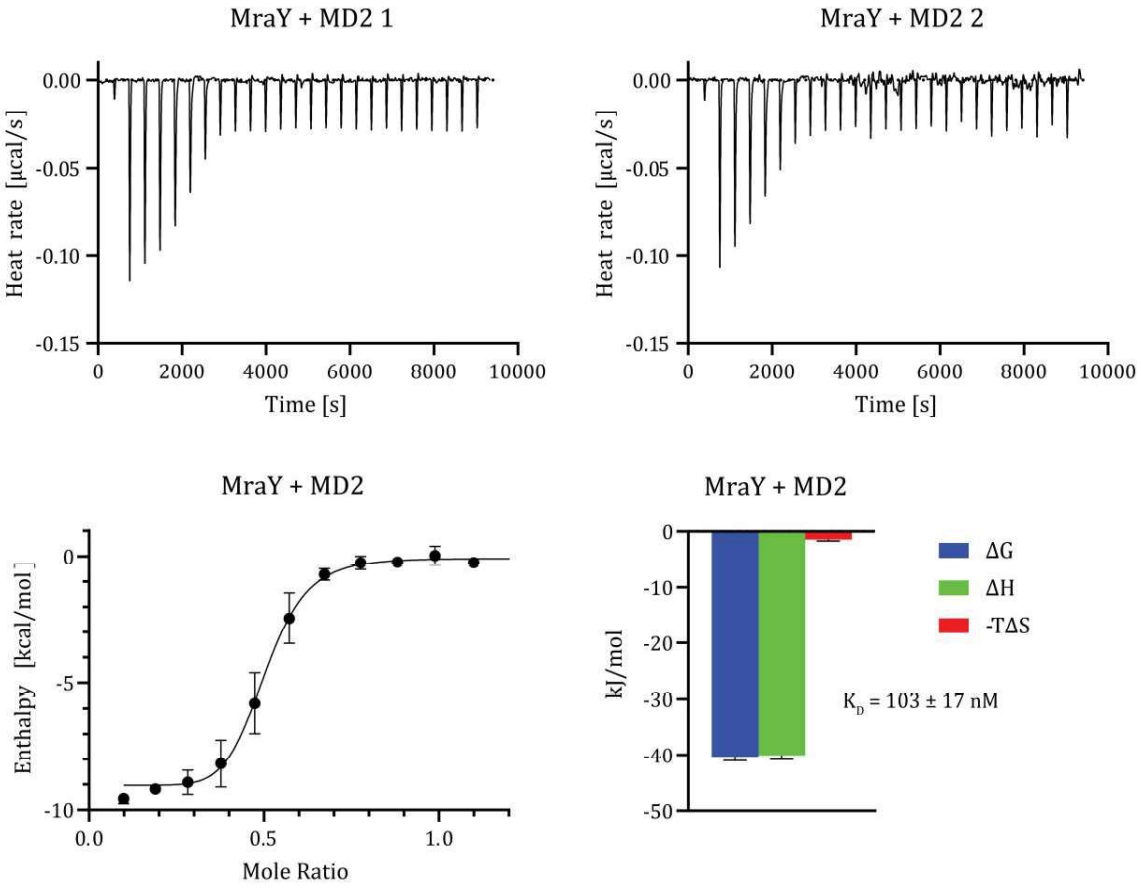


Fig. 6-2 *Duplicate ITC measurements of purified $MraY_{AA}$ in DM micelles with muraymycin D2 MD2. A K_D value of 103 ± 17 nM was obtained.*

magnesium ions. Therefore, protein concentrations must be increased when $MgCl_2$ containing buffers are used to generate good signal-to-noise ratios.

Muraymycin D2 binding to $MraY_{AA}$ was determined to have a K_D value of 103 ± 17 nM (see **Fig. 6-2**), which is in reasonable agreement with inhibitory activity found in fluorescence-based activity assays (IC_{50} = 46 ± 13 nM, from Publication **B Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**) and the K_D value found by CHUNG et al. of ~20 nM[77]. Small differences in binding affinities may arise from partial protein degradation during lengthy ITC experiments. Thermodynamic analysis further suggests that the interaction between $MraY_{AA}$ and MD2 is largely enthalpy-driven.

ITC results for muraymycin B6 with $MraY_{AA}$ are rather unexpected. The enthalpic share of the interaction is smaller compared to MD2, which is compensated by an entropic

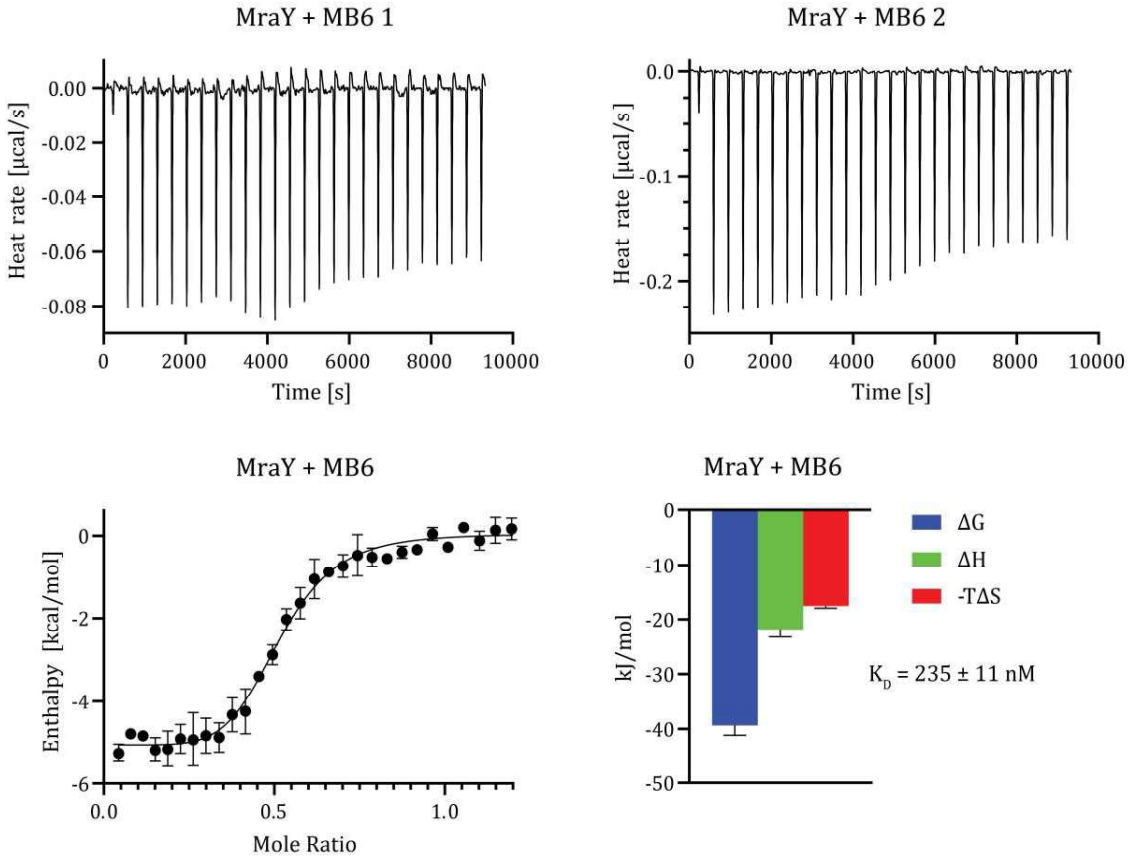


Fig. 6-3 *Duplicate ITC measurements of purified $MraY_{AA}$ in DM micelles with muraymycin B6 MB6. A K_D value of 235 ± 11 nM was obtained.*

portion to yield a K_D value of 235 ± 11 nM (see **Fig. 6-3**). However, the IC_{50} value of MB6 against solubilized $MraY_{AA}$ is lower compared to MD2 by a factor of > 240 (0.19 nM for MB6 vs 49 nM for MD2, Publication **B Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**). Therefore, a significantly lower binding constant for MB6 was expected. A closer look at the control titration of MB6 into DM-containing buffer (see **Fig. 6-4**) revealed an interaction between MB6 and the buffer itself (most likely empty DM micelles). Therefore, the apparent K_D value is higher than expected.

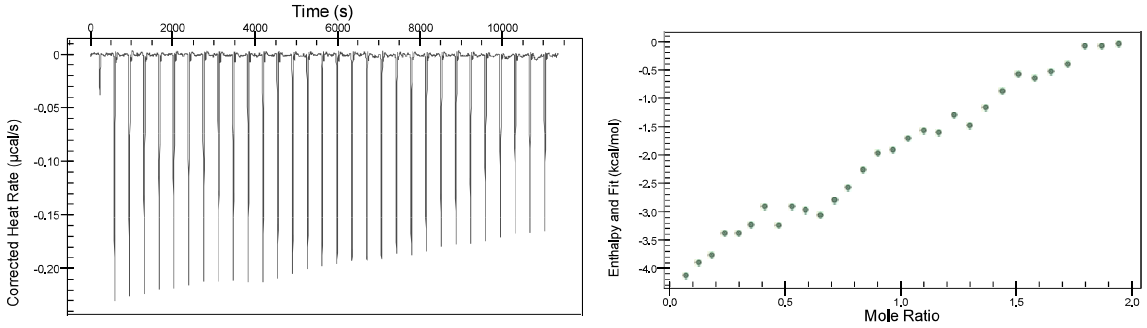


Fig. 6-4 *Titration of muraymycin B6 into DM-containing buffer. Change in heat rate indicates interaction between molecule and buffer.*

The results emphasize the difficulty in determining ligand-protein interactions of membrane proteins solubilized in micelles with more complex ligands such as muraymycin B6, which can interact with detergent molecules through their fatty acyl chains.

In addition to ITC experiments, detergent-solubilized $MraY_{AA}$ was used to study interactions and temperature-dependent aggregation using microscale thermophoresis (MST). In these experiments, His-MBP-tagged MraY was fluorescently labeled using a Ni-NTA-based red fluorescence probe. Binding to muraymycin D2 was then studied at varying temperatures, revealing a temperature dependent aggregation of MraY at temperatures > 30 °C (see **Fig. 6-5**). Based on these results, K_D values were determined at 27 °C and compared between muraymycin D2 and muraymycin B6. The K_D value determined for MB6 against $MraY_{AA}$ in DM micelles was still larger than expected (see **Fig. 6-6**). However, since the fluorescent labeling required the protein to be His-MBP

tagged, it is not guaranteed that the observed effects partially arise from interference with the protein tag.

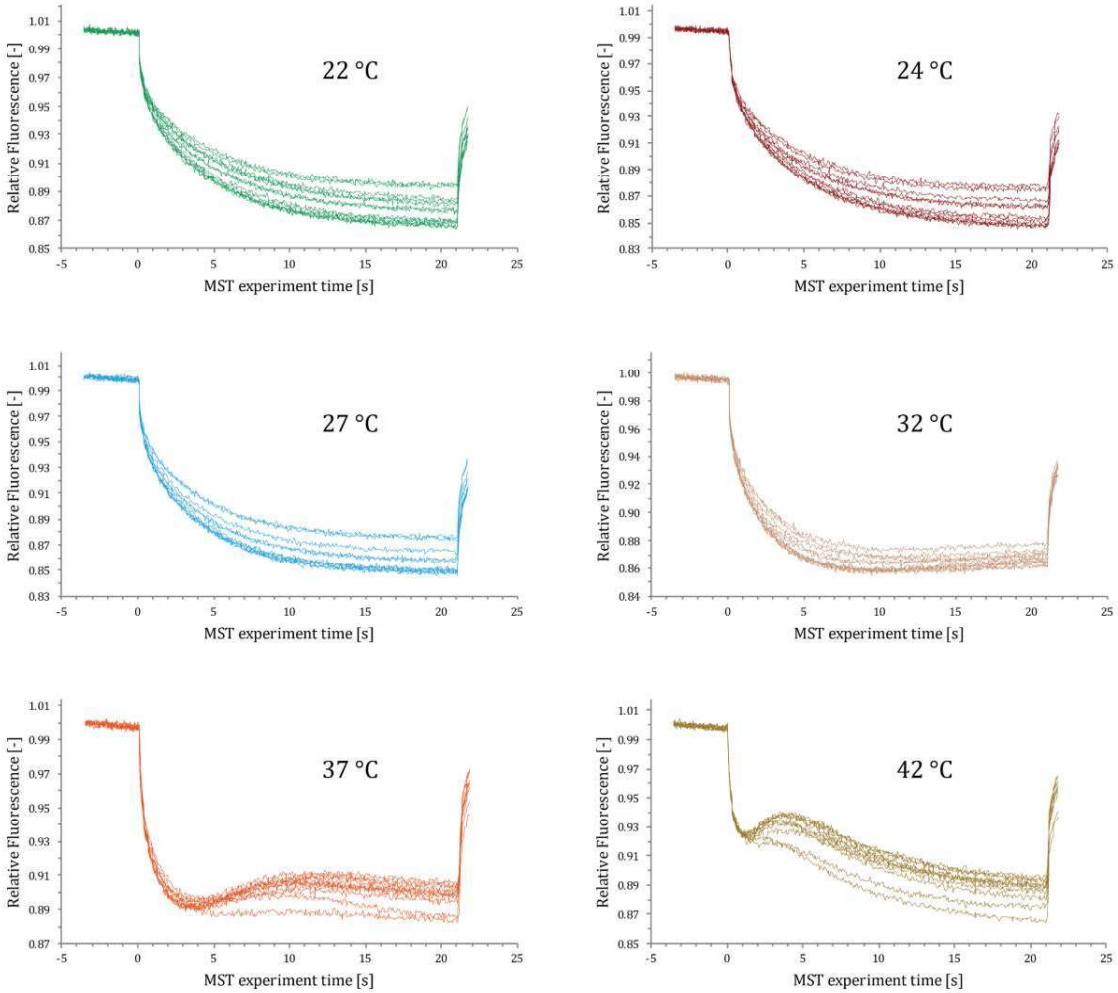


Fig. 6-5 *MST experiments with fluorescently labeled His-MBP MraY in DM micelles and muraymycin D2 at varying temperatures. Exponential decay of fluorescent signal is only observed at 22-27 °C, while variations in fluorescence signals at higher temperatures indicate protein aggregation.*

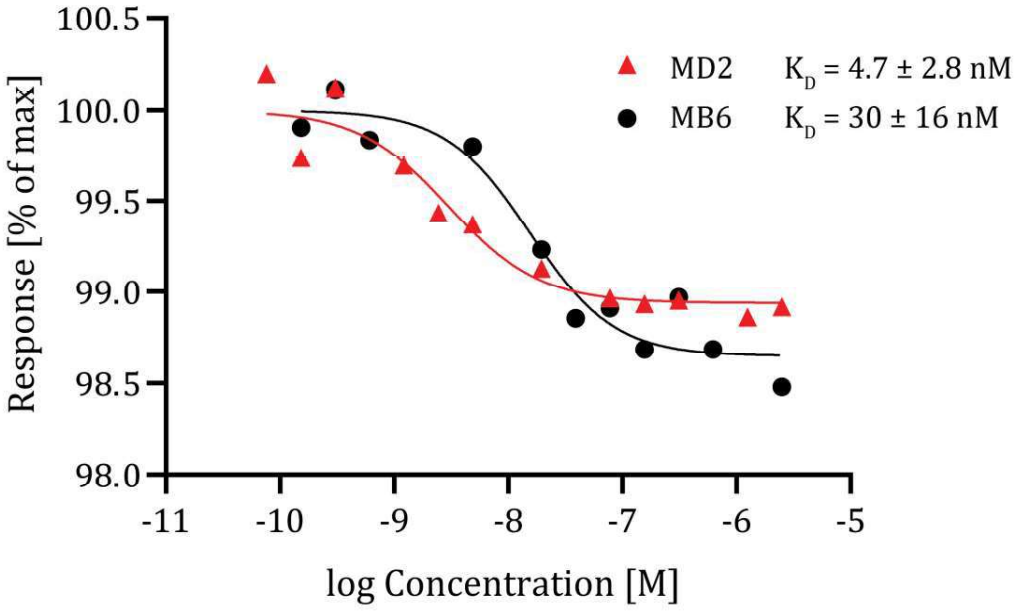


Fig. 6-6 *K_D values of MD2 and MB6 against fluorescently labeled His-MBP MraY in DM micelles at 27 °C obtained from MST measurements.*

6.1.1.2 MraY in phospholipid nanodiscs

This preparation method aimed to provide a compromise between the two advantages that crude membranes and micelles offered: the purity of an isolated protein, but in a near-native lipid environment. Nanodiscs represent a purification system for membrane proteins that consists of a phospholipid bilayer, held together by a helical scaffold protein called membrane scaffold protein (MSP). The formation of suitable nanodiscs largely depends on the composition of the lipid bilayer as well as its size, which is a consequence of the length of the used MSP. Considering that the functional form of MraY is a homodimer with a molecular weight of ~ 80 kDa, we opted for the use of MSP1E3D1, which is capable of forming nanodiscs of ~12.9 nm diameter.[366] Membranes of *E. coli* typically consist of a phospholipid mixture containing 78 % phosphatidylethanolamine (PE), 12 % phosphatidylglycerol (PG), 6 % cardiolipin (CL) and <3 % phosphatidic acid (PA).[367] To keep the phospholipid composition of the MraY nanodiscs similar to the natural *E. coli* membrane environment, a mixture of 85 % 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE) and 15 % 1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-*rac*-

glycerol) (DMPG) was used. Nanodiscs were formed by combination of DM-solubilized $MraY_{AA}$, MSP1E3D1 and lipid mixture and detergent removal with Bio-Beads SM-2 Resin. The success of MraY-nanodisc formation was confirmed by size exclusion chromatography followed by SDS-PAGE analysis, transmission electron microscopy (see **Fig. 6-7**) and the fluorescence-based *in vitro* activity assay mentioned above (see **Fig. 6-8**).

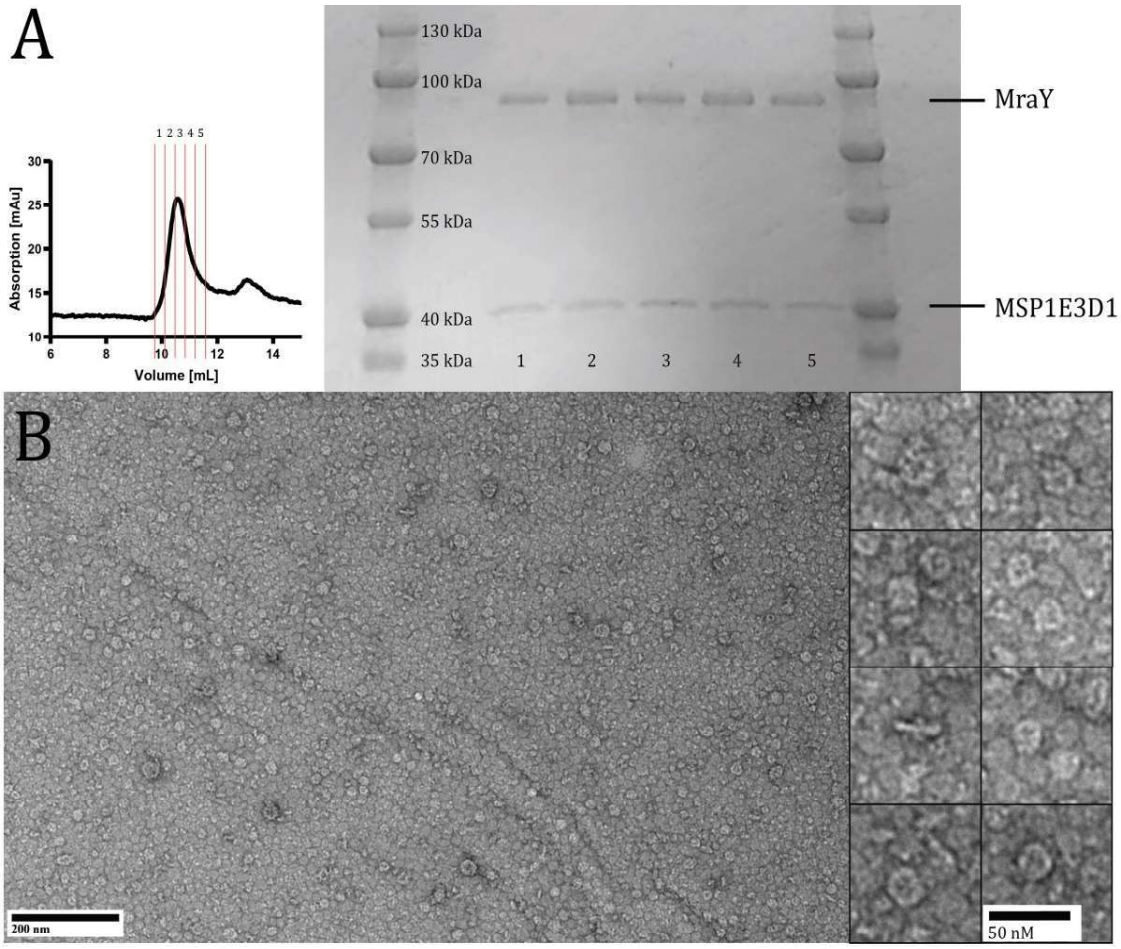


Fig. 6-7 A. *SDS-PAGE analysis of the $MraY_{AA}$-nanodisc-containing fractions. Coeluted protein from gel filtration using a Superdex™ 200 Increase 5/150 GL consists of MraY and MSP1E3D1, therefore confirming the complex (nanodisc).* ***B.*** *Negative stain images of $MraY_{AA}$-nanodiscs using a JEOL JEM 1400 Transmission Electron Microscope (TEM).*

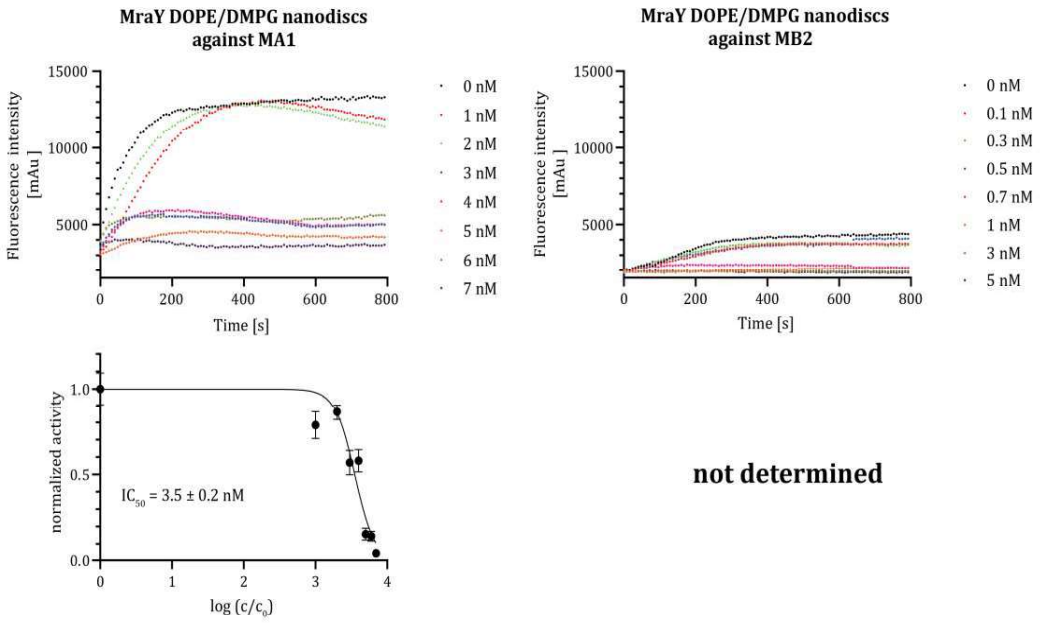


Fig. 6-8 *IC_{50} determination of muraymycins against $MraY_{AA}$ in DOPE/DMPG nanodiscs. Sufficient enzymatic activity was only found in MraY preparations used for the determination of the IC_{50} value of MA1, as can be seen by comparing slopes of uninhibited protein used for MA1 (left) and MB2 (right).*

Active nanodiscs were then tested for inhibitory potency of one member of each muraymycin series (muraymycins A1, B2, C1 and D2). Of these experiments, only the IC_{50} value determined for muraymycin A1 is reliable, because the enzymatic activity of MraY DOPE/DMPG nanodiscs rapidly decreased over time (see **Fig. 6-8**).

Compared with crude membrane preparations of $MraY_{CB}$, the inhibitory potency of muraymycin A1 dramatically decreased by a factor of ~200. Compared to $MraY_{CB}$ in DM micelles, the IC_{50} reduction was less pronounced, but still observable by a factor of ~21 (IC_{50} of muraymycin A1 against $MraY_{CB}$ in crude membranes = 0.017 nM and in DM micelles = 0.16 nM, from **Publication B Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**). Comparing $MraY_{AA}$ inhibition in nanodiscs to $MraY_{AA}$ inhibition in DM micelles, to exclude species-dependent effects, the inhibitory potency decrease was still pronounced by a factor of ~32 (IC_{50} of muraymycin A1 against $MraY_{AA}$ in DM micelles = 0.11 nM, from **Publication B Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**). Again, the necessity for higher inhibitory concentrations could be a result

of the presence of catalytically inactive protein that is still able to bind ligands. Another explanation could be that the active conformation of the enzyme is stabilized in the presence of a lipid bilayer, thus hindering inhibition. However, one would expect overexpressed MraY in crude membranes to show similar effects in this case. Thirdly, the existence of interacting proteins present in crude membranes could facilitate

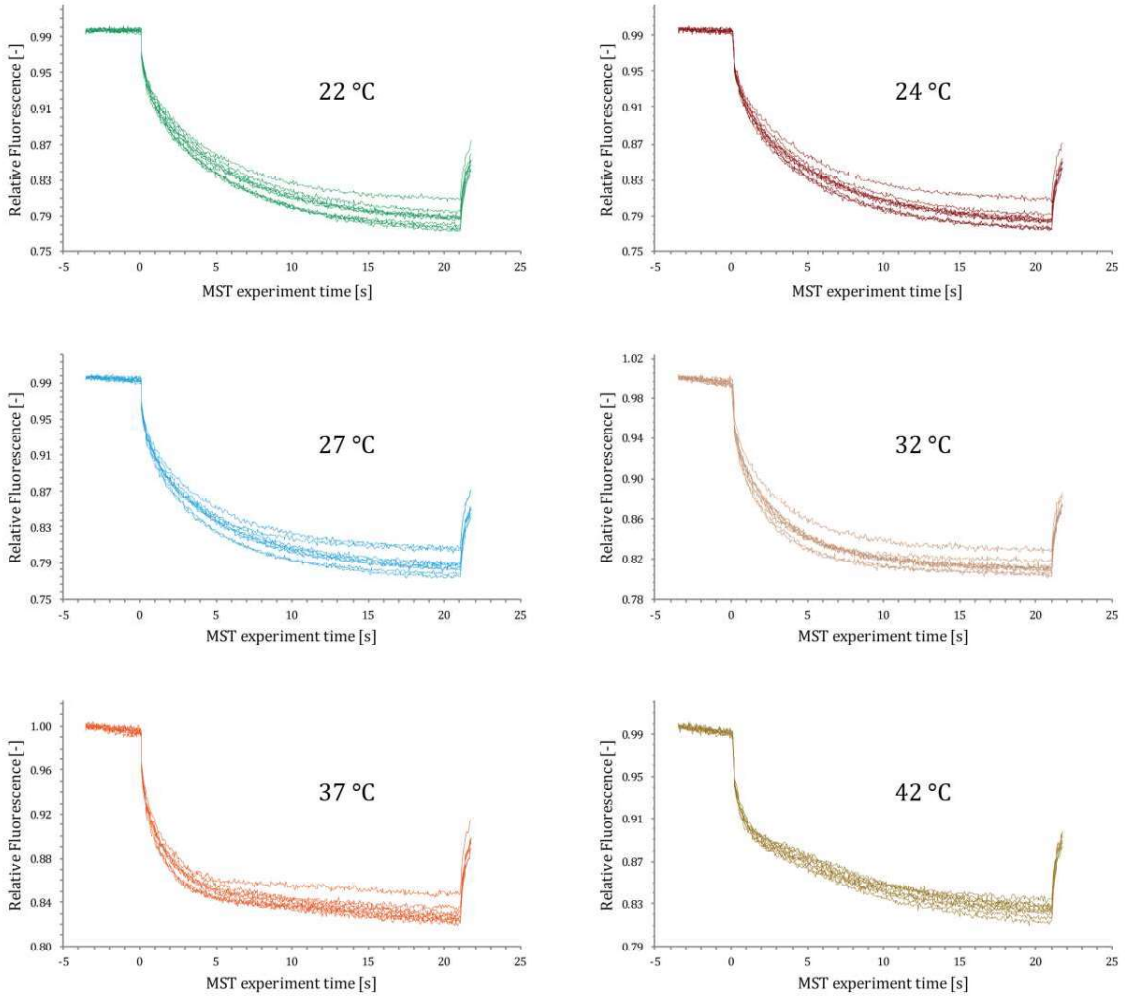


Fig. 6-9 *MST experiments with fluorescently labeled His-MBP MraY in DOPE/DMPG nanodiscs and muraymycin D2 at varying temperatures. Exponential decay of fluorescent signals is only observed at 22-37 °C, while a variation in fluorescence signal is only seen at 42 °C.*

conformational rearrangements, thus allowing inhibitors to cause changes from the active conformation more easily and therefore inhibiting the enzyme more effectively.
Like solubilized protein in micelles, $MraY_{AA}$ in nanodiscs was studied using MST. Notably, the observed temperature-dependent aggregation was less severe compared to micellar MraY and did not occur until temperatures of 42 °C.
These results suggest that, at least in MST experiments, nanodiscs are less prone to aggregation at elevated temperature than micelles.
Overall, the performed experiments highlighted the differences between MraY preparations in micelles and nanodiscs. It was shown, that MraY-containing nanodiscs can readily be prepared and that they could present a better tool than micelles to explore interactions between MraY and lipophilic inhibitors due to the absence of cross-reacting detergents.

6.1.2 Resistant MraY mutations

6.1.2.1 Relevance of D193

CHUNG et al.[77] and MASHALIDIS et al.[79] showed the importance of D193 as part of the HS1 hot spot adjacent to the uridine binding pocket of MraY. Moreover, the work of CHUNG et al.[77] described a point mutation (D193N) that retains enzymatic activity, while becoming resistant to inhibition by muraymycin D2. The X-ray co-crystal structure shows the importance of D193 in the binding of the 5'-aminoribosyl moiety, suggesting the necessity of this specific interaction for inhibitory activity.
In this work, this theory was investigated by creating a plasmid coding for $MraY_{AA}$ D193N, expressing and purifying this mutant in DM micelles and measuring inhibitory activities of muraymycin D2 as well as a 5'-defunctionalized analogue (MWP447) against wild type and mutant $MraY_{AA}$ (see **Fig. 6-10**). In accordance to the work of CHUNG et al.[77], 300 nM MD2 was tested against both wild type and D193N $MraY_{AA}$ and compared to identical treatment with the 5'-defunctionalized analogue MWP447. Unpublished results in our group determined an IC_{50} value of 4.0 nM for MWP447 against $MraY_{SA}$ in crude membranes, therefore being less active than MD2 by a factor of > 10 (see **Publication B Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside**

Antibiotics). Therefore, discrepancies between these two compounds observed at 300 nM are considered significant.

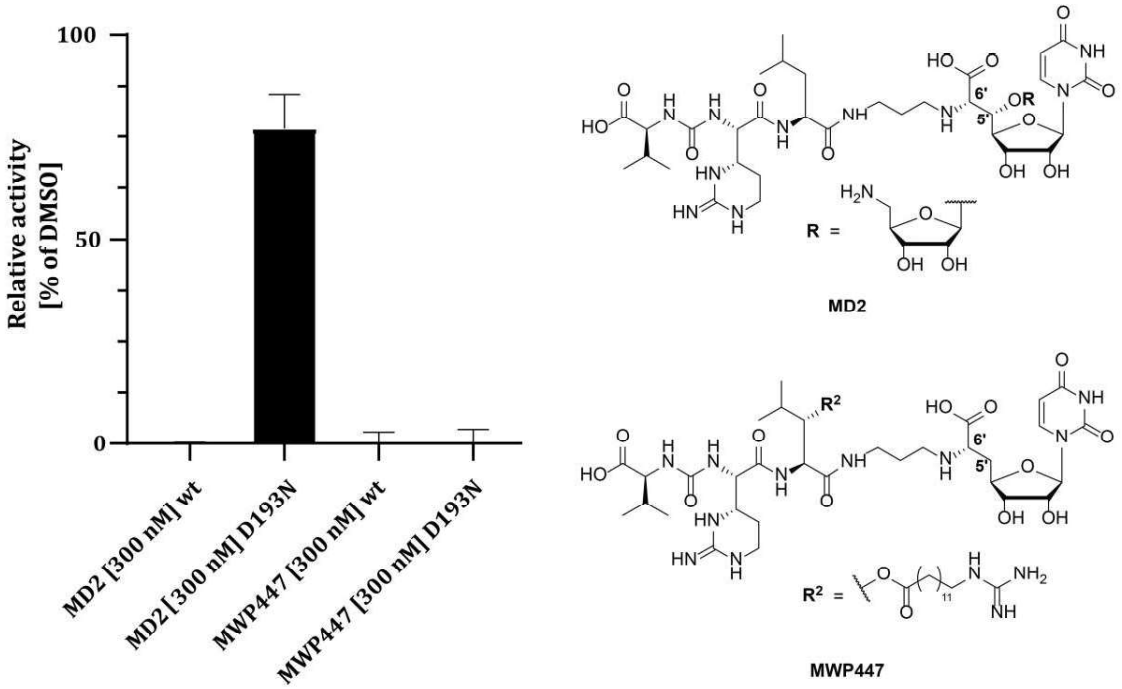


Fig. 6-10 *Relative activities of $MraY_{AA}$ wild type and D193N in DM micelles against muraymycin D2 (MD2) and MWP447 at a fixed concentration of 300 nM.*

The results from CHUNG et al.[77] were confirmed: while completely inhibiting wild type $MraY_{AA}$, ~ 75 % of enzymatic activity was restored for the D193N mutant in the presence of 300 nM muraymycin D2. On the other hand, the same concentration of MWP447 was capable of inhibiting both wild type and mutant protein equally.

These findings emphasize the importance of the interaction between the 5'-aminoribosyl moiety and D193. With the only difference being the presence of the guanidinylated fatty acyl chain, the conclusion would be that interactions of this group can compensate for the loss of inhibitory potency due to the lack of the 5'-aminoribosyl moiety. Also, the results demonstrate that 5'-defunctionalized muraymycin analogues such as MWP447 might be more robust towards the development of resistances mediated by the D193N mutation in MraY.

6.1.2.2 Relevance of $MraY_{AA}$ H324

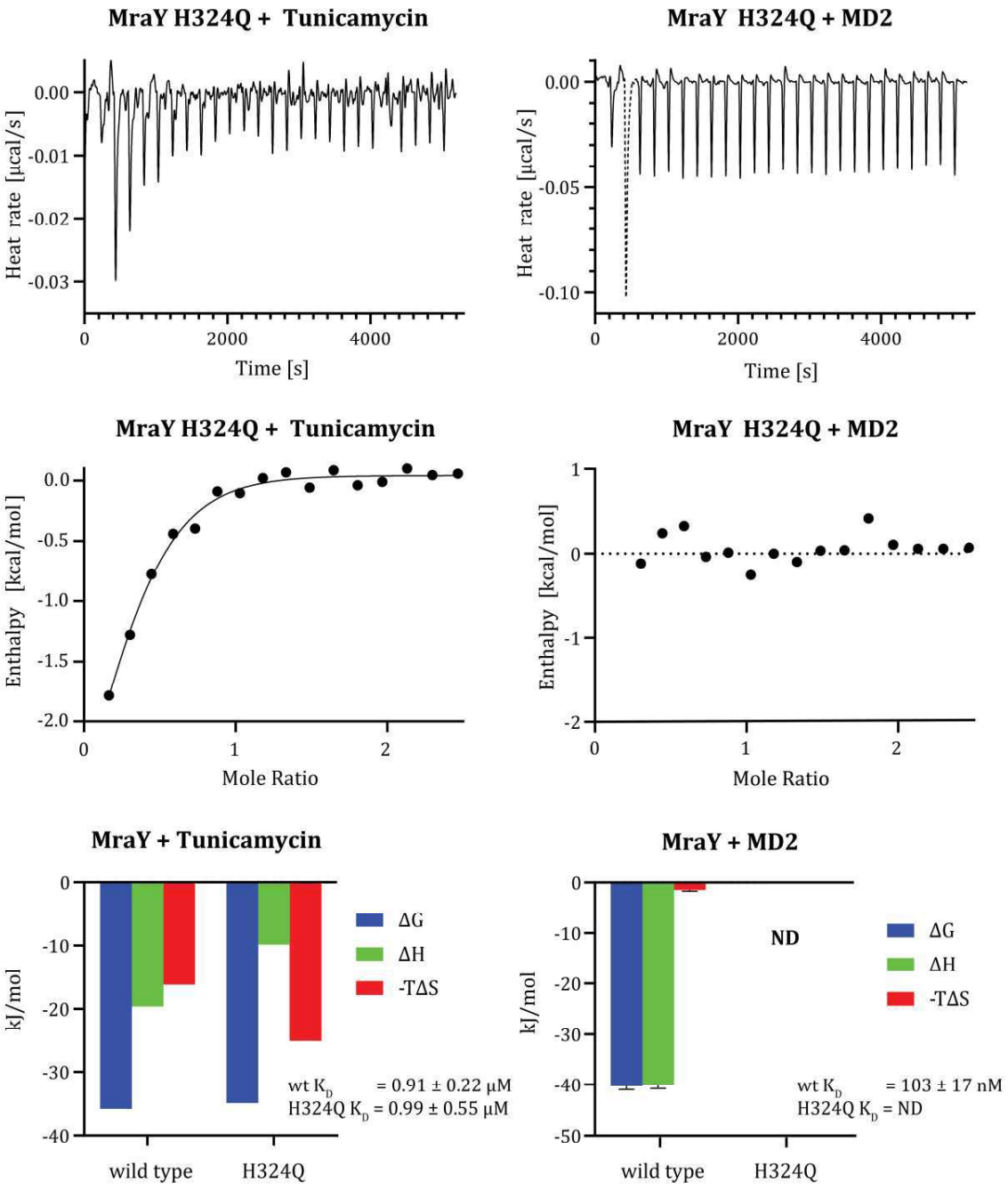


Fig. 6-11 *ITC measurements of $MraY_{AA}$ H324Q with tunicamycin and MD2. Tunicamycin binds to the H324Q mutant with the same affinities as to wild type protein. MD2 does not bind to H324Q, while it did bind wild type protein with a dissociation constant of ~ 100 nM. ND = not determined.*

Alignments of apo-MraY$_{AA}$ and MraY$_{AA}$ bound to muraymycin D2 (shown in **Fig. 2-5 B** and **C**) reveal the conformational flexibility and importance of H324 in MraY$_{AA}$. This observation leads to the suggestion that a single point mutation from histidine to glutamine, as observed in MraY of *Streptomyces sp.*, might compromise the conformational transition of MraY required to bind nucleoside antibiotics such as muraymycin D2, whilst retaining enzymatic activity.

Tunicamycin on the other hand inhibits MraY without obvious interactions with this specific residue.[78] Analysis of the tunicamycin biosynthetic gene cluster TunM in *Streptomyces chartreusis* revealed the existence of a putative SAM-dependent methyltransferase encoded in this cluster, suggesting that intracellular tunicamycin methylation might be used as a mechanism of self-resistance.[368] Hence, tunicamycin might potentially serve as a positive control to evaluate mutants and protein isoforms purified from *Streptomyces sp*.

For this purpose, a plasmid coding for MraY$_{AA}$ H324Q was cloned, expressed and purified in DM micelles. Binding analysis of tunicamycin and muraymycin D2 was then performed using ITC and compared between wild type MraY and MraY$_{AA}$ H324Q (see **Fig. 6-11**).

Tunicamycin, as shown in **Fig. 6-1**, binds to wild type MraY$_{AA}$ with a binding constant of 0.91 ± 0.22 μM. The binding to MraY$_{AA}$ H324Q was identical within the margin of experimental error, with a K_D value of 0.99 ± 0.55 μM (see **Fig. 6-11**). For muraymycin D2 on the other hand, the binding previously measured with a K_D value against wild type MraY$_{AA}$ of 103 ± 17 nM was completely abolished against MraY$_{AA}$ H324Q.

Recently, CUI et al.[112] reported the role of covalently modified (i.e. phosphorylated and adenylated) precursors in self-resistance. These modifications lead to notable reductions of IC_{50} values determined by MraY activity assays. However, even though reduced, these precursors still possessed inhibitory activity, suggesting additional mechanisms of self-resistance. The results presented here suggest that MraY from *Streptomyces sp.* retains enzymatic activity whilst rendering inhibitory effects of self-produced muraymycins ineffective. Thus, a plausible mechanism for self-resistance would be a combination of these two effects.

To gain further insights into the role of H324, constructs coding for MraY$_{SA}$ H287Q, MraY from *Streptomyces platensis* (MraY$_{SP}$) and MraY$_{SP}$ Q316H were cloned into pET28 vectors.

MraY isoforms (i.e. $MraY_{SA}$ wild type and $MraY_{CB}$ wild type) coded in this vector are used routinely in our group to prepare overexpressed MraY in crude membranes. Direct comparison of crude membrane fractions with overexpressed $MraY_{SA}$ wild type, $MraY_{SA}$ H287Q, $MraY_{SP}$ wild type and $MraY_{SP}$ Q316H will provide detailed activity analysis of muraymycins against these proteins. Furthermore, constructs coding for $MraY_{AA}$ H324Q, $MraY_{SP}$ wild type and $MraY_{SP}$ Q316H in pET26 vectors with N-terminal His-MBP tags were cloned, as this vector has been previously used for structure determination by X-ray crystallography[76,77,79]. An effort will be made to solve these structures to obtain information about the structural implications of this mutation.

6.1.3 Protein-protein interactions

6.1.3.1 MraY interacts with MurF

To test whether MraY is a part of a multienzyme complex in the biosynthesis of peptidoglycan, experiments were carried out to test its interaction with another enzyme in the pathway: MurF. UDP-*N*-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase (MurF) was investigated because it is the preceding enzyme that produces the substrate for MraY, i.e. Park's nucleotide. The codon-optimized sequence coding for MurF from *E. coli* was designed in the expression vector pET28a with an N-terminal His-tag that can be cleaved using TEV protease and the construct was ordered from GeneWiz. The plasmid was transformed into BL21(DE3) pLysS competent cells, expressed and purified using Ni-NTA agarose resin.

The protein-protein interaction between MraY and MurF was confirmed using MST experiments with fluorescently labeled His-MBP-$MraY_{AA}$ DOPE/DMPG nanodiscs. The experiment revealed that the two proteins interact with a binding constant (K_D) of 36 µM ± 5 µM (see **Fig. 6-12**).

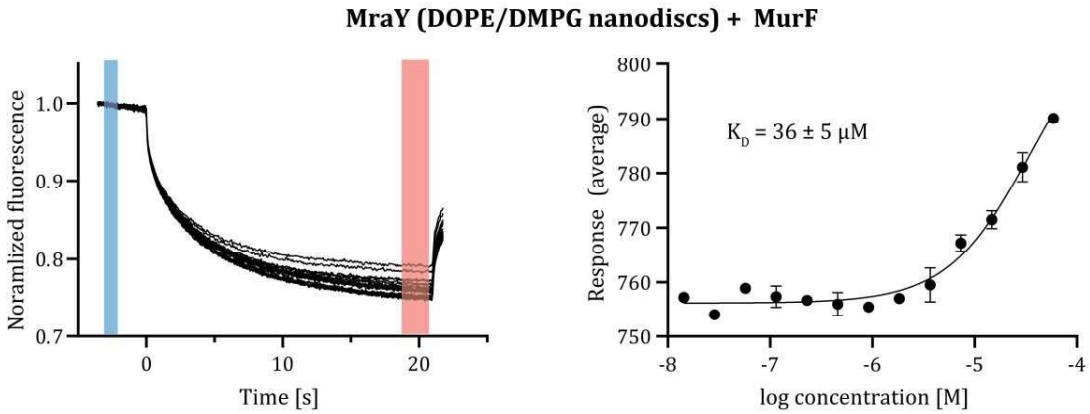


Fig. 6-12 *Interaction between His-MBP-MraY$_{AA}$ nanodiscs (DOPE/DMPG) and MurF. The proteins interact with a K_D value of 36 ± 5 µM.*

To further validate this interaction, the complex of MraY$_{AA}$ in DM micelles with MurF (ratio 1:4) was formed and the mixture of both proteins was loaded on a Superdex 200 Increase 5/150 GL column and eluted. The chromatogram displayed two peaks; according to subsequent SDS-PAGE, the first peak corresponded to both MurF as well as MraY, confirming the existence of a stable protein complex, while the second peak consisted of excess MurF (see **Fig. 6-13**).

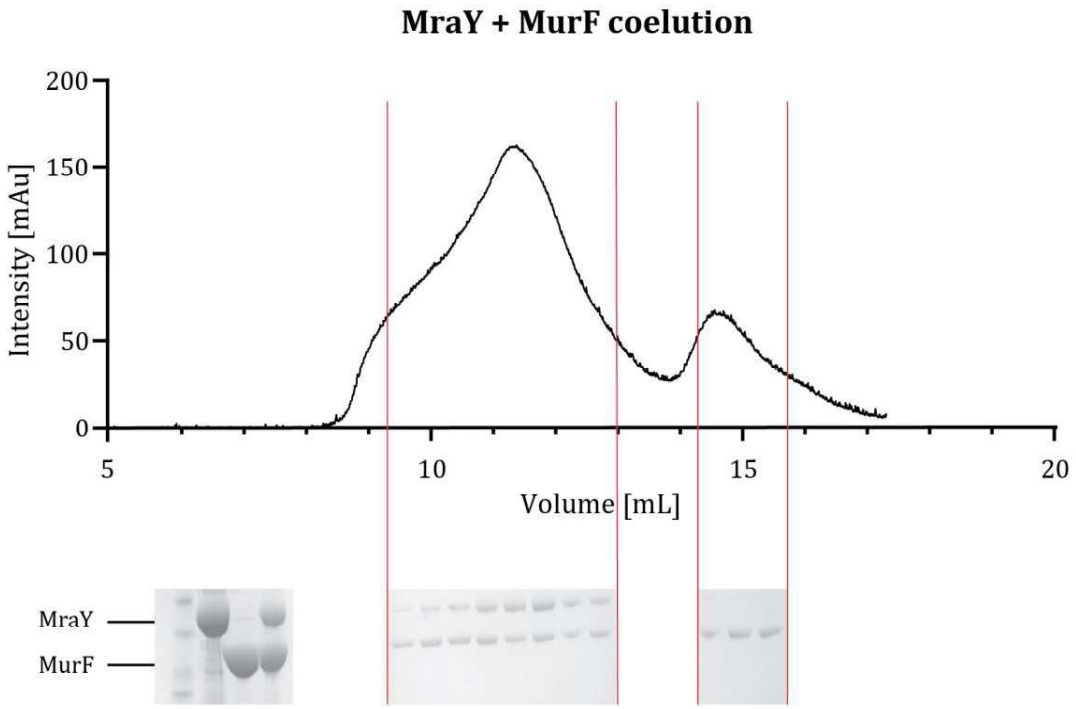


Fig. 6-13 *Coelution of $MraY_{AA}$ and MurF. The chromatogram displays the gel filtration profile of $MraY_{AA}$ in DM micelles as a mixture with MurF on a Superdex 200 Increase 5/150 GL column. Below is the SDS-PAGE analysis of the gel filtration experiment. The first four lanes represent protein ladder, $MraY_{AA}$ alone, MurF alone and a mixture of $MraY_{AA}$ and MurF. The following lanes correspond to the fractions of the size exclusion chromatogram shown above.*

6.1.3.2 Impact of the MraY-MurF interaction on protein activity and inhibition

After the interaction between the soluble protein MurF with the membrane-bound enzyme MraY had been confirmed, the biological effects of this interaction were investigated by assessing the enzyme's activity and inhibition using the fluorescence-based activity assay.

Purified MraY in DOPE/DMPG (85:15) nanodiscs was inhibited using 3 nM muraymycin A1. In the presence of the inhibitor, different concentrations of MurF were used to study the effects on MraY inhibition.

Interestingly, at high concentrations of MurF, no significant effect on MraY inhibition was observed. At concentrations below 8 µM however, MraY activity was restored in a

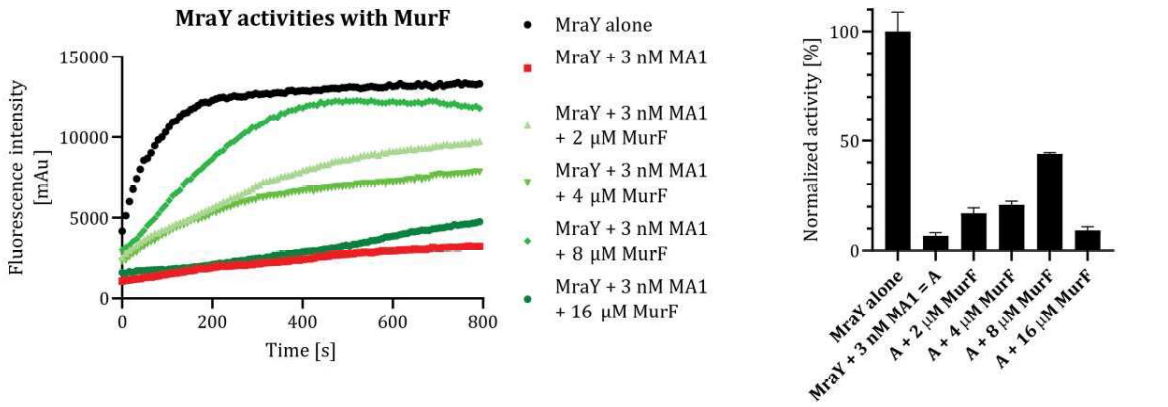


Fig. 6-14 *Influence of increasing MurF concentrations on the inhibition of MraY. The graph on the left shows one of the three MraY DOPE/DMPG nanodisc activity measurements in the absence of inhibitor (black), presence of 3 nM MA1 (red) and in the presence of 3 nM MA1 as well as varying concentrations of MurF (shades of green). Normalized activities as a percentage of uninhibited MraY are shown on the right. 8 µM MurF restore almost 50% of the original MraY activity.*

concentration-dependent manner (see **Fig. 6-14**). The most effective concentration (8 µM) corresponded to a MraY:MurF ratio of ~ 2:1, indicating that one functional MraY dimer is capable of binding one MurF monomer. The decrease at higher MurF concentrations could be a result of competitional binding of MurF to the dansylated Park's nucleotide, as MraY's substrate likely has some affinity to the preceding enzyme in the biosynthetic cascade, MurF. Moreover, the observation that MraY preparations in the presence of 3 nM muraymycin A1 are inhibited by more than 90 % compared to the control suggests that the enzymatic activity of the nanodisc preparation was already reduced when conducting these experiments. Therefore, restauration of enzymatic activity in the presence of 8 µM MurF to almost 50 % of MraY activity obtained from freshly prepared nanodiscs is even more relevant.

The experiments display impressively that the interaction between MraY and MurF is of vital importance for the enzymatic activity of the membrane protein. The results indicate that the dynamic interplay between MraY and MurF might be used as a tool to enable rapid transition between the enzymes involved in the biosynthetic pathway.

6.1.3.3 Attempts towards solving the structure of the MraY-MurF complex

Solving the structure of the MraY-MurF complex would pave the way to a better interpretation of the biochemical data. The fluorescence-based activity assay in the presence of MurF suggests that the protein-protein interaction traps MraY in an active conformation, or at least a conformation that hampers inhibition by naturally occurring antibiotics. These findings might at least partially explain the relatively weak antibacterial activities found in minimal inhibitory concentration (MIC) assays (see **Publication B Insights into the Target Interaction of Naturally Occurring Muraymycin Nucleoside Antibiotics**). In these experiments, micromolar concentrations of inhibitors are needed to achieve any antibacterial effects even for binders in the low picomolar range.

Furthermore, characterizing the protein-protein interface could help in creating a new set of drug candidates that inhibit this interaction. High throughput screens and drug development for this new target could easily be imagined as a source of new antibiotics or in combination with already existing natural nucleoside antibiotics.

To solve the structure of the supposed heterotrimer consisting of two MraY molecules and one MurF enzyme, cryo-EM appears to be the method of choice. A comparison of MraY in the absence and presence of MurF using negative stain electron microscopy revealed an increase in particle size as the complex is formed (see **Fig. 6-15**).

For negative stain EM analysis, $MraY_{AA}$ solubilized in buffer containing 4 mM DM was used. Importantly, samples must be directly diluted from gel filtration fractions obtained after the complex formation. Prior concentration of the complex using centrifugal filters led to detergent accumulation and formation of detergent aggregates in the negative stained samples (data not shown).

Fig. 6-15 shows the comparison of negative stain EM images of $MraY_{AA}$ (**Fig. 6-15 A**), MurF (**Fig. 6-15 B**) and the MraY-MurF complex (**Fig. 6-15 C**) at a concentration of 0.05 mg/mL. Protein particles observed in the complex-containing sample are mainly uniform and present an ideal starting point to solve the structure of the complex.

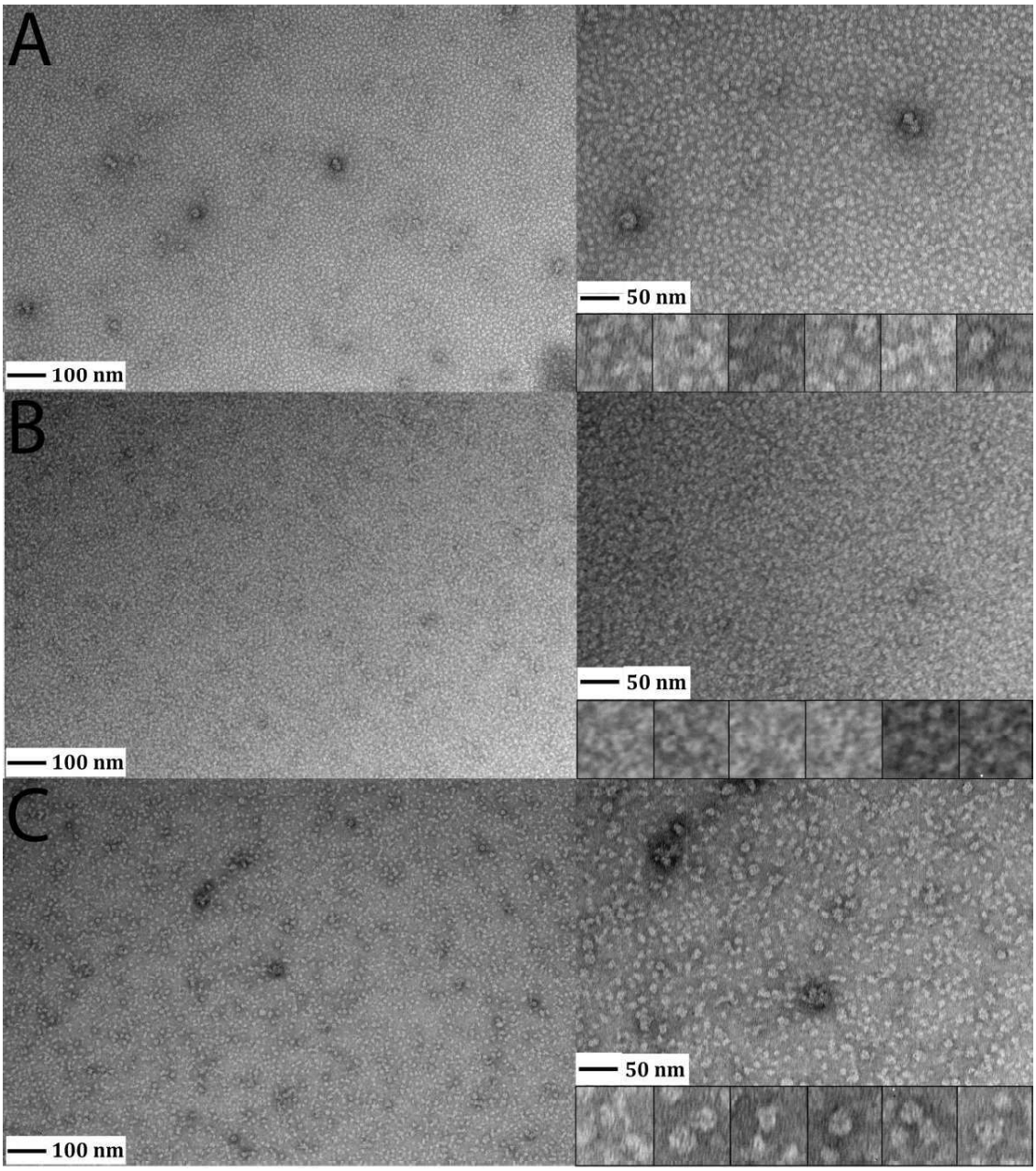


Fig. 6-15 *Negative stain EM images of MraY, MurF and the MraY-MurF complex.* ***A.*** *40k magnified and 80k magnified $MraY_{AA}$ at 4 nM DM (~ 2 x CMC) and zoom in on individual particles.* ***B.*** *40k magnified and 80k magnified MurF at 4 nM DM (~ 2 x CMC) and zoom in on individual particles.* ***C.*** *40k magnified and 80k magnified $MraY_{AA}$-MurF complex at 4 nM DM (~ 2 x CMC) and zoom in on individual particles.*

6.1.4 Discussion

The investigation of the interactions that nucleoside antibiotics exhibit with MraY revealed that the results greatly depend on the preparation method for MraY. The two methods that use purified MraY (DM micelles and phospholipid nanodiscs) mainly differ in the presence of detergent in the case of micelles. While empty micelles have been shown to be interfering with the binding of compounds that harbor fatty acid side chains using ITC, these restrictions are not true for diluted MraY preparations for fluorescence-based *in vitro* activity assays. Notably, MraY nanodisc preparations require higher concentrations of muraymycin A1 for inhibition compared to their DM micelle counterparts. The possibility of a stabilizing effect of the active MraY conformation due to the phospholipid bilayer is a likely explanation for this observation. On the other hand, the use of crude membrane fractions with overexpressed MraY presents the advantage of including biological molecules from bacterial membranes that might be involved in MraY physiology. However, these preparations have the disadvantage of yielding an unidentified amount of MraY and generally suffer from potential impurities that might interfere with inhibitors.

In the second project, it was established that single point mutations are capable of rendering MraY resistant towards inhibition by certain inhibitors, while retaining enzymatic activity. It was further shown that some structural motifs (i.e. the aminoribosyl unit for the inhibition of MraY D193N) might be vulnerable to emerging resistances and that this resistance can be overcome by altering said elements. However, further work needs to be done to furnish a more detailed picture of these preliminary results. Moreover, initial results indicate that the possibility of single point mutations as a resistance mechanism might be involved in self-resistance of *Streptomyces sp.* Mutation of His324 to Gln abolished binding of muraymycin D2 to $MraY_{AA}$ but retained affinity to tunicamycin. Experiments using purified MraY from *Streptomyces platensis* ($MraY_{SP}$) are in preparation, as plasmids coding for wild type $MraY_{SP}$ and $MraY_{SP}$ Q316H were cloned as part of this work.

Lastly, this thesis laid the ground for the investigation of protein-protein interactions that involve MraY. Biochemical, biophysical and functional studies proved the existence of an

interaction between MraY and its biosynthetically preceding enzyme, the soluble protein MurF. First efforts towards solving the structure of this protein complex have been made by analyzing the complex with negative stain electron microscopy. Further work will aim to provide a detailed picture of this interaction at atomic resolution to identify its influence on MraY activity and inhibition.

6.2 NFAT

6.2.1 The transactivation domain of NFAT

To understand the purpose of multisite phosphorylation of NFAT, the role of individual phosphorylation in an isolated environment was studied. The intrinsically disordered transactivation and regulatory domains of NFAT are known to be heavily phosphorylated. While first information about two phosphorylation sites of the regulatory domain by PKA and their effects on nuclear localization were obtained in **Publication A ^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins**, phosphorylation of the transactivation domain remains largely unstudied.

The first reported PTM is a phosphorylation of S23, immediately N-terminal of the binding site for coactivators and scaffold protein such as the KIX domains of Med15 and CBP/p300 or Homer3. To study the role of this phosphorylation, a plasmid coding for His-GB1-NFAT$_{1\text{-}130}$ was cloned to express isotopically labeled protein and to study it using solution-NMR.

6.2.1.1 Interactions of NFAT$_{1\text{-}130}$

First, ^{13}C-^{15}N-labeled NFAT$_{1\text{-}130}$ was expressed and purified for triple-resonance backbone NMR experiments. Thereby, the assignment of peaks in NFAT$_{1\text{-}130}$ ^{15}N-^{1}H-HSQC spectra was accomplished for 100 % of the protein's sequence (excluding prolines), thus enabling the study of individual residues by solution-NMR. Using ^{15}N-^{1}H-HSQC experiments, unlabeled interacting proteins (i.e. Med15 KIX, CBP/p300 KIX and Homer3) were titrated to ^{15}N-labeled NFAT$_{1\text{-}130}$ to map the binding sites of these proteins (see **Fig. 6-16**). The binding of the two KIX domains (CBP/p300 KIX and Med15 KIX) is similar at the 9aaTAD, but an additional binding site between amino acids 60-70 can be identified in the case of CBP/p300 KIX. The NMR experiments further revealed an overlapping binding site of Homer3 with the previously mentioned KIX domains, as well as an additional binding site around amino acid 90. Although previous publications state that Homer3 binding is competitive with calcineurin binding[167], these findings locate the

second Homer-binding site away from the conserved SPRIEIT motif, thus excluding overlapping binding sites for the two proteins.

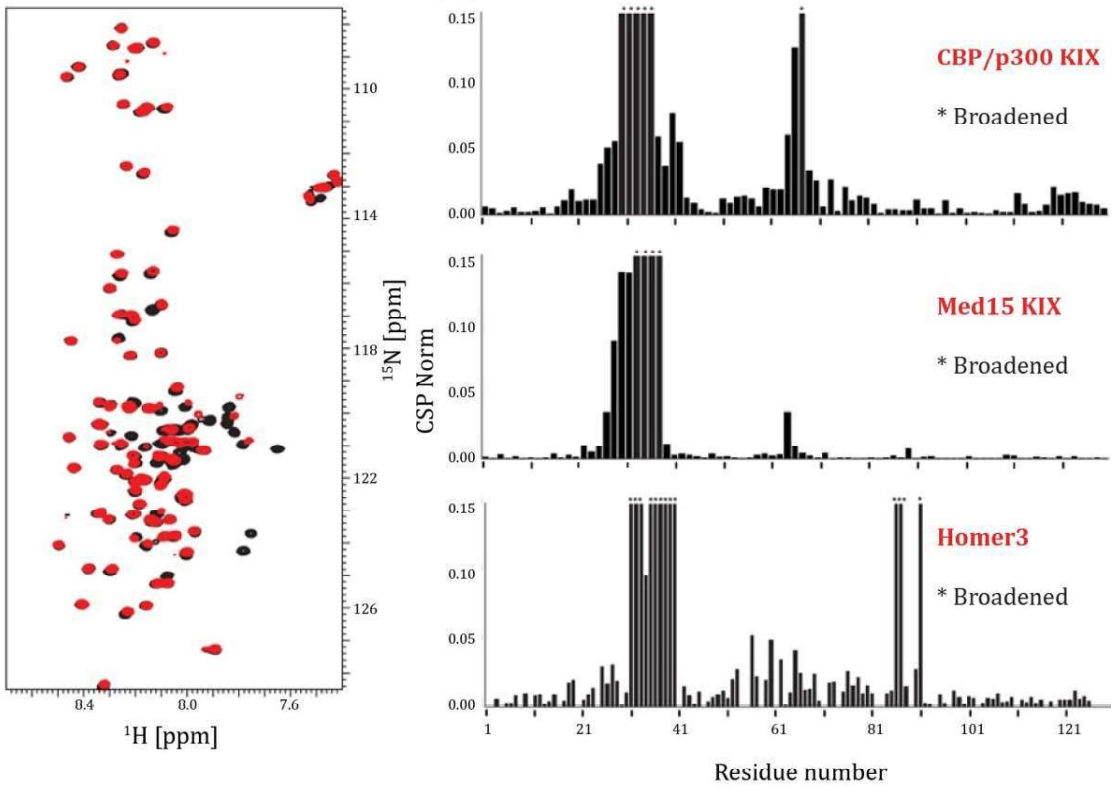


Fig. 6-16 A. *^{15}N-^{1}H-HSQC spectra of NFAT$_{1-130}$ in the absence (black) and presence (red) of CBP/p300 KIX. Chemical shift perturbations (CSPs) and line broadening are observed.* ***B./C./D.*** *CSPs as a function of residue number for the interaction between NFAT$_{1-130}$ with CBP/p300 KIX, Med15 KIX and Homer3.*

6.2.1.2 Phosphorylation of S23 in NFAT$_{1-130}$

It was reported in 2003 that NFAT is a target of p38 MAP kinase by Wu et al.[342] It was shown that p38 is involved in T cell activation and that inhibition of the kinase interferes with NFAT activation. The responsible mechanisms were identified as activation of the NFAT promoter, stabilization of NFAT mRNA, translation of NFAT mRNA and binding of NFAT to CBP. However, mutational analysis of potential p38 phosphorylation sites on NFAT did not yield any changes in binding of the two proteins.[342]

To verify the phosphorylation of NFAT in position 23 and study the effects, we produced ^{15}N-labeled NFAT$_{1-130}$ and treated it with recombinantly purified activated p38 MAP kinase. Over an experimental time of 3 days, phosphorylation of two serines was

observed by a change in ^{1}H chemical shifts of serine backbone NH bonds of 1-2 ppm downfield (see **Fig. 6-17**). Using triple resonance NMR experiments with ^{13}C-^{15}N $NFAT_{1-130}$, the two phosphorylation sites were identified as S23 and S110.

In order to study the effect of phosphorylated S23 in isolation, a S110A mutant was created.

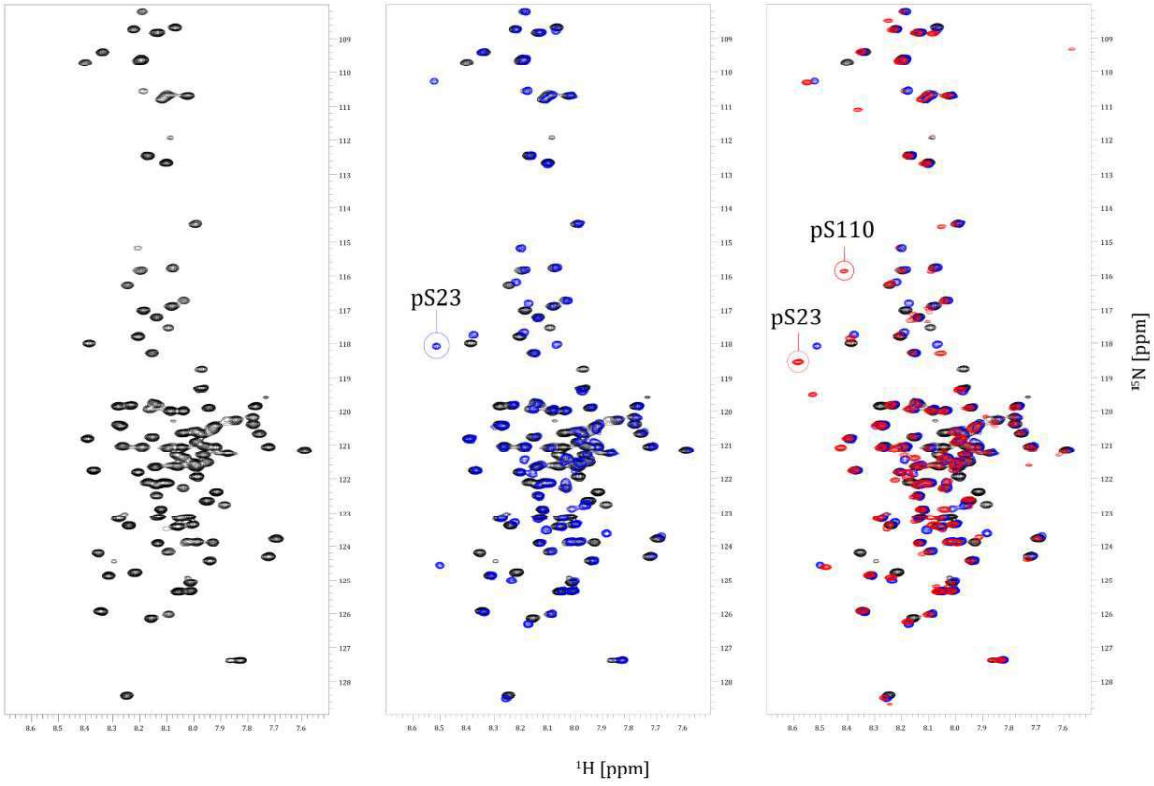


Fig. 6-17 *^{15}N-^{1}H HSQC spectra of $NFAT_{1-130}$ with p38 MAPK over 72 h. Black = $NFAT_{1-130}$ before addition of p38 MAPK, blue = $NFAT_{1-130}$ after adding p38 MAPK and incubation for 24 h, red = $NFAT_{1-130}$ after adding p38 MAPK and incubation for 72 h.*

In addition to the NFAT S110A mutant, several different NFAT mutations were cloned, which are listed in **Table 1.**

mutant	purpose
S110A	Study single phosphorylation site (pS23)
S23A	Study single phosphorylation site (pS110)
D14C S110A	PRE experiments
D35C S110A	PRE experiments
L64C S110A	PRE experiments
S99C S110A	PRE experiments
S110A A124C	PRE experiments
D35C S99C S110A	EPR experiments

Table 1 *$NFAT_{1\text{-}130}$ mutants for NMR experiments*

6.2.1.3 Dynamics of $NFAT_{1\text{-}130}$ S110A and $NFAT_{1\text{-}130}$ S110A pS23 in the ps-ns time scale

To assess the dynamic behavior of NFAT's transactivation domain, relaxation experiments on ^{15}N-labeled NFAT were carried out and dynamics of individual amino acids were analyzed.

The fastest motions in a protein are faster than the average tumbling time (correlation time, τ_m), which is typically > 3 ns. So-called heteronuclear spin-relaxation parameters dictate these motions in NMR. The local field experienced by individual nuclei is influenced by motion of the vector of a bond (i.e. N-H bond) with respect to the external magnetic field B_0. Nuclear relaxation processes can be enhanced by such fluctuations, if they occur at the right frequencies. Dipole-dipole (DD) interactions of the non-proton nucleus (S, i.e. nitrogen) with the proton (I) and chemical shift anisotropy (CSA) of the S nucleus are the primary sources for these effects. Protein bond vectors can reorientate on the ps-ns timescale to contribute to longitudinal (R^1) and transverse (R^2) exponential relaxation of the non-proton spin and to the rate of heteronuclear cross-relaxation (σ_{SH}), as described by equations 1-3;[369]

$$R^1 = \left(\frac{d^2}{4}\right)[J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)] + c^2 J(\omega_S) \qquad (1)$$

$$R^2 = \left(\frac{d^2}{8}\right)[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)]$$

$$+ \left(\frac{c^2}{6}\right)[4J(0) + 3J(\omega_S)] + R_{ex} \quad (2)$$

$$\sigma_{SH} = \left(\frac{d^2}{4}\right)[6J(\omega_H + \omega_S) - J(\omega_H - \omega_S)] \quad (3)$$

with d being the magnitude of DD and CSA interactions; $d = (\mu 0 h\gamma_I\gamma_S/8\pi^2)r_{IS}^{-3}$, $c = \Delta\sigma_S/\sqrt{-3}$, μ_0 being the permeability of free space, h being Planck's constant, γ being the gyromagnetic ratio, r_{IS} being the I–S bond length, ω being the Larmor frequency and Δσ being the CSA of the S spin. Axial symmetry is assumed for simplicity. R_{ex} is the exchange rate, an additional parameter that influences the transverse relaxation rate for slower motions (chemical or conformational exchange). When R_{ex} is assumed to be zero, only motions in the ps-ns timescale are observed. The relaxation rates mentioned in the equations above are measured through NMR experiments known as spectral density mapping.

The spin-lattice relaxation rate R^1 is determined as the reciprocal spin-lattice relaxation time T_1. This can be determined by applying a 90 ° pulse and wait for the magnetization to return to the z-axis. Transient magnetic fields due to molecular motion at the Larmor frequency drive T_1 relaxation processes. An experiment to probe this parameter is the inversion recovery pulse sequence: A 180 ° pulse is applied which converts the magnetization vector from +z to -z. After a varying waiting period τ, a 90 ° pulse is applied, followed by acquisition. This yields a signal which is directly proportional to the relaxing magnetization along z, described by equation 4:

$$I_\tau = I_\infty(1 - 2^{e\left(\frac{-\tau}{T_1}\right)}) \quad (4)$$

When τ is varied from small to large values, T_1 can be obtained by nonlinear regression. The spin-spin relaxation rate R^2 is determined as the reciprocal of the spin-spin relaxation time T_2. It describes the mechanism by which the magnetization vector decays in the xy plane. A CPMG (Carr-Purcell-Meiboom-Gill) experiment[370,371] can be employed. In this experiment, which makes use of a "spin-echo" pulse sequence, a 90 ° pulse along x is applied, followed by an evolution time Δ, a 180 ° pulse along y and another evolution time Δ followed by acquisition. This yields intensity for a signal which can be calculated from equation 5:

$$I_\tau = I_\infty(e^{\frac{-\tau}{T_2}}) \quad (5)$$

where τ is the total evolution time, or 2Δ. Spin-spin coupling is not refocused by this pulse sequence. Therefore, multiplets require a different sequence that will not be discussed here. Peak heights plotted against varying mixing times (τ) result in exponential decay, which can be fitted to yield the relaxation rate R^2.

A process called the nuclear Overhauser effect (NOE)[372] is described by the cross-relaxation σ_{SH} between a non-proton spin S attached to a proton spin I. It can be probed by an experiment called ^{15}N-^{1}H heteronuclear NOE. Decreased NOEs compared to the average of all residues are indicative for bond vectors that undergo motions faster than the overall tumbling. A single refocused 2D Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) experiment can be used to obtain cross-relaxation rates. Before the initial 90 ° pulse on the S spin (i.e. nitrogen), a relaxation delay T (typically 3-4 s) results

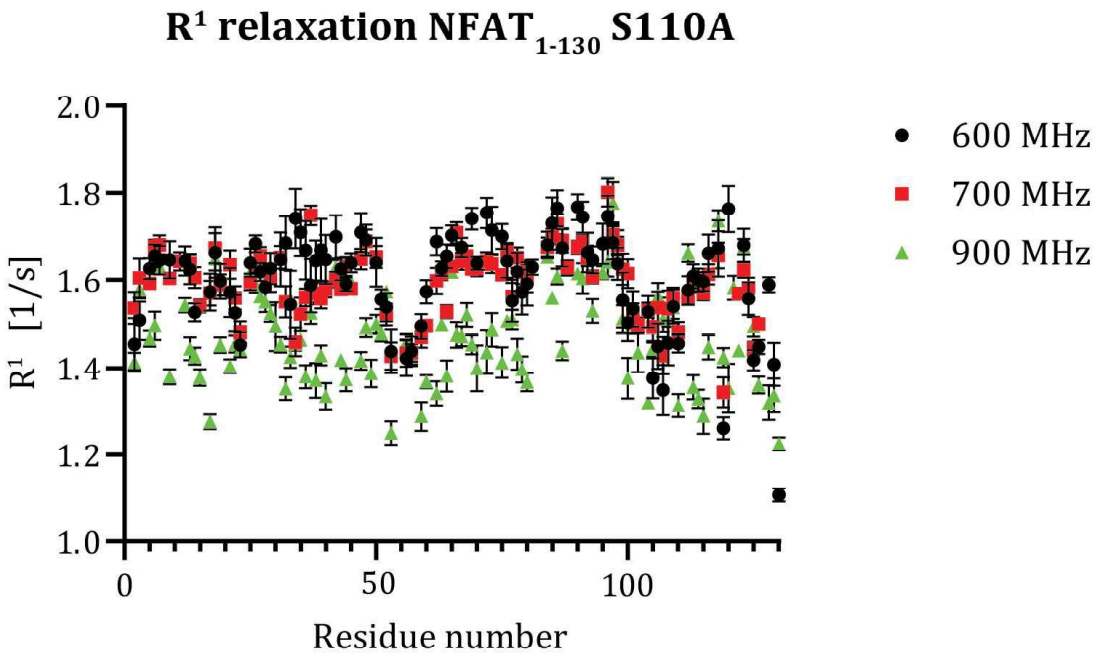


Fig. 6-18 *Field-dependent R^1 relaxation rates of $NFAT_{1-130}$ S110A.*

in proton saturation. Saturated and equilibrium spectra are recorded in an interleaved fashion. NOEs can then by calculated from resulting peak intensities according to equation 6:

$$NOE = \left(\frac{I_{saturated}}{I_{equilibrium}}\right) - 1 \quad (6)$$

Dynamics of $NFAT_{1\text{-}130}$ S110A and $NFAT_{1\text{-}130}$ pS23 S110A in the ps-ns timescale were assessed using T_1, T_2 and hetNOE relaxation experiments at different magnetic field strengths.

Minima in R^1 relaxation rates at amino acids 50-60 and 100-115 indicate that these residues might contribute to longitudinal relaxation by being more capable of transferring energy to the environment (see **Fig. 6-18**). No significant difference is observed between 600 and 700 MHz, while R^1 rates are lower throughout the entire sequence at 900 MHz.

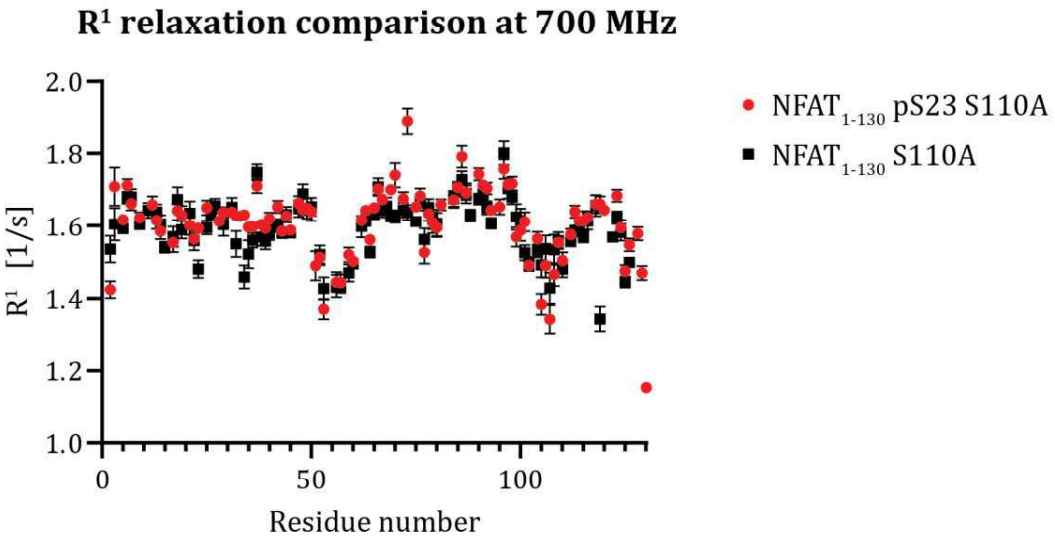


Fig. 6-19 *R^1 relaxation rates for $NFAT_{1\text{-}130}$ S110A and $NFAT_{1\text{-}130}$ pS23 S110A at 700 MHz.*

Next, we evaluated changes to longitudinal relaxation rates by phosphorylation at position S23 at 700 MHz. As can be seen in **Fig. 6-19**, no significant changes to R^1 relaxation rates were caused by phosphorylation.

A different picture was drawn when analyzing contribution to the transverse relaxation rate R^2 (see **Fig. 6-20**).

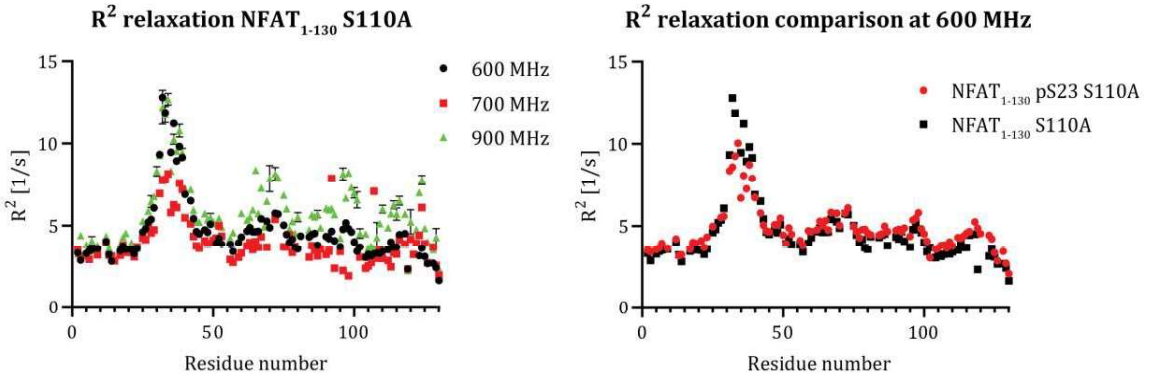


Fig. 6-20 *Field-dependent R^2 relaxation rates of $NFAT_{1-130}$ S110A and R^2 relaxation rate comparison of $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A at 600 MHz.*

Strikingly, high transverse relaxation rates were observed between amino acids ~ 30-45. This peak overlaps with the 9aaTAD and is located exactly were coactivator and mediator proteins such as Med15 and CBP KIX are binding. Also, it is located immediately after the phosphorylation site at position 23. Furthermore, other regions of interest include amino acids 65-75, 90-100 and 115-125. These regions show increased R^2 rates as well, but not as pronounced as the 9aaTAD.

Compared to unphosphorylated NFAT, NFAT pS23 had significantly decreased R^2 relaxation rates in the transactivation domain. This change is a result of decreased flexibility of this region. Where high flexibility resulted in the elevated R^2 rates within the 9aaTAD, rigidification was noted upon phosphorylation. These observations point towards the possibility of altered flexibility of the TAD to vary protein-protein interactions.

Next, the influence of phosphorylation on fast dynamics affecting the N-H bond vector was investigated by recording hetNOE experiments (see **Fig. 6-21**).

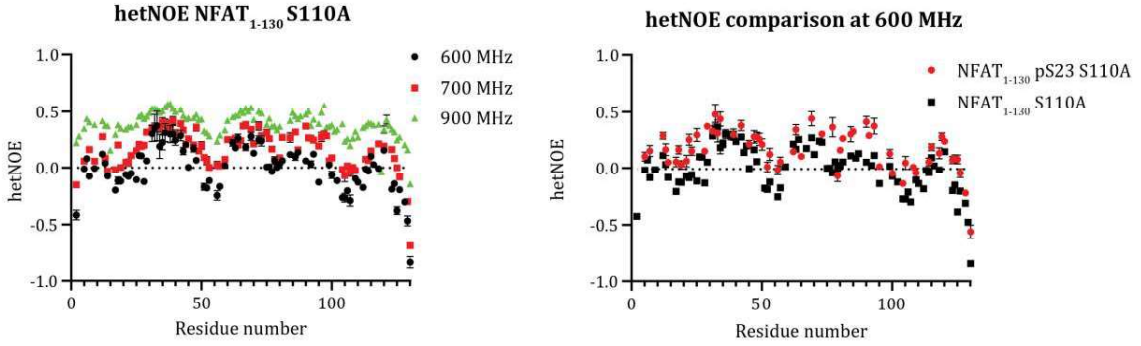


Fig. 6-21 *Field-dependent hetNOE of NFAT$_{1-130}$ S110A and hetNOE comparison of NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ pS23 S110A at 600 MHz.*

Fast motions affecting the N-H bond vector result in low hetNOE values. The fastest motions are observed at the C-terminus. However, the regions between amino acids 50-60 and 100-110 experience fast dynamics within the backbone vectors as well. The field dependency is quite pronounced. Interestingly, the observation for the regions between amino acids 30-40, 60-70 and 90-100 is the opposite compared to what was seen from the R^2 relaxation data. These results indicate slower motions within the N-H bond vectors of these regions compared to the rest of the protein.

Generally, the fast motions observed in N-H bond vectors observed in unphosphorylated NFAT$_{1-130}$ slowed upon phosphorylation. Notably, the global increase in hetNOE values was least pronounced for the TAD region between amino acids 35-45, in contrast to the changes seen in R^2 values.

Taken together, the dynamic analysis of motions in the ps-ns timescale revealed rigidification of NFAT$_{1-130}$ upon phosphorylation.

6.2.1.4 Dynamics of NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ S110A pS23 in the µs-ms time scale

To probe if conformational exchange contributes to transverse relaxation, we performed CPMG R_{ex} experiments on NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ S110A pS23.

Contributions to protein dynamics in the µs-ms timescale provide immense insights into biological functions of proteins, since most biological processes occur in the same timescale, as well as chemical exchange. Chemical exchange between two states A and B can generally be described as a dynamic equilibrium dictated by the rate constants k_1 for the reaction of A to B and k_{-1} for the reverse reaction of B to A and an exchange rate $k_{ex} = k_1+k_{-1}$. This can also be applied to binding events, where A and B represent unbound and bound forms of the protein. Effects of conformational exchange on R^2 can be obtained from CPMG experiments. The refocused "spin-echo" described before is perturbed by the existence of chemical exchange which alters the magnetic environment of spins. A 90 ° pulse followed by a $(\tau-180°-\tau)_n$ pulse train will suppress these effects, thus creating the

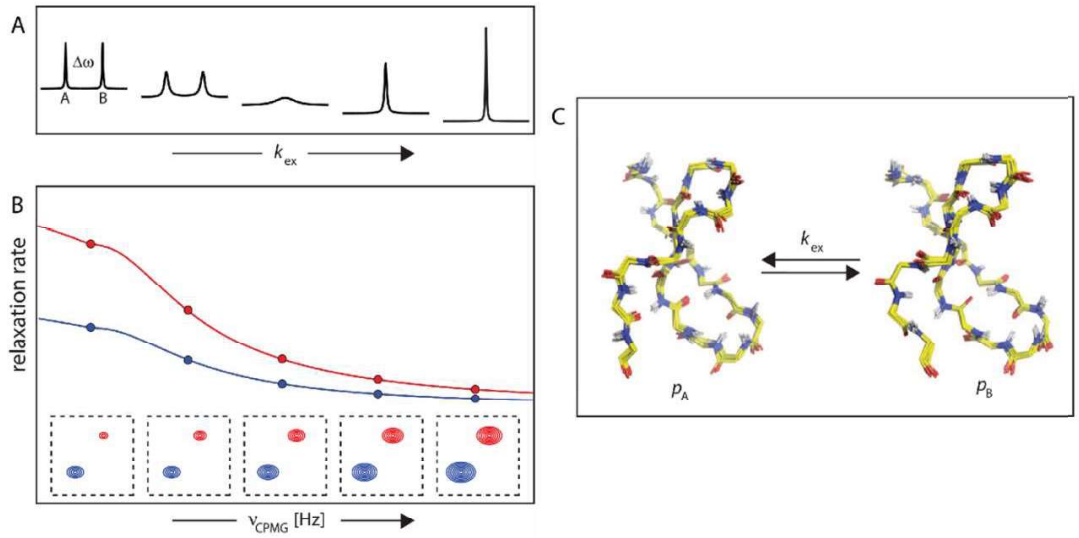


Fig. 6-22 A. *Peak distribution between two exchanging states A and B. Individual peaks for each state are visible for slow exchange regimes. They coalesce into one peak when k_{ex} increases and a fast exchange regime is obtained. Between slow and fast, at intermediate exchange regimes, line broadening is observed.* ***B.*** *Resonance intensities are measured at varying CPMG frequencies. From resulting peaks, relaxation dispersion profiles are obtained.* ***C.*** *Analysis of the relaxation dispersion profiles yield information about kinetic, thermodynamic and structural parameters of protein dynamics. Figure adapted from S. Grutsch, S. Brüschweiler, M. Tollinger,* NMR Methods to Study Dynamic Allostery. PLOS Comput. Biol. ***2016***, 12, *e1004620.*

opportunity to measure them by varying the amount of CPMG sequences n. These experiments are called relaxation dispersion experiments.[373] Depending on the

exchange constant (k_{ex}) relative to the difference between the two chemical shifts in state A and B ($\Delta\omega$), one can distinguish three different exchange rates:

$k_{ex} > \Delta\omega$ fast exchange (single peak at the population-weighted average of A and B)

$k_{ex} \approx \Delta\omega$ intermediate exchange (intensity loss due to line broadening)

$k_{ex} < \Delta\omega$ slow exchange (two peaks for both A and B state)

The use of a series of refocusing pulses with variable pulse spacing (CPMG) yields relaxation dispersion profiles that are characteristic for the underlying dynamics of the protein.[374] When transverse relaxation rates are plotted as a function of CPMG frequencies, dynamics related to chemical exchange can be identified by decreasing relaxation rates at higher ν_{CPMG} (see **Fig. 6-22**).

When these experiments where applied to $NFAT_{1\text{-}130}$ S110A and $NFAT_{1\text{-}130}$ pS23 S110A, amino acids outside of the regions of interest noticed in the T_2 experiments experienced no significant conformational exchange contribution. This is illustrated by transverse relaxation rates independent of CPMG frequency (see His18 as an example, **Fig. 6-23**).

On the contrary, conformational exchange was measured at regions with elevated transverse relaxation rated (i.e. amino acids 30-45, 65-75, 90-100 and 115-125). Especially the TAD revealed significant contributions from conformational exchange for unphosphorylated NFAT. In agreement with our observations from the T_2 experiments, phosphorylation largely abolished dynamics (see selected residues for these regions, **Fig. 6-23**).

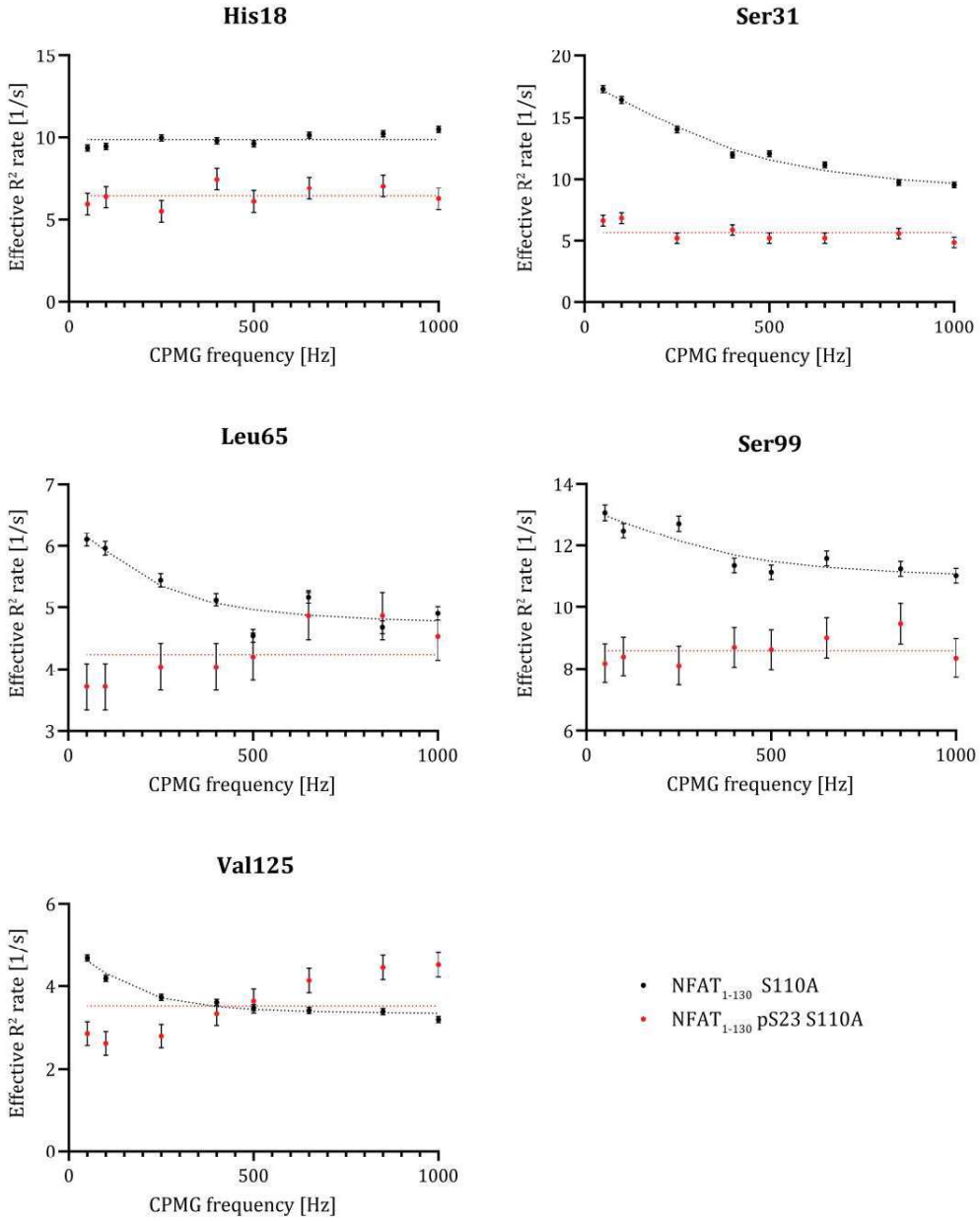


Fig. 6-23 *R_{ex} plots observed for amino acid regions 30-45, 65-75, 90-100 and 115-125 and His18 as an example for a residue without R_{ex}.*

6.2.1.5 Long-range interactions within the transactivation domain of $NFAT_{1-130}$ using paramagnetic relaxation enhancement (PRE)

In order to analyze intramolecular long-range interactions within the transactivation domain of NFAT, paramagnetic relaxation enhancement (PRE) tools were used to obtain structural insights into NFAT phosphorylation. In these experiments, site specific cysteine labeling with paramagnetic spin labels (i.e. 2,5-dihydro-2,2,5,5-tetramethyl-3-

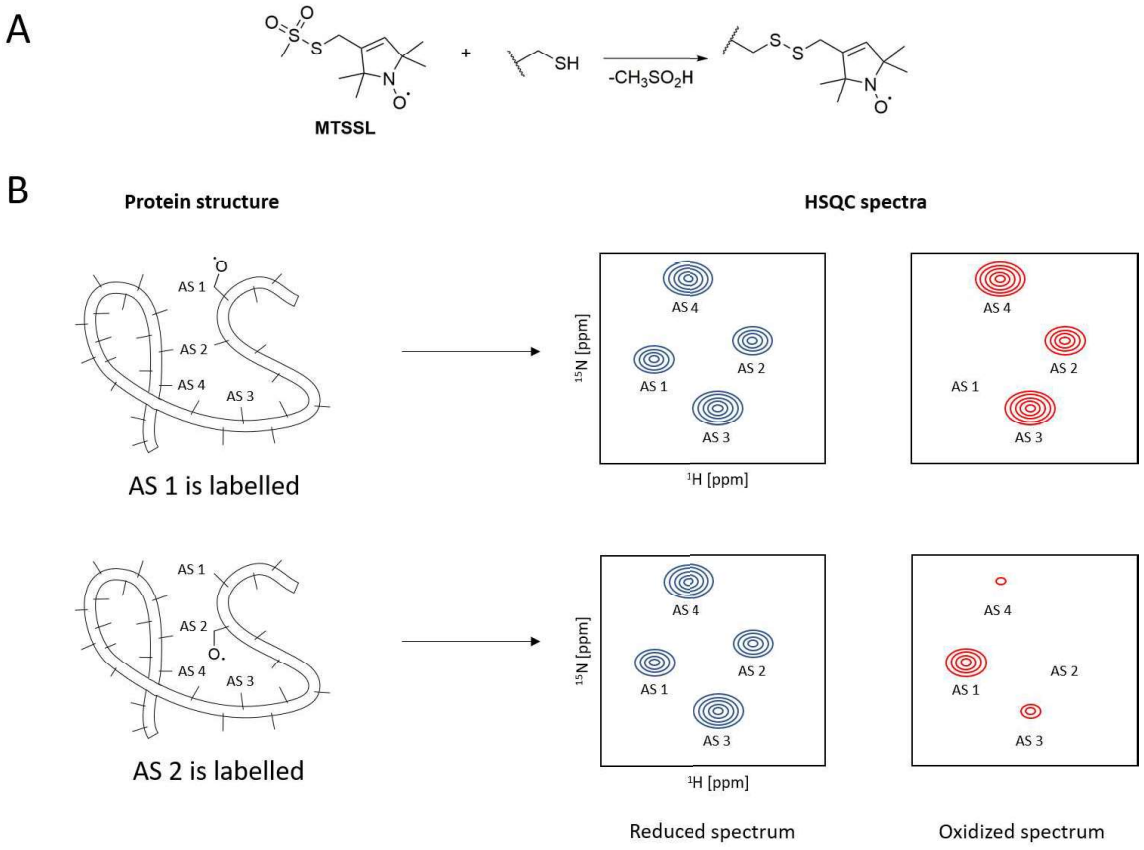


Fig. 6-24 *Principle of PRE experiments.* ***A.*** *MTSSL reacts with the free thiol group of cysteines to disulfide bonds. This results in spin-labeled cysteine residues on the protein.* ***B.*** *When an amino acid that is not involved in long-range interactions is labelled, the resulting ^{15}N-^{1}H-HSQC spectrum does not result in reduced peak intensities for other, non-labelled amino acid peaks (top red). When an amino acid participating in long-range interactions is labelled, neighboring amino acid peaks experience distance-dependent enhanced relaxation, resulting in induced intensities (bottom red). Intensity ratios between these spectra and reduced spectra can be generated by the addition of reducing agents such as ascorbic acid (blue spectra).*

[[(methylsulfonyl)thio]methyl]-1*H*-pyrrol-1-yloxy [MTSSL], see **Fig. 6-24 A**) is used to study the environment in proximity to a residue of interest.

^{15}N-^{1}H-HSQC spectra are recorded for the MTSSL labeled protein (oxidized spectrum) and after the addition of ascorbic acid as a reductant to reduce the oxygen radical (reduced spectrum). Amino acids that are in proximity to the spin label will experience enhanced relaxation due to the presence of the paramagnetic radical electron, resulting in a reduction of peak intensity in the HSQC spectrum (see **Fig. 6-24 B**). Plotting the ratio of peak intensity for oxidized and reduced spectra as a function of protein residue numbers will yield insights into long-range interactions that are present in the protein by "dips" of reduced intensity appearing at interacting locations.

The cysteine mutants listed in **Table 1** at different positions of the TAD of NFAT were cloned and expressed to probe long range interactions. Results were compared between unphosphorylated and phosphorylated proteins.

The positions for the spin labels were chosen based on interesting dynamic regions that were observed in the relaxation experiments. D14 was chosen as a control.

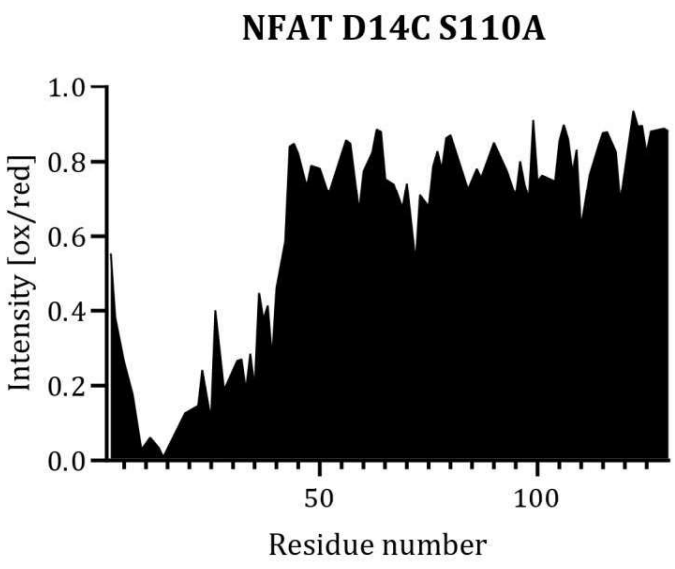


Fig. 6-25 *PRE plot of NFAT$_{1-130}$ D14C S110A. Intensity reduction is observed around the position of the spin label. No long-range interactions are observed.*

As can be seen in **Fig. 6-25**, the spin label in position 14 does not recognize significant long-range interactions. Reductions in peak intensity can be observed in proximity to C14, but not extending to the C-terminus.

Next, we analyzed spin labels at regions of NFAT that experience dynamics and compared results with and without phosphorylation.

Spin labels at positions 35, 65, 99 and 124 revealed interesting insights: the 9aaTAD seems to be in proximity to the rest of the protein,

including the C-terminus. Intensity reduction of ~ 50 % is seen throughout the full sequence of NFAT$_{1\text{-}130}$ when an MTSSL label is placed at position 35 (see **Fig. 6-26**). In accordance, spin labels at positions 65, 99 and 124 show peak reductions of ~ 30 % in the region between amino acids 30-40 in the case of unphosphorylated NFAT. In addition, interactions between all individual regions that exhibit enhanced dynamics through transverse relaxation are observed.

For phosphorylated NFAT, these long-range interactions change dramatically. Especially in the case of D35C, the overall reduction is completely abolished upon phosphorylation. Accordingly, an increase in peak intensity is observed for phosphorylated protein with spin labels in the other positions.

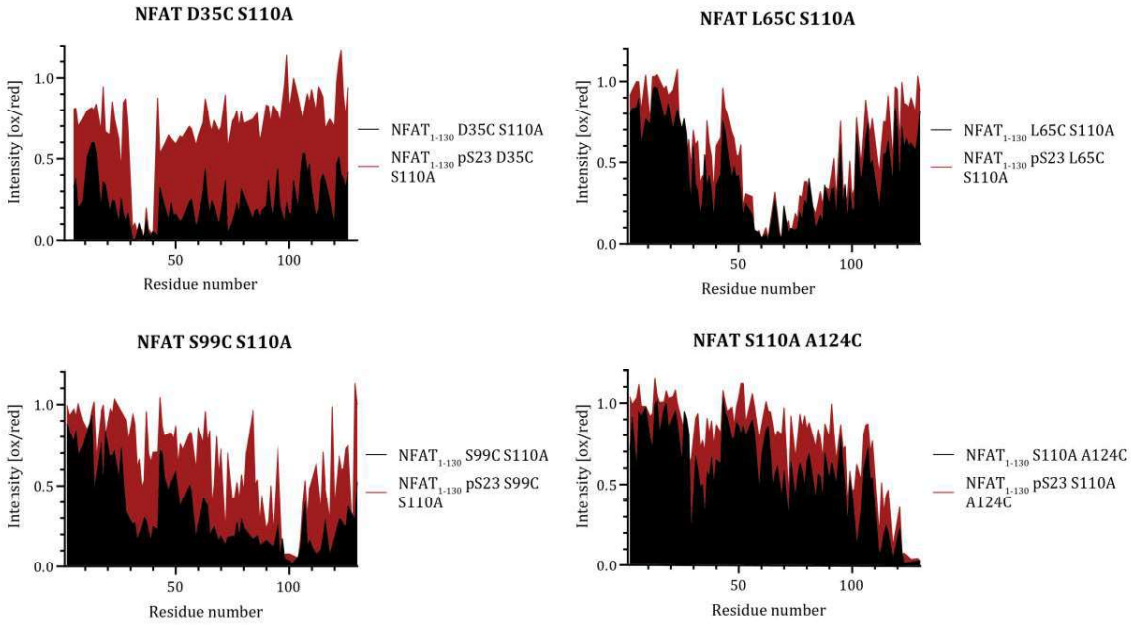


Fig. 6-26 *PRE plots of NFAT$_{1\text{-}130}$ D35C S110A, NFAT$_{1\text{-}130}$ L65C S110A, NFAT$_{1\text{-}130}$ S99C S110A and NFAT$_{1\text{-}130}$ S110A A124C in their unphosphorylated and phosphorylated forms.*

6.2.1.6 Electron paramagnetic resonance (EPR) using the double electron-electron resonance (DEER) technique

To further analyze the distance distributions for unphosphorylated and phosphorylated NFAT, electron paramagnetic resonance (EPR) was used with NFAT$_{1\text{-}130}$ D35C S99C S110A and NFAT$_{1\text{-}130}$ pS23 D35C S99C S110A. For these experiments, two MTSSL tags were attached to cysteines of the aforementioned proteins. The double electron-electron

resonance (DEER) experiment is used to measure distances between two different electron spins (in our case, the spin labels at amino acids 35 and 99). The EPR experiments were performed by our collaborators in the lab of DANIELLA GOLDFARB at the Weizmann Institute of Science. The data showed relatively similar distance distributions for NFAT$_{1\text{-}130}$ D35C S99C S110A and NFAT$_{1\text{-}130}$ pS23 D35C S99C S110A (see **Fig. 6-27**). Two major distances are observed: a major population ~ 3.3 nm and a smaller population at ~5-5.5 nm. It can be noted that the population for the longer distance interaction decreases upon phosphorylation towards shorter distances. However, the differences are subtle, hence the data will only be viewed as preliminary.

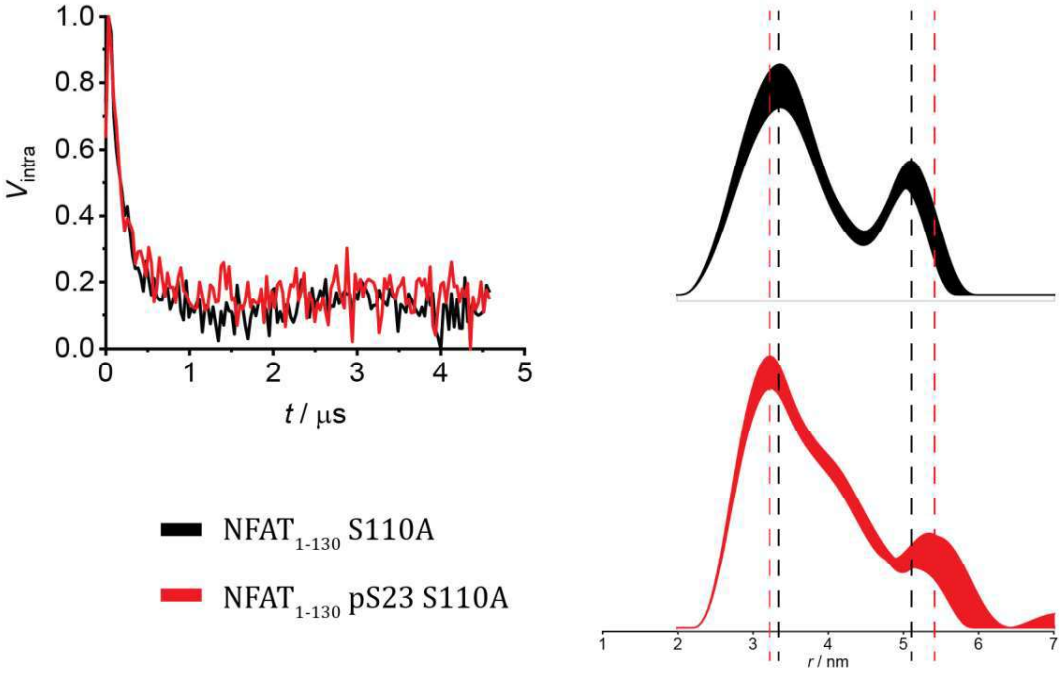


Fig. 6-27 *DEER measurement of NFAT$_{1\text{-}130}$ D35C S99C S110A and NFAT$_{1\text{-}130}$ pS23 D35C S99C S110A. The left graph shows normalized, background-corrected DEER data in the time domain. On the right, this data is translated into distance distributions.*

6.2.1.7 Phosphorylation influences the binding of CBP KIX to NFAT$_{1\text{-}130}$

After analyzing the dynamics of the transactivation domain of NFAT, we wanted to test if phosphorylation had effects on the binding of coactivator/mediator proteins. ^{15}N-^{1}H-

HSQC spectra were recorded with ^{15}N-labeled NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ S110A pS23 and the unlabeled KIX domain of CBP/p300. The interaction interface was compared with data obtained from protein dynamics (see **Fig. 6-28**).

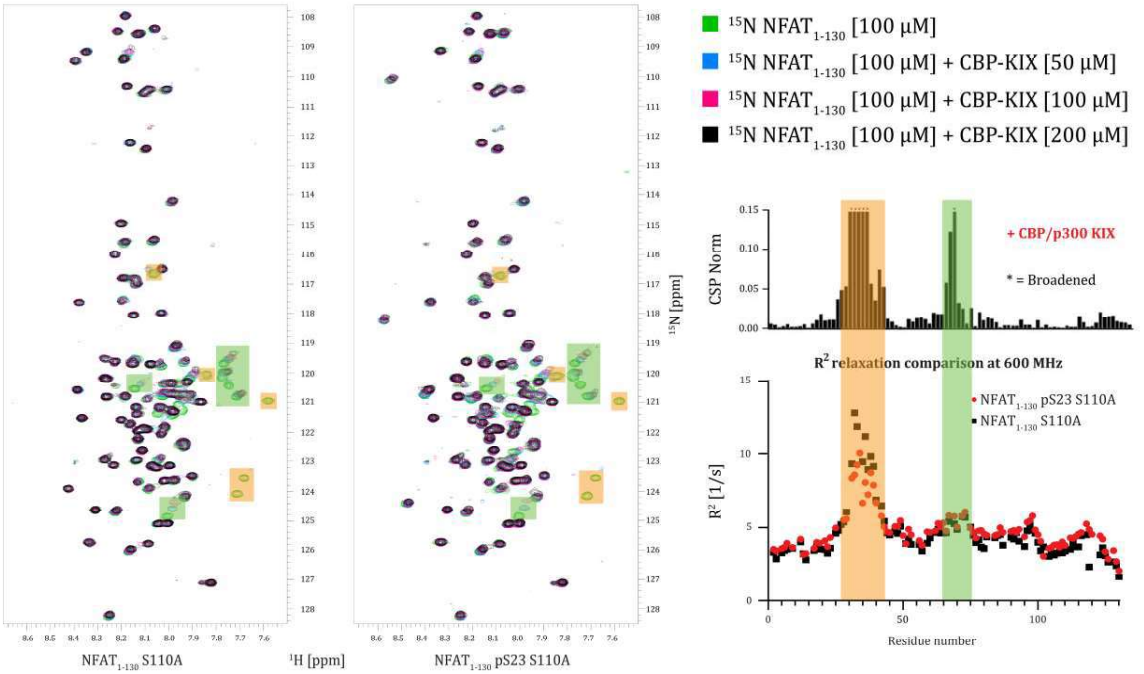


Fig. 6-28 *Interaction between NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ pS23 S110A and CBP/p300 KIX. ^{15}N ^{1}H HSQC titrations of ^{15}N-labeled NFAT with varying concentrations of CBP/p300 KIX were performed. The interaction interface, determined by CSP Norm plotted against residue number, aligns with residues that show reduced R^2 values.*

The NMR experiments were carried out with ^{15}N NFAT alone (green spectrum) that was then titrated with varying amounts of CBP KIX (0.5 equivalents = blue, 1 equivalent = red, 2 equivalents = black). As can be seen in the spectra, most of the interacting region of NFAT is in an intermediate exchange regime with the KIX domain. This manifests itself in the disappearance of peaks around amino acids 30-40 (highlighted in orange). In addition, another region of NFAT that participates in the binding of CBP KIX is seen around amino acids 60-70 (highlighted in green). These peaks are in a fast exchange regime with its interaction partner; therefore, their position moves and chemical shift perturbations (CSPs) can be plotted. Interestingly, both regions also exhibit dynamics in the ps-ns timescale (see **Fig. 6-28**). Although visually, binding of CBP KIX to phosphorylated looks tighter (peak intensity of interacting residues is lower in the

phosphorylated form), the difference in binding is hard to quantify from the NMR experiments. ITC measurements were attempted both with NFAT as the component in the cell and CBP KIX as a ligand in the syringe as well as CBP KIX in the cell and NFAT as the ligand. Both experiments did not result in heat production that could be used to interpret binding data, possibly due to the disordered nature of NFAT. Therefore, MST was carried out as an alternative method to quantify the protein-protein interaction.

The MST titration of phosphorylated NFAT with CBP KIX and unphosphorylated NFAT with CBP KIX revealed a ~ 3-fold increase in binding affinity of the KIX domain to NFAT when it is phosphorylated (see **Fig. 6-29**). This observation is in line with the fact that p38 inhibition was shown to inactivate NFAT in T cells and that p38 MAPK promotes the interaction of NFAT to CBP/p300.[342]

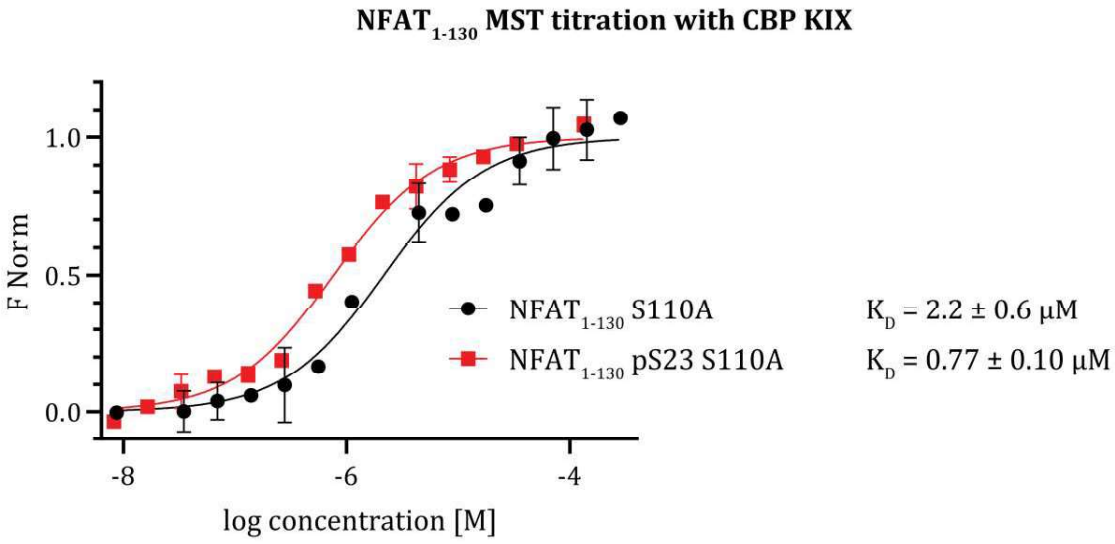


Fig. 6-29 *K_D values of $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A with CBP KIX determined by MST.*

6.2.1.8 Biological influence of NFAT phosphorylation at Ser23

Next, the influence of NFAT phosphorylation at position 23 was examined on a cellular level. Therefore, we obtained NFAT1 (NFATC2) (NM_173091) Human Tagged ORF Clone from OriGene as a plasmid to overexpress full-length GFP-tagged NFAT in mammalian cells. Cloning was performed to generate the plasmids listed in **Table 2.**

MUTANT	PURPOSE
GFP-NFAT1 S23A	Phospho-null mutant to evaluate absence of S23 phosphorylation
GFP-NFAT1 S23D	Phosphomimetic mutant to study effects of S23 phosphorylation
GFP-NFAT1 S110A	Phospho-null mutant to evaluate absence of S110 phosphorylation
GFP-NFAT1 S110D	Phosphomimetic mutant to study effects of S110 phosphorylation

Table 2 *NFAT mutants cloned to study the effect of phosphorylation in cells.*

GFP-tagged NFAT mutants were transfected into HEK293T cells and fluorescence microscopy was used to study cellular localization of NFAT. To evaluate effects on transcriptional activity, transfected HEK293T cells were co-transfected with pGL3-NFAT luciferase, a plasmid containing a triple NFAT binding sequence upstream of a luciferase reporter. After NFAT-pathway activation with PMA/ionomycin for 2 hours, transcription of IL-2 was compared between NFAT mutants by readout of luciferase-induced signals.

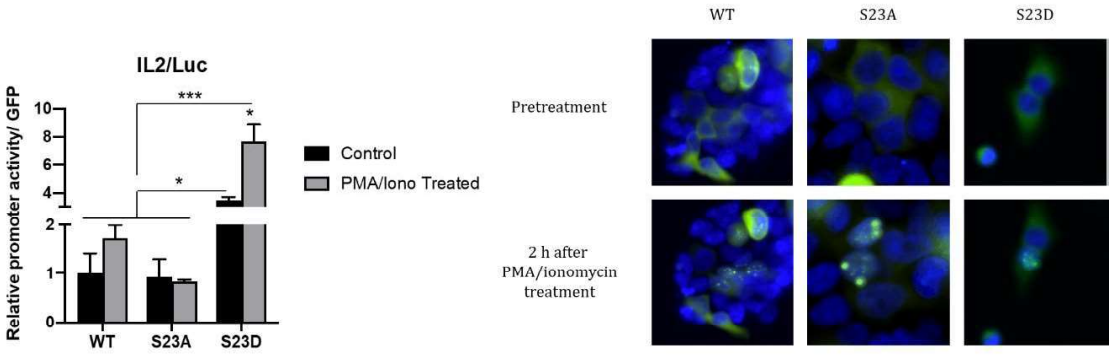


Fig. 6-30 *IL-2 activation capability and cellular localization of NFAT S23X mutants. While promoter activity for wild type NFAT increase ~ 2-fold, no increase is observed for NFAT S23A. NFAT S23D on the other hand shows significant promoter activity in resting cells, which increases ~ 2-fold upon PMA/ionomycin treatment. Activity levels are adjusted for transfection efficiency by normalizing luciferase readouts to GFP levels. Fluorescence microscopy reveals no influence on cellular localization of either NFAT S23A or NFAT S23D.*

Fig. 6-30 shows IL-2 promoter activity for GFP-NFAT before and after treatment with PMA/ionomycin. Both wild type NFAT and the NFAT S23D mutant show a ~ 2-fold increase of IL-2 promoter activity after activation of the NFAT pathway. NFAT S23A does not show any increasing IL-2 promoter activity after PMA/ionomycin treatment, suggesting that the loss of phosphorylation in position 23 decreases transcriptional activity of NFAT. Luciferase signals are normalized as a function of GFP signal to compensate for different overexpression rates of the individual plasmids. NFAT S23D shows 4-fold increased activity in untreated cells relative to wild type NFAT, underlining the importance of phosphorylation at S23 in activating NFAT target genes. Fluorescence microscopy images of NFAT before and after treatment with PMA/ionomycin reveal the cellular localization of NFAT. Nuclei are stained blue with Hoeachst33342 and NFAT appears in green due to the GFP tag. Nuclear translocalization of NFAT can be observed by green foci forming in the nuclei. As can be seen, both the S23A as well as the S23D mutant display unaltered nuclear translocalization relative to wild type NFAT after 2 hours of treatment with PMA/ionomycin. Therefore, translocalization effects contributing to loss of transcriptional activity can be excluded. No significant effects were observed for either GFP-NFAT S110A or S110D (data not shown).

In conclusion, phosphorylation of NFAT at Ser23 alters protein dynamics, interactions with coactivators/mediators and ultimately influences transcriptional activity. The phosphorylation appears to induce a conformational switch to unravel the transactivation domain, which is masked by interactions with the rest of the N-terminus of NFAT1. These dynamics increase binding of CBP/KIX which, as the cellular data suggests, is crucial for activating downstream genes such as IL-2.

It remains to be seen how this phosphorylation affects other binding partners of NFAT such as the KIX domain of Med15 or the scaffold protein Homer. Furthermore, future focus will be on the second phosphorylation addressed by p38 MAPK, S110, which is located just before the calcineurin binding site PRIEIT.

6.2.2 NFAT as a model construct for protein engineering

NFAT was used as a model to introduce new domains into intrinsically disordered domains of proteins to test whether biological pathways can be altered without affecting

the function of the protein. The idea was to replace the canonical calcineurin binding site PxIxIT (PRIEIT in NFAT1) with binding sites of the ubiquitously expressed phosphatase PP2A to generate a variant of NFAT that does not rely on calcium influx for activation.

6.2.2.1 PP2A-B56 binding motif from the *Ebola* virus nucleoprotein

It has been known for a long time that the transcriptional activator VP30 is required for the transcription of the *Ebola* genome.[375] More recently, it has been elucidated that regulation of VP30 is achieved through reversible phosphorylation.[376] The molecular details of this phosphorylation cycle have long been unknown. In 2018, KRUSE et al.[377] identified a conserved motif in *filovirus* species that are required for the dephosphorylation of VP30. The LxxIxE sequence, which was established to be a docking motif of the phosphatase PP2A earlier[378], was found to be necessary for VP30 interaction with PP2A-B56 and subsequent dephosphorylation mediated by the phosphatase. The full sequence, PRMLTPINEEAD, should serve as a recruiting motif for PP2A-B56 phosphatase. Therefore, this study aims to test the possibility of introducing this sequence as a replacement for the calcineurin binding motif PRIEIT in NFAT1 to create a constitutively active NFAT version that does not require calcium influx for transcriptional activation. This will be investigated by observation of cellular localization of GFP-tagged NFAT using fluorescence microscopy and cellular readout of downstream gene activation of NFAT (i.e. IL-2).

A different sequence, PRMATPANAEAD, will be incorporated in the same position in place of the sequence discussed before. The three conserved amino acids from the LxxIxE motif are mutated to alanines, resulting in a significant decrease in PP2A binding affinity (unpublished results from the DAVEY lab). The construct containing this sequence will serve as a control plasmid to compare results obtained from the high-affinity PP2A motif.

6.2.2.2 PP2A-B56 binding motif from BubR1

BubR1 is a mitotic checkpoint kinase involved in the spindle assembly checkpoint (SAC) and necessary for proper kinetochore–microtubule interactions.[379–383] It has been shown that phosphorylation of outer kinetochore proteins is required for the detachment from microtubules to free up the kinetochore to form new interactions.[384,385] PP2A has been identified to dephosphorylate those proteins through

its interaction with BubR1.[386] The PP2A docking motif, IKKLSPIIEDSR, contains the canonical LxxIxE sequence mentioned before. Interestingly, it was shown that phosphorylation of the SP motif in this sequence by cyclin-dependent kinase 1 (CDK1) or polo-like kinase 1 (Plk1) is required for binding to PP2A.[386] This sequence is included in these studies to create a tunable NFAT version, whose constitutive nuclear localization can be interfered with using inhibitors for either CDK1 or Plk1.

6.2.2.3 Cloning of NFAT constructs to engineer function

IVA cloning was used to create five constructs; two plasmids in which the PP2A binding motif from Ebola virus was inserted into the N-terminus of full length GFP-NFAT1 and three plasmids in which PP2A motifs were cloned replacing the SPRIEIT motif in full length GFP-NFAT1 (see **Table 3**).

The plasmids that contained both PP2A motifs as well as calcineurin binding motifs were prepared, but not yet used within this study. In future efforts, they will be used to study the effect of two control mechanism within the same construct.

No	Construct	Includes PRIEIT	Purpose
1	$_{6}$**PRMLTPINEEAD**$_{17}$	yes	Investigation of calcineurin binding in addition to PP2A dephosphorylation by the Ebola virus motif
2	$_{130}$**PRMLTPINEEAD**$_{141}$	yes	Investigation of calcineurin binding in addition to PP2A dephosphorylation by the Ebola virus motif
3	$_{110}$**PRMLTPINEEAD**$_{121}$	no	Effects of high affinity PP2A binding motif from Ebola virus
4	$_{110}$**PRMATPANAEAD**$_{121}$	no	Effects of low affinity PP2A binding motif from Ebola virus
5	$_{110}$**IKKLSPIIEDSR**$_{121}$	no	Effects of PP2A binding motif from BubR1 under phosphorylation control

Table 3 *NFAT constructs with inserted PP2A binding motifs.*

6.2.2.4 PP2A motifs from the *Ebolavirus* nucleoprotein

HEK293T cells were transfected with plasmids of constructs that contain high and low affinity PP2A motifs (see **Table 3**) and cellular localization of NFAT was analyzed using fluorescence microscopy to detect the GFP-tag and nuclei stained with Hoechst33342. Wild type NFAT was used as a control to follow nuclear translocalization after PMA/ionomycin treatment (see **Fig. 6-31**).

Pre PMA/ionomycin treatment, NFAT (green due to the GFP-tag) localizes mainly within the cytoplasm. Co-localization with nuclei occurs 2 hours after activation of the NFAT pathway with PMA/ionomycin, observed by green foci that co-stain with Hoechst33342.

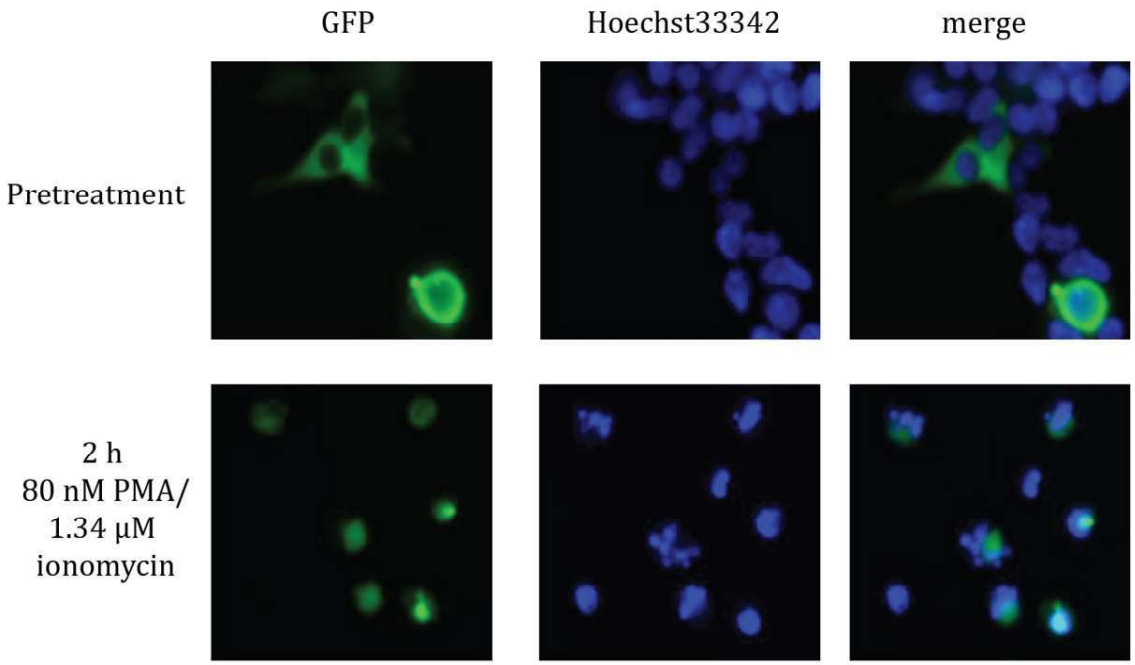


Fig. 6-31 *Cellular localization of wild type GFP-NFAT before and after activation of the T cell receptor pathway.*

Next, the same experiment was conducted using construct no 3. As expected, NFAT plasmids containing the high affinity PP2A binding motif PRMLTPINEEAD express constitutively active NFAT, which translocates to the nucleus without calcium influx stimulation (see **Fig. 6-32**).

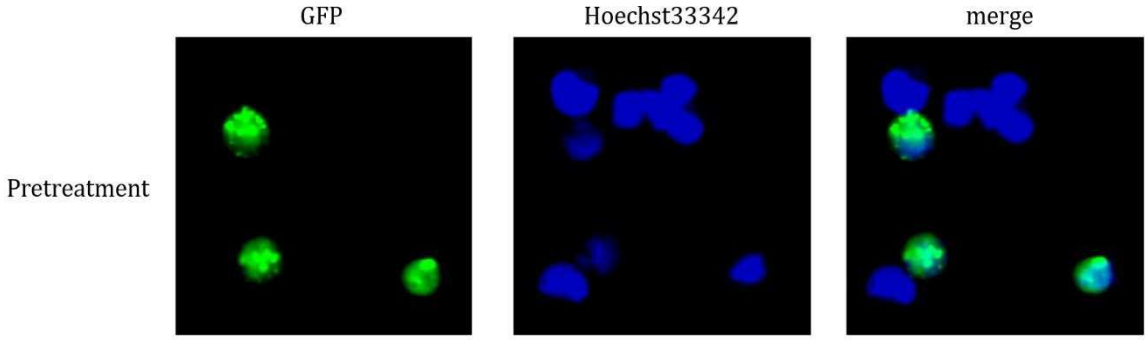


Fig. 6-32 *Cellular localization of construct no 3 including a high affinity PP2A binding motif from the Ebola virus.*

When the PP2A motif is mutated to PRMATPANEEAD, a sequence with lower affinity, cytosolic NFAT can be detected alongside NFAT localized within the nucleus (see **Fig.**

6-33). This data suggests that NFAT dephosphorylation and subsequent nuclear translocalization is less likely to occur with the weaker PP2A binding sequence.

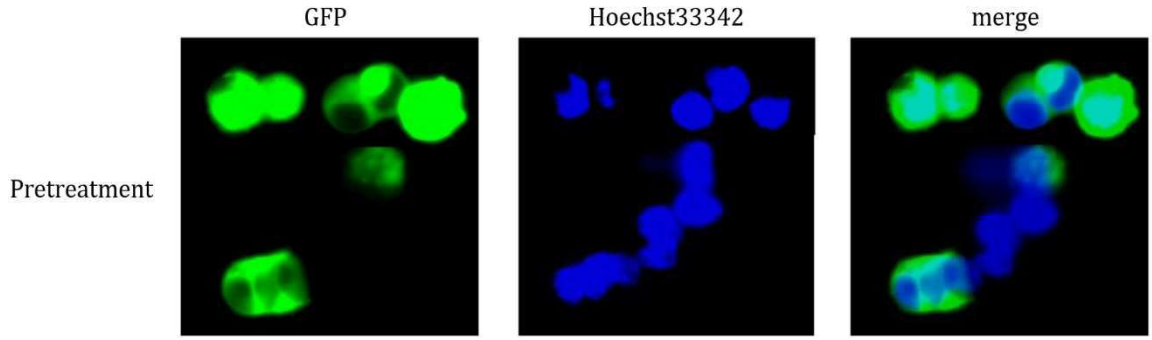


Fig. 6-33 *Cellular localization of construct no 4 including a low affinity PP2A binding motif from the Ebola virus.*

6.2.2.5 PP2A binding motif from the kinase BubR1

Lastly, the PP2A binding motif IKKLSPIIEDSR from BubR1 in which the SP motif needs to be phosphorylated by CDK1 was inserted into GFP-NFAT. Cellular localization of NFAT

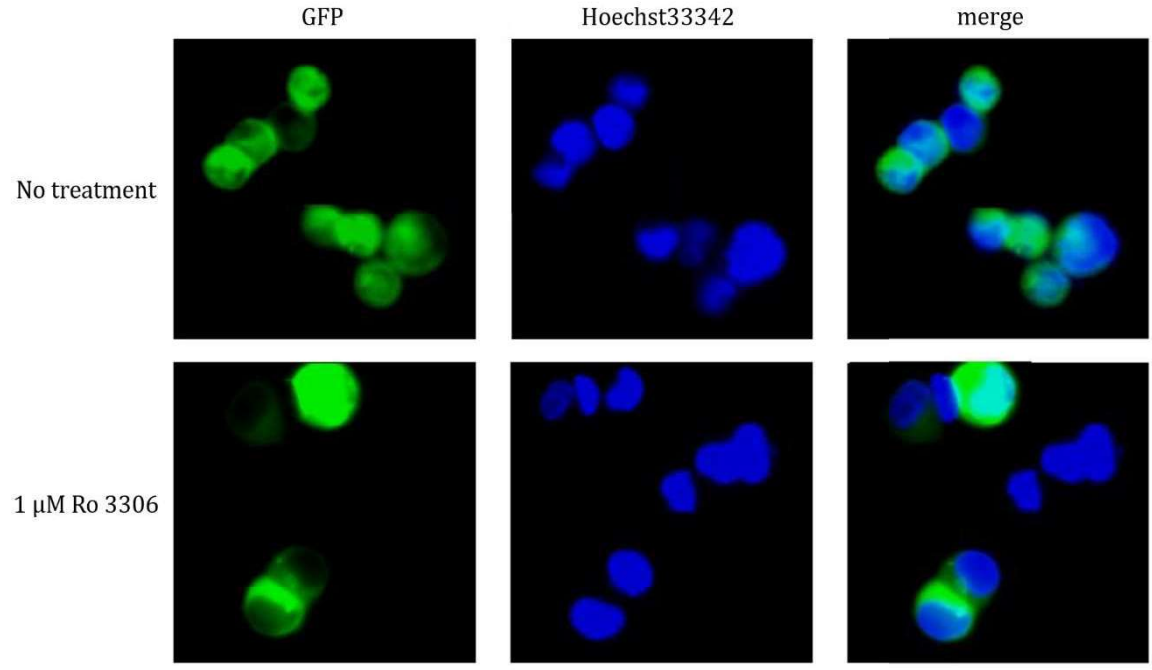


Fig. 6-34 *Cellular localization of construct no 5 including the PP2A binding motif from BubR1. The top panel shows cellular localization of protein in the absence of a CDK1 inhibitor, the bottom shows cellular localization after treatment with 1 µM Ro 3306, a selective CDK1 inhibitor.*

was observed in HEK293T cells in the presence and absence of 1 µM Ro 3306, a selective CDK1 inhibitor (see **Fig. 6-34**).[387]

Whereas NFAT predominantly localizes in nuclei in the absence of CDK1 inhibitors, inhibition of CDK1 with Ro 3306 leads to most cells displaying NFAT localized in the cytosol. This data suggests that phosphorylation of the PP2A binding motif by CDK1 is required for binding of the phosphatase and subsequent translocalization of NFAT to the nucleus.

6.2.3 Discussion

$NFAT_{1\text{-}130}$ was successfully phosphorylated at Ser23 as the single phosphorylation site by the creation of the $NFAT_{1\text{-}130}$ S110A mutant. NMR and MST experiments shed light on the dynamic behavior of the protein in its phosphorylated and dephosphorylated forms, as well as provided insights into altered interaction affinity to the mediator protein CBP/p300 KIX. Overall, flexible regions in the 9aaTAD of $NFAT_{1\text{-}130}$ as well as in other places within the sequence were identified and rigidization of the 9aaTAD was observed upon phosphorylation. Furthermore, PRE and EPR experiments provided first insights into structural organization of the intrinsically disordered N-terminal domain of NFAT. While long-range interactions are found in unphosphorylated NFAT in both PRE and EPR experiments, these interactions are largely abolished upon phosphorylation. Hypothetically, phosphorylation induces a conformational switch in $NFAT_{1\text{-}130}$, by which the 9aaTAD sequesters from the rest of the domain and becomes more available for binding of other proteins. This hypothesis is emphasized by the results found in transcriptional activities based on IL-2 level monitoring performed on NFAT and NFAT mutants. While phosphomimetic (S23D) mutants were able to induce IL-2 production upon T cell receptor activation, phosphor-null (S23A) mutants failed to do so. However, there are likely more factors adding to the conformational rearrangement of NFAT induced by single-site phosphorylation that need to be studied in the future.

Furthermore, NFAT was used as a model protein for the inclusion of protein domains into intrinsically disordered regions to alter the protein's functions. In conclusion, these techniques appear to be working. Manipulation of stimuli that lead to nuclear translocalization of NFAT was successfully achieved in full-length NFAT. Future

endeavors will aim to analyze if transcriptional activity of NFAT is retained and constitutive in these constructs by using the IL-2 luciferase reporter gene assay described in **6.2.1.8.**

6.3 eIF4E

6.3.1 Expression and purification of eIF4E wild type, S209D and pS209

Plasmids expressing wild type eIF4E, eIF4E S209D and eIF4E S209X (X = any amino acid) were created as GB1-TEV constructs for ITC and NMR experiments (GB1 acts as a solubility and stability tag[388]) as well as Δ_{26}-eIF4E constructs for crystallography. The first 26 amino acids of the protein are disordered, hence the constructs for crystallographic purposes lack this region. eIF4E and eIF4E S209D were recombinantly expressed in *E. coli* BL21 (DE3), while eIF4E pS209 was produced using cotransformation of an orthogonal aminoacyl-tRNA synthetase (aaRS) for phosphoserine into *E. coli* BL21ΔserB. Expression of phosphorylated eIF4E worked, as was confirmed by enzymatic digestion and subsequent mass spectrometry (100 % SG(pS)TTK was detected in trypsin digests and 100 % DTATKSG(pS)TTKNRFVV was detected in AspN digests) as well as SDS-PAGE gel analysis (see **Fig. 6-35**). The increase in apparent molecular weight seen on the SDS-PAGE is expected for phosphorylated proteins.[389]

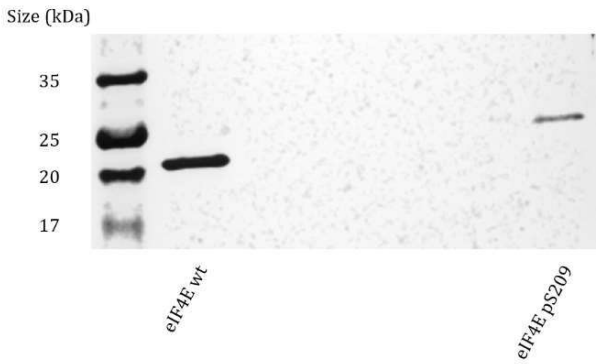
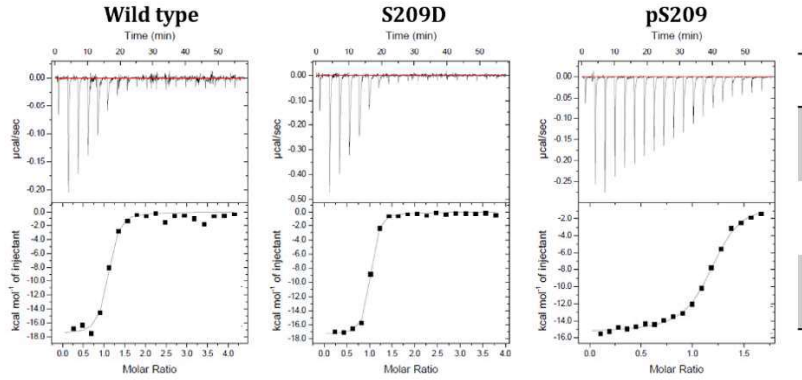


Fig. 6-35 *SDS-PAGE of wild type eIF4E and eIF4E pS209. eIF4E pS209 experiences a so-called phosphorylation dependent mobility shift (PDMS) due to decreased binding to SDS.*

6.3.2 Cap binding affinity comparison for eIF4E variants

To confirm that phosphorylated eIF4E exhibits decreased cap binding affinity and to assess if this is true for the phosphomimetic S209D mutant, ITC experiments using cap-free GB1-eIF4E with m^{7}GTP were performed.



Protein	K_D [nM]
GB1-eIF4E wild type	61 ± 25
GB1-eIF4E S209D	79 ± 14
GB1-eIF4E pS209	174 ± 14

Fig. 6-36 *ITC of GB1-tagged eIF4E, eIF4E S209D and eIF4E pS209 with m^{7}GTP.*

ITC experiments revealed that binding of m^{7}GTP to phosphorylated eIF4E was ~ 3-fold weaker compared to wild type eIF4E, which agrees with previously published results.[59,220] Interestingly, the phosphomimetic eIF4E S209D bound m^{7}GTP with the same affinity as wild type eIF4E, within the margin of error. The same phenomenon was observed by SCHEPER et al.[220] Therefore, the following ITC experiments focused on comparison between eIF4E wild type and eIF4E pS209 only. All three eIF4E variants bound m^{7}GTP enthalpy-driven with negative ΔH values being comparable.

After confirming literature-reported decrease in cap-binding affinity, the influence of mRNA cap variation on the first transcribed nucleotide was investigated. To exclude effects coming from unspecific interactions, these ITC experiments were performed using Δ_{26}-eIF4E and Δ_{26}-eIF4E pS209 without the GB1 tag. The used cap-analogues were; (i) m^{7}GG (*N*^{7}-methyl-guanosine-5'-triphosphate-5'-guanosine) as an unmodified dinucleotide, (ii) m^{7}G(m)G(m)G (m^7(3'OMeG)(5')ppp(5')(2'OMeG)pG) as a cap 1 analogue used in synthetic cap structures, (iii) m^{7}GA(m)G (m^{7}G(5')ppp(5')(2'OMeA)pG) as a cap 1 analogue with A being the first transcribed nucleotide and (iv) m7GG(m)G (m^{7}G(5')ppp(5')(2'OMeG)pG) as a cap 1 analogue with G being the first transcribed nucleotide. These structures are shown in **Fig. 6-37**.

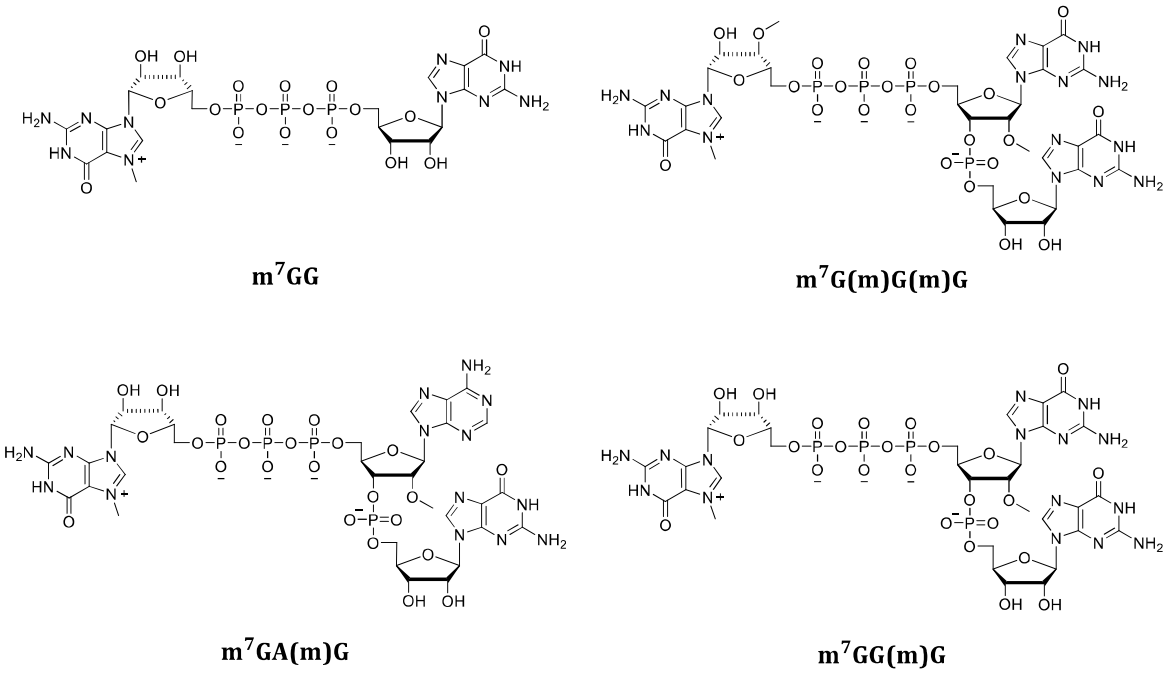


Fig. 6-37 *Structures of the four cap analogues used for ITC studies with eIF4E wild type and eIF4E pS209*

For the unmodified dinucleotide cap analogue m^7GG, a difference of ~ 3-fold between unphosphorylated and phosphorylated eIF4E was found (see **Fig. 6-38**). Furthermore, both titrations reveal a binding stoichiometry of 1:2 m^7GG to eIF4E. This is a different stoichiometry compared to a 1:1 binding found for the simpler m^7GTP cap analogue.

The interaction between eIF4E and the synthetically used $m^7G(m)G(m)G$ cap analogue is the most error prone of all measured interactions, due to low signals induced by ligand binding (see **Fig. 6-39**). $m^7G(m)G(m)G$ is methylated at the 3' oxygen of the m^7G moiety at the 5' end of the trinucleotide. These modifications, so-called anti-reverse cap analogues (ARCAs), don't occur naturally, but are used when generating semisynthetic mRNAs, because they are exclusively incorporated in forward orientation.[390] If m^7G-di- or oligonucleotides without this modification are used, the m^7G end is also incorporated towards the 3'-end of the mRNA. Interestingly, while the binding affinity between wild type and phosphorylated eIF4E is not affected significantly, the entropic binding

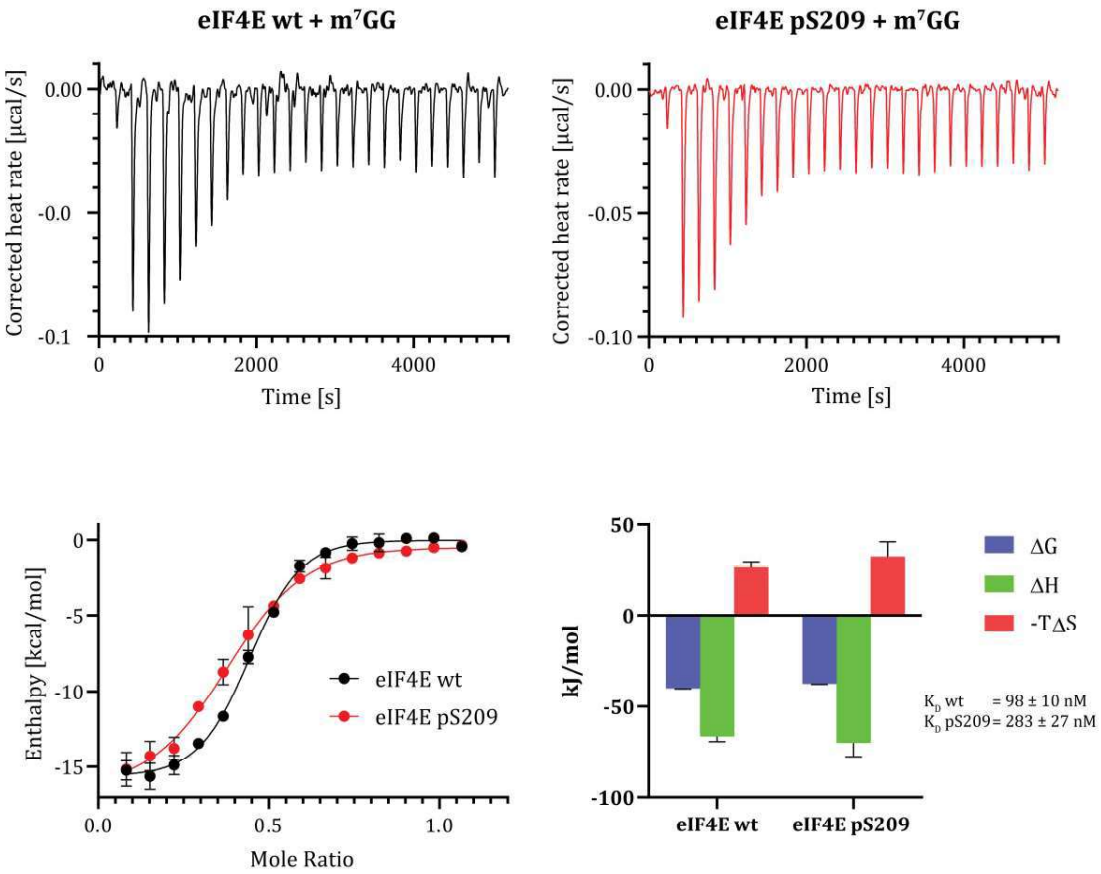


Fig. 6-38 *ITC titration of eIF4E wt and eIF4E pS209 with m7GG. m7GG binds to eIF4E wt with a dissociation constant of 98 nM and ~ 3-fold weaker to eIF4E pS209 with a dissociation constant of 283 nM. The interaction is enthalpy-driven in both cases.*

component is much more pronounced for eIF4E pS209.

The trinucleotide m7GA(m)G represents a so-called cap 1 structure. These cap modifications are employed by higher eukaryotes (including insects and mammals) and some viruses, for which the first transcribed nucleotide undergoes 2'-O-methylation in the nucleus. 2'-O-methylation on cellular mRNAs is used to detect viral RNAs, which are recognized by sensors such as MDA5 and RIG-1.[391–394] Recent studies have found that specifically 2'-O-methylation on the first transcribed nucleotide is important for blocking both exoribonuclease and decapping activities of DXO enzymes. This leads to specific

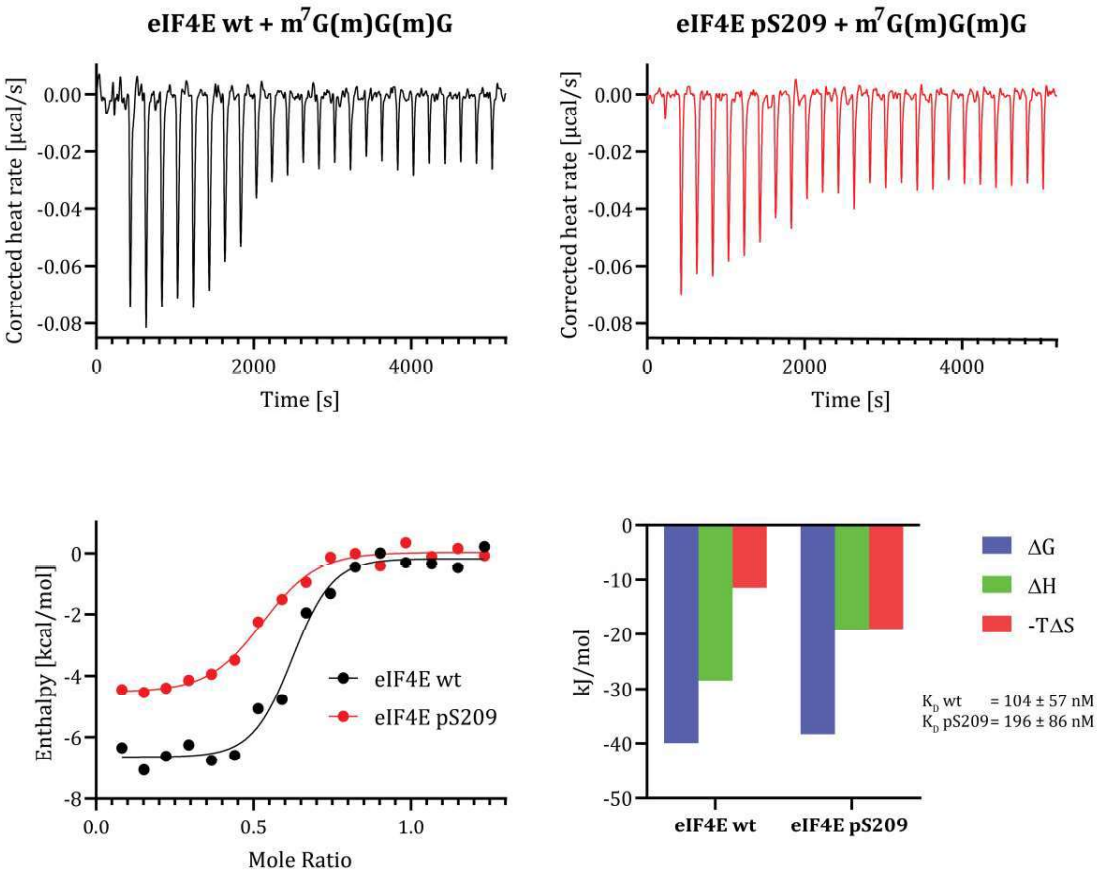


Fig. 6-39 *ITC titration of eIF4E wt and eIF4E pS209 with m⁷G(m)G(m)G. m⁷G(m)G(m)G binds to eIF4E wt with a dissociation constant of 104 nM and ~ 2-fold weaker to eIF4E pS209 with a dissociation constant of 196 nM. The low heat created from this interaction causes low signal to noise ratios and therefore a high error which arises from the sigmoidal fit. Unlike m⁷GG, m⁷G(m)G(m)G has an entropic contribution to the binding event.*

degradation of incompletely capped cellular mRNA transcripts.[395]

The difference and binding affinities for phosphorylated and unphosphorylated eIF4E to m7GA(m) is relatively low, with eIF4E pS209 binding this cap analogues slightly weaker (see **Fig. 6-40**). For both protein isoforms, the binding event is mostly enthalpy-driven.

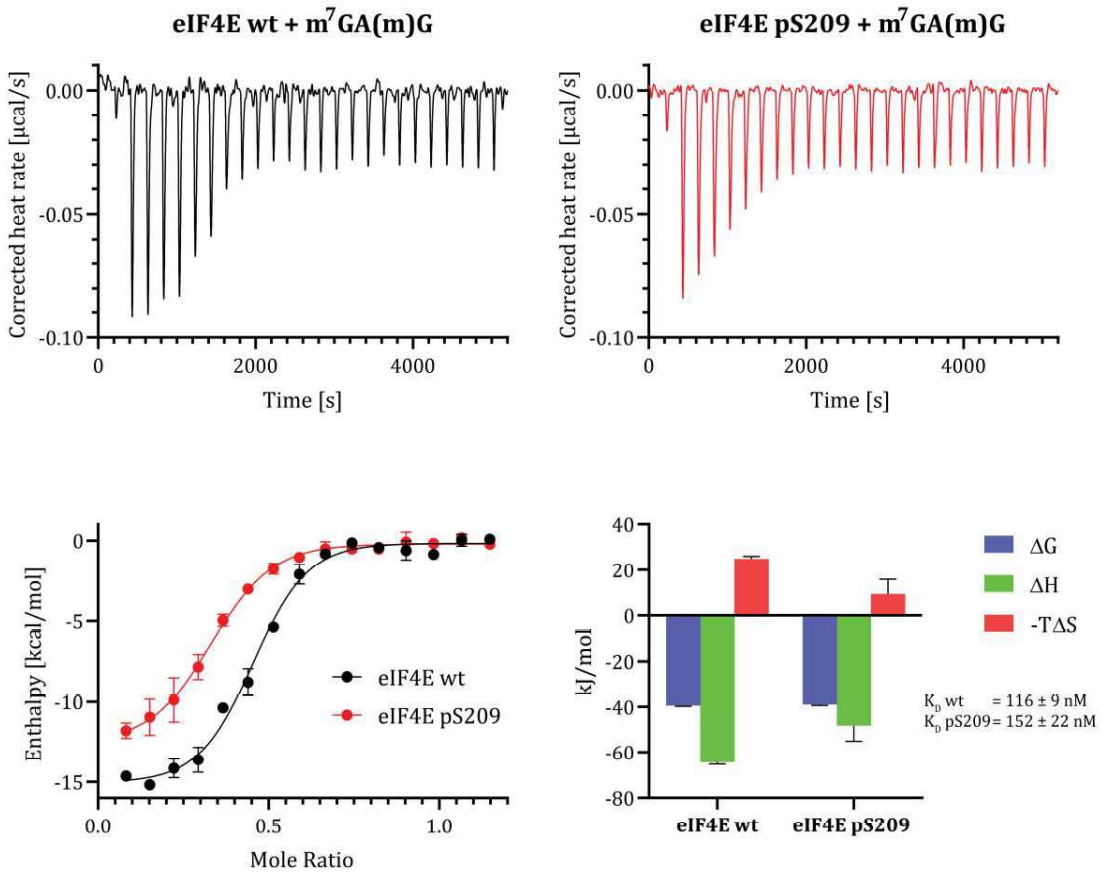


Fig. 6-40 *ITC titration of eIF4E wt and eIF4E pS209 with m⁷GA(m)G. m⁷GA(m)G binds to eIF4E wt with a dissociation constant of 116 nM and only slightly weaker to eIF4E pS209 with a dissociation constant of 152 nM. The interaction is enthalpy-driven in both cases.*

The most dramatic difference is observed for m⁷GG(m)G. Like m⁷GA(m)G, m⁷GG(m)G is a cap 1 analogue, with the only difference being the identity of the first transcribed nucleotide. In this case, the binding affinity differs by a factor of ~ 6 (see **Fig. 6-41**). These results suggest that phosphorylated eIF4E is capable of distinguishing between differently capped mRNA nucleotides, depending on the first transcribed nucleotide. However, not much is known about the differences of this change on a biological level.

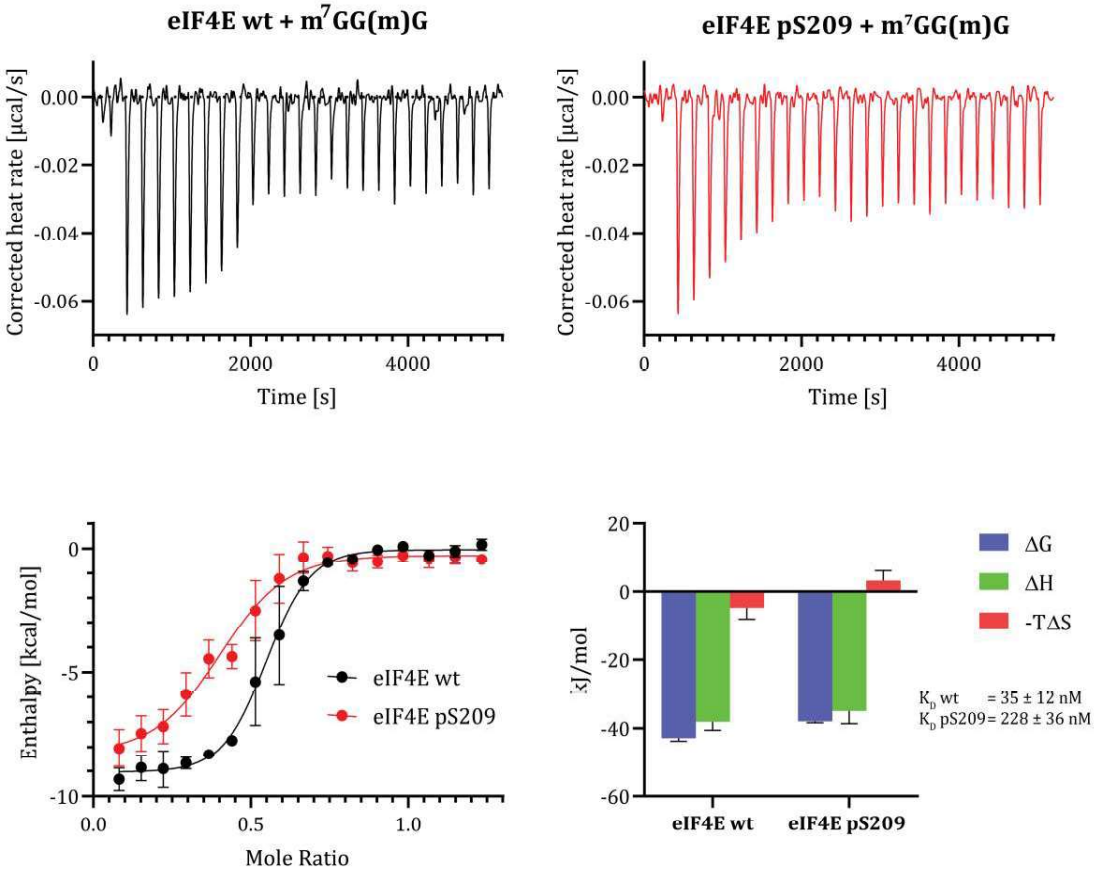


Fig. 6-41 *ITC titration of eIF4E wt and eIF4E pS209 with m⁷GG(m)G. m⁷GG(m)G binds to eIF4E wt with a dissociation constant of 35 nM and ~ 6-fold weaker to eIF4E pS209 with a dissociation constant of 228 nM. The interaction is enthalpy-driven in both cases.*

In conclusion, mRNA caps bind weaker once eIF4E is phosphorylated. The effect is most pronounced for m⁷GG(m)G, which experiences decreased binding by a factor of ~ 6. This is especially interesting as binding for the other cap 1 representing analogue m⁷GA(m)G to eIF4E pS209 is decreased only by a factor of less than 2. These results are summarized in **Fig. 6-42**. Only the differences in binding of m⁷GG and m⁷GG(m)G between unphosphorylated and phosphorylated eIF4E are statistically significant. Furthermore, all binding events are observed at molar ratios of ~ 0.5, meaning that two eIF4E

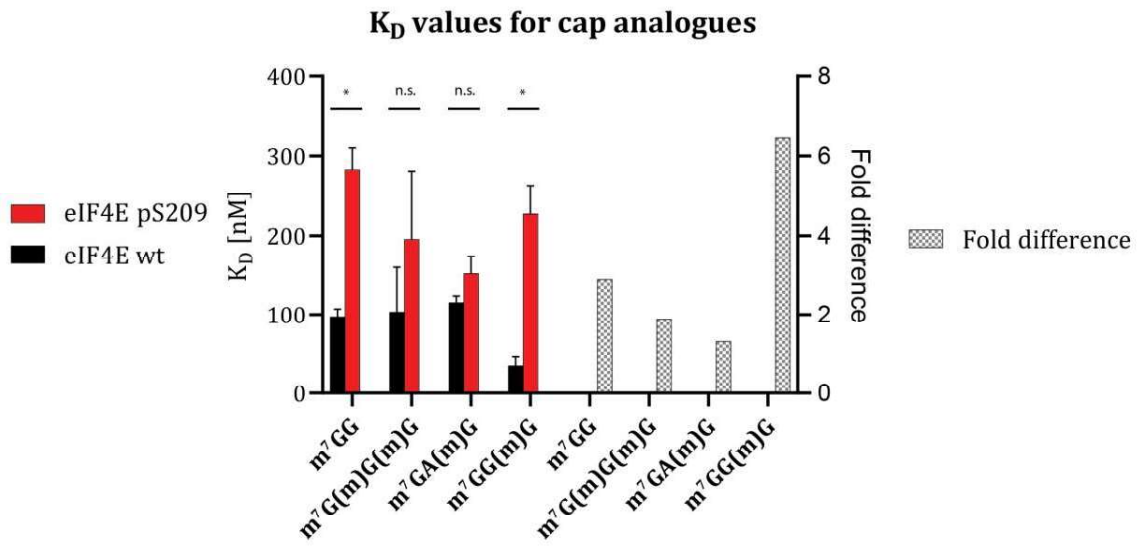


Fig. 6-42 *Comparison of K_D values for cap analogues against eIF4E wild type and eIF4E pS209. K_D values are generally higher for eIF4E pS209, with an ~ 6-fold decrease for $m^7GG(m)G$.*

molecules are necessary to bind one ligand molecule. Consistently, the molar ratio is lower in the case of eIF4E pS209 compared to eIF4E wild type in all four cases. However, more work needs to be done to draw decisive conclusions from this observation.

6.3.3 NMR spectroscopy of eIF4E variants

After assessing cap binding affinities for eIF4E, isotopically labeled proteins were produced for NMR experiments. ^{15}N-labeled GB1-tagged eIF4E wild type, S209D and pS209 were purified, and ^{15}N-^{1}H-HSQC spectra were recorded and overlaid to compare chemical shifts.

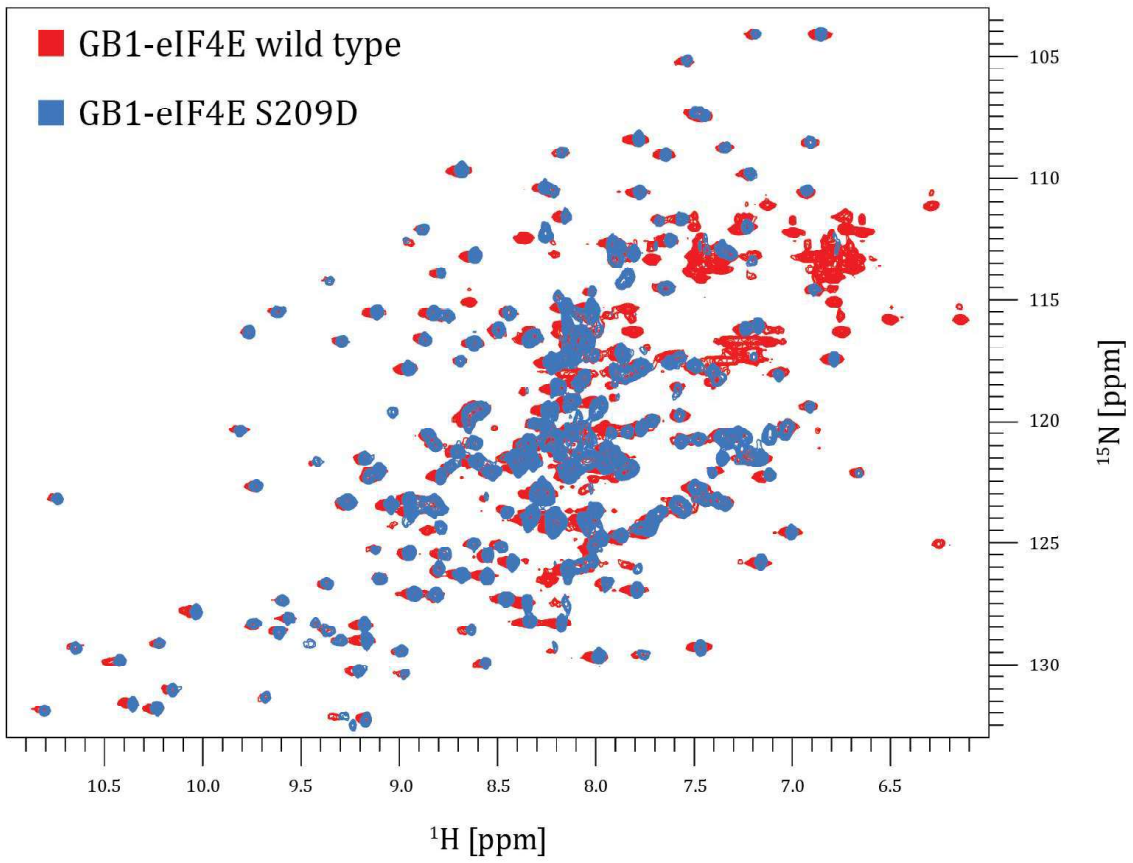


Fig. 6-43 *^{15}N-^{1}H-HSQC overlay of GB1-eIF4E wild type (red) and GB1-eIF4E S209D (blue). The lack of side chain residues in the blue spectra is due to the nature of the experiment, as the HSQC spectrum for GB1-eIF4E S209D was recorded with TROSY selection, which does not give rise to side chain peaks.*

The NMR data suggests that, structurally, only minor changes upon introduction of a phosphomimetic mutation occur. CSPs are found mainly around the mutation site at residue 209.

Next, chemical shift positions were compared for phosphorylated eIF4E.

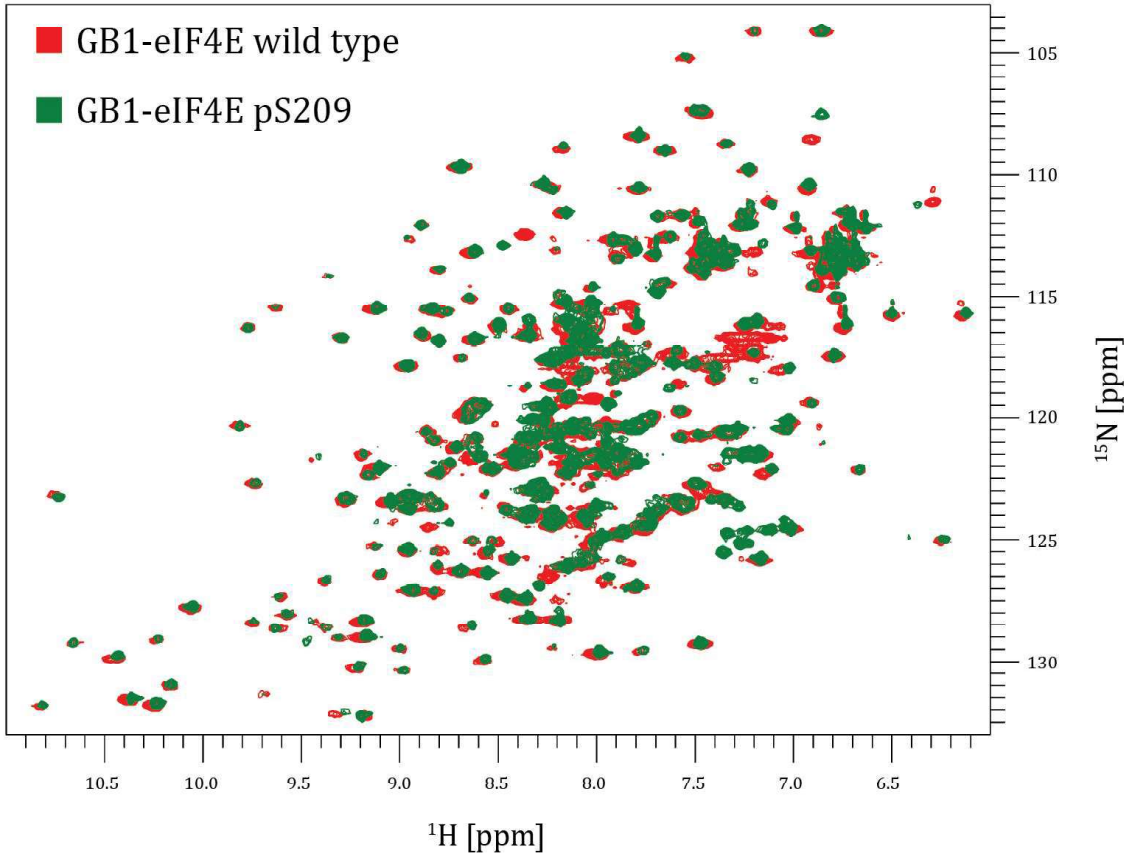


Fig. 6-44 *^{15}N-^{1}H-HSQC overlay of GB1-eIF4E wild type (red) and GB1-eIF4E pS209 (green).*

A similar picture is drawn when comparing wild type eIF4E with eIF4E pS209 using solution-NMR. Chemical shift perturbations are mainly observed near the phosphorylation site, while peak positions in other regions are less affected. Therefore, future efforts will focus on probing dynamics of eIF4E isoforms using NMR spectroscopy.

6.3.4 Structure of eIF4E S209D

The X-ray crystal structure of eIF4E S209D was solved to a resolution of 1.93 Å with 4 protein molecules in the unit cell (see **Fig. 6-45 A**) in collaboration with the crystallography core of SIRANO DHE-PAGANON. Interestingly, electron density for m^{7}GTP is missing in one of the chains.

The β7β8 loop was well resolved in all chains. Notably, there is no available structure of wild type eIF4E only bound to m^{7}GTP in which the β7β8 loop is resolved. Of the 16 existing structures of eIF4E bound to m^{7}GTP, the loop is resolved in only 6. However, all structures include either eIF4E inhibitors or interacting proteins.

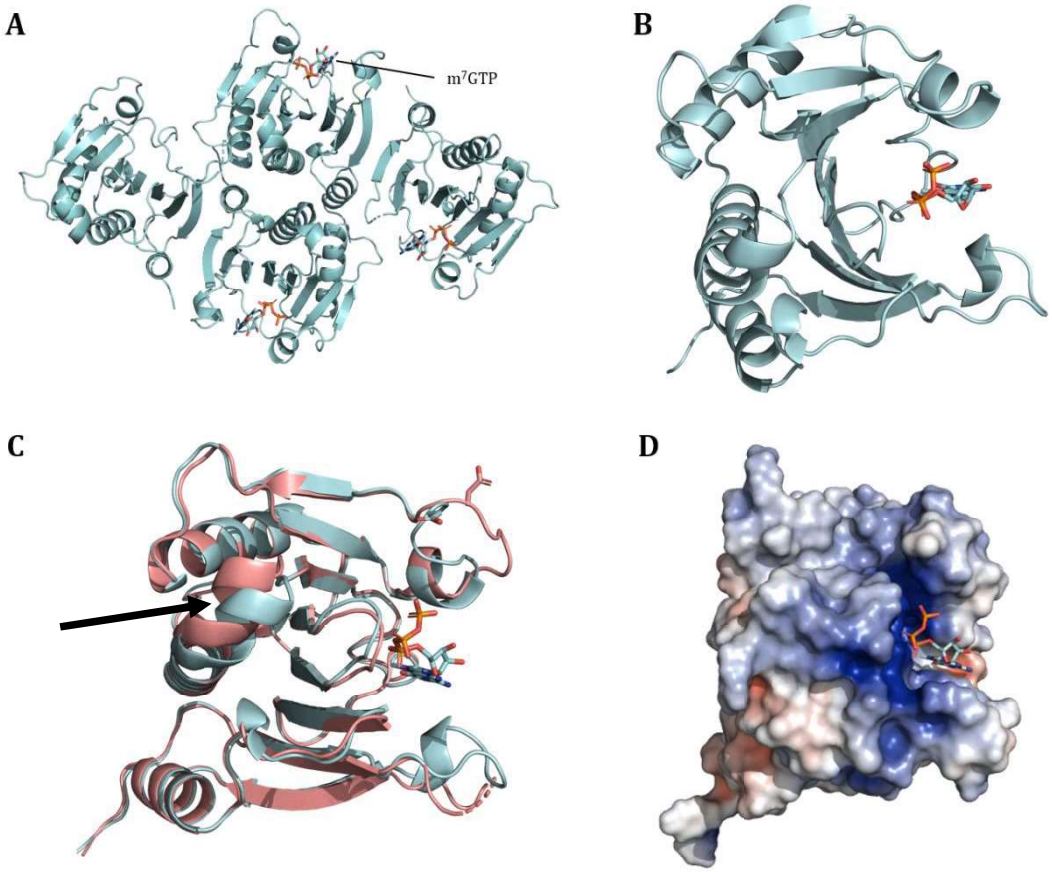


Fig. 6-45 *X-ray crystal structure of eIF4E S209D.* ***A.*** *eIF4E S209D crystallized as a tetramer, with m⁷GTP bound to three of the four molecules in the unit cell.* ***B.*** *The structure of eIF4E S209D, like wild type eIF4E, resembles a hand with β-strands constituting the palm and α-helices representing the dorsal portion of the palm. The m⁷GTP cap binds tightly to the palm-area stabilized by interactions with four tryptophan side chains.* ***C.*** *Alignment of m⁷GTP-bound eIF4E S209D chain A (blue) with cap-free eIF4E S209D chain D (salmon). A kink in helix α2 (arrow) is observed in the structure without m⁷GTP.* ***D.*** *Electrostatic surface of eIF4E S209D. The binding site of the cap is comprised of mainly negatively charged amino acids.*

The β5β6 loop on the other site of the cap binding interface was not resolved in the chain that did not show density for m⁷GTP (see **Fig. 6-45 C**). It was also not resolved for one of the chains that did display the cap**.** The kink in the last two turns of the α2 helix is in a slightly different position compared to the capped chain, which is true for other eIF4E structures as well**.**

Comparison of eIF4E S209D with m^{7}GTP bound with eIF4E wild type structures bound to m^{7}GTP in which the β7β8 loop was resolved (PDB codes: 4TPW, 4TQB, 3AM7 and 2W97 chain A) revealed little to no structural changes due to the mutation (see **Fig. 6-46**).

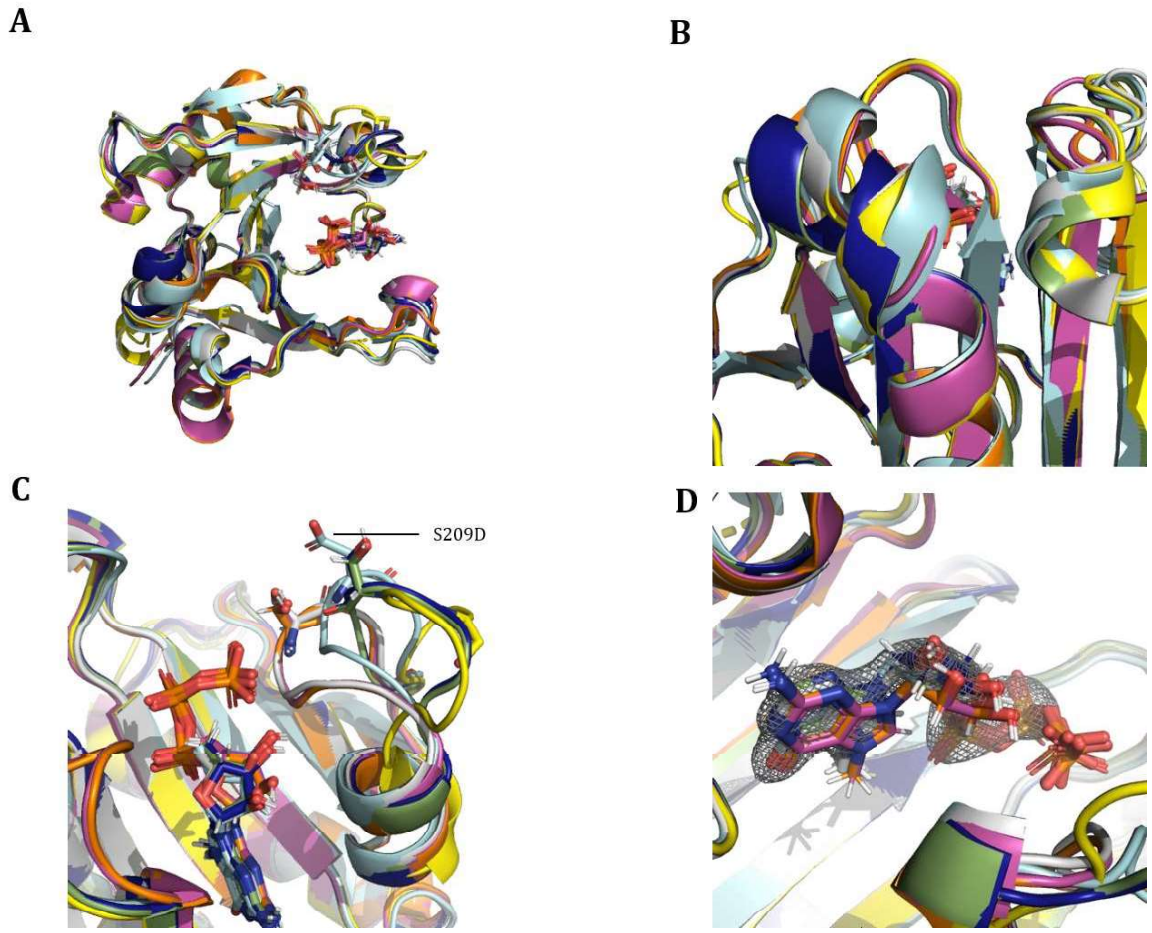


Fig. 6-46 A. *Overlay of cap-bound eIF4E S209D chain A (light blue) with 4TPW, 4TQB, 3AM7 and 2W97 chain A.* ***B.*** *Zoom on the kink of helix α2. Chain A of 4TQB (dark blue) and chain B of 4TPW (green) are tilted to a higher degree than the other structures.* ***C.*** *Zoom on the β7β8 loop. The β7β8 loop of eIF4E S209D chain A (light blue) is positioned between those of other structures. The loops in 2W97 chain A (grey), 4TPW chain A (orange) and 4TQB chain B (purple) are tilted towards the cap, while the loops in 4TQB chain A (dark blue), 4TPW chain B (green) and 3AM7 (yellow) are tilted away from the cap.* ***D.*** *Electron density for the cap overlays well with m^{7}GTP positions of all structures.*

Fig. 6-46 B shows a magnification of the kink in helix α2**.** The degree of tilting of the last two turns varies among m^{7}GTP-bound eIF4E structures**.** While chain A of 4TQB and chain

B of 4TPW are tilted most (comparable to cap-free eIF4E S209D), 4TQB chain B, 4TPW chain A, 3AM7 and 2W97 are tilted to the same degree as eIF4E S209D bound to m^7GTP. The overlay of the β7β8 loop reveals flexibility of the loop, with the loop from eIF4E S209D being positioned between loops of other m7GTP-bound structures (see **Fig. 6-46 C**). The loop is tilted towards the cap binding site in the structures from 2W97 chain A, 4TPW chain A and 4TQB chain B, while being tilted away from the binding site in the other structures. Electron density for the cap and overlay of m^7GTP structures is shown in **Fig. 6-46 D**. Density for the last phosphate is fading, possibly due to partial hydrolysis during crystallization. All cap structures align well.

Next, the cap-free chain was aligned with the three structures of cap-free eIF4E in which the β7β8 loop was resolved (PDB codes: 3U7X and 5ZML). The structures look similar, although the β5β6 loop was not resolved only in cap-free eIF4E S209D (see **Fig. 6-47 B**). The kink seen in the last two turns of helix α2 is also pronounced in cap-free S209D, but not in the other three structures (see **Fig. 6-47 C**). The β7β8 loop is in a similar conformation in all four structure, tilted away from the cap-binding site (see **Fig. 6-47 D**).

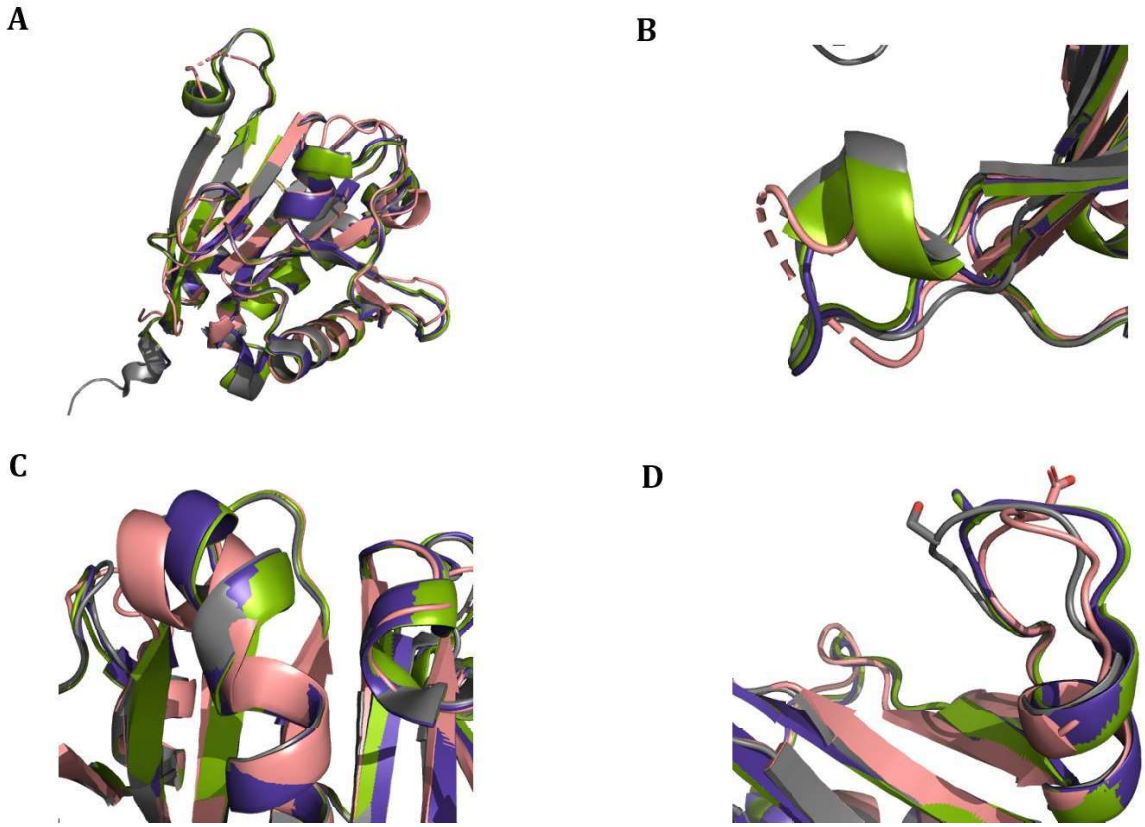


Fig. 6-47 A. *Overlay of cap-free eIF4E S209D chain D (salmon) with 3U7X and 5ZML.* ***B.*** *Zoom on the β5β6 loop. It is not resolved in the eIF4E S209D structure.* ***C.*** *Zoom on the kink in helix α2. It is tilted only for eIF4E S209D.* ***D.*** *Zoom on the β7β8 loop. In all four structures, the loop is tilted away from the cap-binding site and in similar conformations.*

6.3.5 Structure of eIF4E pS209

Recombinantly expressed eIF4E pS209 using the amber codon strategy was crystalized to furnish the X-ray crystal structure at a resolution of 2.18 Å with 4 molecules in the unit cell (see **Fig. 6-48 A**) in collaboration with the crystallography core of SIRANO DHE-PAGANON. Electron density for m^7GTP was only observed in two of the four chains.

The β7β8 loop was well resolved in all structures. Interestingly, the β5β6 loop was not resolved in the structures lacking m^7GTP density, as had been observed for eIF4E S209D. Like the S209D mutant, phosphorylated eIF4E exhibited the kink in the α2 helix only in the cap-free chain C (see **Fig. 6-48 C**). Chain D did not display the same degree of tilting that can be seen in chain C.

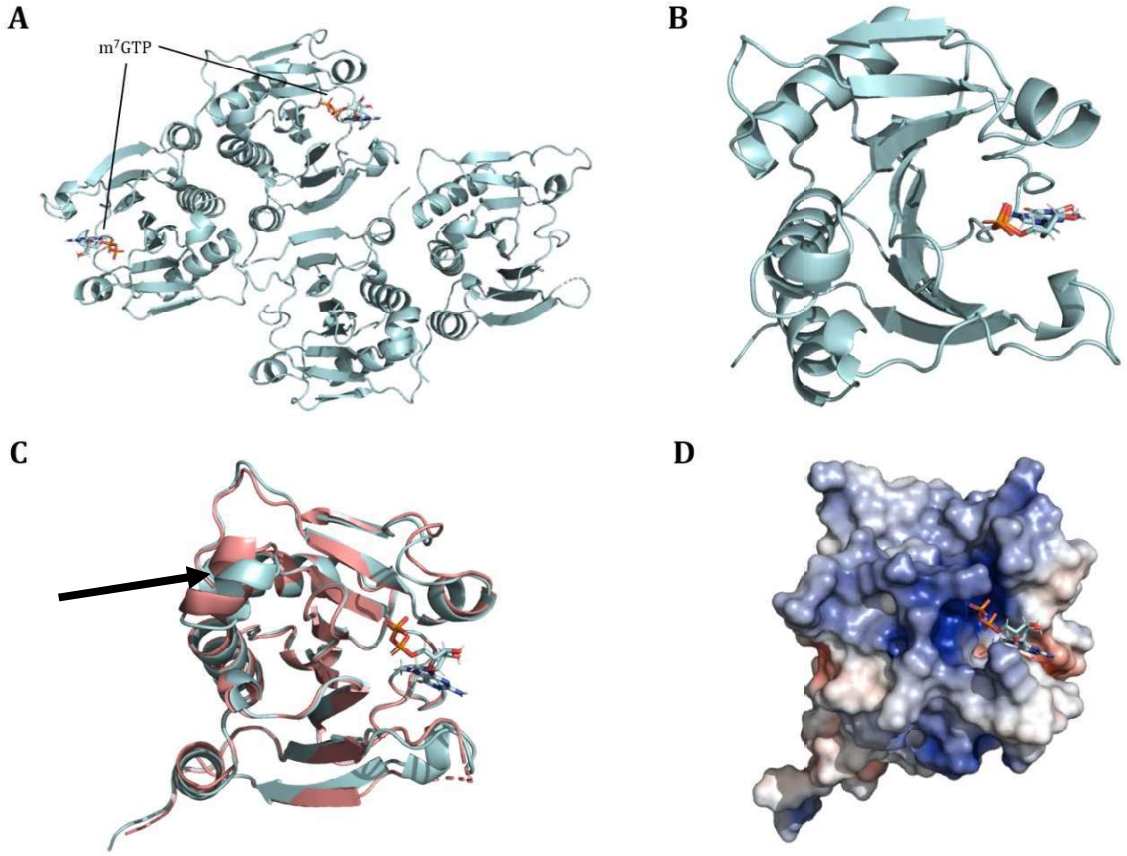


Fig. 6-48 *X-ray crystal structure of eIF4E pS209.* ***A.*** *eIF4E pS209 crystallized as a tetramer, with m⁷GTP bound to two of the four molecules in the unit cell.* ***B.*** *The structure of eIF4E pS209, like wild type eIF4E and eIF4E S209D, resembles a hand with β-strands constituting the palm and α-helices representing the dorsal portion of the palm. The m⁷GTP cap binds tightly to the palm-area stabilized by interactions with four tryptophan side chains.* ***C.*** *Alignment of m⁷GTP-bound eIF4E pS209 chain A (blue) with cap-free eIF4E pS209 chain C (salmon). A kink in helix α2 (arrow) is observed in the structure without m⁷GTP.* ***D.*** *Electrostatic surface of eIF4E pS209. The binding site of the cap is comprised of mainly negatively charged amino acids.*

Similar protein structures are again observed when overlaying m⁷GTP-bound with 4TPW, 4TQB, 3AM7 and 2W97 chain A (see **Fig. 6-49 A**). **Fig. 6-49 B** shows the overlay of the kink in helix α2**.** Like for capped eIF4E S209D, it is not observed in capped eIF4E pS209**.** The β7β8 loop is tilted far towards the cap (see **Fig. 6-49 C**), bringing the phosphate group in proximity to the terminal m⁷GTP phosphate**.** Electron density for m⁷GTP in chain A overlaid with m⁷GTP structures from the other PDB files is shown in

Fig. 6-49 D. Density for the last phosphate is fading, possibly due to partial hydrolysis during crystallization.

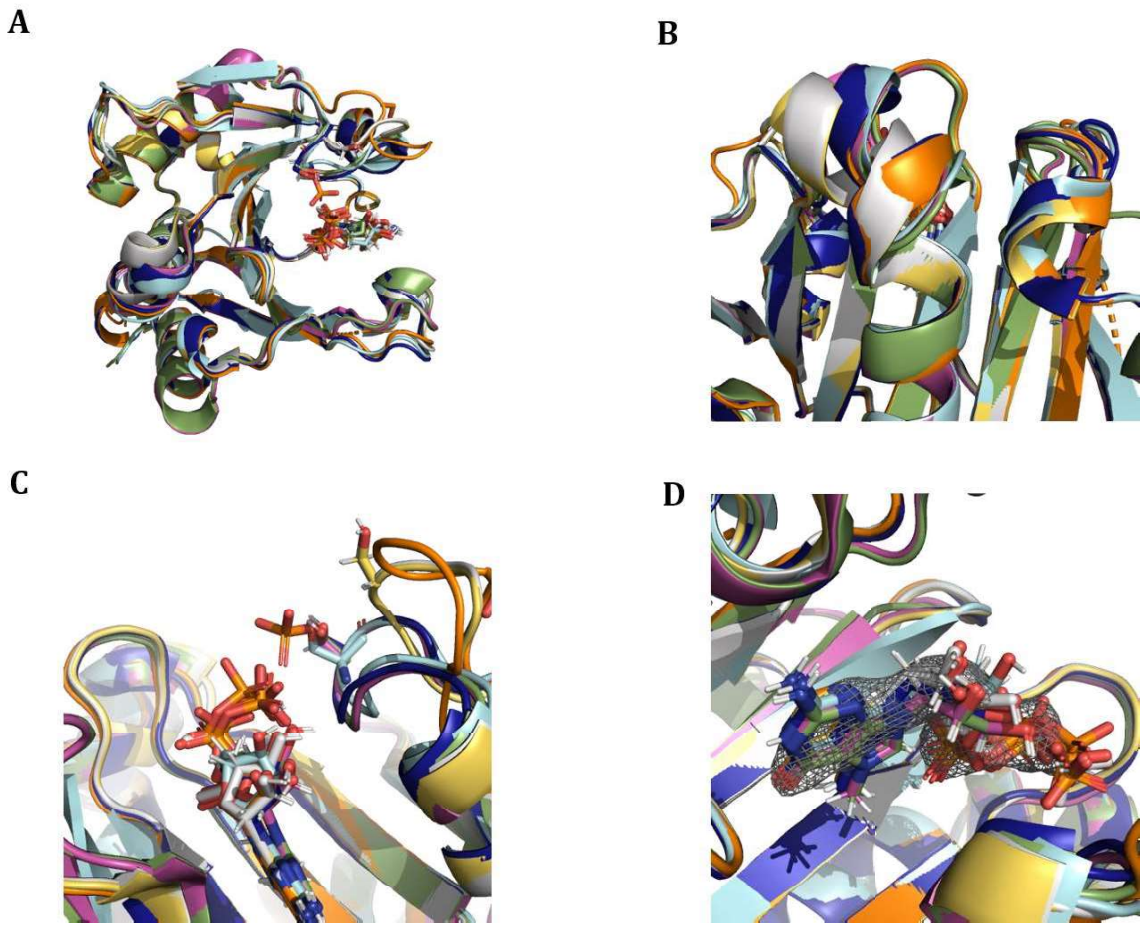


Fig. 6-49 A. *Overlay of cap-bound eIF4E pS209 chain A (light blue) with 4TPW, 4TQB, 3AM7 and 2W97 chain A.* ***B.*** *Zoom on the kink of helix α2. Chain A of 4TQB (grey) and chain B of 4TPW (orange) are tilted to a higher degree than the other structures.* ***C.*** *Zoom on the β7β8 loop. The β7β8 loop of eIF4E pS209 chain A (light blue) is tilted towards the cap. The loops in 2W97 chain A (dark blue), 4TPW chain A (purple) and 4TQB chain B (green) are tilted towards the cap, while the loops in 4TQB chain A (grey), 4TPW chain B (orange) and 3AM7 (yellow) are tilted away from the cap.* ***D.*** *Electron density for the cap overlays well with m^7GTP positions of all structures.*

Fig. 6-50 A shows the overlay of the uncapped chain D with the three structures of uncapped eIF4E in which the β7β8 loop was resolved (PDB codes: 3U7X and 5ZML)**.** Like for uncapped eIF4E S209D, the β5β6 loop is not resolved in phosphorylated eIF4E (see **Fig. 6-50 B**)**.** The kink in helix α2 is not as pronounced as in the other chain that does not show density for m^7GTP, so the helix aligns with those from the other three structures

(see **Fig. 6-50 C**). Loop β7β8 is in a similar conformation compared to 3U7X and 5ZML, with the phosphoserine facing outward. Thus, the theory of a salt-bridge[219] between this residue and one of the basic amino acids on the β5β6 loop (R157, K159 or K162) is unlikely.

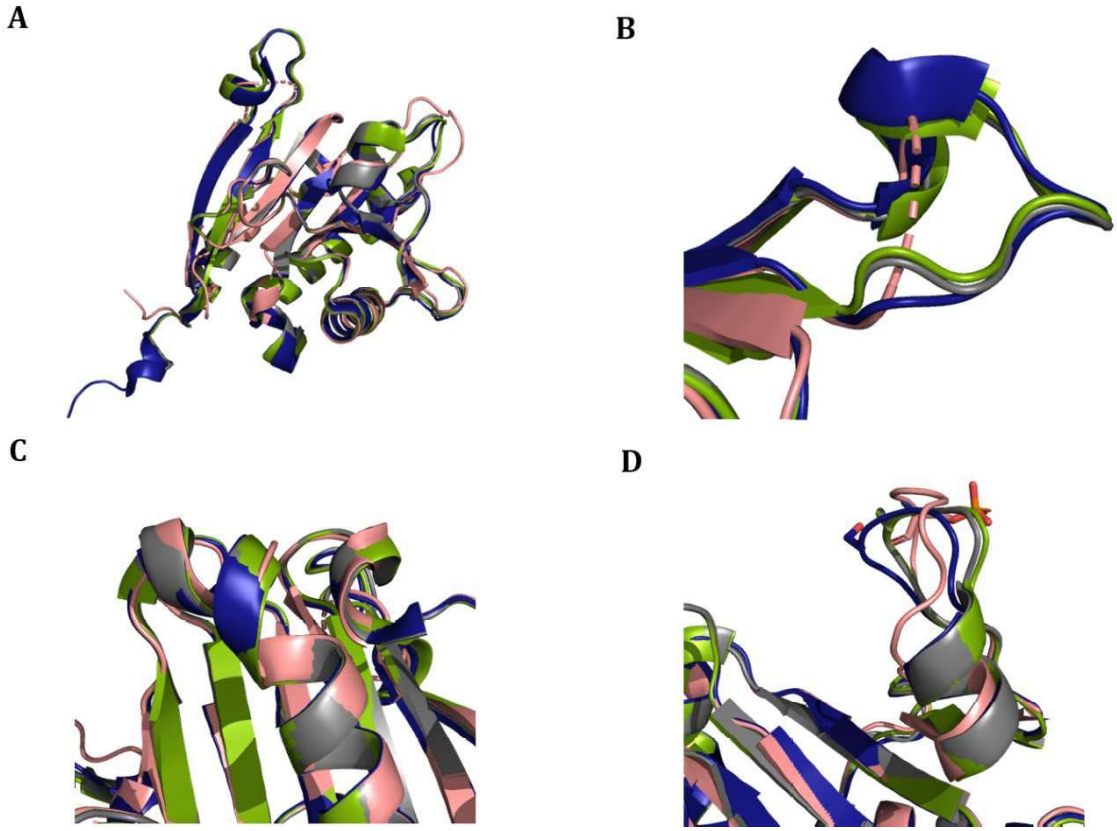


Fig. 6-50 A. *Overlay of cap-free eIF4E pS209 chain D (salmon) with 3U7X and 5ZML.* ***B.*** *Zoom on the β5β6 loop. It is not resolved in the eIF4E pS209 chain D structure.* ***C.*** *Zoom on the kink in helix α2. It is not pronounced in the structure of eIF4E pS209 chain D and aligns with the other three structures.* ***D.*** *Zoom on the β7β8 loop. In all four structures, the loop is tilted away from the cap-binding site and in similar conformations.*

6.3.6 Electron densities for eIF4E S209D and pS209

To additionally confirm that the solved structures are not comprised of wild type eIF4E, a closer look at the electron density at the side chains in position 209 is presented in **Fig. 6-51**. Clear density of the side chains in loop β7β8 confirm the presence of aspartate in the eIF4E S209D structure and phosphoserine in the eIF4E pS209 structure.

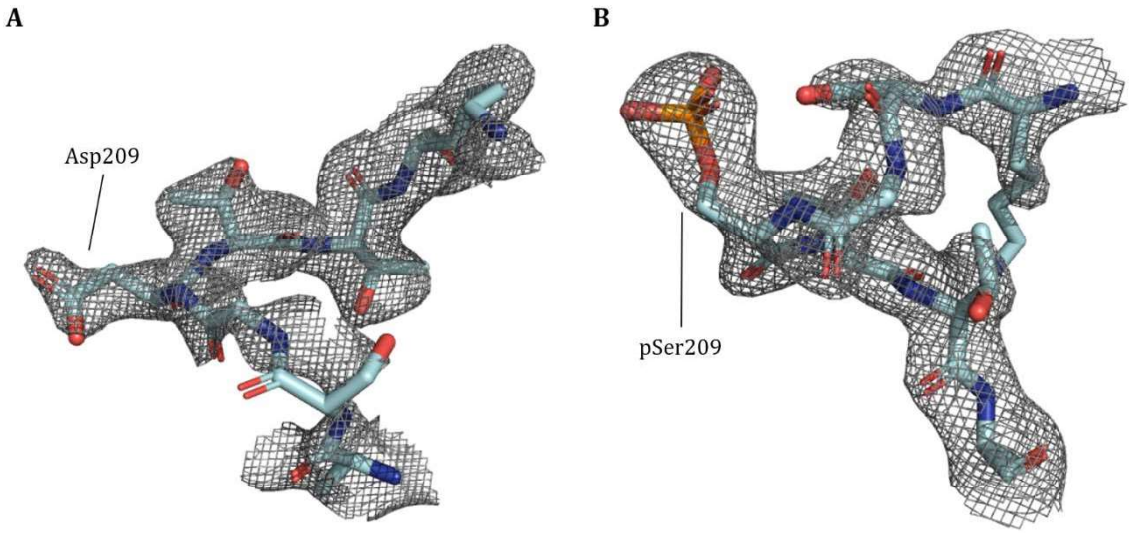


Fig. 6-51 A. *Electron density for the β7β8 loops of eIF4E S209D chain A.* ***B.*** *Electron density for the β7β8 loops of eIF4E pS209 chain A.*

6.3.7 Discussion

In addition to the biophysical investigation of the effects of eIF4E phosphorylation using ITC, X-ray crystal structures for both S209D and phosphorylated eIF4E were solved at high resolution. The structures revealed no major structural changes due to phosphorylation or mutation. However, it is possible that the timing of events is of importance. If phosphorylation occurs prior to cap binding, salt bridge formation between phosphorylated Ser209 and either of the residues Arg157, Lys159 or Lys162 is possible. To test this theory, attempts to solve the structure of cap-free eIF4E pS209 will be carried out. If eIF4E pS209 is purified without the use of m^7GTP, cap-bound eIF4E pS209 will never be formed in these experiments, therefore recreating the correctly timed events if phosphorylation occurs prior to cap binding. It is also possible that binding kinetics of m^7GTP to phosphorylated eIF4E are altered in a way that favor the dissociation of the cap, which might explain the lack of density in two of the four chains. Repulsion between the phosphate of the phosphoserine and triphosphate of m^7GTP would explain this observation. This theory is emphasized by the ITC data comparing eIF4E wild type with eIF4E pS209, where the ligand-protein ratio was lower for eIF4E pS209.

Additionally, effects of phosphorylation on protein half-life will be evaluated using phosphomimetic (S to D) and phosphor-null (S to A) mutants for mammalian expression. These constructs, in combination with Mnk inhibitors or on their own, will provide insights into the effects that phosphorylation has on protein stability, when analyzed using Western Blots.

Furthermore, it will be investigated if phosphorylation of eIF4E affects binding to binding partners such as 4E-BPs or eIF4G by pull-down experiments and subsequent Western Blot analysis.

6.4 ACP

6.4.1 Expression and assignment of apo-ACP and loaded ACPs

ACP was expressed with ^{13}C and ^{15}N labeling to create double labeled samples for triple resonance backbone solution NMR experiments. 100 % of the protein's amino acids (excluding Pro55) were assigned to the peaks in the ^{15}N-^{1}H-HSQC spectrum (see **Fig. 6-52**). To create loaded ACPs, apo-ACP was combined with recombinantly expressed phosphopantetheinyl transferase Sfp and uncleavable thioether analogues of CoA-substrates (3-oxo-C6i-CoA as a substrate for EsaI and C8i-CoA for BmaI1). Sfp catalyzed

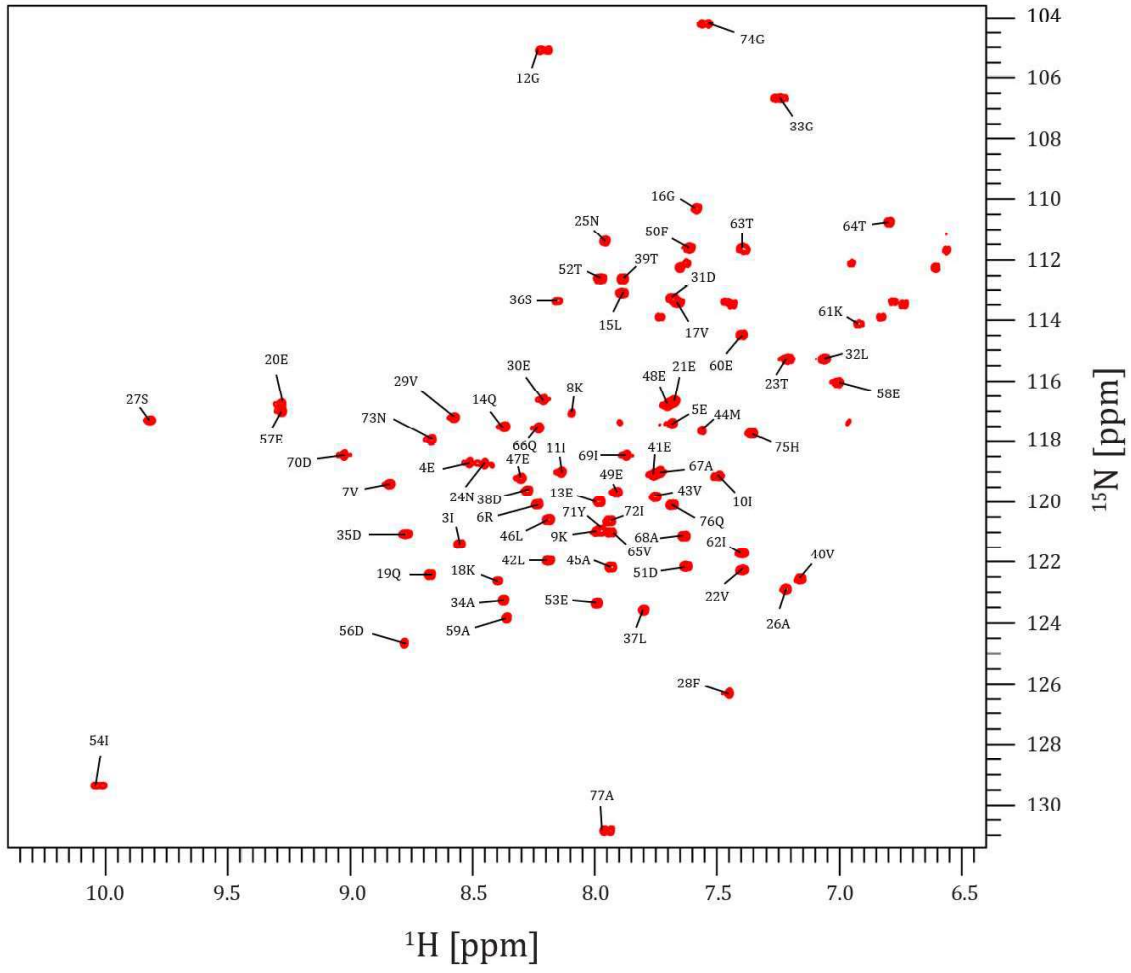


Fig. 6-52 *^{15}N-^{1}H-HSQC with assignments of apo-ACP.*

the transfer of the substrate to Ser36 of ACP. The reactions were monitored using ^{15}N-^{1}H-HSQC spectra and were followed by backbone assignment after completed loading.

6.4.1.1 Assignment of 3-oxo-C6i-ACP

Fig. 6-53 A shows an overlay of HSQC spectra from apo-ACP and 3-oxo-C6i-ACP. In addition to the region around the covalent attachment of the 3-oxo-C6 chain (Ser36), significant CSPs can be observed towards the C-terminus of the protein (~ amino acids

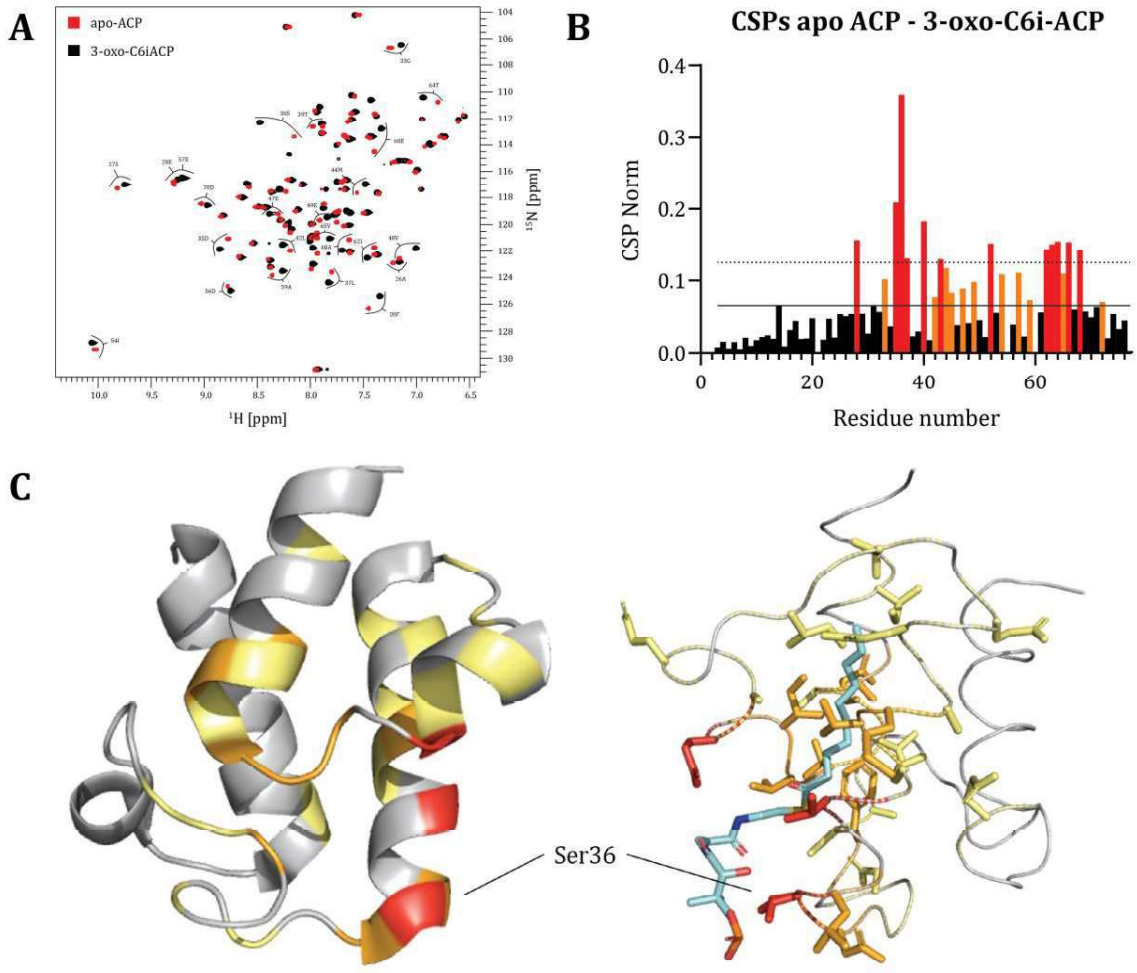


Fig. 6-53 A. *^{15}N-^{1}H-HSQC overlay of apo ACP (red) and 3-oxo-C6i-ACP (black).* ***B.*** *Chemical shift perturbations (CSPs) mapped as a function of residue number. Residues with CSPs between average CSP (solid line) and average CSP + standard deviation (dotted line) are highlighted in orange, residues with CSPs higher than average CSP + standard deviation are highlighted in red.* ***C.*** *CSPs mapped on the structure of ACP. Residues with significant CSPs are highlighted in yellow (> 1 standard deviation), orange (> 2 standard deviations) and red (> 3 standard deviations).*

60-70) (see **Fig. 6-53 B**). **Fig. 6-53 C** shows significant CSPs highlighted on the structure of apo-ACP (PDB code: 1T8K) and the same structure displayed as a ribbon with the side chains of residues that experience significant CSPs highlighted. The structure from ACP in complex with decanoyl-CoA (PDB code: 2FAE) was aligned and the ligand is represented in blue. The mapping of CSPs onto the apo-ACP structure agrees with the structure of a loaded ACP, in which orange and red highlighted residues directly participate in interactions, while yellow highlighted residues likely undergo conformational changes upon binding.

6.4.1.2 Assignment of C8i-ACP

Similar regions of ACP change upon loading with C8i-CoA (see **Fig. 6-54**). Compared to 3-oxo-C6i-ACP however, C8i-ACP displays even more significant chemical shift perturbations towards the C-terminus. This could be due to the longer cargo that is the substrate for BmaI1. **Fig. 6-54 C** shows significant CSPs highlighted on the structure of apo-ACP (PDB code: 1T8K) and the same structure displayed as a ribbon with the side

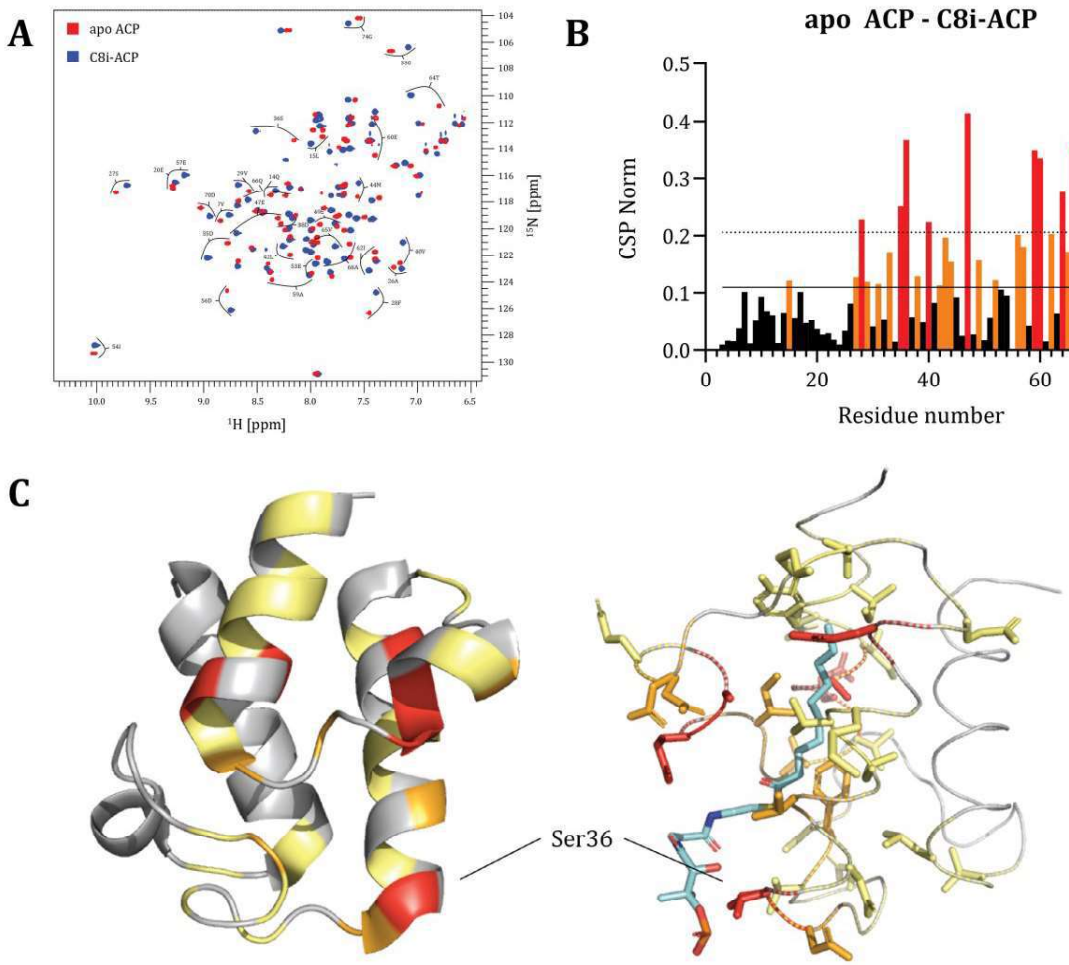


Fig. 6-54 A. *^{15}N-^{1}H-HSQC overlay of apo ACP (red) and C8i-ACP (blue).* ***B.*** *Chemical shift perturbations (CSPs) mapped as a function of residue number. Residues with CSPs between average CSP (solid line) and average CSP + standard deviation (dotted line) are highlighted in orange, residues with CSPs higher than average CSP + standard deviation are highlighted in red.* ***C.*** *CSPs mapped on the structure of ACP. Residues with significant CSPs are highlighted in yellow (> 1 standard deviation), orange (> 2 standard deviations) and red (> 3 standard deviations).*

chains of residues that experience significant CSPs highlighted. The structure from ACP in complex with decanoyl-CoA (PDB code: 2FAE) was aligned and the ligand is represented in blue. Compared to the structure of 3-oxo-C6i-ACP, C8i-ACP experiences more significant CSPs towards the center of the protein, likely due to the longer chain length of C8i-CoA that reaches more residues of the α3 helix.

6.4.2 Monitoring the ACP/AHL synthase interaction with NMR spectroscopy

Using ^{15}N-^{1}H-HSQC experiments with ^{15}N-labeled loaded ACPs and the unlabeled AHL synthases EsaI and BmaI1, differences between the two interaction interfaces were analyzed. Initial experiments were performed at 25 °C in ACP reaction buffer (50 mM MES pH = 6.8, 10 mM $MgCl_2$, 1 mM TCEP-HCl), which does not contain NaCl. These conditions resulted in interactions in an intermediate exchange regime, leading to peak intensity loss for amino acids that experience a change in magnetic environment upon binding (either due to direct interaction or allosteric conformational changes, see **Fig. 6-55** and **Fig. 6-56**).

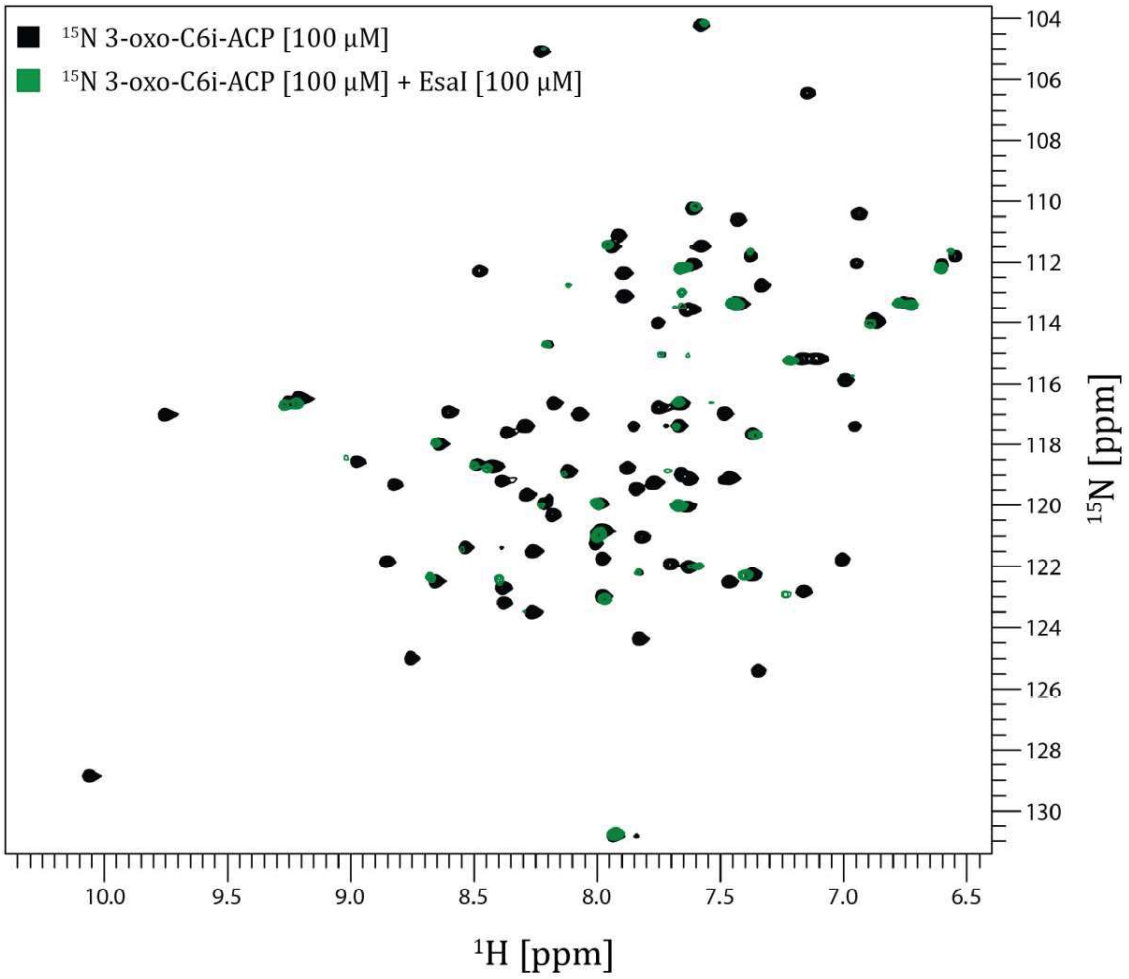


Fig. 6-55 *^{15}N-^{1}H-HSQC overlay of 3-oxo-C6i-ACP (black) and 3-oxo-C6i-ACP with 1 equivalent EsaI (green). Peak broadening is observed for most amino acids involved in the interaction, indicating an intermediate exchange regime.*

Interestingly, more peaks disappear upon interaction of 3-oxo-C6i-ACP with EsaI compared to those from the interaction of C8i-ACP with BmaI1 (Compare **Fig. 6-55** and **Fig. 6-56**). However, intermediate exchange regimes are not ideal for analysis of interaction interfaces, as the NMR experiments that probe the difference between CSPs caused by direct binding and conformational changes (cross-saturation transfer experiments) rely on the monitoring of peak intensity reduction of bound proteins. Therefore, an effort was made to create conditions in which binding in either a fast or slow exchange regime was observed.

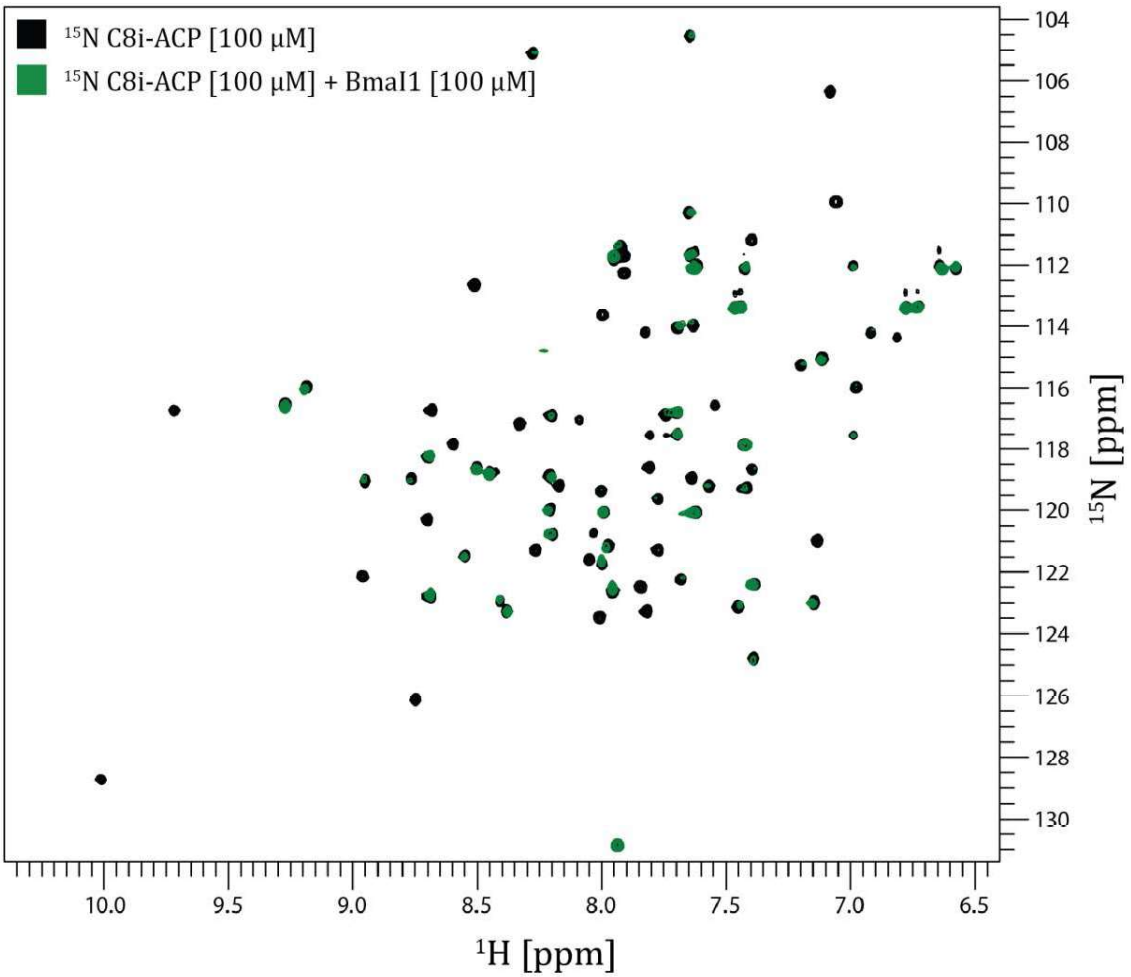


Fig. 6-56 *^{15}N-^{1}H-HSQC overlay of C8i-ACP (black) and C8i-ACP with 1 equivalent BmaI1 (green). Peak broadening is observed for most amino acids involved in the interaction, indicating an intermediate exchange regime.*

6.4.3 ACP / AHL-synthase interaction in fast and slow exchange regimes

To alter the interaction kinetics, changing the ionic strength of the buffer to weaken the protein-protein interaction and shift the interaction profile to a different exchange regime was attempted. As a proof of concept, 100 mM NaCl was added to the interacting proteins and another HSQC spectrum was acquired. For the interaction between C8i-ACP with BmaI1 this change resulted in a change to a slow exchange regime (two peaks are

observed; one for the unbound protein in the original position and one for the bound protein in a new position, see **Fig. 6-57**).

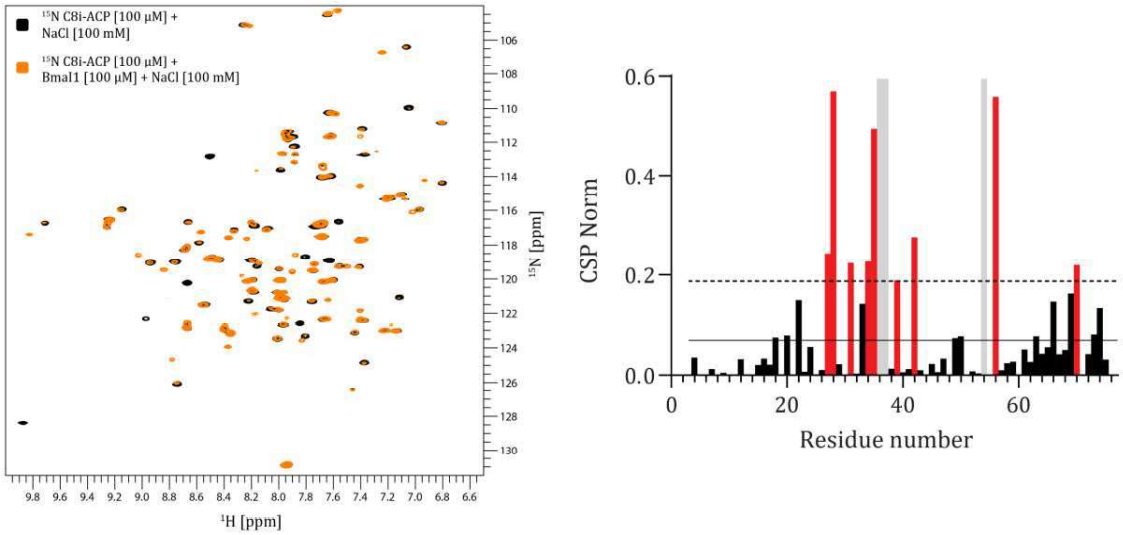


Fig. 6-57 *^{15}N-^{1}H-HSQC overlay of C8i-ACP in buffer containing 100 mM NaCl (black) and C8i-ACP with 1 equivalent BmaI1 in buffer containing 100 mM NaCl (orange). Most residues experience a slow exchange regime. CSPs calculated from the position of the peak of the residue in the bound form are plotted as a function of residue number. Residues with CSPS higher than the sum of the average CSP (solid line) and standard deviation (dotted line) are highlighted in red. Broadened peaks from residues which are still in intermediate exchange regimes are highlighted in grey.*

In case of the 3-oxo-C6i-ACP interaction with EsaI, increasing the ionic strength did not alter the exchange regime. However, increasing the experimental temperature to 37 °C and reducing the concentration of both proteins resulted in more residues experiencing a fast exchange regime (see **Fig. 6-58**).

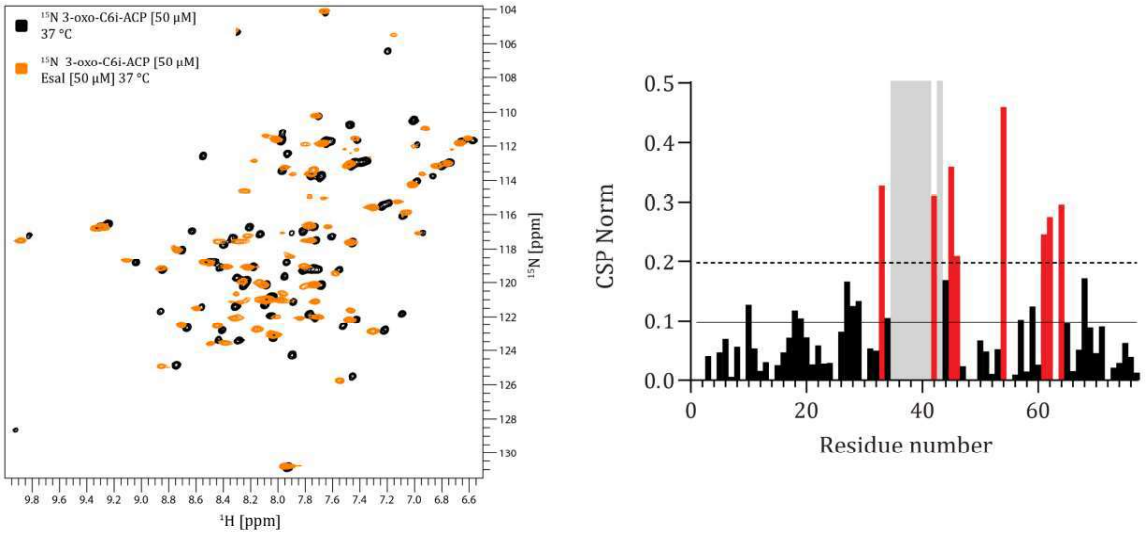


Fig. 6-58 *^{15}N-^{1}H-HSQC overlay of 3-oxo-C6i-ACP at 37 °C (black) and 3-oxo-C6i-ACP with 1 equivalent EsaI at 37 °C (orange). More residues experience a fast exchange regime. CSPs are plotted as a function of residue number. Residues with CSPS higher than the sum of the average CSP (solid line) and standard deviation (dotted line) are highlighted in red. Broadened peaks from residues which are still in intermediate exchange regimes are highlighted in grey.*

6.4.4 Cross saturation transfer to map the direct binding interface

Since chemical shift perturbations (CSPs) obtained from the titration experiments are the result of a perturbed magnetic environment of the nuclei, these effects can be contributed to either direct involvement in binding or allosteric conformational changes. To distinguish CSPs caused by conformational changes from those at the binding interface, additional experiments need to be performed. The cross-saturation transfer (CST)

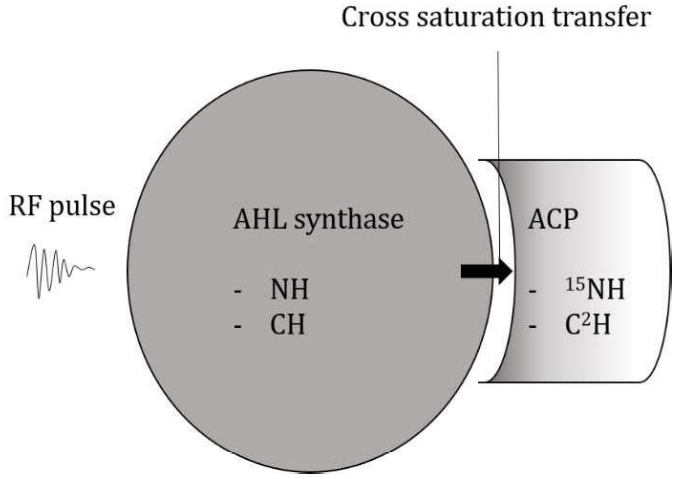


Fig. 6-59 *Schematic description of a CST experiment: ^{1}H nuclei are saturated by an RF pulse, followed by transfer of this saturation onto other ^{1}H nuclei in proximity. Figure adapted from H. Takahashi, T. Nakanishi, K. Kami, Y. Arata, I. Shimada,* A novel NMR method for determining the interfaces of large protein–protein complexes. Nat. Struct. Biol. ***2000****, 7, 4.*

experiment is such a method (see **Fig. 6-59**).[396,397] In principle, the observed protein (^{15}N-labeled loaded acyl-ACP) is expressed as a perdeuterated protein using minimal media in D_2O with ^{15}N NH_4Cl as the only nitrogen and D-Glucose-d7 as the only carbon source.[398–400] Amide protons will back-exchange from ND to NH in aqueous buffer, while CD bonds will remain intact. This protein will have proton resonances in the amide region of a 1D NMR spectrum, ~ 6-10 ppm, but not in the aliphatic CH region. Unlabeled AHL-synthase introduced to such a sample will give rise to additional peaks in the aliphatic CH region of the spectrum, spanning from ~ -1-5 ppm. An RF pulse applied to this region of the spectrum (~ 0 ppm) is used to saturate aliphatic protons. Through spin diffusion, this saturation rapidly spreads through aromatic and amide protons to saturate the entire protein. ^{15}N-^{2}H-ACP is not affected by this saturation directly, but protons located directly at the binding interface will experience transferred saturation by a phenomenon called cross-relaxation.[397] HSQC spectra of ACP recorded without and with saturation of aliphatic protons will result in a difference of intensity of peaks partaking in binding, indicating the direct binding interface.

Chemical shift perturbations for both interactions are observed in the middle region of ACP (~ 25-40 for C8i-ACP with BmaI1, ~35-50 for 3-oxo-C6i-ACP with EsaI) and towards the C-terminus. While it is likely that the CSPs from the middle region arise from direct protein-protein interactions and the CSPs from the C-terminus are due to conformational changes, CST experiments were carried out on the 3-oxo-C6i-ACP interaction with EsaI to confirm this theory.

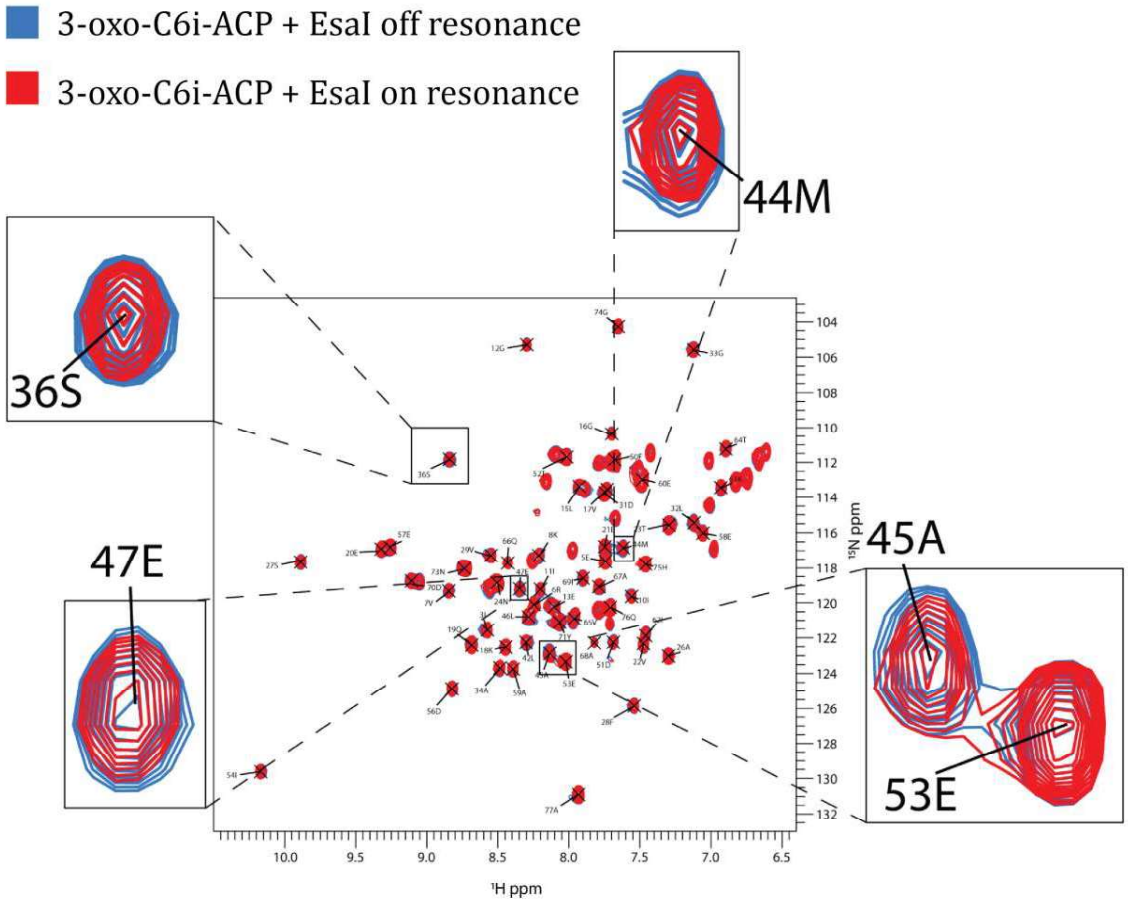


Fig. 6-60 *Cross saturation transfer experiment of 3-oxo-C6i-ACP with EsaI. The spectra show 50 µM ^{15}N-^{2}H-labeled 3-oxo-C6i-ACP + 50 µM EsaI recorded without saturation transfer (blue, off resonance) and 50 µM ^{15}N-^{2}H-labeled 3-oxo-C6i-ACP + 50 µM EsaI recorded with saturation transfer (red, on resonance). Intensity reduction is observed for some, but not all residues (see zoom on 45A and 53E).*

Spectra recorded without saturation transfer (off resonance) and with saturation transfer (on resonance) revealed amino acids directly involved in the binding interface (see **Fig. 6-60**). While significant reduction of peak intensities is observed for residues in the middle region of the protein, residues towards the C-terminus do not experience saturation transfer (see **Fig. 6-61**).

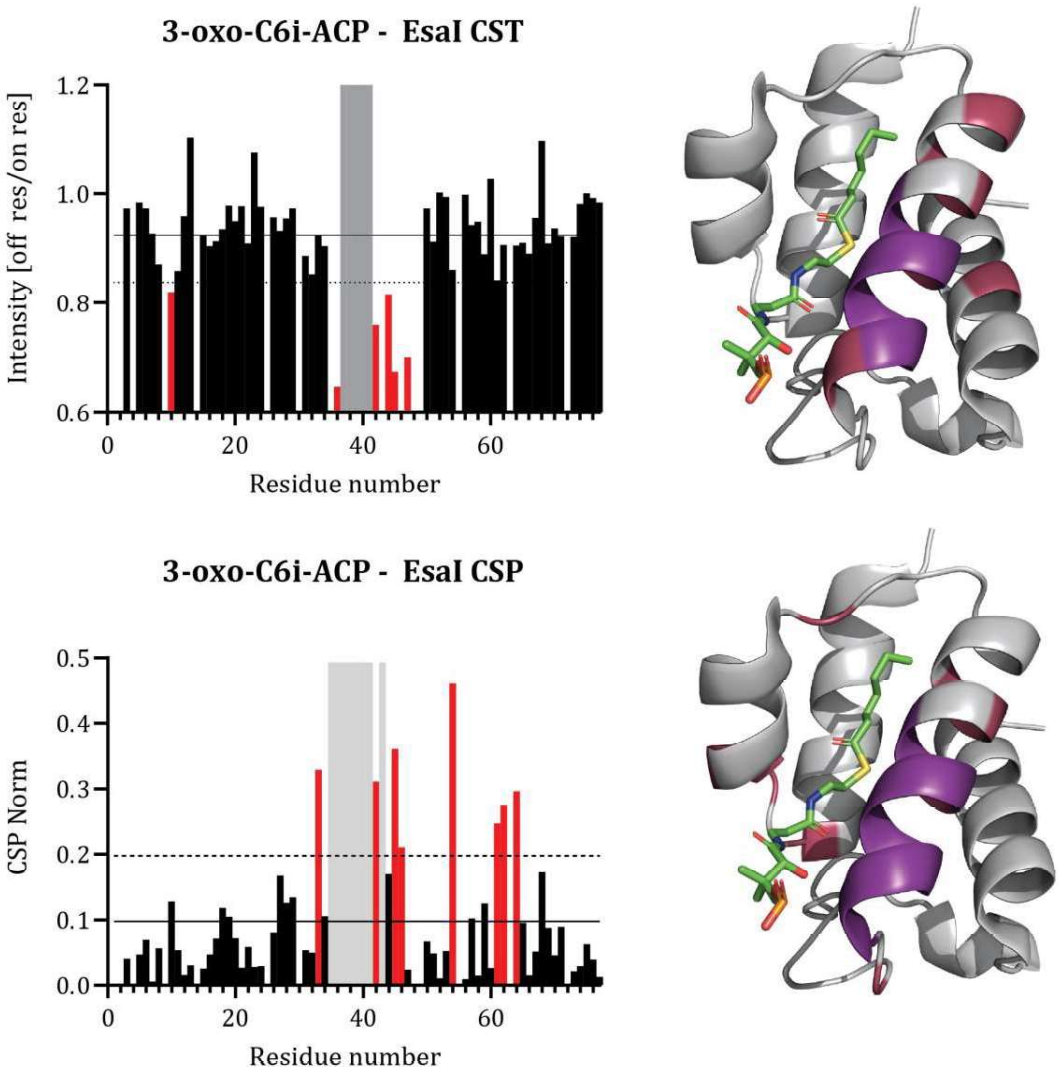


Fig. 6-61 *Top panel: CST plot from the 3-oxo-C6i-ACP interaction with EsaI with ratios obtained from peak intensities in a spectrum where saturation is not transferred (off resonance [off res]) and where it is transferred (on resonance [on res]) as a function of residue number. Peaks from residues for which the intensity is reduced by more than the average ratio (solid line) plus standard deviation (dotted line) are highlighted in red and plotted onto the structure of ACP bound to hexanoyl in red (PDB code: 2FAC). Broadened peaks from residues in intermediate exchange regimes are highlighted in grey and plotted onto the structure in purple. Bottom panel: CSPs of the 3-oxo-C6i-ACP interaction with EsaI. Residues with CSPs higher than the sum of the average CSP (solid line) and standard deviation (dotted line) are highlighted in red and plotted onto the structure of ACP bound to hexanoyl in red (PDB code: 2FAC). Broadened peaks from residues in intermediate exchange regimes are highlighted in grey and plotted onto the structure in purple.*

Alignment of CST and CSP plots is shown in **Fig. 6-61.** Peaks with significant intensity

reductions highlighted in the structure of hexanoyl-ACP (PDB code: 2FAC) reveal involvement of the entire α3 helix in binding of EsaI (see **Fig. 6-61** top panel), while amino acids not involved in the binding interface that experience CSPs belong to flexible residues in the loop regions connecting the helices (see **Fig. 6-61** bottom panel). Therefore, it was concluded that the middle region of ACP belongs to the binding interface between loaded ACP and AHL synthase.

Interestingly, the binding interface between C8i-ACP and BmaI1 is located at the lower end of the α3 helix and stretches to the loop connecting the α3 and α2 helices (see **Fig. 6-62**). These results are surprising since the cargo for BmaI1 (C8) is longer than the cargo for EsaI (3-oxo-C6) and reaches higher parts of the α3 helix (see **Fig. 6-54**).

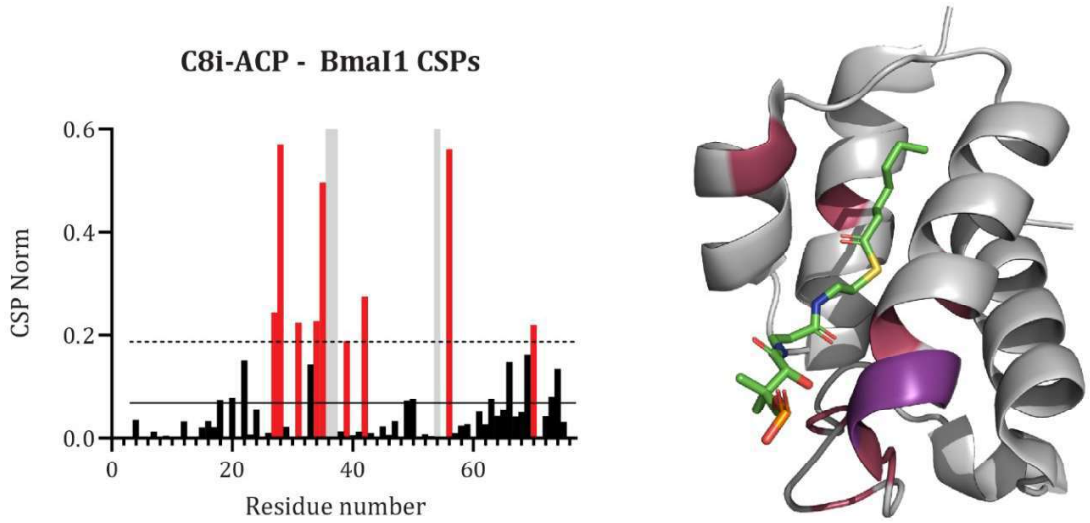


Fig. 6-62 *CSP plot of the interaction between C8i-ACP and BmaI1 as a function of residue number. Residues with CSPs higher than the sum of the average CSP (solid line) and standard deviation (dotted line) are highlighted in red and plotted onto the structure of ACP bound to hexanoyl in red (PDF code: 2FAC). Broadened peaks from residues in intermediate exchange regimes are highlighted in grey and plotted onto the structure in purple.*

6.4.5 Discussion

Chemical shift perturbation experiments in conjunction with cross-saturation transfer NMR enabled the visualization of protein-protein binding interfaces between 3-oxo-C6i-ACP and EsaI as well as C8i-ACP and BmaI1. While EsaI engages 3-oxo-C6i-ACP along the

helix α3, BmaI1 interacts with C8i-ACP from the bottom of helix α3 and the loop region connecting helix α3 and helix α2. Considering the positioning of the acyl-chain, which was analyzed using chemical shift perturbations of apo-ACP with loaded ACPs, ACP must undergo considerable conformational changes to bring the cargo in proximity of the AHL synthase. Therefore, solution NMR experiments probing the dynamics of apo-ACP and loaded ACPs will be carried out in the future to gain understanding into the protein's behavior upon cargo binding, AHL synthase engagement and the involvement of SAM.

6.5 GILZ

6.5.1 Expression and purification of human GILZ

To express human GILZ in *E. coli*, a pET28a plasmid coding for His_6-GB1-TEV-GILZ was designed and purchased from GeneWiz. Transformation of this plasmid into *E. coli* pLysS cells enabled expression of a GILZ fusion protein that is purifiable by Ni-NTA affinity chromatography. Subsequent treatment with TEV protease cleaved the His_6-GB1 tag and yielded recombinantly expressed GILZ.

After being subjected to gel filtration purification, GILZ eluted on a Superdex 200 Increase 10/300 GL at an elution volume of ~ 46 mL, which corresponds to a molecular weight which is considerably higher than the expected 30 kDa of a GILZ homodimer. To exclude protein denaturation as the cause of this observation, the eluted GILZ sample was analyzed using circular dichroism spectroscopy and compared to profiles of typical protein folds[401](see **Fig. 6-63**). Alpha-helical secondary structures were confirmed by the CD experiments, thereby excluding denaturation as the cause for early elution during chromatography. The results indicated that folded GILZ does not, as previously thought, dimerize, but in fact oligomerize to form a high molecular weight complex with α-helical secondary structures.

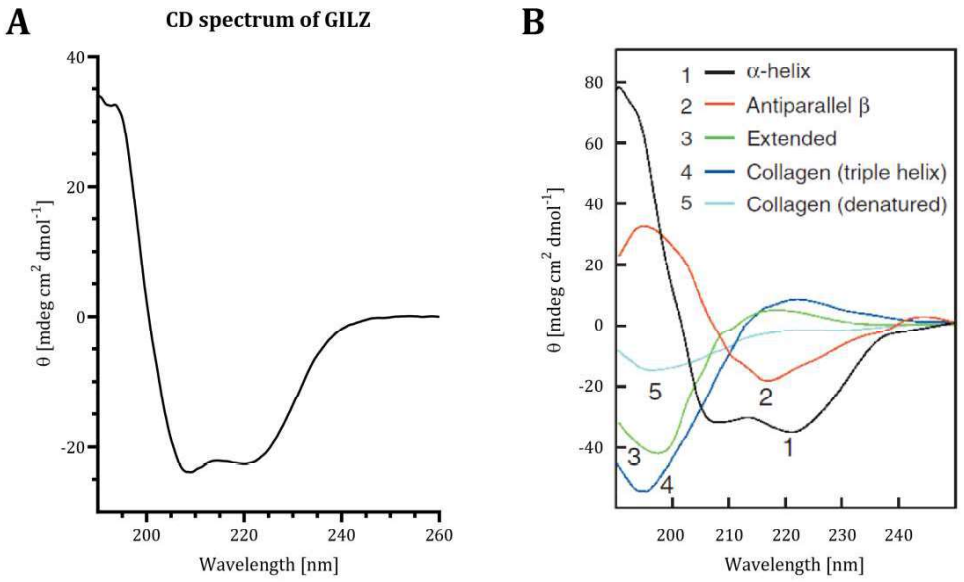


Fig. 6-63 A. *CD spectrum of GILZ in 20 mM sodium phosphate, pH = 6.5 + 500 mM NaF.* ***B.*** *CD spectra of poly-L-lysine at pH 11.1 in the (1, black) α-helical and (2, red) antiparallel β-sheet conformations and at pH 5.7 in the (3, green) extended conformations as well as placental collagen in its (4, blue) native triple-helical and (5, cyan) denatured forms, adapted from N. J. Greenfield,* Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. ***2006***, 1, *2876–2890.*

6.5.2 Cloning and expression of GILZ variants to investigate its oligomeric status

To investigate how individual protein domains and amino acids influence GILZ oligomerization, the constructs summarized in **Table 4** were cloned for bacterial expression. MARCO et al. had described the importance of the leucine zipper domain and specific amino acids on GILZ-GILZ interactions, from which they concluded GILZ homodimerization.[299] Furthermore, they investigated the functional relevance of GILZ "dimerization" by finding that monomeric GILZ is no longer able to interact with NF-κB to fulfill its biological functions. Their experiments indicated that mutation of four leucines of the leucine zipper domain (L76, L83, L90 and L97) to alanines is necessary to abolish GILZ-GILZ interaction. A construct coding for GILZ in which one of these leucines (L83) was deleted was created to study the importance of an intact leucine zipper heptad motif on GILZ oligomerization. Moreover, MARCO et al. reported a single point mutation that abrogated GILZ-GILZ interaction, N87D.[299]

Sequence	Mutation	Purpose
$GILZ_{1-98}$	Deletion of C-terminal PER	Exclusion of C-terminal PER domain to test domain influence on oligomerization
$GILZ_{1-75}$	Deletion of LZ and C-terminal PER	Exclusion of leucine zipper and C-terminal PER domain to test domain influence on oligomerization[299]
$GILZ_{1-134}$ ΔL83	Deletion of Leu83	Influence of Leu83 on oligomerization[299]
$GILZ_{1-134}$ N87D	Point mutation of Asn87 to Asp	Influence of Asn87 on oligomerization[299]

Table 4 *Constructs to study the influence of domains and amino acids on oligomerization of GILZ.*

Constructs excluding the C-terminal PER domain and the leucine zipper domain ($GILZ_{1-98}$ and $GILZ_{1-75}$) were also created to study the relevance of these domains on GILZ oligomerization.

Surprisingly, all constructs, including $GILZ_{1-75}$ and $GILZ_{1-134}$ N87D, which MARCO et al. had shown to be monomeric[299], displayed the same behavior as full length GILZ when subjected to the a Superdex 200 Increase 10/300 GL (data not shown). To gain better understanding, $GILZ_{1-134}$ N87D was expressed as a ^{15}N-labeled protein and analyzed using ^{15}N-^{1}H-HSQC NMR (see **Fig. 6-64**). The NMR spectrum only displayed 15 peaks that are not related to side chain NHs (i.e. not in the region from 6.6 - 7.4 ppm in the proton and 112 – 114 ppm in the nitrogen dimension), 5 of which showed decreased intensity. However, the spectrum of full length monomeric GILZ is expected to display at least 120 peaks, considering the absence of prolines in ^{15}N-^{1}H-HSQC spectra. These results indicate line broadening of protein resonances due to increased molecular weight and corresponding slower tumbling with increased correlation times (see reviews by TUGARINOV et al.[402] and TZAKOS et al.[403] for more information

about NMR spectroscopy on high-molecular weight proteins). Therefore, it can be concluded that $GILZ_{1\text{-}134}$ N87D is still forming homooligomeric complexes.

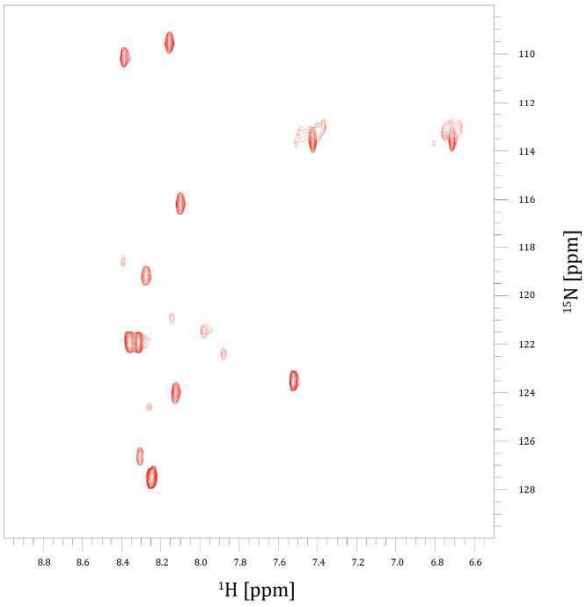


Fig. 6-64 *^{15}N ^{1}H HSQC spectrum of ^{15}N-labeled $GILZ_{1\text{-}134}$ N87D. The spectrum of monomeric GILZ is expected to display ~ 126 peaks (of each residue from the 134 amino acid containing sequence minus 8 prolines). Clearly visible are 10 non-side chain peaks and 5 peaks show decreased intensities, indicating enhanced relaxation due to high molecular weight and corresponding high correlation time.*

Next, the oligomeric nature of GILZ was investigated using negative stain electron microscopy. Full length GILZ was expressed and purified and analyzed using uranyl formate negative stain and transmission electron microscopy.

At a concentration of 0.05 mg/mL, negative stain grids of GILZ displayed good dispersion of molecules that revealed protein complexes of 16 – 41 nm (160 – 400 Å) size (see **Fig. 6-65**). Assuming a perfect spherical size, a protein's minimal radius (R_{min}) can be calculated using equation 7.

$$R_{min} = (\frac{3V}{4\pi})^{1/3} \quad (7)$$

If R_{min} is given in nm and molecular weight (M) is given in Da, it follows:

$$R_{min}[nm] = 0.066M^{1/3} \quad (8)$$

Therefore, a 5 kDa protein would have a minimal radius of 1.1 nM, while a 500 kDa protein would have a minimal radius of 5.2 nm. Examples for globular proteins analyzed by ERICKSON confirm these assumptions.[404] For example, ovalbumin is a protein with a molecular weight of 43 kDa and its dimensions are 7.0 × 3.6 × 3.0 nm, while phosphofructokinase, a 345 kDa protein, displayed dimensions of 14 × 9 × 9 nm (calculated from PDB files). Therefore, particle sizes of > 16 nm for GILZ indicate molecular weight complexes far beyond dimeric GILZ. Considering the distribution of particle sizes ranging from 16 to 41 nm, elongated oligomeric structures are more

likely than a globular shape. Examples for elongated proteins analyzed by ERICKSON are the tenascin fragments TNfn1–5, TNfn1-8 and TNfnALL with protein sizes of 50, 79 and 148 kDa, respectively. Their calculated dimensions are 14.7 × 1.7 × 2.8 nm (TNfn1-5), 24.6 × 1.7 × 2.8 nm (TNfn1-8) and 47.9 × 1.7 × 2.8 nm (TNfnALL).[404]

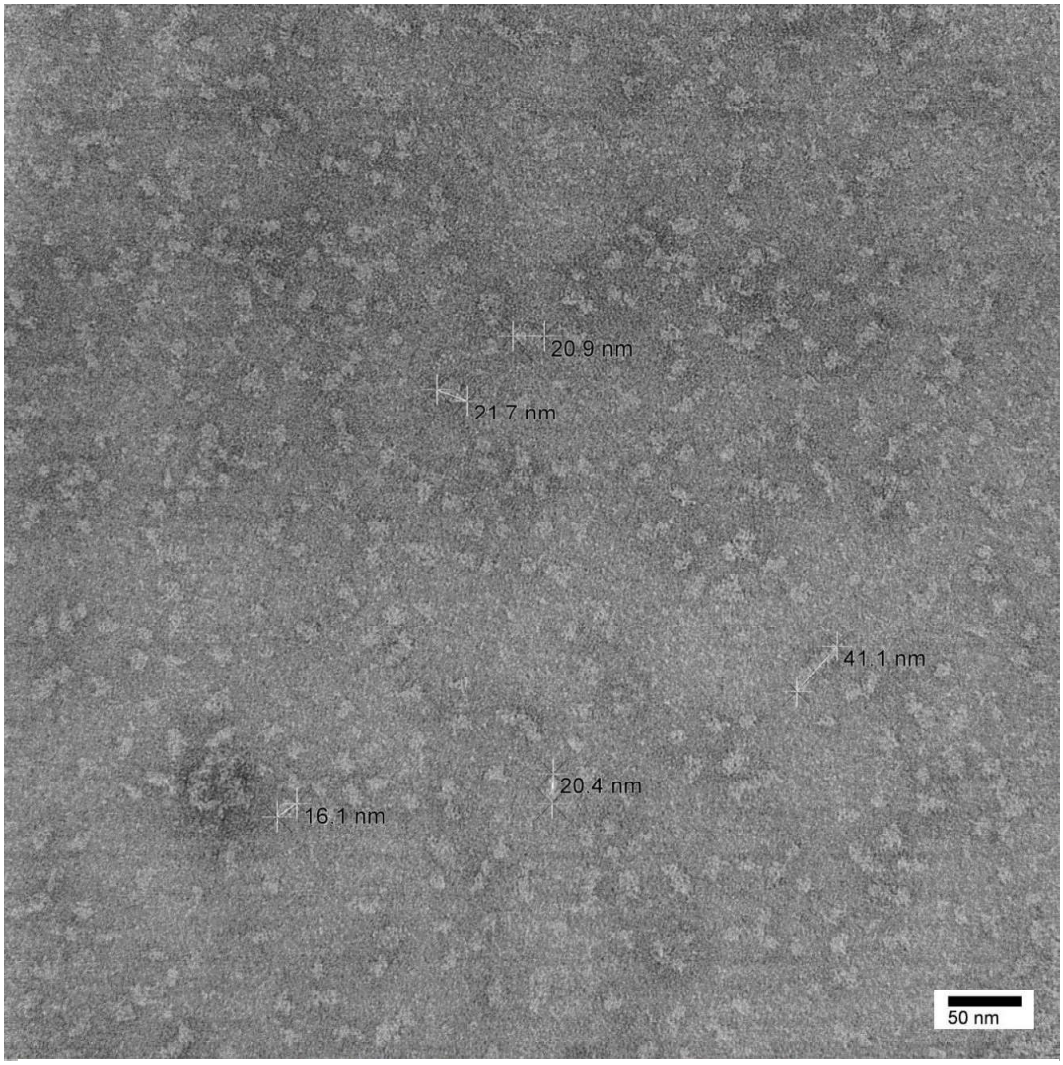


Fig. 6-65 *Negative stain transmission electron microscopy image of $GILZ_{1-134}$. Monodisperse particles of GILZ oligomers without aggregation are observed.*

Therefore, drawing conclusions from particle size to the number of GILZ molecules in the oligomeric complex is not trivial. However, oligomerization beyond dimers is likely, considering data obtained from gel filtration, NMR and EM experiments.

6.5.3 Investigation of a potential GILZ-HDAC7 interaction

The interaction between GILZ and HDAC7 found by collaborators of the lab of ALEXANDRA K. KIEMER was investigated using Western Blot analysis. A plasmid coding for Glutathione *S*-transferase (GST) tagged GILZ was subcloned, expressed and purified. Purified GST-GILZ was then incubated with glutathione sepharose beads and HEK293T lysates transfected with HDAC7-FLAG. Subsequent pull down and Western Blot analysis using FLAG antibodies allowed the identification of protein which interacted with GST-GILZ. To narrow down the binding interface, the truncated plasmids summarized in **Table 5** were cloned.

Construct	Purpose
GST-GILZ$_{1\text{-}134}$	Pull-down for interacting protein with full length GILZ
GST-GILZ$_{1\text{-}98}$	Pull-down for interacting protein with GILZ excluding the C-terminal PER domain
GST-GILZ$_{1\text{-}75}$	Pull-down for interacting protein with GILZ excluding the leucine zipper and C-terminal PER domain
HDAC7-FLAG$_{1\text{-}952}$	Mammalian expression of full length HDAC7 to test interaction with GST-GILZ
HDAC7-FLAG$_{219\text{-}952}$	Mammalian expression of HDAC7 excluding the disordered N-terminal domain to test interaction with GST-GILZ
HDAC7-FLAG$_{269\text{-}952}$	Mammalian expression of HDAC7 excluding transcription repression domain 1 to test interaction with GST-GILZ
HDAC7-FLAG$_{480\text{-}952}$	Mammalian expression of HDAC7 excluding the N-terminal region before the histone deacetylase domain to test interaction with GST-GILZ
HDAC7-FLAG$_{875\text{-}952}$	Mammalian expression of the C-terminal domain of HDAC7 to test interaction with GST-GILZ

Table 5 *Constructs to determine the approximate binding interface between GILZ and HDAC7.*

GST pull-down of full length GST-GILZ with full length HDAC7-FLAG and the interaction of the two proteins was confirmed (see **Fig. 6-66**).

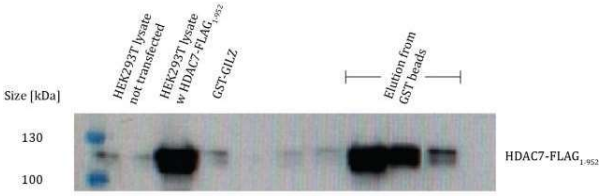


Fig. 6-66 *Western Blot of extracts from HEK293T cells using anti-FLAG antibody. HDAC7-FLAG$_{1-952}$ is overexpressed after transfection. GST-GILZ does not stain for FLAG. GST beads were incubated with GST-GILZ and extracts, washed and then eluted three times with reduced glutathione. Eluted GST beads stain for HDAC7-FLAG$_{1-952}$.*

Next, the truncated HDAC7-FLAG constructs listed in **Table 5** were tested to narrow down the interaction interface. Of the five HDAC7-FLAG constructs, only full length, 269-952 and 480-952 expressed in HEK293T cells. Elution of GST beads incubated with purified GST or GST-GILZ confirmed interaction under the experimental conditions with both GST and GST-GILZ (see **Fig. 6-67**). Experimental conditions need to be improved to assure that HDAC7 is not being pulled down by the tag, but by GILZ itself. Interestingly, truncation of HDAC7 led to the loss of interaction. Therefore, it was concluded that the N-terminal 268 amino acids of HDAC7 are necessary for GILZ interaction (see **Fig. 6-67** top Blot). To further narrow down the interaction interface, the plasmid coding for HDAC7-FLAG$_{219-952}$ was re-purified and transfected again. This time, overexpression was observed. However, GST-GILZ was not able to pull down HDAC7-FLAG$_{219-952}$, placing the interacting region of HDAC7 within the first 218 amino acids (see **Fig. 6-67** bottom Blot). To test whether the N-terminus of HDAC7 alone is sufficient to be pulled down with GST-GILZ, a plasmid coding for HDAC7-FLAG$_{1-129}$ was cloned. Experiments confirming the interaction between HDAC7$_{1-129}$ and GILZ will be carried out in the future.

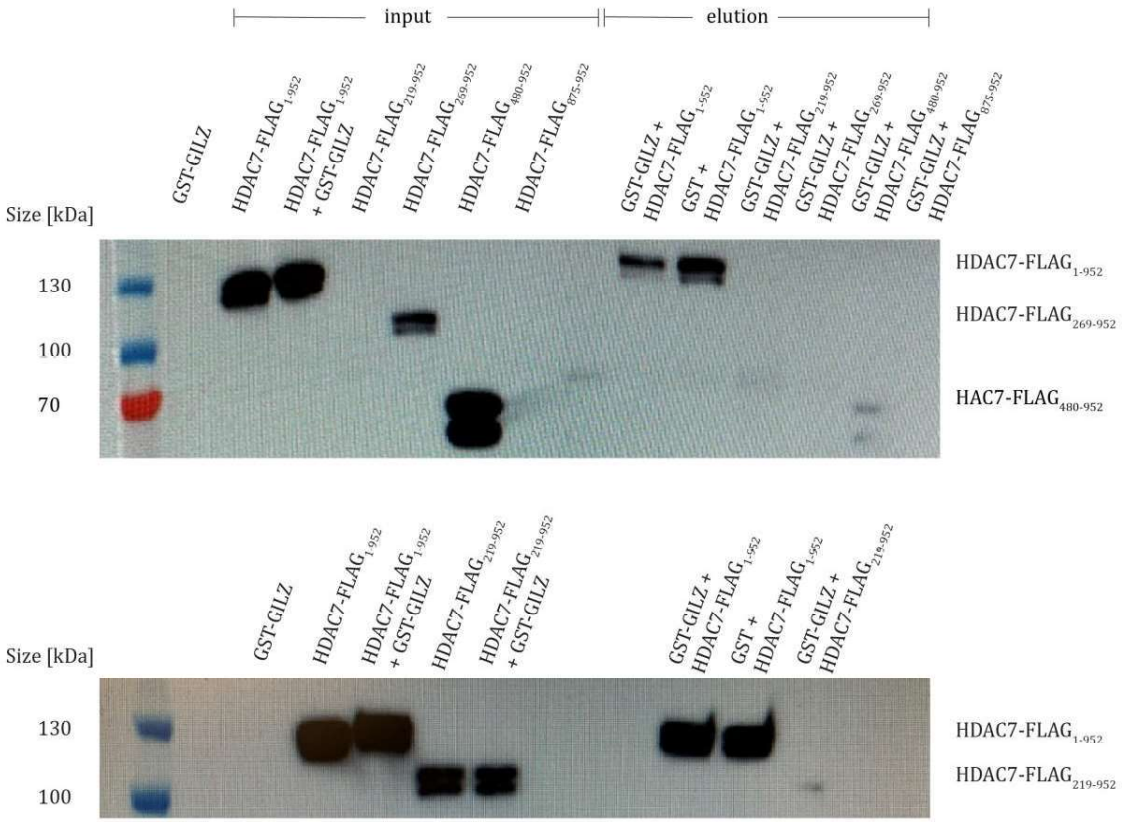


Fig. 6-67 *Top Blot: Western Blot of extracts from HEK293T cells using anti-FLAG antibody. HDAC7-FLAG$_{1-952}$, HDAC7-FLAG$_{269-952}$ and HDAC7-FLAG$_{480-952}$ are overexpressed after transfection. GST beads were incubated with GST-GILZ or GST and extracts, washed and then eluted with reduced glutathione (30 mM). Eluted GST beads stain for only HDAC7-FLAG$_{1-952}$. Bottom Blot: Western Blot of extracts from HEK293T cells using anti-FLAG antibody. HDAC7-FLAG$_{1-952}$ and HDAC7-FLAG$_{219-952}$ are overexpressed after transfection. GST beads were incubated with GST-GILZ or GST and extracts, washed and then eluted with reduced glutathione (30 mM).*

6.5.4 Discussion

NMR and TEM analyses of GILZ and GILZ mutants confirmed the oligomeric nature of the protein. Future endeavors using cryo-EM will be made to solve the structure of GILZ to gain deeper understanding of the implications of GILZ oligomerization in its biological functions.

Furthermore, preliminary results seem to confirm the interaction between GILZ and HDAC7. Truncating mutations of HDAC7 have further narrowed down the interaction site to the N-terminal 218 amino acids of HDAC7. However, experimental conditions need to

be improved to discard false potential positive pull-down results. Since others have used GST pull-down assays with HDAC7 before without noticing HDAC7-GST interaction, repetition of the GST-GILZ and GST pull-downs will be carried out using the experimental conditions described by the authors.[405]

6.6 Keap1

6.6.1 Crystallization and ligand-soaking with iKeap inhibitors

To accompany the data obtained in our publication on Keap1 inhibitors (see **C An open-source drug discovery platform enables ultra-large virtual screens**), an effort was made to generate X-ray co crystal structures of Keap1 with some of the best performing inhibitors from the publication (i.e. iKeap1, iKeap2, iKeap7 and iKeap8).

Although difficulties on ligand soaking with wild type Keap1 crystals have been reported, the first attempt within this work was made using wild type Keap1 in collaboration with the crystallography core of SIRANO DHE-PAGANON. Keap1 crystallized in rhombic crystals (see **Fig. 6-68 A**) and ligand soaking was performed with different concentrations of either iKeap1, iKeap2, iKeap7 or iKeap8. The crystals showed diffraction patterns ranging from ~1.4 to 2 Å (see **Fig. 6-68 B** for representative crystals and their diffraction) and data processing furnished 12 X-ray crystal structures between 1.3 and 1.5 Å.

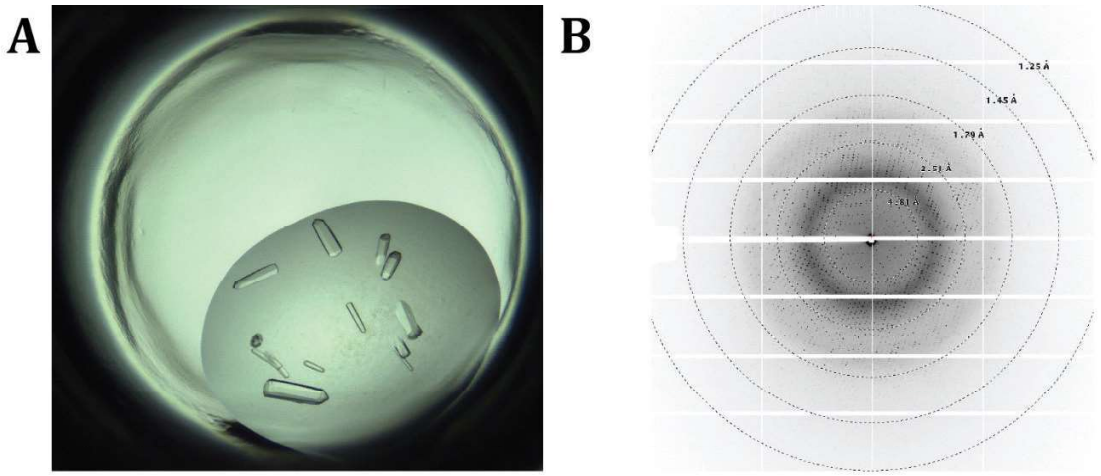


Fig. 6-68 A. *Crystals obtained of wild type Keap1, crystallized in orthorhombic shapes.* ***B.*** *Diffraction pattern of one dataset obtained from wild type Keap1 crystals soaked with iKeap1.*

The obtained crystal structures of Keap1 soaked with iKeap1 ligand were aligned and electron density maps were analyzed for ligand density at the expected binding region

(see **Fig. 6-69**). Sadly, none of the solved structures displayed density for small molecules in the binding cavity.

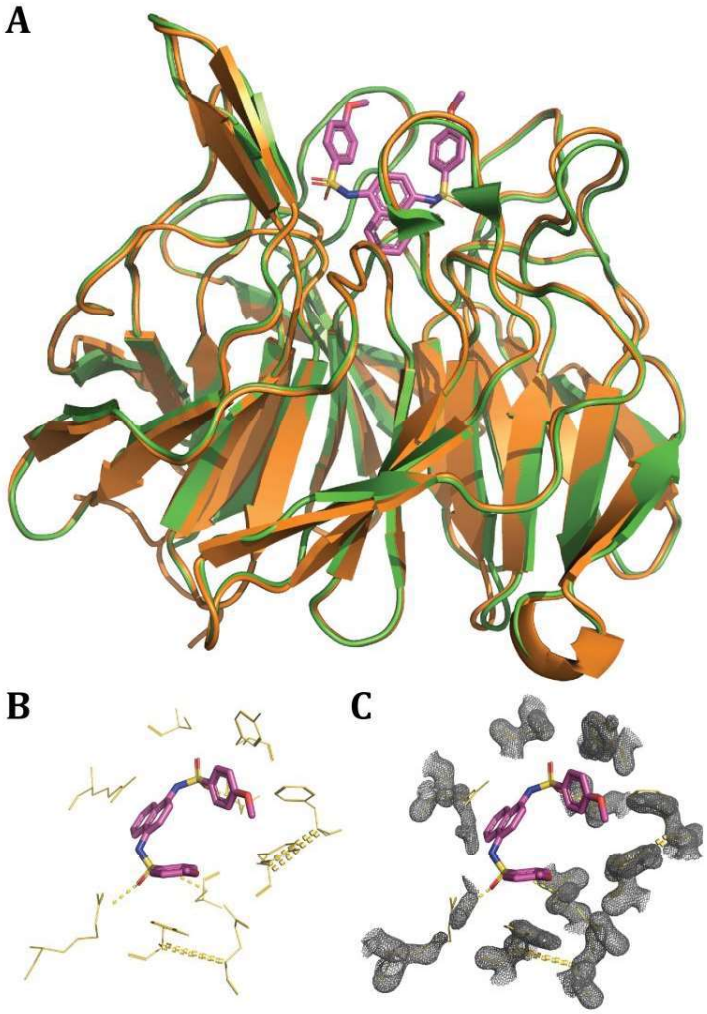


Fig. 6-69 *X-ray Crystal structure of Keap1 soaked with iKeap1, at a resolution of 1.4 Å.* ***A.*** *Alignment of the crystal structure of Keap1 soaked with iKeap1 and Keap1 bound to C17 (PDB ID: 4IQK). Keap 1 soaked with iKeap1 is presented in orange, 4IQK is presented in green and the ligand C17 is presented in magenta. The fold of both structures is identical.* ***B.*** *Stick representation of the ligand-protein interaction site between C17 (magenta) and interacting side chains of Keap1 (from the solved structure of Keap1 soaked with iKeap1, aligned to 4IQK, shown in yellow).* ***C.*** *Electron density mapping of the ligand-protein interaction site obtained from the structure of Keap1 soaked with iKeap1. No electron density is found at the expected binding site.*

6.6.2 Preformation of wild type Keap1 – iKeap1 complexes

Next, an approach alternative to ligand soaking with apo-Keap1 crystals was employed by forming Keap1-ligand complexes before crystallization. To accomplish this, Keap1 was concentrated and combined with iKeap1 in a 1:2 ratio. The complex was centrifuged at 10,000 x g for 15 min at 4 °C to remove precipitate and then subjected to crystallization. As for wild type Keap1, formation of rhombic crystals was observed (see **Fig. 6-70**). As these initial results look promising, synchrotron measurements of the

crystals shown in **Fig. 6-70** will be carried out in the near future to solve the structure of the supposed protein complexes.

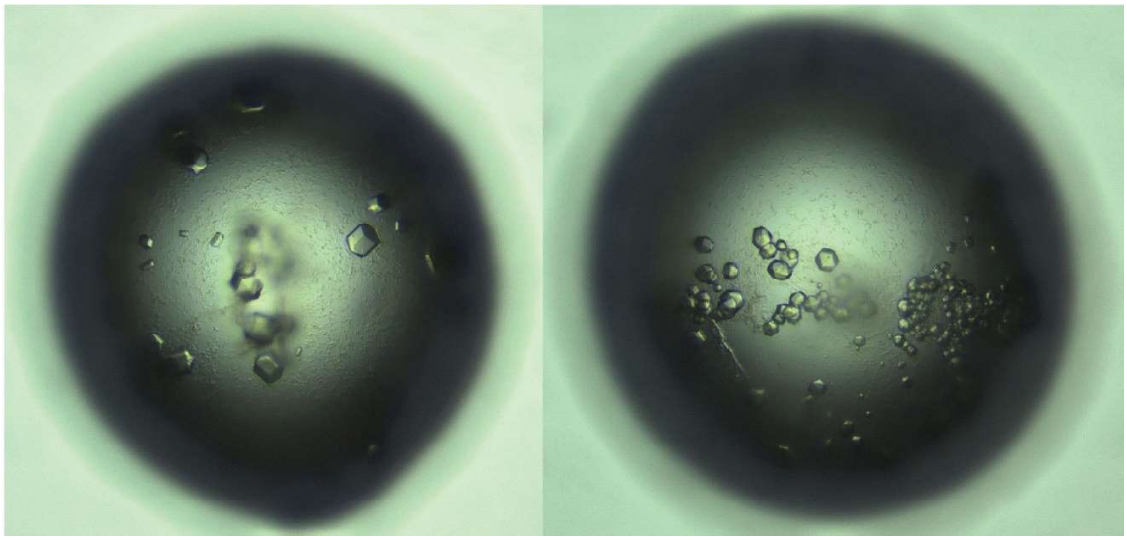


Fig. 6-70 *Crystals obtained from wild type Keap1 formed from a complex of wild type Keap1 and iKeap1 crystallize in orthorhombic shapes.*

6.6.3 Crystal preparation of Keap1 E540A E542A

To circumvent the problem that crystal packing of wild type Keap1 causes with respect to ligand soaking, a double mutation of Keap1 (E540A, E542A) was cloned to create a construct which crystallizes under different conditions.[340] Initial results revealed that the double mutant indeed does not crystallize under the same conditions as wild type Keap1. However, more optimization of the new crystallization conditions for Keap1 E540A E542A needs to be performed, as crystals obtained so far are of poor quality (see **Fig. 6-71**).

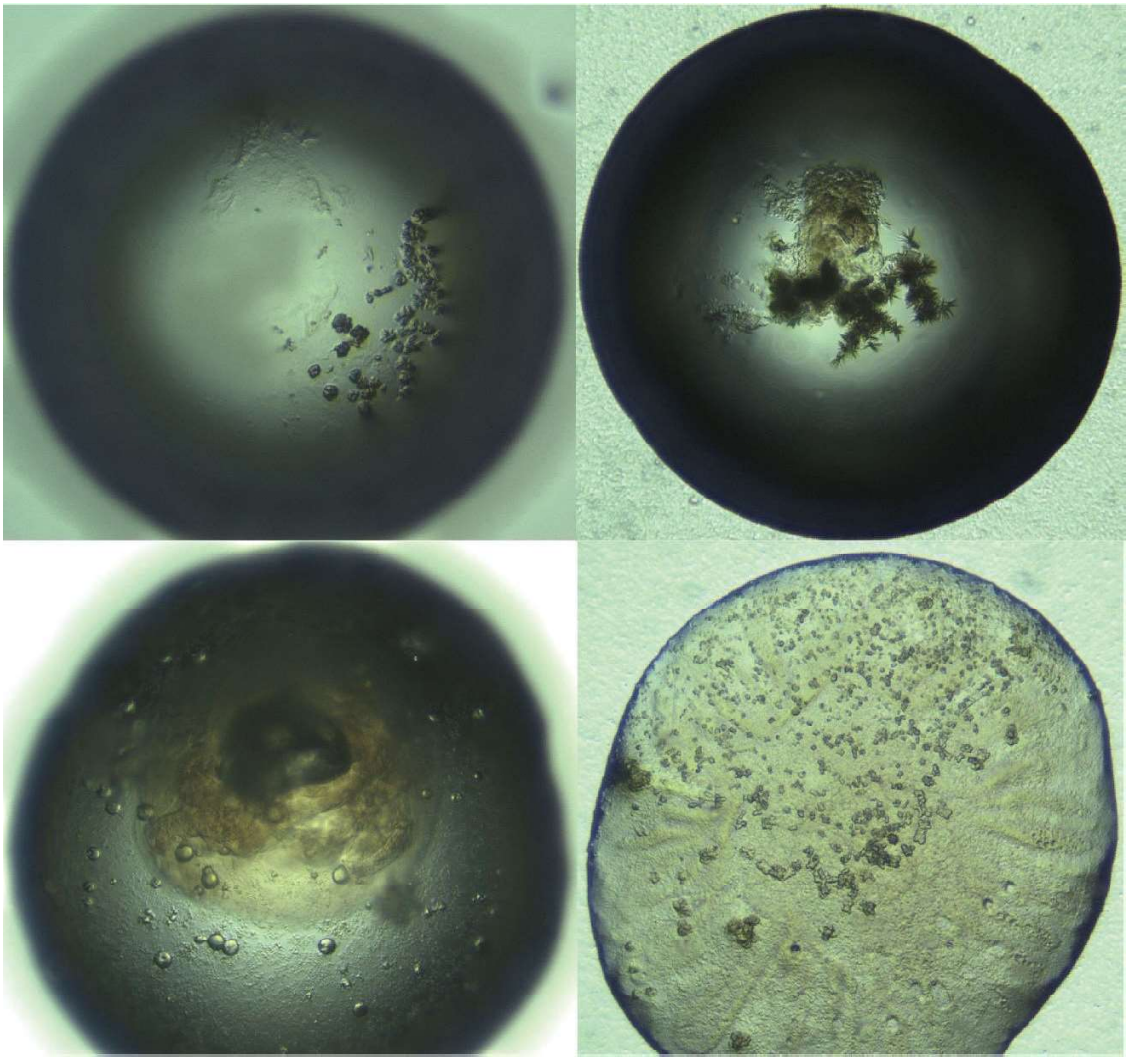


Fig. 6-71 *Initial crystal hits for Keap1 E540A E542A. Obtained protein crystals form in different conditions compared to wild type Keap1 and crystal forms look dissimilar.*

6.6.4 Discussion

Optimization of Keap1 crystallization experiments revealed that wild type Keap1 crystals are indeed, as previously stated[340], not suitable for ligand-soaking. First progress was made by creating preformed Keap1-iKeap1 complexes which produced orthorhombic crystals similar to apo-Keap1. Structure determination will be attempted for these crystals and molecular replacement will be performed to locate iKeap1 densities within the protein structure. Additionally, the Keap1 double mutant E540A E542A was successfully cloned and purified to create Keap1 crystals suitable for ligand soaking. First

results indicate different crystallization conditions and crystal forms compared to wild type Keap1. However, further optimization is needed and will be performed in the future to obtain high quality crystals to be soaked with iKeap inhibitors.

7 Experimental section

7.1 General protocols

7.1.1 Plasmid transformation into chemically competent *E. coli*

25 µL of varying strains of chemically competent *E. coli* purchased from New England BioLabs were thawed on ice for 10 min. Then, ~ 50 ng of one of the purified plasmids was added and the mixture was incubated for 30 min. A heat shock was performed at 42 °C for 45 s to transform the plasmid into the bacteria. After an incubation period of 2 min on ice, 300 µL SOC medium was added and the mixture was incubated in a shaking incubator at 37 °C for 1 h. Then, 50 µL of the mixture was plated on LB agar plates supplemented with either 50 µg/mL kanamycin ($MraY_{AA}$, $MraY_{AA}$ H324Q, $MraY_{AA}$ D193N, $MraY_{CB}$, Homer3, Δ_{26}-eIF4E, Δ_{26}-eIF4E S209D, GB1-eIF4E, GB1-eIF4E S209D, Sfp, GST-Keap1 and Keap1 E540A E542A), 50 µg/mL kanamycin + 30 µg/mL chloramphenicol (MurF, GILZ, $GILZ_{1\text{-}75}$, $GILZ_{1\text{-}98}$, GILZ N87D, GST-GILZ, $MraY_{CB}$, $MraY_{SA}$, $MraY_{SA}$ H287Q, $MraY_{SP}$ and $MraY_{SP}$ Q316H, 100 µg/mL carbenicillin ($NFAT_{1\text{-}130}$ S110A, $NFAT_{1\text{-}130}$ D14C S110A, $NFAT_{1\text{-}130}$ D35C S110A, $NFAT_{1\text{-}130}$ L65C S110A, $NFAT_{1\text{-}130}$ S99C S110A, $NFAT_{1\text{-}130}$ S110A A124C, $NFAT_{1\text{-}130}$ D35C S99C S110A, CBP/p300 KIX, Med15 KIX and EsaI), 100 µg/mL streptomycin sulfate (BmaI1) or 100 µg/mL streptomycin sulfate + 100 µg/mL carbenicillin (MKK6-EE/GST-p38α) and incubated at 37 °C overnight.

7.1.2 Recombinant expression of unlabeled proteins in *E. coli*

All proteins in this thesis were expressed in *E. coli*. For protein expression without specific isotope labeling, cells were grown in Luria broth (LB) Miller.

7.1.2.1 Procedure A: Expression in *E. coli* BL21(DE3)

Procedure A was used to produce unlabeled $NFAT_{1\text{-}130}$ S110A, $NFAT_{1\text{-}130}$ D35C S99C S110A, MKK6-EE/GST-p38α, CBP/p300 KIX, Med15 KIX, Homer3, Δ_{26}-eIF4E, Δ_{26}-eIF4E S209D, GB1-eIF4E, GB1-eIF4E S209D, Sfp, BmaI1, EsaI, GST-Keap1 and Keap1 E540A E542A.

10 mL LB media per liter large culture supplemented with the appropriate antibiotic was inoculated with 1-5 colonies from the LB agar plates and incubated in a shaking incubator at 37 °C overnight. The following morning, 1 L of LB supplemented with the appropriate antibiotic was inoculated with 10 mL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C to an OD_{600} of ~ 0.4, after which the temperature was dropped to 20 °C. Protein expression was induced at an OD_{600} of 0.7-0.9 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of either 1 mM ($NFAT_{1-130}$ S110A, $NFAT_{1-130}$ D35C S99C S110A, CBP/p300 KIX, Med15 KIX, Homer3, GST-Keap1 and Keap1 E540A E542A), 0.5 mM (MKK6-EE/GST-p38α, Sfp, BmaI1 and EsaI) or 0.2 mM (Δ_{26}-eIF4E, Δ_{26}-eIF4E S209D, GB1-eIF4E and GB1-eIF4E S209D) and grown at 20 °C overnight. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

7.1.2.2 Procedure B: Expression in *E. coli* C41(DE3)

Procedure B was used to produce $MraY_{AA}$, $MraY_{AA}$ H324Q, $MraY_{AA}$ D193N and $MraY_{CB}$.

10 mL LB media per liter large culture supplemented with 50 µg/mL kanamycin was inoculated with 1-5 colonies from the LB agar plates and incubated in a shaking incubator at 37 °C overnight. The following morning, 1 L of LB supplemented with 50 µg/mL kanamycin was inoculated with 10 mL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C to an OD_{600} of ~ 0.4, after which the temperature was dropped to 20 °C. Protein expression was induced at an OD_{600} of 0.7-0.9 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM and grown at 20 °C overnight. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

7.1.2.3 Procedure C: Expression in *E. coli* Lemo21(DE3)

Procedure C was used to produce overexpressed $MraY_{CB}$, $MraY_{SA}$, $MraY_{SA}$ H287Q, $MraY_{SP}$, $MraY_{SP}$ Q316H.

10 mL LB media per 500 mL large culture supplemented with 50 µg/mL kanamycin and 30 µg/mL chloramphenicol was inoculated with 1-5 colonies from the LB agar plates and

incubated in a shaking incubator at 37 °C overnight. The following morning, 500 mL of LB supplemented with 50 µg/mL kanamycin, 30 µg/mL chloramphenicol and 1 mM L-rhamnose was inoculated with 500 µL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C and protein expression was induced at an OD_{600} of 0.6 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Cells were harvested after 4 h at 37 °C by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

7.1.2.4 Procedure D: Expression in *E. coli* BL21(DE3)pLysS

Procedure D was used for the expression of TEV protease, 3C protease, MurF, GILZ, $GILZ_{1-75}$, $GILZ_{1-98}$, GILZ ΔL83, GILZ N87D and GST-GILZ.

10 mL LB media per liter large culture supplemented with 50 µg/mL kanamycin and 30 µg/mL chloramphenicol was inoculated with 1-5 colonies from the LB agar plates and incubated in a shaking incubator at 37 °C overnight. The following morning, 1 L of LB supplemented with 50 µg/mL kanamycin was inoculated with 10 mL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C to an OD_{600} of ~ 0.4, after which the temperature was dropped to 20 °C. Protein expression was induced at an OD_{600} of 0.7-0.9 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.5 mM and grown at 20 °C overnight. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

7.1.3 Expression of U-^{15}N and U-^{15}N^{13}C labeled proteins

Modified M9 minimal media for expression of U-^{15}N^{13}C labeled proteins was prepared as follows: 6 g Na_2HPO_4, 3 g KH_2PO_4 and 0.5 g NaCl were dissolved in 1 L of double deionized water and autoclaved. Then, the appropriate antibiotic was added, as well as sterile filtered solutions of $MgSO_4$ (final concentration = 2.5 mM), $CaCl_2$ (final concentration = 0.1 mM), 2 g D-Glucose-$^{13}C_6$, 1 g $^{15}NH_4Cl$ and 100 µL of a 10,000 x vitamin mix (5 g riboflavin, 5 g niacinamide, 5 g pyridoxine monohydrate and 5 g thiamine dissolved in ethanol). For the expression of U-^{15}N labeled proteins, 4 g D-Glucose was used in an otherwise identical medium.

10 mL LB media per liter large culture supplemented with the appropriate antibiotic was inoculated with 1-5 colonies from the LB agar plates and incubated in a shaking incubator at 37 °C overnight. The following morning, 1 L of the modified M9 media described above was inoculated with 10 mL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C to an OD_{600} of ~ 0.4, after which the temperature was dropped to 20 °C. Protein expression was induced at an OD_{600} of 0.7-0.9 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of either 1 mM ($NFAT_{1-130}$ S110A, $NFAT_{1-130}$ D14C S110A, $NFAT_{1-130}$ D35C S110A, $NFAT_{1-130}$ L65C S110A, $NFAT_{1-130}$ S99C S110A, $NFAT_{1-130}$ S110A A124C), 0.5 mM (GILZ N87D) or 0.2 mM (GB1-eIF4E and GB1-eIF4E S209D) and grown at 20 °C overnight. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

7.1.4 Purification of His_6-tagged proteins

This procedure was used for the purification of MurF, $NFAT_{1-130}$, $NFAT_{1-130}$ S110A, $NFAT_{1-130}$ D14C S110A, $NFAT_{1-130}$ D35C S110A, $NFAT_{1-130}$ D35C S99C S110A, $NFAT_{1-130}$ L65C S110A, $NFAT_{1-130}$ S99C S110A, $NFAT_{1-130}$ S110A A124C, MSP1E3D1, Homer3, Sfp, BmaI1, EsaI, GILZ, $GILZ_{1-75}$, $GILZ_{1-98}$, GILZ ΔL83 and GILZ N87D.

Bacterial cell pellets were resuspended in ~ 15 mL resuspension buffer (50 mM Tris-HCl, pH = 8.0, 350 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanole [BME], 1 cOmplete™ Protease Inhibitor Cocktail tablet) per liter culture. Resuspended cells were then lysed by sonication (30 % amplitude, 5 min sonication time, 2 s on pulse, 4 s off pulse). Cell lysates were pelleted by centrifugation at 30,000 x g for 40 min at 4 °C. Cleared lysates were incubated with ~ 3 mL Ni-NTA agarose resin, preequilibrated with wash buffer (50 mM Tris-HCl, pH = 8.0, 350 mM NaCl, 10 mM imidazole, 2 mM BME) per liter culture for 2-16 h at 4 °C. The resin was washed with wash buffer until 10 µL of eluent no longer stained blue with 50 µL of Bradford reagent (Coomassie blue G-250). Bound protein was eluted with elution buffer (50 mM Tris-HCl, pH = 8.0, 350 mM NaCl, 350 mM imidazole, 2 mM BME). Further purification was performed according to individual protocols.

7.1.5 Purification of GST-tagged proteins

This procedure was used for the purification of CBP/p300 KIX, Med15 KIX, MKK6-EE/GST-p38α, GST-GILZ, Keap1 and Keap1 E540A E542A.

Bacterial cell pellets were resuspended in ~ 15 mL resuspension buffer (50 mM Tris-HCl, pH = 7.5, 350 mM NaCl, 2 mM β-mercaptoethanole [BME], 1 mM ethylenediaminetetraacetic acid [EDTA], 1 cOmplete™ Protease Inhibitor Cocktail tablet) per liter culture. Resuspended cells were then lysed by sonication (30 % amplitude, 5 min sonication time, 2 s on pulse, 4 s off pulse). Cell lysates were pelleted by centrifugation at 30,000 x g for 40 min at 4 °C. Cleared lysates were incubated with ~ 3 mL Ni-NTA agarose resin, preequilibrated with wash buffer (50 mM Tris-HCl, pH = 7.5, 350 mM NaCl, 2 mM BME, 1 mM EDTA) per liter culture for 4-16 h at 4 °C. The resin was washed with wash buffer until 10 µL of eluent no longer stained blue with 50 µL of Bradford reagent (Coomassie blue G-250). Bound protein was eluted with elution buffer (50 mM Tris-HCl, pH = 7.5, 350 mM NaCl, 2 mM BME, 1 mM EDTA, 30 mM reduced glutathione). Further purification was performed according to individual protocols.

7.1.6 Cloning protocols using IVA cloning

All cloning procedures have been performed using *in vivo* assembly (IVA) cloning, according to the general protocol by García-Nafría et al.[406] The general procedure for each cloning type is described below, followed by specific conditions for the designed plasmids individually.

7.1.6.1 Single point mutations

Fig. 7-1 shows the design and principle of mutagenesis experiments: The forward primer is a long primer that consists of a homologous region (orange) and a template binding region (green). The replacement codon is positioned between these two parts and is not taken into consideration for primer lengths and melting temperature (T_m). It has been shown that colony-forming units per plate (CFU/plate) increased linearly with increasing T_m of the homologous region.[406] However, to avoid primer dimerization, it is favorable to limit the T_m value to a maximum of 52 °C. Similarly, the CFU/plate value increases linearly with increasing primer length, reaching saturation at 15 base pairs. It should be noted that, while increasing the length of the homologous region to 20 base pairs yielded no additional CFU/plate, reduction to 10 base pairs decreased the CFU/plate-value by nearly half. It is therefore advisable to aim for a homologous region length of at least 15 base pairs, enduring slightly higher T_m values in case of high GC content. After the replacement codon, a template binding region with a T_m value of 60 °C is added. The reverse primer consists of a template binding region with a T_m value of 60 °C, which should include the complementary homologous region to the forward primer towards its 5' end.

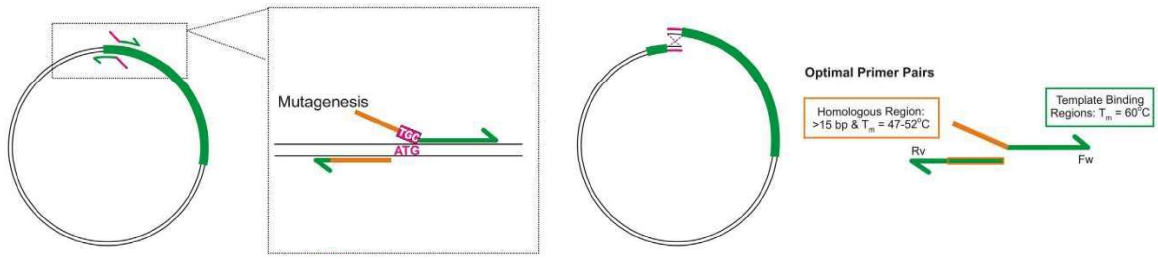


Fig. 7-1 *Principle of single point mutagenesis using IVA cloning. Highlighted in green is the gene of interest. Primers are comprised of a homologous region (orange) and a template binding region (green). Taken from J. García-Nafría, J. F. Watson, I. H. Greger,* IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Sci. Rep. ***2016***, 6, *27459.*

Unless otherwise stated, PCR reaction were set up as follows: Phusion High-Fidelity PCR Master Mix with GC Buffer was combined with 3-5 ng template DNA, 1 µL DMSO and 200 nM primers to a final volume of 25 µL. PCR was then performed using the following protocol:

1. Initial denaturation: 98 °C, 30 s.
2. Denaturation: 98 °C, 10 s.
3. Annealing: 60 °C, 30 s.
4. Extension: 72 °C, 20 s/kb.
   Repeat from 2, 19 times
5. Final extension: 72 °C, 5 min.

$MraY_{AA}$ H324Q

Fwd primer: 5'-AAACGTGCCCCGTTTCAACATCACCTGGAACTGAATGGTCT-3', rev primer: 5'-AAACGGGGCACGTTTGAAC-3'.

$MraY_{AA}$ D193N

Fwd primer: 5'-GCCGTCAATCTGACCAACGGGCTGGATGGCCTGG-3', rev primer: 5'-GGTCAGATTGACGGCATTGG-3'.

$MraY_{SP}$ Q317H

Fwd primer: 5'-CGTATGGCGCCTCTCCATCATCACTTCGAACTGAAAGGTTGG-3', rev primer: 5'-GAGAGGCGCCATACGGAAC-3'.

$MraY_{SP}$ 357stop

Fwd primer: 5'-GCAAGCCGCGAAATGATCTGGTTCGGAAAACCTTAACTCG-3', rev primer: 5'-TTTCGCGGCTTGCCAACC-3'.

$MraY_{SA}$ H287Q

Fwd primer: 5'-GTATATTTAAAATGAGTCCGATTCAACATCATTTTGAATTGATAGGTTGGAGTGA-3', rev primer: 5'-AATCGGACTCATTTTAAATATACGCTTTCC-3'.

$NFAT_{1\text{-}130}$ S110A

Fwd primer: 5'-GCCTCGGGCCTGGCCCCTCGGATCGAGATCACTCCG-3', rev primer: 5'-CAGGCCCGAGGCCCC-3'.

$NFAT_{1\text{-}130}$ D14C S110A

Fwd primer: 5'-GACGGCGGGTGCGCCCCAGGCCAC-3', rev primer: 5'-GGGTTGGGGCTGCCG-3'.

$NFAT_{1\text{-}130}$ D35C S110A

Fwd primer: 5'-GACTTCTCCATCCTCTTCTGCTATGAGTATTTGAATCCGAACGAAGAAGA-3', rev primer: 5'-GAAGAGGATGGAGAAGTCAAGCTC-3'.

NFAT$_{1\text{-}130}$ L65C S110A

Fwd primer: 5'-CATACCCCGATGATGTCTGTGACTATGGCCTCAAGCCATACAG-3', rev primer: 5'-GACATCATCGGGGTATGCGG-3'.

NFAT$_{1\text{-}130}$ S99C S110A

Fwd primer: 5'-GCCGCAGAAGTTTCTGTGCGCGGCCAAGCCAGCA-3', rev primer: 5'-CAGAAACTTCTGCGGCCCTA-3'.

NFAT$_{1\text{-}130}$ S110A A124C

Fwd primer: 5'-CCACGAACTGATCCAGTGTGTGGGGCCCCTCCGC-3', rev primer: 5'-TGGATCAGTTCGTGGGACGG-3'.

GFP-NFAT S23A

Fwd primer: 5'-CCTGGGGGCGCCCCCCAAGACGAGCTTG-3', rev primer: 5'-GCCCCCAGGCTCGTG-3'.

GFP-NFAT S23D

Fwd primer: 5'-GAGCCTGGGGGCGACCCCCAAGACGAGCTTGACTTC-3', rev primer: 5'-GCCCCCAGGCTCGTG-3'.

GILZ N87D

Fwd primer: 5'-CGAACTGGTGGAGAAAGATAGCCAGCTGGAGCGC-3', rev primer: 5'-TTTCTCCACCAGTTCGCGG-3'.

Keap1 E540A

Fwd primer: 5'-TGAACGTTACGACGTGGCAACCGAAACCTGGACCTTCG-3', rev primer: 5'-CACGTCGTAACGTTCAACGC-3'.

Keap1 E542A

Fwd primer: 5'-TACGACGTGGCAACCGCAACCTGGACCTTCGTTGCG-3', rev primer: 5'-GGTTGCCACGTCGTAACGT-3'.

7.1.6.2 Insertions and deletions

The insertion is created ideally by using the new sequence to act as the homologous region in the forward and reverse primers (pictured in magenta). The forward primer will continue at the desired position after the insertion of the used plasmid for additional bases until a template binding region with a T_m value of 60 °C is achieved. The reverse primer contains a template binding region with a T_m value of 60 °C, complementary to

the region before the desired position of the insertion. For deletions, the reverse primer contains a homologous region according to the rules mentioned before (> 15 bp, T_m 47-52 °C) starting complementary to the region before the desired deletion. The reverse primer is extended to reach a T_m value of 60 °C to act as a template binding region. The forward primer starts with the homologous region complementary to the homologous region in the reverse primer, followed by a template binding region with a T_m value of 60 °C at the position after the desired deletion.

PCR reactions were set up as described in **7.1.6.1**.

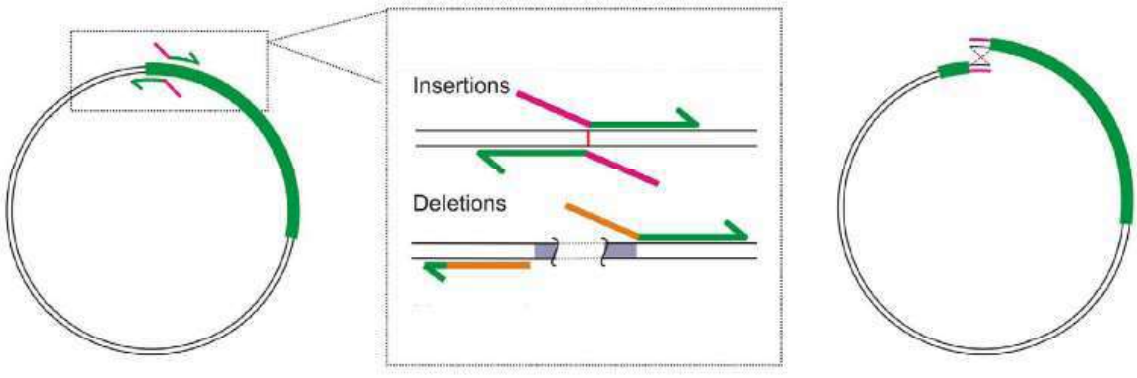


Fig. 7-2 *Principle of insertions and deletions using IVA cloning. Highlighted in green is the gene of interest. For insertions, the inserted DNA sequence simultaneously serves as a homologous region (magenta). For deletions, homologous regions are depicted in orange. Template binding regions complete the primers (green). Taken from J. García-Nafría, J. F. Watson, I. H. Greger,* IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Sci. Rep. ***2016***, 6, *27459.*

MraY$_{AA}$ ΔMBP

Fwd primer: 5'-TAACTTTAAGAAGGAGATATACATATGCTGTATCAACTGGCTCTGC-3', rev primer: 5'-ATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCT-3'.

GFP-NFAT-$_6$PRMLTPINEEAD$_{17}$

Fwd primer: 5'-CCTAGAATGCTGACACCTATCAACGAGGAGGCCGACCGGCAGCCCCAACCC-3', rev primer: 5'-GTCGGCCTCCTCGTTGATAGGTGTCAGCATTCTAGGCTCGGGGGCGTTCATGG-3'.

GFP-NFAT-$_{130}$PRMLTPINEEAD$_{141}$

Fwd primer: 5'-CCTAGAATGCTGACACCTATCAACGAGGAGGCCGACATGAGAGACGCGGGCCT-3', rev primer: 5'-GTCGGCCTCCTCGTTGATAGGTGTCAGCATTCTAGGGCGGAGGGGCCCCA-3'.

GFP-NFAT-$_{110}$PRMLTPINEEAD$_{121}$

Fwd primer: 5'-CCTAGAATGCTGACACCTATCAACGAGGAGGCCGACCCGTCCCACGAACTGATCCA-3', rev primer: 5'-GTCGGCCTCCTCGTTGATAGGTGTCAGCATTCTAGGCAGGCCCGAGGCCCC-3'

GFP-NFAT-$_{110}$PRMATPANAEAD$_{121}$

Fwd primer: 5'-CCTAGAATGGCCACACCTGCCAACGCCGAGGCCGACCCGTCCCACGAACTGATCCA-3', rev primer: 5'-GTCGGCCTCGGCGTTGGCAGGTGTGGCCATTCTAGGCAGGCCCGAGGCCCC-3'.

GFP-NFAT-$_{110}$IKKLSPIIEDSR$_{121}$

Fwd primer: 5'-ATCAAGAAGCTGAGCCCTATCATCGAGGACAGCAGACCGTCCCACGAACTGATCCA-3', rev primer: 5'-TCTGCTGTCCTCGATGATAGGGCTCAGCTTCTTGATCAGGCCCGAGGCCCC-3'.

GILZ ΔL83

Fwd primer: 5'-TCTGCTGTCCTCGATGATAGGGCTCAGCTTCTTGATCAGGCCCGAGGCCCC-3', rev primer: 5'-TTCGCGGATCTGCTCCTTTAA-3'.

GILZ$_{1\text{-}98}$

Fwd primer: 5'-CGCGAAAACACTTTACTGTAAACTTTAGCCAGCCCGGAAC-3', rev primer: 5'-CAGTAAAGTGTTTTCGCGCTCC-3'.

GILZ$_{1\text{-}75}$

Fwd primer: 5'-GCGAAGAAGTGGAAATTTGAAAGGAGCAGATCCGCGAAC-3', rev primer: 5'-AATTTCCACTTCTTCGCGAACG-3'.

HDAC7-FLAG$_{219\text{-}952}$

Fwd primer: 5'-CGGTGGAGCCCATGCCCAATCCCATCCTGGGC-3', rev primer: 5'-CATGGGCTCCACCGAGC-3'.

HDAC7-FLAG$_{269\text{-}952}$

Fwd primer: 5'-CGGTGGAGCCCATGAGGACCCATCCGACTCTGG-3', rev primer: 5'-CATGGGCTCCACCGAGC-3'.

HDAC7-FLAG$_{480-952}$

Fwd primer: 5'-CGGTGGAGCCCATGCCTCTGTCCCGGGCTCA-3', rev primer: 5'-CATGGGCTCCACCGAGC-3'.

HDAC7-FLAG$_{875-952}$

Fwd primer: 5'-CGGTGGAGCCCATGAAACCCAACCTCAATGCCATC-3', rev primer: 5'-CATGGGCTCCACCGAGC-3'.

HDAC7-FLAG$_{1-218}$

Fwd primer: 5'-CAGCGAGCACGGCTCGACTGTGCCTTCTAGTTGC-3', rev primer: 5'-GCCGTGCTCGCTGTCATT-3'.

7.1.6.3 Subcloning

Subcloning approaches using IVA cloning methods make use of the simultaneous amplification of the vector in which the gene of interest is supposed to be cloned into (either empty or containing another gene) and the insert (gene of interest, ideally in another vector). The amplified insert contains homologous regions to the amplified vector at the 5' and 3' ends of the gene of interest. Both polymerase chain reactions (PCRs) (insert and vector) can be performed in a single tube or in separate tubes followed

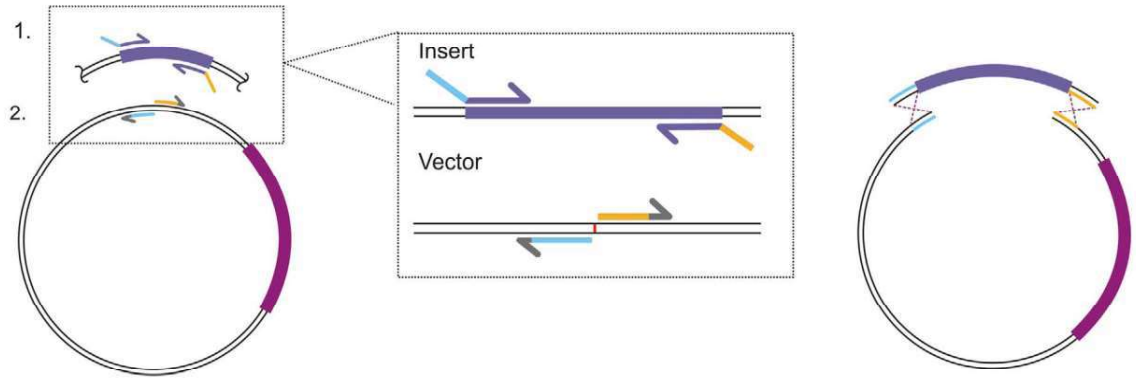


Fig. 7-3 *Principle of subcloning using the IVA method. Four primers are designed: an insert forward primer including the homologous region complementary to the vector reverse primer (light blue) and the template binding region (purple), an insert reverse primer including the homologous region complementary to the vector forward primer (orange) and the template binding region (purple), as well as two template binding vector primers (grey/orange and grey/light blue). Taken from J. García-Nafría, J. F. Watson, I. H. Greger,* IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Sci. Rep. ***2016***, 6, *27459.*

by the combination of the two amplified products later. The forward primer for the insert contains a template binding region (T_m = 60 °C) at the 5' end of the gene of interest, preceded by a nucleotide sequence which is complementary to the homologous region of the reverse primer for the vector amplification (same rules as before apply; > 15 bp and T_m = 47-52 °C). The reverse primer for the insert contains the complementary sequence from the 3' end of the gene of interest as a template binding region (T_m = 60 °C), followed by a nucleotide sequence which is complementary to the homologous region of the forward primer for the vector amplification. The forward primer for the vector amplification begins at the region where the 3' end of the gene of interest is supposed to be inserted and contains a template binding region with a T_m value of 60 °C. The reverse primer for the vector amplification begins at the region where the 5' end of the gene of interest is supposed to be inserted and contains a template binding region from 3' to 5' with a T_m value of 60 °C. Both amplified PCR products can be transformed into *E. coli* without further purification to produce the vector including the gene of interest *in vivo*. PCR reactions for each amplification were set up as described in **7.1.6.1**.

$MraY_{SP}$ into $MraY_{AA}$ pET26

The sequence coding for $MraY_{SP}$ was cloned into pET26 using $MraY_{AA}$ pET26, replacing the sequence coding for $MraY_{AA}$. Vector sequence: 5'-

TTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGC
AGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTC
TCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGAT
TTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG
CCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGA
CTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGG
GATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGA
ATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTT
AGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCA
TATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGAT
GGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATT
TCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGT

GAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTC
GTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGAC
GAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGG
AACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAA
TGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAAT
GCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTA
ACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCA
TACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATAT
AAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATG
GCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAA
ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC
TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC
AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCA
GCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGT
GGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCG
GTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTG
ATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCC
TGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATA
ACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCG
AGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC
GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAA
CACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA
CCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGC
TGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCG
TCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTT
AAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGG

GGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACAT
GCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAG
AAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTA
GCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTT
TCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGT
TTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAG
GCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGC
CGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGA
AGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCG
CTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGT
TGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAA
GGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGA
GCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGT
GGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGA
GAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGG
TTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC
GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTG
GCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACC
GGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGAT
ATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGC
GCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGG
GAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATT
AGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCC
CACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTT
CTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAA
TTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTG
CCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACT
TTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATA
AGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGA
ATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGT

CCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTG
AGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCC
CCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCG
AGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCC
GGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAAT
ACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTA
ACTTTAAGAAGGAGATATACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGC
TCCTCGCTGCCCAGCCGGCGATGGCCATGGATATCGGAATTAATTCGGATCCGAATTCGAGC
TCACATCACCATCACCATCACCATCACCATCACCCCATGAAAATCGAAGAAGGTAAACTGGT
AATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGA
AAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAG
GTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTAC
GCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCC
GTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAG
CGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGA
ACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACG
GCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTC
CTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGC
TGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCG
ACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAA
CCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAA
AGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAAC
CGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCC
GCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTT
CTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAG
CCCTGAAAGACGCGCAGACTCCGGGTAGCCTGGAAGTTCTGTTCCAGGGGCCCGCGGTACCTA
GGATGCTGTATCAACTGGCTCTGCTGCTGAAAGATTATTGGTTCGCCTTTAACGTCCTGAAA
TATATCACCTTCCGCTCGTTTACTGCCGTTCTGATCGCCTTCTTCCTGACACTGGTTCTGAGT
CCGTCCTTTATTAACCGTCTGCGCAAAATCCAACGCCTGTTTGGTGGCTATGTTCGTGAGTAT
ACCCCGGAAAGCCATGAAGTGAAAAAATATACGCCGACCATGGGTGGTATTGTGATCCTGAT

TGTGGTTACCCTGAGCACACTGCTGCTGATGCGCTGGGATATCAAATATACCTGGGTGGTTC
TGCTGTCTTTTCTGAGCTTTGGCACCATCGGTTTTTGGGACGACTATGTGAAACTGAAAAAC
AAAAAAGGCATCTCCATCAAAACCAAATTTCTGCTGCAGGTGCTGTCAGCATCTCTGATTAG
CGTGCTGATCTATTATTGGGCCGACATCGACACGATTCTGTATTTCCCGTTCTTCAAAGAGCT
GTATGTGGATCTGGGAGTCCTGTATCTGCCGTTCGCTGTGTTTGTGATTGTTGGTAGCGCCA
ATGCCGTCAATCTGACCGACGGGCTGGATGGCCTGGCAATTGGACCGGCAATGACTACTGCT
ACTGCCCTGGGTGTTGTTGCTTATGCCGTGGGACATTCGAAAATCGCCCAGTATCTGAATAT
TCCGTATGTTCCATATGCTGGTGAACTGACCGTGTTCTGCTTTGCTCTGGTGGGTGCTGGTCT
GGGGTTTCTGTGGTTCAACTCCTTCCCGGCTCAAATGTTTATGGGGGATGTGGGTAGCCTGA
GTATTGGAGCGTCACTGGCAACTGTGGCACTGCTGACCAAAAGCGAGTTCATCTTCGCCGTT
GCCGCTGGAGTGTTCGTTTTTGAAACCATCTCCGTGATTCTGCAAATCATCTATTTCCGCTGG
ACAGGTGGAAAACGTCTGTTCAAACGTGCCCCGTTTCATCATCACCTGGAACTGAATGGTCT
GCCTGAACCGAAAATTGTTGTCCGTATGTGGATTATTAGCATCCTGCTGGCCATTATCGCCA
TTTCAATGCTGAAACTGCGTTAACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTA
ACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCC
CTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGA-3’.
Fwd primer $MraY_{SP}$ insert: 5’-GGCCCGCGGTACCTAGGATGAATTCCATGAAACAAATCCTGTTCA-3’, rev primer $MraY_{SP}$ insert: 5’-GGTGGTGGTGCTCGAGTTAAGGTTTTCCGAACCAGAGCC-3’, Fwd primer $MraY_{AA}$ pET26 vector: 5’-TAACTCGAGCACCACCACC-3’, rev primer $MraY_{AA}$ pET26 vector: 5’-CCTAGGTACCGCGGGCC-3’.

$MraY_{SP}$ into $MraY_{SA}$ pET28a

The sequence coding for $MraY_{SP}$ was cloned into pET28 using $MraY_{SA}$ pET28a, replacing the sequence coding for $MraY_{SA}$. Vector sequence: 5’-ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGG
GGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTT
AGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGATGCACTCCAATCCATAAACCGA
TTAAACCTGAAATCAGACCAACAGCCCAAAATACTGTAACTACTTTCCATTCACTCCAACCT
ATCAATTCAAAATGATGATGAATCGGACTCATTTTAAATATACGCTTTCCAGTCAATTTAAA
GCTAGCGACTTGTAACATAACAGATAATGTTTCAATTACGAATACTAAACCTATAAAAATTA
ATGATAATTCCTGATTAAGCATGATTGAAATCGTAGCAAATATACCACCTAAAGCTAAGCTA

CCTGTATCTCCCATAAACACTTTAGCAGGGTTAATGTTATATGGTAAAAATCCTAAAAGTGC
AAACAACATAATGATACAGAAAATACCAATCGCCGTTTCTCCTAACACAAAGCTCATGATGG
CATACATTGTAAATCCGATAATTGACAGTCCAGTTGCTAATCCATCTAAACCATCTGTTAAA
TTTACCGCATTAGAAAAACCTACTTGCCAAAAAACAATGAAAATAACATATGCAAATGATAG
TGGGATTGCTACATTCGTAAATGGAATATGTATGCTCGTAGAAAAATTCACCAAATGAAATA
CATTACTTAAAACAAAGAAAATAATCGCAATACCAATTTGCGCCAAAAACTTCTGTTTACTT
GTTAAACCTTGGTTATTCTTTTTAACAACAATAATATAATCATCTATAAAACCAATTAACCC
AAAACCAATCGTCACAAATAATAACAGTATGATTGGATTAGCTTGATCTACAAATATAATAG
CCACCAAAGACGTTATCACAATACTTAATAGAAATGTTAGTCCACCCATCGTTGGTGTACCA
GTCTTCTTCATATGGCTTTGTGGGCCTTCTTCTCGAATACTTTGACCAAATTTCATCCTTTTT
AATGTAGGTATTAAAACAGGTACCAAAACAAATGTAATCACTAGCGCTAATAACGCATATAC
AAAAACCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTA
TCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCT
ACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCG
CCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCG
TGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCAT
TCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGT
CGCATAAGGGAGAGCGTCGAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGGTAT
GGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTAT
ACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCA
GCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATT
CCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCC
AGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTG
GGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGT
GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGG
ATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACC
AGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCAT
CTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCG
CGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGA
ACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGG
GCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCA

TTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAA
GACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAA
ACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTT
GCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCG
CGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGT
GAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCGGGATCTCGACCGATGCC
CTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGC
ACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCA
TTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCG
GAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGA
AGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGA
GGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAG
CGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTT
CGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCC
GGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAG
CGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTA
CCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTC
TCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTG
ACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCT
TCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACG
ACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCT
GACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAA
GCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGT
AGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
CACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA
GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACAT
GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCC
ATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT

CATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAA
CCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA
GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA
CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTACATAA
ACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGA
TTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCA
ATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAAC
ATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACG
GAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTC
ACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGA
AAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATT
GTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGT
TTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAA
AGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCAC
TTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGA
ATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCA
TTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTT
TCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAAT
TGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTA
ACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTG
AGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGG
GCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTT
TGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCT
TGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCG
CTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATG
CGCCGCTACAGGGCGCGTCCCATTCGCCA-3’.

Fwd primer $MraY_{SP}$ insert: 5'-ACTTTAAGAAGGAGATATACCATGAATTCCATGAAACAAATCCTGTTCA-3', rev primer $MraY_{SP}$ insert: 5'-TGTTAGCAGCCGGATCTTTCGCGGCTTGCCAACC-3', Fwd primer $MraY_{SA}$ pET28a vector: 5'-GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAG-3', rev primer $MraY_{SA}$ pET28a vector: 5'-GATCCGGCTGCTAACAAAGCC-3'.

GILZ into Keap1-pGEX-6P-3

The sequence coding for GILZ was cloned into pGEX using Keap1-pGEX-6P-3, replacing the sequence coding for Keap1**.** Vector sequence: 5'-

ACGTTATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGT
ATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTG
GATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGA
CAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAA
ACAGTATTCATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCG
ACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTG
ATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATT
GATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAA
CATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGG
ATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGAT
TTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATA
TTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTT
TATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATT
GAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCA
GGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGAAGTTCTGT
TCCAGGGGCCCCTGGGATCCCCGAAGGTTGGTCGTCTGATCTACACCGCGGGCGGCTACTTTC
GTCAGAGCCTGAGCTACCTGGAGGCGTACAATCCGAGCAACGGCAGCTGGCTGCGTCTGGCG
GATCTGCAGGTTCCGCGTAGCGGTCTGGCGGGTTGCGTGGTTGGTGGCCTGCTGTATGCGGT
GGGTGGCCGTAACAACAGCCCGGATGGTAACACCGACAGCAGCGCGCTGGACTGCTATAACC
CGATGACCAACCAATGGAGCCCGTGCGCGAGCATGAGCGTTCCGCGTAACCGTATCGGTGTG
GGCGTTATCGATGGTCACATTTACGCGGTTGGTGGCAGCCACGGCTGCATTCACCACAGCAG
CGTGGAGCGTTATGAGCCGGAACGTGACGAATGGCACCTGGTTGCGCCGATGCTGACCCGTC
GTATTGGCGTGGGCGTTGCGGTGCTGAACCGTCTGCTGTACGCGGTGGGTGGCTTCGATGGT

ACCAACCGTCTGAACAGCGCGGAGTGCTACTATCCGGAGCGTAACGAATGGCGTATGATCAC
CCCGATGAACACCATTCGTAGCGGTGCGGGCGTTTGCGTGCTGCACAACTGCATTTACGCGGC
GGGTGGCTATGATGGCCAGGACCAACTGAACAGCGTTGAACGTTACGACGTGGAAACCGAAA
CCTGGACCTTCGTTGCGCCGATGCGTCACCACCGTAGCGCGCTGGGTATCACCGTTCACCAGG
GCAAGATTTACGTGCTGGGTGGCTATGATGGTCACACCTTTCTGGACAGCGTGGAGTGCTAT
GACCCGGATAGCGACACCTGGAGCGAAGTTACCCGTATGACCAGCGGTCGTAGCGGCGTGGG
CGTGGCGGTTACCATGGAGCCGTGCCGCAAACAGATTGACCAACAGAACTGCACCTGCTAAC
TCGAGCGGCCGCATCGTGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAA
AACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAG
CAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCC
AGTCACGTAGCGATAGCGGAGTGTATAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTA
TTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGG
AAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCA
TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCA
ACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCC
AGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCG
AACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATG
ATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGA
GCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAG
AAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGT
GATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTT
TTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAG
CCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAA
CTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATA
AATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG
CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAG
ACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACT
CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATC
CTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGA
CCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT

TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACT
CTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTA
GCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAAT
CCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGC
TTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCAC
CTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGC
CAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCC
TGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCG
CCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGC
GGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAAATTCCGACACCATCGAAT
GGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTG
AATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGT
TTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGG
CGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCG
TTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCG
ATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGG
CGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCA
TTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCG
GCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGAC
GGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGC
GGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCG
CAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAAC
AAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAG
ATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCG
GTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTCAACCACCATCAA
ACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCA
GGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGC
CCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGG

TTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTA
GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGAT
AACAATTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACA
ACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGC
ACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACA
GTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGG
AAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAG
ATGCACGGTTACGATGCGCCCATCTACACCAACGTAACCTATCCCATTACGGTCAATCCGCCG
TTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTG
GCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTGGAATT-3'.

Fwd primer GILZ insert: 5'-GTTCCAGGGGCCCATGAACACCGAAATGTACCAGACC -3', rev primer GILZ insert: 5'-ATCACCGAAACGCGCTTACACGGCGCTACCACC-3', Fwd primer Keap1-pGEX vector: 5'-TGACTGACGATCTGCCTCGC-3', rev primer: 5'-GGGCCCCTGGAACAGAACTT-3'.

7.2 Expression and purification of ACP

ACP from *E. coli* was expressed using a construct and bacterial strain designed by CRONAN and THOMAS.[407] They used the *E. coli* K-12 strain DK574[350], which is strain SJ16[408], bearing a pJT94 plasmid encoding holo-ACP under *lac* promoter control.

Unlabeled apo-ACP was expressed by growing a culture of the aforementioned strain in 10 mL LB per liter large culture, supplemented with 50 µg/mL kanamycin and 100 µg/mL spectinomycin in a shaking incubator at 37 °C overnight. The following morning, 1 L LB supplemented with 50 µg/mL kanamycin and 100 µg/mL spectinomycin was inoculated with 10 mL of the overnight culture. The culture was grown in a shaking incubator at 37 °C until an OD_{600} of ~ 0.4, after which the temperature was reduced to 20 °C. At an OD600 of 0.7-0.9, IPTG was added to a final concentration of 1 mM and the protein was expressed at 20 °C overnight. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

Modified M9 minimal media for expression of U-^{15}N^{13}C labeled ACP was prepared as follows: 6 g Na_2HPO_4, 3 g KH_2PO_4 and 0.5 g NaCl were dissolved in 1 L of double deionized water and autoclaved. Then, the appropriate antibiotic was added, as well as sterile filtered solutions of $MgSO_4$ (final concentration = 2.5 mM), $CaCl_2$ (final concentration = 0.1 mM), 2 g D-Glucose-$^{13}C_6$, 1 g $^{15}NH_4Cl$, 100 µL of a 10,000 x vitamin mix (5 g riboflavin, 5 g niacinamide, 5 g pyridoxine monohydrate and 5 g thiamine dissolved in ethanol), 50 mg U-^{15}N L-methionine and β-alanine at a final concentration of 8 µM. For the expression of U-^{15}N labeled proteins, 4 g D-Glucose was used in an otherwise identical medium.

10 mL LB media per liter large culture supplemented with 50 µg/mL kanamycin and 100 µg/mL spectinomycin was inoculated with *E. coli* K-12 strain DK574 bearing the pJT94 plasmid and incubated in a shaking incubator at 37 °C overnight. The following morning, 1 L of the modified M9 media described above was inoculated with 10 mL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C to an OD_{600} of ~ 0.4, after which the temperature was dropped to 20 °C. Protein expression was induced at an OD_{600} of 0.7-0.9 with isopropyl β-D-1-thiogalactopyranosid (IPTG) at a final concentration of 1 mM. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

Differently modified M9 minimal media for expression of U-^{15}N^{2}H labeled ACP was prepared as follows: 6 g Na_2HPO_4, 3 g KH_2PO_4 and 0.5 g NaCl were dissolved in 1 L of D_2O in a dry autoclaved 2 L glass beaker. Then, 50 mg kanamycin and 100 mg spectinomycin were added, as well as sterile filtered solutions of $MgSO_4$ (final concentration = 2.5 mM), $CaCl_2$ (final concentration = 0.1 mM) in D_2O, 2 g D-Glucose-(1,2,3,4,5,6,6-d7), 1 g $^{15}NH_4Cl$, 100 µL of a 10,000 x vitamin mix (5 g riboflavin, 5 g niacinamide, 5 g pyridoxine monohydrate and 5 g thiamine dissolved in ethanol), 50 mg U-^{15}N L-methionine and β-alanine at a final concentration of 8 µM. The prepared medium was sterile filtered using a 0.2 µm filter.

25 mL LB media in D_2O per liter large culture were prepared by dissolving 0.625 g LB powder in 25 mL D_2O, supplemented with 50 µg/mL kanamycin and 100 µg/mL spectinomycin. The medium was sterile filtered using a 0.2 µm filter and inoculated with

E. coli K-12 strain DK574 bearing the pJT94 plasmid and incubated in a shaking incubator at 37 °C for 10 h. The culture was centrifuged at 4,300 x g for 15 min and the bacterial cell pellet was gently resuspended in 50 mL of the modified M9 media in D_2O described above. This culture was grown in an Erlenmeyer flask at 37 °C overnight. The next morning, 1 L of the modified M9 media in D_2O described above was inoculated with 10 mL of the overnight culture. The bacteria were grown in a shaking incubator at 37 °C to an OD_{600} of ~ 0.4, after which the temperature was dropped to 20 °C. At this point, 1 g of Celtone Base Powder (D, 97 %+; ^{15}N, 98 %+, Cambridge Isotope Laboratories, Inc.) was added to the culture. Protein expression was induced at an OD_{600} of 0.7-0.9 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. The next day, the bacteria were harvested by centrifugation at 6,000 x g for 15 min and the cell pellets were either stored at -80 °C or immediately prepared for purification.

Cell pellets were resuspended in ~ 30 mL resuspension buffer (50 mM Tris-HCl, pH = 7.5, 350 mM NaCl, 2 mM β-mercaptoethanole [BME], 1 mM ethylenediaminetetraacetic acid [EDTA], 1 cOmplete™ Protease Inhibitor Cocktail tablet) per liter culture. Resuspended cells were then lysed by sonication (30 % amplitude, 5 min sonication time, 2 s on pulse, 4 s off pulse). Cell lysates were pelleted by centrifugation at 20,000 x g for 30 min at 4 °C. To transform holo-ACP into apo-ACP, cleared lysates were then incubated with 25 mM $MgCl_2$ and 1.2 mM $MnSO_4$ at 37 °C for 4 h. After this incubation period, protein was precipitated by addition of isopropanol to the cleared lysates to a final concentration of 50 % (vol/vol). Precipitated protein was pelleted using centrifugation at 20,000 x g for 30 min at 4 °C, and the supernatant was loaded onto diethylaminoethyl (DEAE)–Sepharose beads (~ 5 mL per liter culture), preequilibrated with equilibration buffer (10 mM MES, pH = 6.1). Protein was incubated with resin at 4 °C for 16 h. The resin was washed with wash buffer (10 mM MES, pH = 6.1, 25 mM LiCl) until 10 µL of eluent no longer stained blue with 50 µL of Bradford reagent (Coomassie blue G-250). Bound protein was eluted with elution buffer (10 mM MES, pH = 6.1, 500 mM LiCl). The eluted protein was concentrated to a final volume of ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with ACP buffer (50 mM MES, pH = 6.8, 10 mM $MgCl_2$, 1 mM TCEP-HCl).

7.3 Preparation of crude membranes with overexpressed MraY

Crude membranes with overexpressed $MraY_{SA}$, $MraY_{SA}$ H287Q, $MraY_{CB}$, $MraY_{SP}$, $MraY_{SP}$ Q316H were prepared as described before.[409]

7.4 Purification of MraY isoforms in DM micelles

This procedure was used to purify $MraY_{AA}$, $MraY_{AA}$ H324Q, $MraY_{AA}$ D193N and $MraY_{CB}$. Bacterial cell pellets were resuspended in ~ 15 mL resuspension buffer (50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanole [BME], 1 cOmplete™ Protease Inhibitor Cocktail tablet) per liter culture. Resuspended cells were then lysed by sonication (30 % amplitude, 5 min sonication time, 2 s on pulse, 4 s off pulse). Cell lysates were incubated with 30 mM *n*-dodecyl β-D-maltoside (DDM) and stirred vigorously for 2 h at 4 °C. Lysates were then pelleted by centrifugation at 30,000 x g for 40 min at 4 °C. Cleared lysates were incubated with ~ 3 mL Ni-NTA agarose resin, preequilibrated with wash buffer (50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 10 mM imidazole, 2 mM BME, 1 mM DDM) per liter culture for 16 h at 4 °C. The resin was washed with wash buffer until 10 µL of eluent no longer stained blue with 50 µL of Bradford reagent (Coomassie blue G-250). Bound protein was eluted with elution buffer (50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 1 mM DDM, 350 mM imidazole). EDTA and dithiothreitol (DTT) were added to eluted protein to a final concentration of 1 mM each. Cleavage of MBP tags was performed for 16 h at 4 °C by addition of His_6-3C protease (~ 0.5 mg per liter culture). Proteins were concentrated to ~ 5 mL and subjected to size exclusion chromatography using a Superdex 200 increase 10/300 GL column preequilibrated with MraY buffer (50 mM Tris-HCl, 150 mM NaCl, 2mM DTT, 4 mM DM).

7.5 Purification of NFAT isoforms

This procedure was used to purify $NFAT_{1-130}$, $NFAT_{1-130}$ S110A, $NFAT_{1-130}$ D14C S110A, $NFAT_{1-130}$ D35C S110A, $NFAT_{1-130}$ D35C S99C S110A, $NFAT_{1-130}$ L65C S110A, $NFAT_{1-130}$ S99C S110A, $NFAT_{1-130}$ S110A A124C after His-tag purification described in **7.1.4**.
Eluted proteins were dialyzed against 2 L dialysis buffer (20 mM sodium phosphate, pH = 6.8, 150 mM NaCl, 2 mM DTT, 1 mM EDTA) after addition of TEV protease (~0.2 g purified TEV per liter culture) to cleave GB1-His_6-tags for 16 h at 4 °C. Cleaved protein was concentrated to ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with NFAT NMR-phosphorylation buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT).

7.6 Purification of eIF4E isoforms

This procedure was used to purify Δ_{26}-eIF4E, Δ_{26}-eIF4E S209D, Δ_{26}-eIF4E pS209, GB1-eIF4E, GB1-eIF4E S209D and GB1-eIF4E pS209.
Bacterial cell pellets were resuspended in ~ 15 mL resuspension buffer (10 mM sodium phosphate, pH = 6.5, 100 mM NaCl, 2 mM BME, 1 cOmplete™ Protease Inhibitor Cocktail tablet) per liter culture. Resuspended cells were then lysed by sonication (30 % amplitude, 5 min sonication time, 2 s on pulse, 4 s off pulse). Cell lysates were pelleted by centrifugation at 30,000 x g for 40 min at 4 °C. Cleared lysates were incubated with ~ 500 µL modified m^7-GDP agarose resin (prepared using previously described methods[410]), preequilibrated with resuspension buffer, for 45 min at 4 °C. The resin was washed with resuspension buffer until 10 µL of eluent no longer stained blue with 50 µL of Bradford reagent (Coomassie blue G-250). Bound protein was eluted with either elution buffer A (10 mM sodium phosphate, pH = 6.5, 100 mM NaCl, 2 mM BME, 0.2 mM m^7-GTP) or elution buffer B (10 mM sodium phosphate, pH = 6.5, 500 mM NaCl, 2 mM BME). Elution using elution buffer A produced cap-bound eIF4E, while elution with elution buffer B produced cap-free eIF4E. Purified protein was concentrated to ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with either eIF4E ITC buffer (10 mM HEPES, pH = 7.5, 125 mM

NaCl, 2 mM DTT) or eIF4E NMR buffer (20 mM sodium phosphate, pH = 6.5, 100 mM KCl, 2 mM DTT).

7.7 Purification of GILZ isoforms

This procedure was used to purify $GILZ_{1-75}$, $GILZ_{1-98}$, $GILZ_{1-134}$ and $GILZ_{1-134}$ N87D after either His-tag or GST purification described in **7.1.4** and **7.1.5**.

Eluted proteins were dialyzed against 2 L dialysis buffer (20 MES, pH = 5.5, 150 mM NaCl, 2 mM DTT, 1 mM EDTA) after addition of TEV protease (~0.2 g purified TEV per liter culture) to cleave GB1-His_6-tags or 3C protease (~0.2 g purified 3C per liter culture) to cleave GST tags for 16 h at 4 °C. Cleaved protein was then buffer exchanged to anion exchange buffer (20 mM MES, pH = 5.5, 2 mM DTT). Protein was subjected to anion exchange chromatography using a HiTrap Q HP anion exchange chromatography column preequilibrated with anion exchange buffer. GILZ was eluted using anion exchange buffer with a NaCl gradient from 0 to 1 M final concentration. Eluted GILZ fractions were analyzed using SDS-PAGE and concentrated to a final volume of ~ 5 mL and subjected to size exclusion chromatography using a Superdex 200 increase 10/300 GL column preequilibrated with GILZ buffer (PBS, pH = 7.0, 125 mM NaCl, 2 mM DTT).

7.8 Purification of Keap1 isoforms

This procedure was used for the purification of Keap1 and Keap1 E540A E542A after GST purification described in **7.1.5**.

Eluted proteins were dialyzed against 2 L dialysis buffer (20 Tris-HCl, pH = 7.5, 150 mM NaCl, 2 mM DTT, 1 mM EDTA) after addition of 3C protease (~0.2 g purified 3C per liter culture) to cleave GST tags for 16 h at 4 °C. Cleaved proteins were concentrated to a final volume of ~ 5 mL and subjected to size exclusion chromatography using a Superdex 200 increase 10/300 GL column preequilibrated with Keap1 buffer (20 mM Tris-HCl, pH = 8.0, 50 mM NaCl, 10 mM DTT).

7.9 Purification of MurF

His-tagged MurF was purified as described in **7.1.4.**

Eluted protein was concentrated to a final volume of ~ 5 mL and subjected to size exclusion chromatography using a Superdex 200 increase 10/300 GL column preequilibrated with MurF buffer (50 mM Tris-HCl, pH = 7.5, 150 mM NaCl, 1 mM DTT).

7.10 Purification of Homer3

His-GB1-tagged Homer3 was purified as described in **7.1.4.**

Eluted proteins were dialyzed against 2 L dialysis buffer (20 Tris-HCl, pH = 7.5, 150 mM NaCl, 2 mM DTT, 1 mM EDTA) after addition of TEV protease (~0.2 g purified TEV per liter culture) to cleave GB1-His_6-tags for 16 h at 4 °C. Cleaved protein was concentrated to a final volume of ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT).

7.11 Purification of Med15 and CBP/p300 KIX

GST-tagged proteins were purified as described in **7.1.5.**

Eluted proteins were dialyzed against 2 L dialysis buffer (20 Tris-HCl, pH = 7.5, 150 mM NaCl, 2 mM DTT, 1 mM EDTA) after addition of 3C protease (~0.2 g purified 3C per liter culture) to cleave GST tags for 16 h at 4 °C. Cleaved proteins were concentrated to a final volume of ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with Med15 buffer (PBS, pH = 6.5).

7.12 Purification of MS1E3D1

His_6-tagged MS1E3D1 was purified as described in **7.1.4.**

Eluted protein was dialyzed against 2 L of dialysis buffer (50 mM sodium phosphate, pH = 7.0, 150 mM NaCl, 1 mM EDTA, 2 mM DTT) after addition of TEV protease (~0.2 g purified TEV per liter culture) to cleave His_6-tags for 16 h at 4 °C. Cleaved protein was

concentrated to ~ 5 mL and subjected to size exclusion chromatography using a Superdex 200 increase 10/300 GL column preequilibrated with MSP buffer (20 mM Tris-HCl, pH = 7.5, 100 mM NaCl, 0.5 mM EDTA). Fractions containing MSP1E3D1 were concentrated to ~ 700 µM and aliquots were stored at -80 °C.

7.13 Purification of *Bacillus subtilis* Sfp

His_6-tagged Sfp was purified as described in **7.1.4**.
Eluted protein was concentrated to ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with Sfp buffer (50 mM MES, pH = 6.8, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT).

7.14 Purification of BmaI1

His_6-tagged BmaI1 was purified as described in **7.1.4**.
Eluted protein was concentrated to ~ 15 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with BmaI1 buffer (50 mM MES, pH = 6.8, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT).

7.15 Purification of EsaI

His_6-tagged EsaI was purified as described in **7.1.4**.
Eluted protein was concentrated to ~ 5 mL and subjected to size exclusion chromatography using a Superdex 75 increase 10/300 GL column preequilibrated with BmaI1 buffer (50 mM MES, pH = 6.8, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT).

7.16 Transmission electron microscopy experiments

TEM images for $MraY_{AA}$ nanodiscs, MraY-MurF complex and GILZ were recorded on a JEOL JEM 1400 Transmission Electron Microscope.
Samples for negative stain experiments were prepared using uranyl formate staining. For this, a 1 % uranyl formate solution was prepared by dissolving 20 mg uranyl formate in 2 mL of hot water, followed by addition of 8 µL of 5 M NaOH while stirring. The solution

was then filtered using a 0.2 µm syringe filter. FCF400-CU, 400 mesh grids were prepared for staining by glow discharge for 30 s at 10 mA. 5 µL sample was applied to the grid's surface and incubated for 45 s. The grid was washed with protein buffer twice. Then, 7 µL of the uranyl formate solution was applied to the grid's surface and incubated for 45 s, followed by removal of the liquids with a filter paper.

7.17 Formation of MraY nanodiscs

DOPE/DMPG nanodiscs of $MraY_{AA}$ were produced using purified $MraY_{AA}$, MSP1E3D1 and a lipid mixture of DOPE/DMPG (8.5/1.5).

To make the lipid mixture, stock solutions of DOPE and DMPG in chloroform were combined in a glass tube to create a mixture of DOPE (51 mM) and DMPG (9 mM) in chloroform. Chloroform was evaporated under nitrogen gas while turning the glass tube to create a thin film of lipids on the tube's walls. The dry mixture was redissolved in cholate buffer (20 mM Tris-HCl, pH = 7.5, 100 mM NaCl, 0.5 mM EDTA, 100 mM sodium cholate) to make a stock solution of DOPE/DMPG (8.5/1.5, 60 mM).

A general protocol for nanodisc formation consisted of a final volume of 0.5 mL. With respect to final $MraY_{AA}$ concentration, MSP1E3D1 was added in 3-fold excess and lipid mixture was added in 100-fold excess. The mixture was incubated at room temperature for 1 h. To this assembly mixture, ~ 0.5 g Biobeads SM-2 (Biorad) were added, which had been washed with water before. Removal of the detergent by Biobeads was performed at room temperature for 16 h with gentle shaking on an orbital shaker. Biobeads were removed by centrifugation and the mixture was subjected to size exclusion chromatography using a Superdex 200 increase 5/150 GL column preequilibrated with nanodisc buffer (50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT). Eluted protein was analyzed by SDS-PAGE and fractions containing $MraY_{AA}$ and MSP1E3D1 were pooled, concentrated and stored at -80 °C.

7.18 Fluorescence-based MraY activity assays

Activity assays for MraY and IC_{50} calculations were performed according to previously described protocols.[409]

7.19 ITC experiments

All ITC experiments were performed using The Affinity ITC instrument from TA® instruments. Unless otherwise stated, titration experiments were performed at 25 °C, 500 µL protein in the cell, 300 µL ligand in the syringe and 24 injections of 2.5 µL each.

7.19.1 ITC of MraY isoforms

Tunicamycin was ordered from Sigma-Aldrich. MD2 and MB6 were kindly provided by STEVEN VAN LANEN.

7.19.1.1 Titration of $MraY_{AA}$ with tunicamycin

Titration without Magnesium: protein and ligand in 50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 2 mM DTT, 4 mM DM. $MraY_{AA}$ (cell) concentration = 20 µM, tunicamycin (syringe) concentration = 200 µM (based on an average molecular weight of 840 g/mol). Titration with Magnesium: protein and ligand in 50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 4 mM DM. $MraY_{AA}$ (cell) concentration = 20 µM, tunicamycin (syringe) concentration = 200 µM (based on an average molecular weight of 840 g/mol). Titrations was performed at 37 °C, with 24 injections of 2.0 µL each.

7.19.1.2 Titration of $MraY_{AA}$ with muraymycin D2

Titration was performed at 37 °C, with 24 injections of 2.0 µL each. Protein and ligand were dissolved in 50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 4 mM DM. One titration was performed with 15 µM $MraY_{AA}$ and 75.3 µM MD2. Two more titrations were performed with 16.7 µM $MraY_{AA}$ and 66.7 µM MD2.

7.19.1.3 Titrations of $MraY_{AA}$ with muraymycin B6

Titration was performed at 37 °C, with 30 injections of 2.0 µL each. Protein and ligand were dissolved in 50 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 4 mM DM. One titration was performed with 15 µM $MraY_{AA}$ and 50 µM MB6. Another titration was performed with 30 µM $MraY_{AA}$ and 150 µM MB6.

7.19.2 ITC of eIF4E isoforms

7.19.2.1 Titration of GB1-eIF4E and GB1-eIF4E S209D with m^7-GTP

Titrations of GB1-eIF4E wild type and GB1-eIF4E S209D were performed at 25 °C, with 20 injections of 2.5 µL each. Protein and ligand were dissolved in 10 mM HEPES, pH = 7.5, 125 mM NaCl, 2 mM DTT. All titrations were performed with 15 µM eIF4E wild type or eIF4E S209D and 150 µM m^7-GTP.

7.19.2.2 Titration of GB1-eIF4E pS209 with m^7-GTP

Titrations of GB1-eIF4E pS209 were performed at 25 °C, with 20 injections of 2.5 µL each. Protein and ligand were dissolved in 10 mM HEPES, pH = 7.5, 125 mM NaCl, 2 mM DTT. All titrations were performed with 15 µM eIF4E pS209 and 90 µM m^7-GTP.

7.19.2.3 Titration of Δ_{26}-eIF4E with cap analogues

Cap analogues were ordered from TriLink Biotechnologies. Protein and ligand were dissolved in 10 mM HEPES, pH = 7.5, 125 mM NaCl, 2 mM DTT. Duplicate titrations were performed with 10 µM Δ_{26}-eIF4E and 50 µM of either m^7GG, m^7GA(m)G or m^7GG(m)G. Duplicate titrations were performed with 20 µM Δ_{26}-eIF4E and 100 µM m^7G(m)G(m)G.

7.19.2.4 Titration of Δ_{26}-eIF4E pS209 with cap analogues

Protein and ligand were dissolved in 10 mM HEPES, pH = 7.5, 125 mM NaCl, 2 mM DTT. One set of titrations was performed with 10 µM Δ_{26}-eIF4E pS209 and 100 µM of either m^7GG, m^7GA(m)G, m^7GG(m)G or m^7G(m)G(m)G. Another set of titrations was performed with 10 µM Δ_{26}-eIF4E pS209 and 50 µM of either m^7GG, m^7GA(m)G or m^7GG(m)G.

7.20 MST experiments

All MST experiments were carried out on a Monolith NT.115pico from NanoTemper Technologies, equipped with a Pico RED detector (excitation wavelength = 600-650 nm).

7.20.1 MST of MraY isoforms

All MST experiments of MraY have been carried out using Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation for His-tag labeling and Monolith NT.115 Premium Capillaries.

7.20.1.1 MST with $MraY_{AA}$ in DM micelles

MST data was measured using 5 nM fluorescently labeled His-MBP-$MraY_{AA}$. For muraymycin D2 and muraymcin B6, experiments were performed with 16 serial

dilutions of ligands ranging from 2.5 μM to 76.3 pM. Datasets were analyzed using MO.Affinity Analysis.

7.20.1.2 MST with $MraY_{AA}$ in nanodiscs

MST data was measured using 10 nM fluorescently labeled His-MBP-$MraY_{AA}$ in DOPE/DMPG nanodiscs. Experiments were performed in duplicates with 16 serial dilutions of MurF ranging from 58.8 μM to 14.3 nM. Datasets were analyzed using MO.Affinity Analysis.

7.21 MST of NFAT

All MST experiments of NFAT were carried out using Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation for His-tag labeling and Monolith NT.115 Capillaries. MST data was measured using 5 nM fluorescently labeled His-CBP/p300 KIX. Experiments were performed in duplicates with either 16 serial dilutions of NFAT1-130 S110A ranging from 285 μM to 8.68 nM or 16 serial dilutions of NFAT1-130 pS23 S110A ranging from 134 μM to 8.18 nM. Datasets were analyzed using MO.Affinity Analysis.

7.22 Phosphorylation of NFAT isoforms

Phosphorylation was carried out with recombinantly expressed GST-p38α (see **7.1.5**). NFAT isoforms were phosphorylated *in vitro* by the addition of ~ 20 μg of purified GST-p38α in NFAT phosphorylation buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT and 5 mM ATP). Completion of *in vitro* phosphorylation was monitored using NMR.

7.23 Spin labeling of NFAT isoforms for PRE and EPR experiments

MTSSL was purchased from Cayman Chemical Company and dissolved in acetonitrile to create a stock solution of 200 mM. Immediately before labeling, NFAT isoforms were buffer exchangeed to labeling buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$) without DTT, and diluted to ~ 1 mL at a concentration between 250-500 μM. MTSSL was

added to a final concentration that was 10 x the concentration of NFAT. The mixture was incubated for 16 h at 4 °C. Then, the protein was diluted in labeling buffer to ~ 15 mL and dialyzed against 2 L of labeling buffer for 16 h at 4 °C. Dialyzed protein was concentrated to ~ 450 µL for NMR experiments and ^{15}N-^{1}H-HSQC spectra were recorded, followed by addition of sodium ascorbate to a final concentration that was 5 x the concentration of NFAT and acquisition of another ^{15}N-^{1}H-HSQC spectrum to create the "reduced" spectrum.

7.24 NMR experiments

7.24.1 Assignment of proteins for NMR

Assignments of peaks in ^{15}N-^{1}H-HSQC spectra were performed on U-^{15}N-^{13}C labeled $NFAT_{1\text{-}130}$, $NFAT^{1\text{-}130}$ S110A, $NFAT_{1\text{-}130}$ pS23 S110A, apo-ACP, 3-oxo-C6i-ACP and C8i-ACP. Assignments for GB1-eIF4E have been obtained from previously published work.[411]
In general, triple resonance backbone experiments were recorded on a Varian (Agilent DD2 600) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 600 MHz proton frequency. NMR experiments for NFAT isoforms were recorded at 288 K, experiments for ACP isoforms were recorded at 298 K. NFAT isoforms were concentrated to a final concentration of 500 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). ACP isoforms were concentrated to 500 µM in ACP NMR buffer (50 mM MES, pH = 6.8, 10 mM $MgCl_2$, 1 mM TCEP-HCl). For backbone resonance assignment, HNCA, HN(CO)CA, HNCACB, HNCO, HN(CA)CO and CCCONH experiments were recorded. All 3D spectra were acquired using non-uniform sampling (NUS) with 12% sampling of the Nyquist grid using Poisson Gap Sampling.[412] The NUS spectra were reconstructed and processed with hmsIST.[413] NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414]

7.24.2 $NFAT_{1-130}$ titration experiments with Med15 KIX, CBP/p300 KIX and Homer3

NMR titration experiments of $NFAT_{1-130}$ were performed at 288 K on a Varian (Agilent DD2 700) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 700 MHz proton frequency. ^{15}N-labelled $NFAT_{1-130}$ was concentrated to a final concentration of 100 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). ^{15}N-^{1}H-HSQC experiments were recorded, once in the absence of unlabeled interacting partners, and in the presence of 25, 50, 75, 100 and 200 µM of either Med15 KIX, CBP/p300 KIX or Homer3. Med15 KIX, CBP/p300 KIX or Homer3 were buffer exchanged into NFAT NMR buffer prior to titration experiments. NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414] Chemical shift perturbations (CSPs) were calculated using the formula $CSP = (\Delta\delta H^2 + [0.14\Delta\delta N]^2)^{1/2}$, where ΔδH and ΔδN are the changes in chemical shifts for proton and nitrogen frequencies, respectively, in ppm.

7.24.3 $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A titration experiments with CBP/p300 KIX

NMR titration experiments on $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A were recorded on a Bruker Avance III 800 MHz spectrometer with a TXO-style cryogenically cooled probe, operating at 800 MHz proton frequency. ^{15}N-labelled $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A were concentrated to a final concentration of 100 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). ^{15}N-^{1}H-HSQC experiments were recorded, once in the absence of unlabeled CBP/p300 KIX, and in the presence of 50, 100 and 200 µM CBP/p300 KIX. CBP/p300 KIX was buffer exchanged into NFAT NMR buffer prior to titration experiments. NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414] Chemical shift perturbations (CSPs) were calculated using the formula $CSP = (\Delta\delta H^2 + [0.14\Delta\delta N]^2)^{1/2}$, where ΔδH and ΔδN are the changes in chemical shifts for proton and nitrogen frequencies, respectively, in ppm.

7.24.4 Relaxation experiments of NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ pS23 S110A in the ps-ns timescale

7.24.4.1 T_1 experiments

For T_1 measurements, ^{15}N-labeled NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ pS23 S110A were concentrated to a final concentration of 270 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). Experiments were recorded at 288 K on a Varian (Agilent DD2 700) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 700 MHz proton frequency, a Varian (Agilent DD2 600) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 600 MHz proton frequency and a Bruker 900 MHz spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 900 MHz proton frequency. T1 experiments were recorded as pseudo-3D experiments, with δH as the direct dimension, δN as the indirect dimension and the waiting period τ as the third dimension. τ was set to 0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.5, 0.8, 1.0, 1.2, 1.5 and 1.8 ms, with repeats at 0.2 and 0.8 ms. NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414]

7.24.4.2 T_2 experiments

For T_2 measurements, ^{15}N labeled NFAT$_{1-130}$ S110A and NFAT$_{1-130}$ pS23 S110A were concentrated to a final concentration of 270 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). Experiments were recorded at 288 K on a Varian (Agilent DD2 700) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 700 MHz proton frequency, a Varian (Agilent DD2 600) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 600 MHz proton frequency and a Bruker 900 MHz spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 900 MHz proton frequency. T2 experiments were recorded as pseudo-3D experiments, with δH as the direct dimension, δN as the indirect dimension and the waiting period τ as the third dimension. τ was set to 0.01, 0.03, 0.05, 0.07, 0.09, 0.13, 0.15 and 0.17 ms, with repeats at 0.5 and 0.13 ms. NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414]

7.24.4.3 hetNOE experiments

For hetNOE measurements, ^{15}N labeled $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A were concentrated to a final concentration of 270 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). Experiments were recorded at 288 K on a Varian (Agilent DD2 700) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 700 MHz proton frequency, a Varian (Agilent DD2 600) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 600 MHz proton frequency and a Bruker 900 MHz spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 900 MHz proton frequency. To obtain NOEs, two INEPT experiments were recorded. The first, or reference experiment was recorded without saturation, followed by a second experiment with a 3.5 s long saturation period. NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414] hetNOE values were obtained by normalizing the ratio between reference spectra and saturated spectra.

7.24.4.4 R_{ex} experiments

For R_{ex} measurements, ^{15}N labeled $NFAT_{1-130}$ S110A and $NFAT_{1-130}$ pS23 S110A were concentrated to a final concentration of 270 µM in NFAT NMR buffer (50 mM MES, pH = 6.8, 140 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 5 % D_2O). Experiments were recorded at 288 K on a Varian (Agilent DD2 600) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 600 MHz proton frequency. CPMG experiments were used at νCPMG frequencies of 50, 100, 250, 400, 500, 650, 850 and 1000 Hz to calculated effective R^2 rates. NMR experiments were processed with nmrPipe and analyzed using the relaxGUI software (version 2.4.3).[415]

7.24.5 ^{15}N-^{1}H-HSQC experiments of eIF4E isoforms

eIF4E isoforms were concentrated to a final concentration of 100 µM in eIF4E NMR buffer (100 sodium phosphate, pH = 6.5 and 2 mM DTT). HSQC experiments were recorded for ^{15}N-labeled GB1-eIF4E, GB1-eIF4E S209D and GB1-eIF4E pS209 without TROSY selection (GB1-eIF4E and GB1-eIF4E pS209) and with TROSY selection (GB1-eIF4E) at 298 K on a Varian (Agilent DD2 600) spectrometer equipped with a TXO-style cryogenically cooled probe, operating at 600 MHz proton frequency. NMR experiments

were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414]

7.24.6 ^{15}N-^{1}H-HSQC experiments of ACP isoforms

Apo-ACP, 3-oxo-C6i-ACP and C8i-ACP were concentrated to a final concentration of 200 µM. HSQC experiments were recorded for ^{15}N-labeled ACP isoforms at 298 K on a were recorded on a Bruker Avance III 800 MHz spectrometer with a TXO-style cryogenically cooled probe, operating at 800 MHz proton frequency.

For titration experiments of 3-oxo-C6i-ACP and C8i-ACP with EsaI and BmaI, respectively, experimental conditions were adjusted as stated in the results section. NMR experiments were processed with nmrPipe and analyzed using the ccpNMR software (version 2.4.1).[414]

7.25 Cell culture and fluorescence imaging

HEK293T cells were grown in 4.5 g/L glucose DMEM supplemented with 10% fetal bovine serum (Gibco 16000-49) and 1% penicillin/streptomycin. Cells were seeded into clear 6-well plates at 600,000 cells per well. 24 h post seeding, cells were transfected with 1 µg purified plasmid, 0.5 µL P3000™ Reagent and 1.5 µL Lipofectamine® 3000 Reagent (Invitrogen) per well. Cells were grown for 48 h, then stained using Hoeachst33342 and imaged using a Widefield inverted Nikon Ti2 fluorescence microscope with incubation chamber. Images were analyzed using the ImageJ software.

7.26 Expression of eIF4E pS209

For the expression of eIF4E pS209, the amber codon technique described in **3.4 Project 3: The eukaryotic translation initiation factor eIF4E** was used. A plasmid with a mutation to the amber codon (TAG) in position 209 had been cloned by a former lab member (SANDEEP CHHABRA) before. This plasmid and a SepOTSλ plasmid (addgene) coding for an orthogonal aminoacyl-tRNA synthetase (aaRS) that only accepts phosphoserine were cotransformed into electrocompetent BL21ΔserB (addgene),

according to PARK et al.[348] Bacterial expression from there on was performed as described in **7.1.2.4 Procedure D: Expression in *E. coli* BL21(DE3)pLysS.**

7.27 Crystallization of eIF4E isoforms

Δ_{26}-eIF4E S209D crystals were obtained from purified protein at a final concentration of 420 µM in 100 mM sodium malate, pH = 5.2, 14 % PEG-4000 and 2 mM TCEP-HCl. The dataset was recorded using an APS_24IDE beamline with 0.2 deg oscillation, 0.2 sec exposure, operating at the following energy level: Se: 12662.0eV (0.9792A). Refined parameters are found in **Table 6.**

Data collection	**eIF4E S209D**
Space group	$P12_11$
Wavelength (Å)	0.94
Cell dimensions	
a, b, c (Å)	39.75, 120.10, 81.12
α, β, γ (°)	90.00, 102.90, 90.00
R-merge	0.066 (0.907)
I/σI	8.8 (0.8)
Completeness	97.8 (83.4)
$CC_{1/2}$	0.997 (0.417)

Table 6 *Dataset parameters for the structure obtained from Δ_{26}-eIF4E S209D bound to m^7-GTP.*

Δ_{26}-eIF4E pS209 crystals were obtained from purified protein at a final concentration of 100 µM in 100 mM sodium citrate, pH = 5.0, 50 mM $(NH_4)_2SO_4$ and 24 % PEG-3350. The dataset was recorded using an APS_24IDE beamline with 0.2 deg oscillation, 0.2 sec exposure, operating at the following energy level: Se: 12662.0eV (0.9792A). Refined parameters are found in **Table 7.**

Data collection	eIF4E pS209
Space group	P121
Wavelength (Å)	0.993
Cell dimensions	
a, b, c (Å)	40.19, 115.35, 81.89
α, β, γ (°)	90.00, 97.88, 90.00
R-merge	0.121 (1.593)
I/σI	8.6 (1.3)
Completeness	97.7 (98.0)
$CC_{1/2}$	0.994 (0.470)

Table 7 *Dataset parameters for the structure obtained from Δ_{26}-eIF4E S209D bound to m^7-GTP.*

7.28 Loading of apo-ACP with 3-oxo-C6i-CoA and C8i-CoA

For acyl-ACP synthesis, apo-ACP was concentrated to concentrations between 100-500 µM in ACP NMR buffer (50 mM MES, pH = 6.8, 10 mM $MgCl_2$, 1 mM TCEP-HCl) and combined with 1 µM purified Sfp and 1.25 equivalents (with respect to apo-ACP) of either 3-oxo-C6i-CoA or C8i-CoA. The reaction mixture was allowed to incubate at 25 °C for 1 h and completeness was confirmed by ^{15}N-^{1}H-HSQC NMR spectra.

7.29 Circular dichroism experiments of GILZ

To probe secondary structure motifs in GILZ using circular dichroism (CD), GILZ was buffer exchanged into a suitable buffer (10 mM sodium phosphate, pH = 6.5, 500 mM NaF, 2 mM DTT). The protein was concentrated to a final concentration of 0.2 mg/mL and analyzed using a 1 mm pathlength quartz cuvette (Hellma 110-1-40, style 110-QS, sigma-Aldrich) in a J-815 CD Spectropolarimeter (JASCO). CD spectra were analyzed using DICHROWEB.[416]

7.30 GST-GILZ pull-down experiments and Western Blots

HEK293T cells were grown in 4.5 g/L glucose DMEM supplemented with 10% fetal bovine serum (Gibco 16000-49) and 1% penicillin/streptomycin. Cells were seeded into clear 6-well plates at 600,000 cells per well. 24 h post seeding, cells were transfected with 1 µg purified FLAG-HDAC7 plasmids, 0.5 µL P3000™ Reagent and 1.5 µL Lipofectamine® 3000 Reagent (Invitrogen) per well. After 24 h, cells were resuspended in 500 µL RIPA buffer (20 mM Tris-HCl, pH = 7.5, 150 mM NaCl, 1 mM EDTA, 1 % NP-40, 1 % sodium deoxycholate) per well and lysed using sonication (30 % amplitude, 1 min sonication time, 1 s on pulse, 1 s off pulse). 50 µL of glutathione sepharose 4 fast flow bead slurry (GE Healthcare) were equilibrated in binding buffer (PBS, pH = 7.3, 3 mM DTT), incubated for 2 h with 0.3 mg of purified GST-GILZ or GST control, and washed in 5 × 50 µL binding buffer to remove unbound protein. Then, washed beads were added to the aforementioned HEK293T cell lysates and incubated for 2 h at 4 °C. The beads were washed 6 times with 1 mL wash buffer (PBS, pH = 7.3, 3 mM DTT, 0.05 % Tween 20, 500 mM NaCl), with centrifugation steps at 500 x g for 3 min in between washes. Proteins were eluted by addition of 25 µL GST elution buffer (50 mM Tris-HCl, pH = 8.0, 10 mM reduced glutathione), followed by 1 h of incubation. The samples were then analyzed by SDS-PAGE and transferred to a polyvinylidene difluoride membrane using a transfer apparatus according to the manufacturer's protocols (Bio-Rad). After incubation with 2 % Non-Fat Dry Milk in TBST (10 mM Tris-HCl, pH = 8.0, 150 mM NaCl, 0.5 % Tween 20) for 60 min, the membrane was washed once with TBST and incubated with antibodies against monoclonal anti-FLAG® M2 antibody (Sigma-Aldrich, 1:1,000) for 16 h at 4 °C. Blots were washed three times in TBST and incubated for 1 h with goat anti-mouse IgG HRP (sigma-Aldrich, 1:50,000). For detection, the Immun-Star™ AP Chemiluminescence Kits (Bio-Rad) was used according to the manufacturer's protocols.

8 Bibliography

[1] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea, et al., *A comprehensive map of molecular drug targets*. *Nat. Rev. Drug Discov.* **2017**, *16*, 19–34.

[2] T. L. Blundell, H. Jhoti, C. Abell, *High-throughput crystallography for lead discovery in drug design*. *Nat. Rev. Drug Discov.* **2002**, *1*, 45–54.

[3] C. W. Murray, T. L. Blundell, *Structural biology in fragment-based drug design*. *Curr. Opin. Struct. Biol.* **2010**, *20*, 497–507.

[4] J. R. Deschamps, *The Role of Crystallography in Drug Design*. *AAPS J.* **2005**, *7*, E813–E819.

[5] M. Almén, K. J. Nordström, R. Fredriksson, H. B. Schiöth, *Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin*. *BMC Biol.* **2009**, *7*, 50.

[6] A. Krogh, B. Larsson, G. von Heijne, E. L. L. Sonnhammer, *Predicting transmembrane protein topology with a hidden markov model: application to complete genomes*. *J. Mol. Biol.* **2001**, *305*, 567–580.

[7] J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *How many drug targets are there? Nat. Rev. Drug Discov.* **2006**, *5*, 993–996.

[8] L. Frey, N.-A. Lakomek, R. Riek, S. Bibow, *Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics*. *Angew. Chem. Int. Ed.* **2017**, *56*, 380–383.

[9] A. Rosenhouse-Dantsker, D. Mehta, I. Levitan, *Regulation of Ion Channels by Membrane Lipids*. *Compr. Physiol.* **2012**, *2*, 31–68.

[10] K. Tsujita, T. Itoh, *Phosphoinositides in the regulation of actin cortex and cell migration*. *Biochim. Biophys. Acta* **2015**, *1851*, 824–831.

[11] B. Chaigne-Delalande, J.-F. Moreau, P. Legembre, *Rewinding the DISC*. *Arch. Immunol. Ther. Exp.* **2008**, *56*, 9–14.

[12] F. Szoka, D. Papahadjopoulos, *Comparative Properties and Methods of Preparation of Lipid Vesicles (Liposomes)*. *Annu. Rev. Biophys. Bioeng.* **1980**, *9*, 467–508.

[13] J.-L. Rigaud, B. Pitard, D. Levy, *Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins*. *Biochim. Biophys. Acta BBA - Bioenerg.* **1995**, *1231*, 223–246.

[14] G. D. Eytan, *Use of liposomes for reconstitution of biological functions*. *Biochim. Biophys. Acta* **1982**, *694*, 185–202.

[15] I. G. Denisov, S. G. Sligar, *Nanodiscs in Membrane Biochemistry and Biophysics*. *Chem. Rev.* **2017**, *117*, 4669–4713.

[16] A. Helenius, K. Simons, *Solubilization of membranes by detergents*. *Biochim. Biophys. Acta BBA - Rev. Biomembr.* **1975**, *415*, 29–79.

[17] D. Lee, K. F. A. Walter, A.-K. Brückner, C. Hilty, S. Becker, C. Griesinger, *Bilayer in Small Bicelles Revealed by Lipid–Protein Interactions Using NMR Spectroscopy*. *J. Am. Chem. Soc.* **2008**, *130*, 13822–13823.

[18] P. S. Chae, S. G. F. Rasmussen, R. R. Rana, K. Gotfryd, A. C. Kruse, A. Manglik, K. H. Cho, S. Nurva, U. Gether, L. Guan, et al., *A New Class of Amphiphiles Bearing Rigid Hydrophobic Groups for Solubilization and Stabilization of Membrane Proteins*. *Chem. - Eur. J.* **2012**, *18*, 9485–9490.

[19] H. Noguchi, *Structure formation in binary mixtures of lipids and detergents: Self-assembly and vesicle division*. *J. Chem. Phys.* **2013**, *138*, 024907.

[20] D. E. Warschawski, A. A. Arnold, M. Beaugrand, A. Gravel, É. Chartrand, I. Marcotte, *Choosing membrane mimetics for NMR structural studies of transmembrane proteins*. *Biochim. Biophys. Acta BBA - Biomembr.* **2011**, *1808*, 1957–1974.

[21] P. T. Duong, H. L. Collins, M. Nickel, S. Lund-Katz, G. H. Rothblat, M. C. Phillips, *Characterization of nascent HDL particles and microparticles formed by ABCA1-mediated efflux of cellular lipids to apoA-I*. *J. Lipid Res.* **2006**, *47*, 832–843.

[22] F. Hagn, M. L. Nasr, G. Wagner, *Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR*. *Nat. Protoc.* **2018**, *13*, 79–98.

[23] P. E. Wright, H. J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation*. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29.

[24] A. Russo, S. Manna, E. Novellino, A. Malfitano, D. Marasco, *Molecular signaling involving intrinsically disordered proteins in prostate cancer*. *Asian J. Androl.* **2016**, *18*, 673.

[25] G. I. Abelev, *Alpha-Fetoprotein in Ontogenesis and its Association with Malignant Tumors*. *Adv. Cancer Res.* **1971**, *14*, 295–358.

[26] K. H. Vousden, X. Lu, *Live or let die: the cell's response to p53*. *Nat. Rev. Cancer* **2002**, *2*, 594–604.

[27] A. R. Venkitaraman, *Cancer Susceptibility and the Functions of BRCA1 and BRCA2*. *Cell* **2002**, *108*, 171–182.

[28] M. Goedert, *Filamentous nerve cell inclusions in neurodegenerative diseases: tauopathies and α-synucleinopathies*. *Philos Trans R Soc Lond B Biol Sci* **1999**, *354*, 1101–1118.

[29] K. Simmons, C. May, J. Tomaselli, E. Rydel, S. Fuson, W. Becker, N. Brems, W. Ying, *Secondary Structure of Amyloid 3 Peptide Correlates with Neurotoxic Activity In Vitro*. *Mol. Pharmacol.* **1994**, *45*, 373–379.

[30] R. P. Harvey, E. Degryse, L. Stefani, F. Schamber, J. P. Cazenave, M. Courtney, P. Tolstoshev, J. P. Lecocq, *Cloning and expression of a cDNA coding for the anticoagulant hirudin from the bloodsucking leech, Hirudo medicinalis. Proc. Natl. Acad. Sci.* USA **1986**, *83*, 1084–1088.

[31] H. M. H. Spronk, J. I. Borissoff, H. ten Cate, *New Insights into Modulation of Thrombin Formation*. *Curr. Atheroscler. Rep.* **2013**, *15*, 363.

[32] E. T. A. S. Jaikaran, C. E. Higham, L. C. Serpell, J. Zurdo, M. Gross, A. Clark, P. E. Fraser, *Identification of a novel human islet amyloid polypeptide β-sheet domain and factors influencing fibrillogenesis*. *J. Mol. Biol.* **2001**, *308*, 515–525.

[33] V. Bandaru, W. Cooper, S. S. Wallace, S. Doublié, *Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that recognizes oxidative DNA damage*. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1142–1144.

[34] C. J. Oldfield, B. Xue, Y.-Y. Van, E. L. Ulrich, J. L. Markley, A. K. Dunker, V. N. Uversky, *Utilization of protein intrinsic disorder knowledge in structural proteomics*. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2013**, *1834*, 487–498.

[35] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins*. *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.

[36] N. Sibille, P. Bernadó, *Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS*. *Biochem. Soc. Trans.* **2012**, *40*, 955–962.

[37] J. A. Marsh, J. D. Forman-Kay, *Ensemble modeling of protein disordered states: Experimental restraint contributions and validation*. *Proteins Struct. Funct. Bioinforma.* **2012**, *80*, 556–572.

[38] S. Frey, R. P. Richter, D. Görlich, *FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties*. *Science* **2006**, *314*, 815–817.

[39] P. Tompa, *The interplay between structure and function in intrinsically unstructured proteins*. *FEBS Lett.* **2005**, *579*, 3346–3354.

[40] M. Guharoy, B. Szabo, S. C. Martos, S. Kosol, P. Tompa, *Intrinsic Structural Disorder in Cytoskeletal Proteins: Intrinsic Structural Disorder*. *Cytoskeleton* **2013**, *70*, 550–571.

[41] V. N. Uversky, *Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder*. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30.

[42] M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, et al., *Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels*. *Cell* **2012**, *149*, 753–767.

[43] S. C. Weber, C. P. Brangwynne, *Getting RNA and Protein in Phase*. *Cell* **2012**, *149*, 1188–1191.

[44] A. H. Millar, J. L. Heazlewood, C. Giglione, M. J. Holdsworth, A. Bachmair, W. X. Schulze, *The Scope, Functions, and Dynamics of Posttranslational Protein Modifications*. *Annu. Rev. Plant Biol.* **2019**, *70*, 119–151.

[45] M. D. Resh, *Covalent lipid modifications of proteins*. *Curr. Biol.* **2013**, *23*, R431–R435.

[46] S. E. J. Bowman, K. L. Bren, *The chemistry and biochemistry of heme c: functional bases for covalent attachment*. *Nat. Prod. Rep.* **2008**, *25*, 1118.

[47] M. T. A. Alexandre, T. Domratcheva, C. Bonetti, L. J. G. W. van Wilderen, R. van Grondelle, M.-L. Groot, K. J. Hellingwerf, J. T. M. Kennis, *Primary Reactions of the LOV2 Domain of Phototropin Studied with Ultrafast Mid-Infrared Spectroscopy and Quantum Chemistry*. *Biophys. J.* **2009**, *97*, 227–237.

[48] A. Drazic, L. M. Myklebust, R. Ree, T. Arnesen, *The world of protein acetylation*. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864*, 1372–1401.

[49] J. Murn, Y. Shi, *The winding path of protein methylation research: milestones and new frontiers*. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 517–527.

[50] F. Ardito, M. Giuliani, D. Perrone, G. Troiano, L. L. Muzio, *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy*. *Int. J. Mol. Med.* **2017**, *40*, 271–280.

[51] G. A. Khoury, R. C. Baliban, C. A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database*. *Sci. Rep.* **2011**, *1*, 90.

[52] B. M. Babior, *NADPH oxidase: an update*. *Blood* **1999**, *93*, 1464–1476.

[53] M. S. Feschenko, K. J. Sweadner, *Phosphorylation of Na,K-ATPase by Protein Kinase C at Ser ^{18} Occurs in Intact Cells but Does Not Result in Direct Inhibition of ATP Hydrolysis*. *J. Biol. Chem.* **1997**, *272*, 17726–17733.

[54] P. C. van Weeren, K. M. T. de Bruyn, A. M. M. de Vries-Smits, J. van Lint, B. M. Th. Burgering, *Essential Role for Protein Kinase B (PKB) in Insulin-induced Glycogen Synthase Kinase 3 Inactivation: Characterization of dominant-negative mutant of PKB*. *J. Biol. Chem.* **1998**, *273*, 13150–13156.

[55] P. Nash, X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, M. Tyers, *Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication*. *Nature* **2001**, *414*, 514–521.

[56] S. Orlicky, X. Tang, A. Willems, M. Tyers, F. Sicheri, *Structural Basis for Phosphodependent Substrate Selection and Orientation by the SCFCdc4 Ubiquitin Ligase*. *Cell* **2003**, *112*, 243–256.

[57] C. Salazar, T. Höfer, *Allosteric Regulation of the Transcription Factor NFAT1 by Multiple Phosphorylation Sites: A Mathematical Analysis*. *J. Mol. Biol.* **2003**, *327*, 31–45.

[58] H. Okamura, J. Aramburu, C. García-Rodríguez, J. P. B. Viola, A. Raghavan, M. Tahiliani, X. Zhang, J. Qin, P. G. Hogan, A. Rao, *Concerted Dephosphorylation of the Transcription Factor NFAT1 Induces a Conformational Switch that Regulates Transcriptional Activity*. *Mol. Cell* **2000**, *6*, 539–550.

[59] J. Zuberek, *Phosphorylation of eIF4E attenuates its interaction with mRNA 5' cap analogs by electrostatic repulsion: Intein-mediated protein ligation strategy to obtain phosphorylated protein*. *RNA* **2003**, *9*, 52–61.

[60] M. Kõivomägi, E. Valk, R. Venta, A. Iofik, M. Lepiku, E. R. M. Balog, S. M. Rubin, D. O. Morgan, M. Loog, *Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase*. *Nature* **2011**, *480*, 128–131.

[61] W. Vollmer, D. Blanot, M. A. De Pedro, *Peptidoglycan structure and architecture*. *FEMS Microbiol. Rev.* **2008**, *32*, 149–167.

[62] C. L. Ventola, *The antibiotic resistance crisis: part 1: causes and threats*. *P T* **2015**, *40*, 277–283.

[63] R. R. Yocum, J. R. Rasmussen, J. L. Strominger, *The mechanism of action of penicillin. Penicillin acylates the active site of Bacillus stearothermophilus D-alanine carboxypeptidase*. *J. Biol. Chem.* **1980**, *255*, 3977–3986.

[64] D. S. Boyle, W. D. Donachie, *mraY Is an Essential Gene for Cell Growth in Escherichia coli*. *J. Bacteriol.* **1998**, *180*, 6429–6432.

[65] A. Bouhss, M. Crouvoisier, D. Blanot, D. Mengin-Lecreulx, *Purification and Characterization of the Bacterial MraY Translocase Catalyzing the First Membrane Step of Peptidoglycan Biosynthesis*. *J. Biol. Chem.* **2004**, *279*, 29974–29980.

[66] Y. Liu, J. P. G. L. M. Rodrigues, A. M. J. J. Bonvin, E. A. Zaal, C. R. Berkers, M. Heger, K. Gawarecka, E. Swiezewska, E. Breukink, M. R. Egmond, *New Insight into the Catalytic Mechanism of Bacterial MraY from Enzyme Kinetics and Docking Studies*. *J. Biol. Chem.* **2016**, *291*, 15057–15068.

[67] M. Ikeda, M. Wachi, H. K. Jung, F. Ishino, M. Matsuhashi, *The Escherichia coli mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-N-acetylmuramoyl-pentapeptide transferase. J. Bacteriol.* **1991**, *173*, 1021–1026.

[68] L. A. McDonald, L. R. Barbieri, G. T. Carter, E. Lenoy, J. Lotvin, P. J. Petersen, M. M. Siegel, G. Singh, R. T. Williamson, *Structures of the Muraymycins, Novel Peptidoglycan Biosynthesis Inhibitors*. *J. Am. Chem. Soc.* **2002**, *124*, 10260–10261.

[69] F. Isono, M. Inukai, *Mureidomycin A, a new inhibitor of bacterial peptidoglycan synthesis. Antimicrob. Agents Chemother.* **1991**, *35*, 234–236.

[70] M. Igarashi, N. Nakagawa, N. Doi, S. Hattori, H. Naganawa, M. Hamada, *Caprazamycin B, a novel anti-tuberculosis antibiotic, from Streptomyces sp*. *J. Antibiot.* **2003**, *56*, 580–583.

[71] D. Wiegmann, S. Koppermann, M. Wirth, G. Niro, K. Leyerer, C. Ducho, *Muraymycin nucleoside-peptide antibiotics: uridine-derived natural products as lead structures for the development of novel antibacterial agents*. *Beilstein J. Org. Chem.* **2016**, *12*, 769–795.

[72] K. Kimura, *Liposidomycin, the first reported nucleoside antibiotic inhibitor of peptidoglycan biosynthesis translocase I: The discovery of liposidomycin and related compounds with a perspective on their application to new antibiotics*. *J. Antibiot.* **2019**, *72*, 877–889.

[73] F. Wiker, N. Hauck, S. Grond, B. Gust, *Caprazamycins: Biosynthesis and structure activity relationship studies*. *Int. J. Med. Microbiol.* **2019**, *309*, 319–324.

[74] T. D. H. Bugg, R. V. Kerr, *Mechanism of action of nucleoside antibacterial natural product antibiotics*. *J. Antibiot.* **2019**, *72*, 865–876.

[75] T. Bugg, A. Lloyd, D. Roper, *Phospho-MurNAc-Pentapeptide Translocase (MraY) as a Target for Antibacterial Agents and Antibacterial Proteins*. *Infect. Disord. - Drug Targets* **2006**, *6*, 85–106.

[76] B. C. Chung, J. Zhao, R. Gillespie, D. Y. Kwon, Z. Guan, J. Hong, P. Zhou, S.-Y. Lee, *Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis*. *Science* **2013**, *341*, 1012–1016.

[77] B. C. Chung, E. H. Mashalidis, T. Tanino, M. Kim, A. Matsuda, J. Hong, S. Ichikawa, S.-Y. Lee, *Structural insights into inhibition of lipid I production in bacterial cell wall synthesis*. *Nature* **2016**, *533*, 557–560.

[78] J. K. Hakulinen, J. Hering, G. Brändén, H. Chen, A. Snijder, M. Ek, P. Johansson, *MraY–antibiotic complex reveals details of tunicamycin mode of action*. *Nat. Chem. Biol.* **2017**, *13*, 265–267.

[79] E. H. Mashalidis, B. Kaeser, Y. Terasawa, A. Katsuyama, D.-Y. Kwon, K. Lee, J. Hong, S. Ichikawa, S.-Y. Lee, *Chemical logic of MraY inhibition by antibacterial nucleoside natural products*. *Nat. Commun.* **2019**, *10*, 2917.

[80] A. Takatsuki, K. Arima, G. Tamura, *Tunicamycin, a new antibiotic. I. Isolation and characterization of tunicamycin*. *J. Antibiot.* **1971**, *24*, 215–223.

[81] S. Tanaka, W. M. Clemons Jr, *Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ΦX174: A mutational analysis of phiX174 gene E mediated cell lysis*. *Mol. Microbiol.* **2012**, *85*, 975–985.

[82] S. Mendel, *Interaction of the transmembrane domain of lysis protein E from bacteriophage X174 with bacterial translocase MraY and peptidyl-prolyl isomerase SlyD*. *Microbiology* **2006**, *152*, 2959–2967.

[83] C. L. White, A. Kitich, J. W. Gober, *Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD: Control of cell shape in bacteria*. *Mol. Microbiol.* **2010**, *76*, 616–633.

[84] F. Laddomada, M. Miyachiro, A. Dessen, *Structural Insights into Protein-Protein Interactions Involved in Bacterial Cell Wall Biogenesis*. *Antibiotics* **2016**, *5*, 14.

[85] L. L. Silver, *Challenges of Antibacterial Discovery*. *Clin. Microbiol. Rev.* **2011**, *24*, 71–109.

[86] A. Gaballah, A. Kloeckner, C. Otten, H.-G. Sahl, B. Henrichfreise, *Functional Analysis of the Cytoskeleton Protein MreB from Chlamydophila pneumoniae*. *PLoS ONE* **2011**, *6*, e25129.

[87] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, et al., *The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible*. *Nucleic Acids Res.* **2017**, *45*, D362–D368.

[88] G. A. H. Buttle, W. H. Gray, D. Stephenson, *Protection of Mice against Strepto-coeeal and Other Infections by p-Amino-benzenesulphonamide and Related Substances*. *Lancet* **1982**, *36*, 184–188.

[89] S. B. Zaman, M. A. Hussain, R. Nye, V. Mehta, K. T. Mamun, N. Hossain, *A Review on Antibiotic Resistance: Alarm Bells are Ringing*. *Cureus* **2017**, *9*, e1403.

[90] E. M. Wise, M. M. Abou-Donia, *Sulfonamide resistance mechanism in Escherichia coli: R plasmids can determine sulfonamide-resistant dihydropteroate synthases. Proc. Natl. Acad. Sci.* USA **1975**, *72*, 2621–2625.

[91] M. Kolář, K. Urbánek, T. Látal, *Antibiotic selective pressure and development of bacterial resistance*. *Int. J. Antimicrob. Agents* **2001**, *17*, 357–363.

[92] R. Laxminarayan, G. M. Brown, *Economics of Antibiotic Resistance: A Theory of Optimal Use*. *J. Environ. Econ. Manag.* **2001**, *42*, 183–206.

[93] D. N. Wilson, *Ribosome-targeting antibiotics and mechanisms of bacterial resistance*. *Nat. Rev. Microbiol.* **2014**, *12*, 35–48.

[94] J. M. Munita, C. A. Arias, *Mechanisms of Antibiotic Resistance*. *Microbiol. Spectr.* **2016**, *4*, 10.1128/microbiolspec.VMBF-0016-2015.

[95] V. M. D'Costa, C. E. King, L. Kalan, M. Morar, W. W. L. Sung, C. Schwarz, D. Froese, G. Zazula, F. Calmels, R. Debruyne, et al., *Antibiotic resistance is ancient*. *Nature* **2011**, *477*, 457–461.

[96] K. Bush, *Proliferation and significance of clinically relevant β-lactamases: β-lactamase overview*. *Ann. N. Y. Acad. Sci.* **2013**, *1277*, 84–90.

[97] K. Bush, *The ABCD's of β-lactamase nomenclature*. *J. Infect. Chemother.* **2013**, *19*, 549–559.

[98] T. J. Beveridge, *Structures of Gram-Negative Cell Walls and Their Derived Membrane Vesicles*. *J. Bacteriol.* **1999**, *181*, 4725–4733.

[99] H. I. Zgurskaya, C. A. López, S. Gnanakaran, *Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It*. *ACS Infect. Dis.* **2015**, *1*, 512–522.

[100] S. I. Miller, *Antibiotic Resistance and Regulation of the Gram-Negative Bacterial Outer Membrane Barrier by Host Innate Immune Molecules*. *mBio* **2016**, *7*, e01541-16.

[101] J. P. Quinn, E. J. Dudek, C. A. DiVincenzo, D. A. Lucks, S. A. Lerner, *Emergence of Resistance to Imipenem During Therapy for Pseudomonas aeruginosa Infections*. *J. Infect. Dis.* **1986**, *154*, 289–294.

[102] U. O. Hasdemir, J. Chevalier, P. Nordmann, J.-M. Pages, *Detection and Prevalence of Active Drug Efflux Mechanism in Various Multidrug-Resistant Klebsiella pneumoniae Strains from Turkey*. *J. Clin. Microbiol.* **2004**, *42*, 2701–2706.

[103] L. J. V. Piddock, *Multidrug-resistance efflux pumps - not just for resistance*. *Nat. Rev. Microbiol.* **2006**, *4*, 629–636.

[104] I. Chopra, M. Roberts, *Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance*. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 232–260.

[105] S. R. Connell, D. M. Tracz, K. H. Nierhaus, D. E. Taylor, *Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance*. *Antimicrob. Agents Chemother.* **2003**, *47*, 3675–3681.

[106] L. Martínez-Martínez, A. Pascual, G. A. Jacoby, *Quinolone resistance from a transferable plasmid*. *The Lancet* **1998**, *351*, 797–799.

[107] J. M. Rodríguez-Martínez, C. Velasco, Á. Pascual, M. E. Cano, L. Martínez-Martínez, L. Martínez-Martínez, Á. Pascual, *Plasmid-mediated quinolone resistance: an update*. *J. Infect. Chemother.* **2011**, *17*, 149–182.

[108] B. Weisblum, *Erythromycin resistance by ribosome modification*. *Antimicrob. Agents Chemother.* **1995**, *39*, 577–585.

[109] R. Leclercq, *Mechanisms of Resistance to Macrolides and Lincosamides: Nature of the Resistance Elements and Their Clinical Implications*. *Clin. Infect. Dis.* **2002**, *34*, 482–492.

[110] J. Flensburg, O. Skold, *Massive overproduction of dihydrofolate reductase in bacteria as a response to the use of trimethoprim*. *Eur. J. Biochem.* **1987**, *162*, 473–476.

[111] H. G. Floss, T.-W. Yu, *RifamycinMode of Action, Resistance, and Biosynthesis*. *Chem. Rev.* **2005**, *105*, 621–632.

[112] Z. Cui, X.-C. Wang, X. Liu, A. Lemke, S. Koppermann, C. Ducho, J. Rohr, J. S. Thorson, S. G. Van Lanen, *Self-Resistance during Muraymycin Biosynthesis: a Complementary Nucleotidyltransferase and Phosphotransferase with Identical Modification Sites and Distinct Temporal Order*. *Antimicrob. Agents Chemother.* **2018**, *62*, e00193-18.

[113] P. Jehl, J. Manguy, D. C. Shields, D. G. Higgins, N. E. Davey, *ProViz—a web-based visualization tool to investigate the functional and evolutionary features of protein sequences*. *Nucleic Acids Res.* **2016**, *44*, W11–W15.

[114] P. Oelschlaeger, *Impact of remote mutations on metallo- -lactamase substrate specificity: Implications for the evolution of antibiotic resistance*. *Protein Sci.* **2005**, *14*, 765–774.

[115] W. Wehrli, *Rifampin: Mechanisms of Action and Resistance*. *Clin. Infect. Dis.* **1983**, *5*, S407–S411.

[116] R. M. Peterson, T. Huang, J. D. Rudolf, M. J. Smanski, B. Shen, *Mechanisms of Self-Resistance in the Platensimycin- and Platencin-Producing Streptomyces platensis MA7327 and MA7339 Strains*. *Chem. Biol.* **2014**, *21*, 389–397.

[117] A. Fabbretti, R. Çapuni, A. M. Giuliodori, L. Cimarelli, A. Miano, V. Napolioni, A. La Teana, R. Spurio, *Characterization of the Self-Resistance Mechanism to Dityromycin in the* Streptomyces *Producer Strain*. *mSphere* **2019**, *4*, e00554-19.

[118] M. Merika, D. Thanos, *Enhanceosomes*. *Curr. Opin. Genet. Dev.* **2001**, *11*, 205–208.

[119] Y. Wu, M. Borde, V. Heissmeyer, M. Feuerer, A. D. Lapan, J. C. Stroud, D. L. Bates, L. Guo, A. Han, S. F. Ziegler, et al., *FOXP3 Controls Regulatory T Cell Function through Cooperation with NFAT*. *Cell* **2006**, *126*, 375–387.

[120] A. Kiani, *Regulation of interferon-γ gene expression by nuclear factor of activated T cells*. *Blood* **2001**, *98*, 1480–1488.

[121] O. Kaminuma, F. Kitamura, N. Kitamura, T. Hiroi, H. Miyoshi, A. Miyawaki, S. Miyatake, *Differential Contribution of NFATc2 and NFATc1 to TNF-alpha Gene Expression in T Cells*. *J. Immunol.* **2008**, *180*, 319–326.

[122] M. R. Müller, A. Rao, *NFAT, immunity and cancer: a transcription factor comes of age*. *Nat. Rev. Immunol.* **2010**, *10*, 645–656.

[123] C. Sommerer, S. Meuer, M. Zeier, T. Giese, *Calcineurin inhibitors and NFAT-regulated gene expression*. *Clin. Chim. Acta* **2012**, *413*, 1379–1386.

[124] N. A. Clipstone, G. R. Crabtree, *Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation*. *Nature* **1992**, *357*, 695–697.

[125] A. Rao, C. Luo, P. G. Hogan, *Transcription factors of the NFAT family: regulation and function*. *Annu. Rev. Immunol.* **1997**, *15*, 707–747.

[126] C. Lopez-Rodriguez, J. Aramburu, A. S. Rakeman, A. Rao, *NFAT5, a constitutively nuclear NFAT protein that does not cooperate with Fos and Jun*. *Proc. Natl. Acad. Sci.* USA **1999**, *96*, 7214–7219.

[127] P. G. Hogan, *Transcriptional regulation by calcium, calcineurin, and NFAT*. *Genes Dev.* **2003**, *17*, 2205–2232.

[128] A. M. Ranger, M. Oukka, J. Rengarajan, L. H. Glimcher, *Inhibitory Function of Two NFAT Family Members in Lymphoid Homeostasis and Th2 Development*. *Immunity* **1998**, *9*, 627–635.

[129] C. Dong, *Defective T Cell Differentiation in the Absence of Jnk1*. *Science* **1998**, *282*, 2092–2095.

[130] C.-W. Chow, M. Rincón, R. J. Davis, *Requirement for Transcription Factor NFAT in Interleukin-2 Expression*. *Mol. Cell. Biol.* **1999**, *19*, 2300–2307.

[131] T. Ishida, K. Kinoshita, *PrDOS: prediction of disordered protein regions from amino acid sequence*. *Nucleic Acids Res.* **2007**, *35*, W460–W464.

[132] C. García-Rodríguez, A. Rao, *Nuclear Factor of Activated T Cells (NFAT)-dependent Transactivation Regulated by the Coactivators p300/CREB-binding Protein (CBP)*. *J. Exp. Med.* **1998**, *187*, 2031–2036.

[133] M. Piskacek, M. Havelka, M. Rezacova, A. Knight, *The 9aaTAD Transactivation Domains: From Gal4 to p53*. *PLOS ONE* **2016**, *11*, e0162842.

[134] J. Aramburu, F. García-Cózar, A. Raghavan, H. Okamura, A. Rao, P. G. Hogan, *Selective Inhibition of NFAT Activation by a Peptide Spanning the Calcineurin Targeting Site of NFAT*. *Mol. Cell* **1998**, *1*, 627–637.

[135] S. Park, M. Uesugi, G. L. Verdine, *A second calcineurin binding site on the NFAT regulatory domain*. *Proc. Natl. Acad. Sci.* USA **2000**, *97*, 7130–7135.

[136] L. Chen, J. N. M. Glover, P. G. Hogan, A. Rao, S. C. Harrison, *Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA*. *Nature* **1998**, *392*, 42–48.

[137] J. C. Stroud, L. Chen, *Structure of NFAT Bound to DNA as a Monomer*. *J. Mol. Biol.* **2003**, *334*, 1009–1022.

[138] S. A. Wolfe, P. Zhou, V. Dötsch, L. Chen, A. You, S. N. Ho, G. R. Crabtree, G. Wagner, G. L. Verdine, *Unusual Rel-like architecture in the DNA-binding domain of the transcription factor NFATc*. *Nature* **1997**, *385*, 172–176.

[139] P. G. del Arco, S. Martínez-Martínez, J. L. Maldonado, I. Ortega-Pérez, J. M. Redondo, *A Role for the p38 MAP Kinase Pathway in the Nuclear Shuttling of NFATp*. *J. Biol. Chem.* **2000**, *275*, 13872–13878.

[140] H. Okamura, C. Garcia-Rodriguez, H. Martinson, J. Qin, D. M. Virshup, A. Rao, *A Conserved Docking Motif for CK1 Binding Controls the Nuclear Localization of NFAT1*. *Mol. Cell. Biol.* **2004**, *24*, 4184–4195.

[141] Y. Gwack, S. Sharma, J. Nardone, B. Tanasa, A. Iuga, S. Srikanth, H. Okamura, D. Bolton, S. Feske, P. G. Hogan, et al., *A genome-wide Drosophila RNAi screen identifies DYRK-family kinases as regulators of NFAT*. *Nature* **2006**, *441*, 646–650.

[142] J. R. Arron, M. M. Winslow, A. Polleri, C.-P. Chang, H. Wu, X. Gao, J. R. Neilson, L. Chen, J. J. Heit, S. K. Kim, et al., *NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21*. *Nature* **2006**, *441*, 595–600.

[143] J. G. Murphy, J. L. Sanderson, J. A. Gorski, J. D. Scott, W. A. Catterall, W. A. Sather, M. L. Dell'Acqua, *AKAP-Anchored PKA Maintains Neuronal L-type Calcium Channel Activity and NFAT Transcriptional Signaling*. *Cell Rep.* **2014**, *7*, 1577–1588.

[144] C. M. Sheridan, E. K. Heist, C. R. Beals, G. R. Crabtree, P. Gardner, *Protein Kinase A Negatively Modulates the Nuclear Accumulation of NF-ATc1 by Priming for Subsequent Phosphorylation by Glycogen Synthase Kinase-3*. *J. Biol. Chem.* **2002**, *277*, 48664–48676.

[145] T. Xiao, J. J. Zhu, S. Huang, C. Peng, S. He, J. Du, R. Hong, X. Chen, A. M. Bode, W. Jiang, et al., *Phosphorylation of NFAT3 by CDK3 induces cell transformation and promotes tumor growth in skin cancer*. *Oncogene* **2017**, *36*, 2835–2845.

[146] J. Deng, E. S. Wang, R. W. Jenkins, S. Li, R. Dries, K. Yates, S. Chhabra, W. Huang, H. Liu, A. R. Aref, et al., *CDK4/6 Inhibition Augments Antitumor Immunity by Enhancing T-cell Activation*. *Cancer Discov.* **2018**, *8*, 216–233.

[147] C. M. Porter, M. A. Havens, N. A. Clipstone, *Identification of Amino Acid Residues and Protein Kinases Involved in the Regulation of NFATc Subcellular Localization*. *J. Biol. Chem.* **2000**, *275*, 3543–3551.

[148] F. Meggio, L. A. Pinna, *One-thousand-and-one substrates of protein kinase CK2? FASEB J.* **2003**, *17*, 349–368.

[149] M. C. Lawrence, N. Borenstein-Auerbach, K. McGlynn, F. Kunnathodi, R. Shahbazov, I. Syed, M. Kanak, M. Takita, M. F. Levy, B. Naziruddin, *NFAT Targets Signaling Molecules to Gene Promoters in Pancreatic β-Cells*. *Mol. Endocrinol.* **2015**, *29*, 274–288.

[150] B. San-Antonio, M. A. Íñiguez, M. Fresno, *Protein Kinase Cζ Phosphorylates Nuclear Factor of Activated T Cells and Regulates Its Transactivating Activity*. *J. Biol. Chem.* **2002**, *277*, 27073–27080.

[151] K. L. Abbott, B. B. Friday, D. Thaloor, T. J. Murphy, G. K. Pavlath, *Activation and Cellular Localization of the Cyclosporine A-sensitive Transcription Factor NF-AT in Skeletal Muscle Cells*. *Mol. Biol. Cell* **1998**, *9*, 2905–2916.

[152] D. Cantrell, *Protein kinase B (Akt) regulation and function in T lymphocytes*. *Semin. Immunol.* **2002**, *14*, 19–26.

[153] R. V. Parry, K. Reif, G. Smith, D. M. Sansom, B. A. Hemmings, S. G. Ward, *Ligation of the T cell co-stimulatory receptor CD28 activates the serine-threonine protein kinase protein kinase B*. *Eur. J. Immunol.* **1997**, *27*, 2495–2501.

[154] D. A. E. Cross, D. R. Alessi, P. Cohen, M. Andjelkovich, B. A. Hemmings, *Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B*. *Nature* **1995**, *378*, 785–789.

[155] W. I. Al-Daraji, K. R. Grant, K. Ryan, A. Saxton, N. J. Reynolds, *Localization of Calcineurin/NFAT in Human Skin and Psoriasis and Inhibition of Calcineurin/NFAT Activation in Human Keratinocytes by Cyclosporin A*. *J. Invest. Dermatol.* **2002**, *118*, 779–788.

[156] L. A. Timmerman, N. A. Clipstone, S. N. Ho, J. P. Northrop, G. R. Crabtree, *Rapid shuttling of NF-AT in discrimination of Ca2+ signals and immunosuppression*. *Nature* **1996**, *383*, 837–840.

[157] P. A. Garrity, D. Chen, E. V. Rothenberg, B. J. Wold, *Interleukin-2 transcription is regulated in vivo at the level of coordinated binding of both constitutive and regulated factors. Mol. Cell. Biol.* **1994**, *14*, 2159–2169.

[158] C. Loh, J. A. Carew, J. Kim, P. G. Hogan, A. Rao, *T-cell receptor stimulation elicits an early phase of activation and a later phase of deactivation of the transcription factor NFAT1. Mol. Cell. Biol.* **1996**, *16*, 3945–3954.

[159] S. Piskacek, M. Gregor, M. Nemethova, M. Grabner, P. Kovarik, M. Piskacek, *Nine-amino-acid transactivation domain: Establishment and prediction utilities*. *Genomics* **2007**, *89*, 756–768.

[160] M. Piskacek, *Common Transactivation Motif 9aaTAD recruits multiple general co-activators TAF9, MED15, CBP and p300*. *Nat. Preced.* **2009**.

[161] H. Li, A. Rao, P. G. Hogan, *Structural Delineation of the Calcineurin–NFAT Interaction and its Parallels to PP1 Targeting Interactions*. *J. Mol. Biol.* **2004**, *342*, 1659–1674.

[162] J. Aramburu, M. B. Yaffe, C. López-Rodríguez, L. C. Cantley, P. G. Hogan, A. Rao, *Affinity-Driven Peptide Selection of an NFAT Inhibitor More Selective Than Cyclosporin A*. *Sci. New Ser.* **1999**, *285*, 2129–2133.

[163] J. O. Liu, B. A. Nacev, J. Xu, S. Bhat, *It Takes Two Binding Sites for Calcineurin and NFAT to Tango*. *Mol. Cell* **2009**, *33*, 676–678.

[164] S. Grigoriu, R. Bond, P. Cossio, J. A. Chen, N. Ly, G. Hummer, R. Page, M. S. Cyert, W. Peti, *The Molecular Mechanism of Substrate Engagement and Immunosuppressant Inhibition of Calcineurin*. *PLoS Biol.* **2013**, *11*, e1001492.

[165] M. Gal, S. Li, R. E. Luna, K. Takeuchi, G. Wagner, *The LxVP and PxIxIT NFAT Motifs Bind Jointly to Overlapping Epitopes on Calcineurin's Catalytic Domain Distant to the Regulatory Domain*. *Structure* **2014**, *22*, 1016–1027.

[166] A. Rodríguez, J. Roy, S. Martínez-Martínez, M. D. López-Maderuelo, P. Niño-Moreno, L. Ortí, D. Pantoja-Uceda, A. Pineda-Lucena, M. S. Cyert, J. M. Redondo, *A Conserved Docking Surface on Calcineurin Mediates Interaction with Substrates and Immunosuppressants*. *Mol. Cell* **2009**, *33*, 616–626.

[167] G. N. Huang, D. L. Huso, S. Bouyain, J. Tu, K. A. McCorkell, M. J. May, Y. Zhu, M. Lutz, S. Collins, M. Dehoff, et al., *NFAT Binding and Regulation of T Cell Activation by the Cytoplasmic Scaffolding Homer Proteins*. *Science* **2008**, *319*, 476–481.

[168] K. T. Shaw, A. M. Ho, A. Raghavan, J. Kim, J. Jain, J. Park, S. Sharma, A. Rao, P. G. Hogan, *Immunosuppressive drugs prevent a rapid dephosphorylation of transcription factor NFAT1 in stimulated immune cells. Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11205–11209.

[169] A. Soni, A. Akcakanat, G. Singh, D. Luyimbazi, Y. Zheng, D. Kim, A. Gonzalez-Angulo, F. Meric-Bernstam, *eIF4E knockdown decreases breast cancer cell growth without activating Akt signaling*. *Mol. Cancer Ther.* **2008**, *7*, 1782–1788.

[170] F. Meric, K. K. Hunt, *Translation Initiation in Cancer: A Novel Target for Therapy*. *Mol. Cancer Ther.* **2002**, *1*, 971–979.

[171] J. D. Gross, N. J. Moerke, T. von der Haar, A. A. Lugovskoy, A. B. Sachs, J. E. G. McCarthy, G. Wagner, *Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E*. *Cell* **2003**, *115*, 739–750.

[172] A. Marintchev, K. A. Edmonds, B. Marintcheva, E. Hendrickson, M. Oberer, C. Suzuki, B. Herdy, N. Sonenberg, G. Wagner, *Topology and Regulation of the Human eIF4A/4G/4H Helicase Complex in Translation Initiation*. *Cell* **2009**, *136*, 447–460.

[173] P. Schutz, M. Bumann, A. E. Oberholzer, C. Bieniossek, H. Trachsel, M. Altmann, U. Baumann, *Crystal structure of the yeast eIF4A-eIF4G complex: An RNA-helicase controlled by protein-protein interactions*. *Proc. Natl. Acad. Sci.* USA **2008**, *105*, 9564–9569.

[174] W. C. Merrick, *eIF4F: A Retrospective*. *J. Biol. Chem.* **2015**, *290*, 24091–24099.

[175] N. Sonenberg, A. G. Hinnebusch, *Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets*. *Cell* **2009**, *136*, 731–745.

[176] G. W. Rogers, N. J. Richter, W. C. Merrick, *Biochemical and Kinetic Characterization of the RNA Helicase Activity of Eukaryotic Initiation Factor 4A*. *J. Biol. Chem.* **1999**, *274*, 12236–12244.

[177] U. Harms, A. Z. Andreou, A. Gubaev, D. Klostermeier, *eIF4B, eIF4G and RNA regulate eIF4A activity in translation initiation by modulating the eIF4A conformational cycle*. *Nucleic Acids Res.* **2014**, *42*, 7911–7922.

[178] A. Z. Andreou, D. Klostermeier, *eIF4B and eIF4G Jointly Stimulate eIF4A ATPase and Unwinding Activities by Modulation of the eIF4A Conformational Cycle*. *J. Mol. Biol.* **2014**, *426*, 51–61.

[179] G. W. Rogers, N. J. Richter, W. F. Lima, W. C. Merrick, *Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F*. *J. Biol. Chem.* **2001**, *276*, 30914–30922.
[180] Q. Wang, H.-S. Yang, *The role of Pdcd4 in tumor suppression and protein*. *Biol. Cell* **2018**, *110, 169-177*.
[181] H.-S. Yang, A. P. Jansen, A. A. Komar, X. Zheng, W. C. Merrick, S. Costes, S. J. Lockett, N. Sonenberg, N. H. Colburn, *The Transformation Suppressor Pdcd4 Is a Novel Eukaryotic Translation Initiation Factor 4A Binding Protein That Inhibits Translation*. *Mol. Cell. Biol.* **2003**, *23*, 26–37.
[182] A. Pause, G. J. Belsham, A.-C. Gingras, O. Donze, T.-A. Lin, J. C. Lawrence Jr., N. Sonenberg, *Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5' -cap function*. *Nature* **1994**, *371*, 762–767.
[183] A.-C. Gingras, B. Raught, N. Sonenberg, *eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation*. *Annu. Rev. Biochem.* **1999**, *68*, 913–963.
[184] F. Poulin, A.-C. Gingras, H. Olsen, S. Chevalier, N. Sonenberg, *4E-BP3, a New Member of the Eukaryotic Initiation Factor 4E-binding Protein Family*. *J. Biol. Chem.* **1998**, *273*, 14002–14007.
[185] S. Mader, H. Lee, A. Pause, N. Sonenberg, *The Translation Initiation Factor eIF-4E Binds to a Common Motif Shared by the Translation Factor eIF-4 gamma and the Translational Repressors 4E-Binding Proteins*. *Mol. Cell. Biol.* **1995**, *15*, 4990–4997.
[186] X. M. Ma, J. Blenis, *Molecular mechanisms of mTOR-mediated translational control*. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 307–318.
[187] N. Villa, A. Do, J. W. B. Hershey, C. S. Fraser, *Human Eukaryotic Initiation Factor 4G (eIF4G) Protein Binds to eIF3c, -d, and -e to Promote mRNA Recruitment to the Ribosome*. *J. Biol. Chem.* **2013**, *288*, 32932–32940.
[188] H. Imataka, *A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation*. *EMBO J.* **1998**, *17*, 7480–7489.
[189] S. Pyronnet, *Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E*. *EMBO J.* **1999**, *18*, 270–279.
[190] J. Pelletier, J. Graff, D. Ruggero, N. Sonenberg, *Targeting the eIF4F Translation Initiation Complex: A Critical Nexus for Cancer Development*. *Cancer Res.* **2015**, *75*, 250–263.
[191] S. R. Datta, A. Brunet, M. E. Greenberg, *Cellular survival: a play in three Akts*. *Genes Dev.* **1999**, *13*, 2905–2927.
[192] R. J. Shaw, L. C. Cantley, *Ras, PI(3)K and mTOR signalling controls tumour cell growth*. *Nature* **2006**, *441*, 424–430.
[193] P. E. Burnett, R. K. Barrow, N. A. Cohen, S. H. Snyder, D. M. Sabatini, *RAFT1 phosphorylation of the translational regulators p70 S6 kinase and 4E-BP1*. *Proc. Natl. Acad. Sci.* USA **1998**, *95*, 1432–1437.
[194] N. Hay, *Upstream and downstream of mTOR*. *Genes Dev.* **2004**, *18*, 1926–1945.
[195] R. E. Rhoads, *eIF4E: New Family Members, New Binding Partners, New Roles*. *J. Biol. Chem.* **2009**, *284*, 16711–16715.
[196] A. Phillips, J. P. Blaydes, *MNK1 and EIF4E are downstream effectors of MEKs in the regulation of the nuclear export of HDM2 mRNA*. *Oncogene* **2008**, *27*, 1645–1649.

[197] C.-J. Lin, R. Cencic, J. R. Mills, F. Robert, J. Pelletier, *c-Myc and eIF4F Are Components of a Feedforward Loop that Links Transcription and Translation*. *Cancer Res.* **2008**, *68*, 5326–5334.

[198] C. Andrieu, D. Taieb, V. Baylot, S. Ettinger, P. Soubeyran, A. De-Thonel, C. Nelson, C. Garrido, A. So, L. Fazli, et al., *Heat shock protein 27 confers resistance to androgen ablation and chemotherapy in prostate cancer cells through eIF4E*. *Oncogene* **2010**, *29*, 1883–1896.

[199] T. Murata, K. Shimotohno, *Ubiquitination and Proteasome-dependent Degradation of Human Eukaryotic Translation Initiation Factor 4E*. *J. Biol. Chem.* **2006**, *281*, 20788–20800.

[200] M. A. Andrei, *A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies*. *RNA* **2005**, *11*, 717–727.

[201] M. J. Osborne, K. L. B. Borden, *The eukaryotic translation initiation factor eIF4E in the nucleus: taking the road less traveled*. *Immunol. Rev.* **2015**, *263*, 210–223.

[202] M. Kulkarni, S. Ozgur, G. Stoecklin, *On track with P-bodies*. *Biochem. Soc. Trans.* **2010**, *38*, 242–251.

[203] M. Brengues, R. Parker, *Accumulation of Polyadenylated mRNA, Pab1p, eIF4E, and eIF4G with P-Bodies in* Saccharomyces cerevisiae. *Mol. Biol. Cell* **2007**, *18*, 2592–2602.

[204] K. Frydryskova, T. Masek, K. Borcin, S. Mrvova, V. Venturi, M. Pospisek, *Distinct recruitment of human eIF4E isoforms to processing bodies and stress granules*. *BMC Mol. Biol.* **2016**, *17*, 21.

[205] P. Anderson, N. Kedersha, *Stress granules: the Tao of RNA triage*. *Trends Biochem. Sci.* **2008**, *33*, 141–150.

[206] S. Jain, J. R. Wheeler, R. W. Walters, A. Agrawal, A. Barsic, R. Parker, *ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure*. *Cell* **2016**, *164*, 487–498.

[207] K. L. B. Borden, *The eukaryotic translation initiation factor eIF4E wears a "cap" for many occasions*. *Translation* **2016**, *4*, e1220899.

[208] B. Culjkovic, I. Topisirovic, L. Skrabanek, M. Ruiz-Gutierrez, K. L. B. Borden, *eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3'UTR*. *J. Cell Biol.* **2005**, *169*, 245–256.

[209] L. Volpon, B. Culjkovic-Kraljacic, H. S. Sohn, A. Blanchet-Cohen, M. J. Osborne, K. L. B. Borden, *A biochemical framework for eIF4E-dependent mRNA export and nuclear recycling of the export machinery*. *RNA* **2017**, *23*, 927–937.

[210] B. Culjkovic-Kraljacic, A. Baguet, L. Volpon, A. Amri, K. L. B. Borden, *The Oncogene eIF4E Reprograms the Nuclear Pore Complex to Promote mRNA Export and Oncogenic Transformation*. *Cell Rep.* **2012**, *2*, 207–215.

[211] F. Martin, S. Barends, S. Jaeger, L. Schaeffer, L. Prongidi-Fix, G. Eriani, *Cap-Assisted Internal Initiation of Translation of Histone H4*. *Mol. Cell* **2011**, *41*, 197–209.

[212] I. Topisirovic, M. Ruiz-Gutierrez, K. L. B. Borden, *Phosphorylation of the Eukaryotic Translation Initiation Factor eIF4E Contributes to Its Transformation and mRNA Transport Activities*. *Cancer Res.* **2004**, *64*, 8639–8642.

[213] H.-G. Wendel, R. L. A. Silva, A. Malina, J. R. Mills, H. Zhu, T. Ueda, R. Watanabe-Fukunaga, R. Fukunaga, J. Teruya-Feldstein, J. Pelletier, et al., *Dissecting eIF4E action in tumorigenesis*. *Genes Dev.* **2007**, *21*, 3232–3237.

[214] L. Furic, L. Rong, O. Larsson, I. H. Koumakpayi, K. Yoshida, A. Brueschke, E. Petroulakis, N. Robichaud, M. Pollak, L. A. Gaboury, et al., *eIF4E phosphorylation promotes tumorigenesis and is associated with prostate cancer progression*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14134–14139.

[215] J. R. Graff, B. W. Konicek, R. L. Lynch, C. A. Dumstorf, M. S. Dowless, A. M. McNulty, S. H. Parsons, L. H. Brail, B. M. Colligan, J. W. Koop, et al., *eIF4E Activation Is Commonly Elevated in Advanced Human Prostate Cancers and Significantly Related to Reduced Patient Survival*. *Cancer Res.* **2009**, *69*, 3866–3873.

[216] N. Robichaud, S. V. del Rincon, B. Huor, T. Alain, L. A. Petruccelli, J. Hearnden, C. Goncalves, S. Grotegut, C. H. Spruck, L. Furic, et al., *Phosphorylation of eIF4E promotes EMT and metastasis via translational control of SNAIL and MMP-3*. *Oncogene* **2015**, *34*, 2032–2042.

[217] J. Hou, T. Teo, M. J. Sykes, S. Wang, *Insights into the Importance of DFD-Motif and Insertion I1 in Stabilizing the DFD-Out Conformation of Mnk2 Kinase*. *ACS Med. Chem. Lett.* **2013**, *4*, 736–741.

[218] L. Bellsolell, P. F. Cho-Park, F. Poulin, N. Sonenberg, S. K. Burley, *Two Structurally Atypical HEAT Domains in the C-Terminal Portion of Human eIF4G Support Binding to eIF4A and Mnk1*. *Structure* **2006**, *14*, 913–923.

[219] D. Lama, C. S. Verma, *Deciphering the mechanistic effects of eIF4E phosphorylation on mRNA-cap recognition*. *Protein Sci.* **2019**, pro.3798.

[220] G. C. Scheper, B. van Kollenburg, J. Hu, Y. Luo, D. J. Goss, C. G. Proud, *Phosphorylation of Eukaryotic Initiation Factor 4E Markedly Reduces Its Affinity for Capped mRNA*. *J. Biol. Chem.* **2002**, *277*, 3303–3309.

[221] L. Volpon, M. J. Osborne, I. Topisirovic, N. Siddiqui, K. L. Borden, *Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands*. *EMBO J.* **2006**, *25*, 5138–5149.

[222] F. Zhou, M. Yan, G. Guo, F. Wang, H. Qiu, F. Zheng, Y. Zhang, Q. Liu, X. Zhu, L. Xia, *Knockdown of eIF4E suppresses cell growth and migration, enhances chemosensitivity and correlates with increase in Bax/Bcl-2 ratio in triple-negative breast cancer cells*. *Med. Oncol.* **2011**, *28*, 1302–1307.

[223] C. H. Choi, J.-S. Lee, S. R. Kim, Y.-Y. Lee, C.-J. Kim, J.-W. Lee, T.-J. Kim, J.-H. Lee, B.-G. Kim, D.-S. Bae, *Direct inhibition of eIF4E reduced cell growth in endometrial adenocarcinoma*. *J. Cancer Res. Clin. Oncol.* **2011**, *137*, 463–469.

[224] J. R. Graff, B. W. Konicek, T. M. Vincent, R. L. Lynch, D. Monteith, S. N. Weir, P. Schwier, A. Capen, R. L. Goode, M. S. Dowless, et al., *Therapeutic suppression of translation initiation factor eIF4E expression reduces tumor growth without toxicity*. *J. Clin. Invest.* **2007**, *117*, 2638–2648.

[225] J. Wan, F. Shi, Z. Xu, M. Zhao, *Knockdown of eIF4E suppresses cell proliferation, invasion and enhances cisplatin cytotoxicity in human ovarian cancer cells*. *Int. J. Oncol.* **2015**, *47*, 2217–2225.

[226] A. G. Duffy, O. V. Makarova-Rusher, S. V. Ulahannan, O. E. Rahma, S. Fioravanti, M. Walker, S. Abdullah, M. Raffeld, V. Anderson, N. Abi-Jaoudeh, et al., *Modulation of tumor eIF4E by antisense inhibition: A phase I/II translational clinical trial of ISIS 183750-an antisense oligonucleotide against eIF4E-in combination with irinotecan in solid tumors and irinotecan-refractory colorectal cancer: Modulation of Tumor eIF4E by Antisense Inhibition*. *Int. J. Cancer* **2016**, *139*, 1648–1657.

[227] F. Liu, X. Wang, J. Li, K. Gu, L. Lv, S. Zhang, D. Che, J. Cao, S. Jin, Y. Yu, *miR-34c-3p functions as a tumour suppressor by inhibiting eIF4E expression in non-small cell lung cancer*. *Cell Prolif.* **2015**, *48*, 582–592.

[228] X. Yang, J. Zang, X. Pan, J. Yin, Q. Xiang, J. Yu, R. Gan, X. Lei, *miR-503 inhibits proliferation making human hepatocellular carcinoma cells susceptible to 5-fluorouracil by targeting EIF4E*. *Oncol. Rep.* **2017**, *37*, 563–570.

[229] S. Bajan, G. Hutvagner, *RNA-Based Therapeutics: From Antisense Oligonucleotides to miRNAs*. *Cells* **2020**, *9*, 137.

[230] N. Sekiyama, H. Arthanari, E. Papadopoulos, R. A. Rodriguez-Mias, G. Wagner, M. Léger-Abraham, *Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1*. *Proc. Natl. Acad. Sci.* USA **2015**, *112*, E4036–E4045.

[231] E. Papadopoulos, S. Jenni, E. Kabha, K. J. Takrouri, T. Yi, N. Salvi, R. E. Luna, E. Gavathiotis, P. Mahalingam, H. Arthanari, et al., *Structure of the eukaryotic translation initiation factor eIF4E in complex with 4EGI-1 reveals an allosteric mechanism for dissociating eIF4G*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E3187–E3195.

[232] R. Cencic, D. R. Hall, F. Robert, Y. Du, J. Min, L. Li, M. Qui, I. Lewis, S. Kurtkaya, R. Dingledine, et al., *Reversing chemoresistance by small molecule inhibition of the translation initiation complex eIF4F*. *Proc. Natl. Acad. Sci.* USA **2011**, *108*, 1046–1051.

[233] S. Assouline, B. Culjkovic, E. Cocolakis, C. Rousseau, N. Beslu, A. Amri, S. Caplan, B. Leber, D.-C. Roy, W. H. Miller, et al., *Molecular targeting of the oncogene eIF4E in acute myeloid leukemia (AML): a proof-of-principle clinical trial with ribavirin*. *Blood* **2009**, *114*, 257–260.

[234] M. Xu, Z. Tao, S. Wang, Y. Jiang, M. Qu, *Suppression of oncogenic protein translation via targeting eukaryotic translation initiation factor 4E overcomes chemo-resistance in nasopharyngeal carcinoma*. *Biochem. Biophys. Res. Commun.* **2019**, *512*, 902–907.

[235] J. Cao, L. He, G. Lin, C. Hu, R. Dong, J. Zhang, H. Zhu, Y. Hu, C. R. Wagner, Q. He, et al., *Cap-dependent translation initiation factor, eIF4E, is the target for Ouabain-mediated inhibition of HIF-1α*. *Biochem. Pharmacol.* **2014**, *89*, 20–30.

[236] J. K. Moy, A. Khoutorsky, M. N. Asiedu, B. J. Black, J. L. Kuhn, P. Barragán-Iglesias, S. Megat, M. D. Burton, C. C. Burgos-Vega, O. K. Melemedjian, et al., *The MNK–eIF4E Signaling Axis Contributes to Injury-Induced Nociceptive Plasticity and the Development of Chronic Pain*. *J. Neurosci.* **2017**, *37*, 7481–7499.

[237] E. M. Kosciuczuk, D. Saleiro, B. Kroczynska, E. M. Beauchamp, F. Eckerdt, G. T. Blyth, S. M. Abedin, F. J. Giles, J. K. Altman, L. C. Platanias, *Merestinib blocks Mnk kinase activity in acute myeloid leukemia progenitors and exhibits antileukemic effects in vitro and in vivo*. *Blood* **2016**, *128*, 410–414.

[238] B. W. Konicek, J. R. Stephens, A. M. McNulty, N. Robichaud, R. B. Peery, C. A. Dumstorf, M. S. Dowless, P. W. Iversen, S. Parsons, K. E. Ellis, et al., *Therapeutic Inhibition of MAP Kinase Interacting Kinase Blocks Eukaryotic Initiation Factor 4E Phosphorylation and Suppresses Outgrowth of Experimental Lung Metastases*. *Cancer Res.* **2011**, *71*, 1849–1857.

[239] S. Lim, T. Y. Saw, M. Zhang, M. R. Janes, K. Nacro, J. Hill, A. Q. Lim, C.-T. Chang, D. A. Fruman, D. A. Rizzieri, et al., *Targeting of the MNK-eIF4E axis in blast crisis chronic myeloid leukemia inhibits leukemia stem cell function*. *Proc. Natl. Acad. Sci.* USA **2013**, *110*, E2298–E2307.

[240] M. Konstantinidou, J. Li, B. Zhang, Z. Wang, S. Shaabani, F. Ter Brake, K. Essa, A. Dömling, *PROTACs– a game-changing technology*. *Expert Opin. Drug Discov.* **2019**, *14*, 1255–1268.

[241] H. Lebraud, D. J. Wright, C. N. Johnson, T. D. Heightman, *Protein Degradation by In-Cell Self-Assembly of Proteolysis Targeting Chimeras*. *ACS Cent. Sci.* **2016**, *2*, 927–934.

[242] R. Kanasty, J. R. Dorkin, A. Vegas, D. Anderson, *Delivery materials for siRNA therapeutics*. *Nat. Mater.* **2013**, *12*, 967–977.

[243] M. Toure, C. M. Crews, *Small-Molecule PROTACS: New Approaches to Protein Degradation*. *Angew. Chem. Int. Ed.* **2016**, *55*, 1966–1973.

[244] L. Xiao-Jie, X. Hui-Ying, K. Zun-Ping, C. Jin-Lian, J. Li-Juan, *CRISPR-Cas9: a new and promising player in gene therapy*. *J. Med. Genet.* **2015**, *52*, 289–296.

[245] A. R. Schneekloth, M. Pucheault, H. S. Tae, C. M. Crews, *Targeted intracellular protein degradation induced by a small molecule: En route to chemical proteomics*. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5904–5908.

[246] X. Sun, H. Gao, Y. Yang, M. He, Y. Wu, Y. Song, Y. Tong, Y. Rao, *PROTACs: great opportunities for academia and industry*. *Signal Transduct. Target. Ther.* **2019**, *4*, 64.

[247] C. Zhang, X.-R. Han, X. Yang, B. Jiang, J. Liu, Y. Xiong, J. Jin, *Proteolysis Targeting Chimeras (PROTACs) of Anaplastic Lymphoma Kinase (ALK)*. *Eur. J. Med. Chem.* **2018**, *151*, 304–314.

[248] C. H. Kang, D. H. Lee, C. O. Lee, J. Du Ha, C. H. Park, J. Y. Hwang, *Induced protein degradation of anaplastic lymphoma kinase (ALK) by proteolysis targeting chimera (PROTAC)*. *Biochem. Biophys. Res. Commun.* **2018**, *505*, 542–547.

[249] N. Shibata, K. Shimokawa, K. Nagai, N. Ohoka, T. Hattori, N. Miyamoto, O. Ujikawa, T. Sameshima, H. Nara, N. Cho, et al., *Pharmacological difference between degrader and inhibitor against oncogenic BCR-ABL kinase*. *Sci. Rep.* **2018**, *8*, 13549.

[250] T. Kaur, A. Menon, A. L. Garner, *Synthesis of 7-benzylguanosine cap-analogue conjugates for eIF4E targeted degradation*. *Eur. J. Med. Chem.* **2019**, *166*, 339–350.

[251] K. Piecyk, P. Pietrow, T. Arnold, R. Worch, N. L. Korneeva, M. Jankowska-Anyszka, *Effect of HIV-1 TAT Peptide Fusion on 5' mRNA Cap Analogs Cell Membrane Permeability and Translation Inhibition*. *Bioconjugate Chem.* **2020**, *31*, 1156–1166.

[252] D. I. Chan, H. J. Vogel, *Current understanding of fatty acid biosynthesis and the acyl carrier protein*. *Biochem. J.* **2010**, *430*, 1–19.

[253] M. R. Parsek, D. L. Val, B. L. Hanzelka, J. E. Cronan, E. P. Greenberg, *Acyl homoserine-lactone quorum-sensing signal generation*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4360–4365.

[254] C. R. H. Raetz, C. M. Reynolds, M. S. Trent, R. E. Bishop, *Lipid A Modification Systems in Gram-Negative Bacteria*. *Annu. Rev. Biochem.* **2007**, *76*, 295–329.
[255] S. T. Rutherford, B. L. Bassler, *Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control*. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a012427.
[256] W. C. Fuqua, S. C. Winans, E. P. Greenberg, *Quorum Sensing in Bacteria: the LuxR-LuxI Family of Cell Density-Responsive Transcriptional Regulatorst*. **1994**, *176*, 269–275.
[257] W.-L. Ng, B. L. Bassler, *Bacterial Quorum-Sensing Network Architectures*. *Annu. Rev. Genet.* **2009**, *43*, 197–222.
[258] P. Williams, M. Cámara, *Quorum sensing and environmental adaptation in Pseudomonas aeruginosa: a tale of regulatory networks and multifunctional signal molecules*. *Curr. Opin. Microbiol.* **2009**, *12*, 182–191.
[259] R. P. Novick, E. Geisinger, *Quorum Sensing in Staphylococci*. *Annu. Rev. Genet.* **2008**, *42*, 541–564.
[260] K. H. Nealson, J. W. Hastings, *Bacterial bioluminescence: its control and ecological significance. Microbiol. Rev.* **1979**, *43*, 496–518.
[261] N. A. Whitehead, A. M. L. Barnard, H. Slater, N. J. L. Simpson, G. P. C. Salmond, *Quorum-sensing in Gram-negative bacteria*. *FEMS Microbiol. Rev.* **2001**, *25*, 365–404.
[262] C. M. Waters, B. L. Bassler, *The Vibrio harveyi quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers*. *Genes Dev.* **2006**, *20*, 2754–2767.
[263] D. G. Lloyd, D. E. Whitworth, *The Myxobacterium Myxococcus xanthus Can Sense and Respond to the Quorum Signals Secreted by Potential Prey Organisms*. *Front. Microbiol.* **2017**, *8*, 1–7.
[264] J. Engebrecht, K. Nealson, M. Silverman, *Bacterial bioluminescence: Isolation and genetic analysis of functions from Vibrio fischeri*. *Cell* **1983**, *32*, 773–781.
[265] B. L. Hanzelka, E. P. Greenberg, *Quorum sensing in Vibrio fischeri: evidence that S-adenosylmethionine is the amino acid substrate for autoinducer synthesis. J. Bacteriol.* **1996**, *178*, 5291–5294.
[266] M. I. More, L. D. Finger, J. L. Stryker, C. Fuqua, A. Eberhard, S. C. Winans, *Enzymatic Synthesis of a Quorum-Sensing Autoinducer Through Use of Defined Substrates*. *Science* **1996**, *272*, 1655–1658.
[267] M. E. A. Churchill, L. Chen, *Structural Basis of Acyl-homoserine Lactone-Dependent Signaling*. *Chem. Rev.* **2011**, *111*, 68–85.
[268] M. E. A. Churchill, H. M. Sibhatu, C. L. Uhlson, in *Quor. Sens.* (Ed.: K.P. Rumbaugh), Humana Press, Totowa, NJ, **2011**, pp. 159–171.
[269] W. T. Watson, T. D. Minogue, D. L. Val, S. B. von Bodman, M. E. A. Churchill, *Structural Basis and Specificity of Acyl-Homoserine Lactone Signal Production in Bacterial Quorum Sensing*. *Mol. Cell* **2002**, *9*, 685–694.
[270] M. J. Bottomley, E. Muraglia, R. Bazzo, A. Carfi, *Molecular Insights into Quorum Sensing in the Human Pathogen* Pseudomonas aeruginosa *from the Structure of the Virulence Regulator LasR Bound to Its Autoinducer*. *J. Biol. Chem.* **2007**, *282*, 13592–13600.
[271] A. Vannini, *The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA*. *EMBO J.* **2002**, *21*, 4393–4401.

[272] C.-S. Tsai, S. C. Winans, *LuxR-type quorum-sensing regulators that are detached from common scents: LuxR-type apoproteins*. *Mol. Microbiol.* **2010**, *77*, 1072–1082.
[273] J. H. Devine, G. S. Shadel, T. O. Baldwin, *Identification of the operator of the lux regulon from the Vibrio fischeri strain ATCC7744. Proc. Natl. Acad. Sci.* USA **1989**, *86*, 5688–5692.
[274] J. Zhu, S. C. Winans, *Autoinducer binding by the quorum-sensing regulator TraR increases affinity for target promoters in vitro and decreases TraR turnover rates in whole cells*. *Proc. Natl. Acad. Sci.* USA **1999**, *96*, 4832–4837.
[275] K. A. Egland, E. P. Greenberg, *Quorum sensing in Vibrio fischeri: elements of the luxI promoter*. *Mol. Microbiol.* **1999**, *31*, 1197–1204.
[276] S. B. von Bodman, D. R. Majerczak, D. L. Coplin, *A negative regulator mediates quorum-sensing control of exopolysaccharide production in Pantoea stewartii subsp. stewartii*. *Proc. Natl. Acad. Sci.* USA **1998**, *95*, 7687–7692.
[277] G. C. Whitlock, D. Mark Estes, A. G. Torres, *Glanders: off to the races with* Burkholderia mallei. *FEMS Microbiol. Lett.* **2007**, *277*, 115–122.
[278] B. A. Duerkop, R. L. Ulrich, E. P. Greenberg, *Octanoyl-Homoserine Lactone Is the Cognate Signal for Burkholderia mallei BmaR1-BmaI1 Quorum Sensing*. *J. Bacteriol.* **2007**, *189*, 5034–5040.
[279] B. A. Duerkop, J. P. Herman, R. L. Ulrich, M. E. A. Churchill, E. P. Greenberg, *The Burkholderia mallei BmaR3-BmaI3 quorum-sensing system produces and responds to N-3-hydroxy-octanoyl homoserine lactone*. *J. Bacteriol.* **2008**, *190*, 5137–5141.
[280] A. N. Montebello, R. M. Brecht, R. D. Turner, M. Ghali, X. Pu, R. Nagarajan, *Acyl-ACP Substrate Recognition in* Burkholderia mallei *BmaI1 Acyl-Homoserine Lactone Synthase*. *Biochemistry* **2014**, *53*, 6231–6242.
[281] C. Strehl, L. Ehlers, T. Gaber, F. Buttgereit, *Glucocorticoids—All-Rounders Tackling the Versatile Players of the Immune System*. *Front. Immunol.* **2019**, *10*, 1744.
[282] V. F. Panoulas, K. M. J. Douglas, A. Stavropoulos-Kalinoglou, G. S. Metsios, P. Nightingale, M. D. Kita, M. S. Elisaf, G. D. Kitas, *Long-term exposure to medium-dose glucocorticoid therapy associates with hypertension in patients with rheumatoid arthritis*. *Rheumatology* **2008**, *47*, 72–75.
[283] T. R. Dykman, O. S. Gluck, W. A. Murphy, T. J. Hahn, B. H. Hahn, *Evaluation of factors associated with glucocorticoid-induced osteopenia in patients with rheumatic diseases*. *Arthritis Rheum.* **1985**, *28*, 361–368.
[284] J. Widdifield, S. Bernatsky, J. M. Paterson, N. Gunraj, J. C. Thorne, J. Pope, A. Cividino, C. Bombardier, *Serious infections in a population-based cohort of 86,039 seniors with rheumatoid arthritis*. *Arthritis Care Res.* **2013**, *65*, 353–361.
[285] J. M. Piper, *Corticosteroid Use and Peptic Ulcer Disease: Role of Nonsteroidal Anti-inflammatory Drugs*. *Ann. Intern. Med.* **1991**, *114*, 735–740.
[286] K. Onda, M. Nagashima, Y. Kawakubo, S. Inoue, T. Hirano, K. Oka, *Mitogen-activated protein kinase kinase 1/extracellular signal-regulated kinase (MEK-1/ERK) inhibitors sensitize reduced glucocorticoid response mediated by TNFα in human epidermal keratinocytes (HaCaT)*. *Biochem. Biophys. Res. Commun.* **2006**, *351*, 266–272.

[287] E. Carceller, M. Ballegeer, J. Deckers, C. Riccardi, S. Bruscoli, T. Hochepied, C. Libert, P. Pérez, *Overexpression of Glucocorticoid-induced Leucine Zipper (GILZ) increases susceptibility to Imiquimod-induced psoriasis and involves cutaneous activation of TGF-β1*. *Sci. Rep.* **2016**, *6*, 38825.

[288] S. Bruscoli, D. Sorcini, S. Flamini, A. Gagliardi, F. Adamo, S. Ronchetti, G. Migliorati, O. Bereshchenko, C. Riccardi, *Glucocorticoid-Induced Leucine Zipper Inhibits Interferon-Gamma Production in B Cells and Suppresses Colitis in Mice*. *Front. Immunol.* **2018**, *9*, 1720.

[289] P. J. Barnes, *Anti-inflammatory Actions of Glucocorticoids: Molecular Mechanisms*. *Clin. Sci.* **1998**, *94*, 557–572.

[290] R. M. Sapolsky, L. M. Romero, A. U. Munck, *How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions*. **2000**, *21*, 55–89.

[291] R. H. Oakley, J. A. Cidlowski, *The biology of the glucocorticoid receptor: New signaling mechanisms in health and disease*. *J. Allergy Clin. Immunol.* **2013**, *132*, 1033–1044.

[292] O. Bereshchenko, G. Migliorati, S. Bruscoli, C. Riccardi, *Glucocorticoid-Induced Leucine Zipper: A Novel Anti-inflammatory Molecule*. *Front. Pharmacol.* **2019**, *10*, 308.

[293] A. E. Coutinho, K. E. Chapman, *The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights*. *Mol. Cell. Endocrinol.* **2011**, *335*, 2–13.

[294] S. R. Lee, H. K. Kim, J. B. Youm, L. A. Dizon, I. S. Song, S. H. Jeong, D. Y. Seo, K. S. Ko, B. D. Rhee, N. Kim, et al., *Non-genomic effect of glucocorticoids on cardiovascular system*. *Pflüg. Arch. - Eur. J. Physiol.* **2012**, *464*, 549–559.

[295] F. D'Adamio, O. Zollo, R. Moraca, E. Ayroldi, S. Bruscoli, A. Bartoli, L. Cannarile, G. Migliorati, C. Riccardi, *A New Dexamethasone-Induced Gene of the Leucine Zipper Family Protects T Lymphocytes from TCR/CD3-Activated Cell Death*. *Immunity* **1997**, *7*, 803–812.

[296] O. Bereshchenko, M. Coppo, S. Bruscoli, M. Biagioli, M. Cimino, T. Frammartino, D. Sorcini, A. Venanzi, M. Di Sante, C. Riccardi, *GILZ Promotes Production of Peripherally Induced Treg Cells and Mediates the Crosstalk between Glucocorticoids and TGF-β Signaling*. *Cell Rep.* **2014**, *7*, 464–475.

[297] S. Ronchetti, G. Migliorati, C. Riccardi, *GILZ as a Mediator of the Anti-Inflammatory Effects of Glucocorticoids*. *Front. Endocrinol.* **2015**, *6*, 170.

[298] E. Ayroldi, A. Macchiarulo, C. Riccardi, *Targeting glucocorticoid side effects: selective glucocorticoid receptor modulator or glucocorticoid-induced leucine zipper? A perspective*. *FASEB J.* **2014**, *28*, 5055–5070.

[299] B. D. Marco, M. Massetti, S. Bruscoli, A. Macchiarulo, R. D. Virgilio, E. Velardi, V. Donato, G. Migliorati, C. Riccardi, *Glucocorticoid-induced leucine zipper (GILZ)/NF- B interaction: role of GILZ homo-dimerization and C-terminal domain*. *Nucleic Acids Res.* **2006**, *35*, 517–528.

[300] P. R. Mittelstadt, J. D. Ashwell, *Inhibition of AP-1 by the Glucocorticoid-inducible Protein GILZ*. *J. Biol. Chem.* **2001**, *276*, 29603–29610.

[301] E. Ayroldi, O. Zollo, A. Macchiarulo, B. Di Marco, C. Marchetti, C. Riccardi, *Glucocorticoid-Induced Leucine Zipper Inhibits the Raf-Extracellular Signal-Regulated Kinase Pathway by Binding to Raf-1*. *Mol. Cell. Biol.* **2002**, *22*, 7929–7941.

[302] M.-L. Asselin-Labat, M. David, A. Biola-Vidamment, D. Lecoeuche, M.-C. Zennaro, J. Bertoglio, M. Pallardy, *GILZ, a new target for the transcription factor FoxO3, protects T lymphocytes from interleukin-2 withdrawal–induced apoptosis*. *Blood* **2004**, *104*, 215–223.

[303] L. Cannarile, D. V. Delfino, S. Adorisio, C. Riccardi, E. Ayroldi, *Implicating the Role of GILZ in Glucocorticoid Modulation of T-Cell Activation*. *Front. Immunol.* **2019**, *10*, 1823.

[304] S. Bruscoli, M. Biagioli, D. Sorcini, T. Frammartino, M. Cimino, P. Sportoletti, E. Mazzon, O. Bereshchenko, C. Riccardi, *Lack of glucocorticoid-induced leucine zipper (GILZ) deregulates B-cell survival and results in B-cell lymphocytosis in mice*. *Blood* **2015**, *126*, 1790–1801.

[305] E. Ayroldi, O. Zollo, A. Bastianelli, C. Marchetti, M. Agostini, R. Di Virgilio, C. Riccardi, *GILZ mediates the antiproliferative activity of glucocorticoids by negative regulation of Ras signaling*. *J. Clin. Invest.* **2007**, *117*, 1605–1615.

[306] E. Ayroldi, G. Migliorati, S. Bruscoli, C. Marchetti, O. Zollo, L. Cannarile, F. D'Adamio, C. Riccardi, *Modulation of T-cell activation by the glucocorticoid-induced leucine zipper factor via inhibition of nuclear factor. Blood*. **2001**, *98*, 743–753.

[307] S. Bruscoli, V. Donato, E. Velardi, M. Di Sante, G. Migliorati, R. Donato, C. Riccardi, *Glucocorticoid-induced Leucine Zipper (GILZ) and Long GILZ Inhibit Myogenic Differentiation and Mediate Anti-myogenic Effects of Glucocorticoids*. *J. Biol. Chem.* **2010**, *285*, 10385–10396.

[308] E. Ayroldi, G. Migliorati, L. Cannarile, R. Moraca, D. V. Delfino, C. Riccardi, *CD2 rescues T cells from T-cell receptor/CD3 apoptosis: a role for the Fas/Fas-L system*. *Blood* **1997**, *89*, 3717–3726.

[309] K. Itoh, T. Chiba, S. Takahashi, T. Ishii, K. Igarashi, Y. Katoh, T. Oyake, N. Hayashi, K. Satoh, I. Hatayama, et al., *An Nrf2/Small Maf Heterodimer Mediates the Induction of Phase II Detoxifying Enzyme Genes through Antioxidant Response Elements*. *Biochem. Biophys. Res. Commun.* **1997**, *236*, 313–322.

[310] H. Okawa, H. Motohashi, A. Kobayashi, H. Aburatani, T. W. Kensler, M. Yamamoto, *Hepatocyte-specific deletion of the keap1 gene activates Nrf2 and confers potent resistance against acute drug toxicity*. *Biochem. Biophys. Res. Commun.* **2006**, *339*, 79–88.

[311] M. S. Yates, Q. T. Tran, P. M. Dolan, W. O. Osburn, S. Shin, C. C. McCulloch, J. B. Silkworth, K. Taguchi, M. Yamamoto, C. R. Williams, et al., *Genetic versus chemoprotective activation of Nrf2 signaling: overlapping yet distinct gene expression profiles between Keap1 knockout and triterpenoid-treated mice*. *Carcinogenesis* **2009**, *30*, 1024–1031.

[312] K. R. Sekhar, X. X. Yan, M. L. Freeman, *Nrf2 degradation by the ubiquitin proteasome pathway is inhibited by KIAA0132, the human homolog to INrf2*. *Oncogene* **2002**, *21*, 6829–6834.

[313] K. Itoh, N. Wakabayashi, Y. Katoh, T. Ishii, K. Igarashi, J. D. Engel, M. Yamamoto, *Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain*. *Genes Dev.* **1999**, *13*, 76–86.

[314] K. R. Sekhar, G. Rachakonda, M. L. Freeman, *Cysteine-based regulation of the CUL3 adaptor protein Keap1*. *Toxicol. Appl. Pharmacol.* **2010**, *244*, 21–26.

[315] N. M. Johnson, P. A. Egner, V. K. Baxter, M. B. Sporn, R. S. Wible, T. R. Sutter, J. D. Groopman, T. W. Kensler, B. D. Roebuck, *Complete Protection against Aflatoxin B1-Induced Liver Cancer with a Triterpenoid: DNA Adduct Dosimetry, Molecular Signature, and Genotoxicity Threshold*. *Cancer Prev. Res.* **2014**, *7*, 658–665.

[316] A. Zehir, R. Benayed, R. H. Shah, A. Syed, S. Middha, H. R. Kim, P. Srinivasan, J. Gao, D. Chakravarty, S. M. Devlin, et al., *Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients*. *Nat. Med.* **2017**, *23*, 703–713.

[317] Y. R. Kim, J. E. Oh, M. S. Kim, M. R. Kang, S. W. Park, J. Y. Han, H. S. Eom, N. J. Yoo, S. H. Lee, *Oncogenic* NRF2 *mutations in squamous cell carcinomas of oesophagus and skin: NRF2 in squamous cell carcinomas*. *J. Pathol.* **2010**, *220*, 446–451.

[318] K. Taguchi, M. Yamamoto, *The KEAP1–NRF2 System in Cancer*. *Front. Oncol.* **2017**, *7*, 85.

[319] A. Uruno, Y. Furusawa, Y. Yagishita, T. Fukutomi, H. Muramatsu, T. Negishi, A. Sugawara, T. W. Kensler, M. Yamamoto, *The Keap1-Nrf2 System Prevents Onset of Diabetes Mellitus*. *Mol. Cell. Biol.* **2013**, *33*, 2996–3010.

[320] A. Uruno, Y. Yagishita, M. Yamamoto, *The Keap1–Nrf2 system and diabetes mellitus*. *Arch. Biochem. Biophys.* **2015**, *566*, 76–84.

[321] K. Yoh, A. Hirayama, K. Ishizaki, A. Yamada, M. Takeuchi, S. Yamagishi, N. Morito, T. Nakano, M. Ojima, H. Shimohata, et al., *Hyperglycemia induces oxidative and nitrosative stress and increases renal functional impairment in Nrf2-deficient mice*. *Genes Cells* **2008**, *13*, 1159–1170.

[322] R. Rodrigo, J. González, F. Paoletto, *The role of oxidative stress in the pathophysiology of hypertension*. *Hypertens. Res.* **2011**, *34*, 431–440.

[323] H.-Y. Cho, S. R. Kleeberger, *Nrf2 protects against airway disorders*. *Toxicol. Appl. Pharmacol.* **2010**, *244*, 43–56.

[324] H.-Y. Cho, S. P. M. Reddy, M. Yamamoto, S. R. Kleeberger, *The transcription factor NRF2 protects against pulmonary fibrosis*. *FASEB J.* **2004**, *18*, 1258–1260.

[325] N. Kikuchi, Y. Ishii, Y. Morishima, Y. Yageta, N. Haraguchi, K. Itoh, M. Yamamoto, N. Hizawa, *Nrf2 protects against pulmonary fibrosis by regulating the lung oxidant level and Th1/Th2 balance*. *Respir. Res.* **2010**, *11*, 31.

[326] M. Pajares, A. Cuadrado, A. I. Rojo, *Modulation of proteostasis by transcription factor NRF2 and impact in neurodegenerative diseases*. *Redox Biol.* **2017**, *11*, 543–553.

[327] D. A. Johnson, J. A. Johnson, *Nrf2—a therapeutic target for the treatment of neurodegenerative diseases*. *Free Radic. Biol. Med.* **2015**, *88*, 253–267.

[328] A. Cuadrado, P. Moreno-Murciano, J. Pedraza-Chaverri, *The transcription factor Nrf2 as a new therapeutic target in Parkinson's disease*. *Expert Opin. Ther. Targets* **2009**, *13*, 319–329.

[329] A. Cuadrado, G. Manda, A. Hassan, M. J. Alcaraz, C. Barbas, A. Daiber, P. Ghezzi, R. León, M. G. López, B. Oliva, et al., *Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach*. *Pharmacol. Rev.* **2018**, *70*, 348–383.

[330] P. Canning, C. D. O. Cooper, T. Krojer, J. W. Murray, A. C. W. Pike, A. Chaikuad, T. Keates, C. Thangaratnarajah, V. Hojzan, B. D. Marsden, et al., *Structural Basis for Cul3 Protein Assembly with the BTB-Kelch Family of E3 Ubiquitin Ligases*. *J. Biol. Chem.* **2013**, *288*, 7803–7814.

[331] A. Cleasby, J. Yon, P. J. Day, C. Richardson, I. J. Tickle, P. A. Williams, J. F. Callahan, R. Carr, N. Concha, J. K. Kerns, et al., *Structure of the BTB Domain of Keap1 and Its Interaction with the Triterpenoid Antagonist CDDO*. *PLoS ONE* **2014**, *9*, e98896.

[332] K. I. Tong, Y. Katoh, H. Kusunoki, K. Itoh, T. Tanaka, M. Yamamoto, *Keap1 Recruits Neh2 through Binding to ETGE and DLG Motifs: Characterization of the Two-Site Molecular Recognition Model*. *Mol. Cell. Biol.* **2006**, *26*, 2887–2900.

[333] S.-C. Lo, X. Li, M. T. Henzl, L. J. Beamer, M. Hannink, *Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling*. *EMBO J.* **2006**, *25*, 3605–3617.

[334] B. Padmanabhan, K. I. Tong, T. Ohta, Y. Nakamura, M. Scharlock, M. Ohtsuji, M.-I. Kang, A. Kobayashi, S. Yokoyama, M. Yamamoto, *Structural Basis for Defects of Keap1 Activity Provoked by Its Point Mutations in Lung Cancer*. *Mol. Cell* **2006**, *21*, 689–700.

[335] X. Li, D. Zhang, M. Hannink, L. J. Beamer, *Crystal Structure of the Kelch Domain of Human Keap1*. *J. Biol. Chem.* **2004**, *279*, 54750–54758.

[336] L. Hu, S. Magesh, L. Chen, L. Wang, T. A. Lewis, Y. Chen, C. Khodier, D. Inoyama, L. J. Beamer, T. J. Emge, et al., *Discovery of a small-molecule inhibitor and cellular probe of Keap1–Nrf2 protein–protein interaction*. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3039–3043.

[337] E. Jnoff, C. Albrecht, J. J. Barker, O. Barker, E. Beaumont, S. Bromidge, F. Brookfield, M. Brooks, C. Bubert, T. Ceska, et al., *Binding Mode and Structure-Activity Relationships around Direct Inhibitors of the Nrf2-Keap1 Complex*. *ChemMedChem* **2014**, *9*, 699–705.

[338] L. Hu, S. Magesh, L. Chen, L. Wang, T. A. Lewis, Y. Chen, C. Khodier, D. Inoyama, L. J. Beamer, T. J. Emge, et al., *Discovery of a small-molecule inhibitor and cellular probe of Keap1–Nrf2 protein–protein interaction*. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3039–3043.

[339] D. Dubey, B. C. Kieseier, H. P. Hartung, B. Hemmer, C. Warnke, T. Menge, W. A. Miller-Little, O. Stuve, *Dimethyl fumarate in relapsing–remitting multiple sclerosis: rationale, mechanisms of action, pharmacokinetics, efficacy and safety*. *Expert Rev. Neurother.* **2015**, *15*, 339–346.

[340] S. Hörer, D. Reinert, K. Ostmann, Y. Hoevels, H. Nar, *Crystal-contact engineering to obtain a crystal form of the Kelch domain of human Keap1 suitable for ligand-soaking experiments*. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **2013**, *69*, 592–596.

[341] J. H. Carter, J. A. Deddens, N. R. Spaulding IV, D. Lucas, B. M. Colligan, T. G. Lewis, E. Hawkins, J. Jones, J. O. Pemberton, L. E. Douglass, et al., *Phosphorylation of eIF4E serine 209 is associated with tumour progression and reduced survival in malignant melanoma*. *Br. J. Cancer* **2016**, *114*, 444–453.

[342] C.-C. Wu, S.-C. Hsu, H. -m. Shih, M.-Z. Lai, *Nuclear Factor of Activated T Cells c Is a Target of p38 Mitogen-Activated Protein Kinase in T Cells*. *Mol. Cell. Biol.* **2003**, *23*, 6442–6454.

[343] C. R. Beals, C. M. Sheridan, C. W. Turck, P. Gardner, G. R. Crabtree, *Nuclear Export of NF-ATc Enhanced by Glycogen Synthase Kinase-3*. *Science* **1997**, *28*, 1930–1934.

[344] A. J. Waskiewicz, *Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2*. *EMBO J.* **1997**, *16*, 1909–1920.

[345] A. Martínez, M. Sesé, J. H. Losa, N. Robichaud, N. Sonenberg, T. Aasen, S. Ramón y Cajal, *Phosphorylation of eIF4E Confers Resistance to Cellular Stress and DNA-Damaging Agents through an Interaction with 4E-T: A Rationale for Novel Therapeutic Approaches*. *PLOS ONE* **2015**, *10*, e0123352.

[346] L. McKendrick, S. J. Morley, V. M. Pain, R. Jagus, B. Joshi, *Phosphorylation of eukaryotic initiation factor 4E (eIF4E) at Ser209 is not required for protein synthesis* in vitro *and* in vivo*: eIF4E phosphorylation not required for translation*. *Eur. J. Biochem.* **2001**, *268*, 5375–5385.

[347] Y. Nakamura, *Codon usage tabulated from international DNA sequence databases: status for the year 2000*. *Nucleic Acids Res.* **2000**, *28*, 292–292.

[348] H.-S. Park, M. J. Hohn, T. Umehara, L.-T. Guo, E. M. Osborne, J. Benner, C. J. Noren, J. Rinehart, D. Soll, *Expanding the Genetic Code of Escherichia coli with Phosphoserine*. *Science* **2011**, *333*, 1151–1154.

[349] D. Shin, C. Gorgulla, M. E. Boursier, N. Rexrode, E. C. Brown, H. Arthanari, H. E. Blackwell, R. Nagarajan, N- *Acyl Homoserine Lactone Analog Modulators of the* Pseudomonas aeruginosa *Rhll Quorum Sensing Signal Synthase*. *ACS Chem. Biol.* **2019**, *14*, 2305–2314.

[350] D. H. Keating, M. R. Carey, J. E. Cronan, *The Unmodified (Apo) Form of* Escherichia coli *Acyl Carrier Protein Is a Potent Inhibitor of Cell Growth*. *J. Biol. Chem.* **1995**, *270*, 22229–22235.

[351] C. O. Rock, J. E. Cronan, *Improved purification of acyl carrier protein*. *Anal. Biochem.* **1980**, *102*, 362–364.

[352] J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel, M. D. Burkart, *The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life*. *Nat. Prod. Rep.* **2014**, *31*, 61–108.

[353] I. Lemmens, S. Lievens, J. Tavernier, in *Protein-Protein Interact.* (Eds.: C.L. Meyerkord, H. Fu), Springer New York, New York, NY, **2015**, pp. 447–455.

[354] R. S. Bohacek, C. McMartin, W. C. Guida, *The art and practice of structure-based drug design: A molecular modeling perspective*. *Med. Res. Rev.* **1996**, *16*, 3–50.

[355] J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O'Meara, T. Che, E. Algaa, K. Tolmachova, et al., *Ultra-large library docking for discovering new chemotypes*. *Nature* **2019**, *566*, 224–229.

[356] R. Sprangers, L. E. Kay, *Quantitative dynamics and binding studies of the 20S proteasome by NMR*. *Nature* **2007**, *445*, 618–622.

[357] V. Tugarinov, L. E. Kay, *Side Chain Assignments of Ile δ1 Methyl Groups in High Molecular Weight Proteins: an Application to a 46 ns Tumbling Molecule*. *J. Am. Chem. Soc.* **2003**, *125*, 5701–5706.

[358] K. H. Gardner, M. K. Rosen, L. E. Kay, *Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR* †. *Biochemistry* **1997**, *36*, 1389–1401.

[359] R. Linser, V. Gelev, F. Hagn, H. Arthanari, S. G. Hyberts, G. Wagner, *Selective Methyl Labeling of Eukaryotic Membrane Proteins Using Cell-Free Expression*. *J. Am. Chem. Soc.* **2014**, *136*, 11308–11310.

[360] N. Sibille, X. Hanoulle, F. Bonachera, D. Verdegem, I. Landrieu, J.-M. Wieruszeski, G. Lippens, *Selective backbone labelling of ILV methyl labelled proteins*. *J. Biomol. NMR* **2009**, *43*, 219–227.

[361] C.-W. Chow, R. J. Davis, *Integration of Calcium and Cyclic AMP Signaling Pathways by 14-3-3*. *Mol. Cell. Biol.* **2000**, *20*, 702–712.

[362] P. Dziekański, K. Grudziąż, P. Jarvoll, W. Koźmiński, A. Zawadzka-Kazimierczuk, *13C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing*. *J. Biomol. NMR* **2015**, *62*, 179–190.

[363] W. Bermel, I. Bertini, I. C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, J. Stanek, *Speeding up sequence specific assignment of IDPs*. *J. Biomol. NMR* **2012**, *53*, 293–301.

[364] Y. Yoshimura, N. V. Kulminskaya, F. A. A. Mulder, *Easy and unambiguous sequential assignments of intrinsically disordered proteins by correlating the backbone 15N or 13C' chemical shifts of multiple contiguous residues in highly resolved 3D spectra*. *J. Biomol. NMR* **2015**, *61*, 109–121.

[365] D. Sahu, M. Bastidas, S. A. Showalter, *Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods*. *Anal. Biochem.* **2014**, *449*, 17–25.

[366] I. G. Denisov, Y. V. Grinkova, A. A. Lazarides, S. G. Sligar, *Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size*. *J. Am. Chem. Soc.* **2004**, *126*, 3477–3487.

[367] C. R. H. Raetz, *Molecular genetics of membrane phospholipid synthesis*. *Annu. Rev. Genet.* **1986**, *20*, 253–295.

[368] D. Widdick, S. F. Royer, H. Wang, N. M. Vior, J. P. Gomez-Escribano, B. G. Davis, M. J. Bibb, *Analysis of the Tunicamycin Biosynthetic Gene Cluster of* Streptomyces chartreusis *Reveals New Insights into Tunicamycin Production and Immunity*. *Antimicrob. Agents Chemother.* **2018**, *62*, e00130-18.

[369] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, L. J. Skelton, Protein NMR Spectroscopy. Principles and Practice. *Protein NMR Spectroscopy. Principles and Practice*, Academic Press, **1996**.

[370] H. Y. Carr, E. M. Purcell, *Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments*. *Phys. Rev.* **1954**, *94*, 630–638.

[371] S. Meiboom, D. Gill, *Modified Spin-Echo Method for Measuring Nuclear Relaxation Times*. *Rev. Sci. Instrum.* **1958**, *29*, 688–691.

[372] I. Solomon, *Relaxation Processes in a System of Two Spins*. *Phys. Rev.* **1955**, *99*, 559–565.

[373] P. Neudecker, P. Lundström, L. E. Kay, *Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding*. *Biophys. J.* **2009**, *96*, 2045–2054.
[374] S. Grutsch, S. Brüschweiler, M. Tollinger, *NMR Methods to Study Dynamic Allostery*. *PLOS Comput. Biol.* **2016**, *12*, e1004620.
[375] E. Mühlberger, M. Weik, V. E. Volchkov, H.-D. Klenk, S. Becker, *Comparison of the Transcription and Replication Strategies of Marburg Virus and Ebola Virus by Using Artificial Replication Systems*. *J. Virol.* **1999**, *73*, 2333–2342.
[376] N. Biedenkopf, C. Lier, S. Becker, *Dynamic Phosphorylation of VP30 Is Essential for Ebola Virus Life Cycle*. *J. Virol.* **2016**, *90*, 4914–4925.
[377] T. Kruse, N. Biedenkopf, E. P. T. Hertz, E. Dietzel, G. Stalmann, B. López-Méndez, N. E. Davey, J. Nilsson, S. Becker, *The Ebola Virus Nucleoprotein Recruits the Host PP2A-B56 Phosphatase to Activate Transcriptional Support Activity of VP30*. *Mol. Cell* **2018**, *69*, 136–145.
[378] E. P. T. Hertz, T. Kruse, N. E. Davey, B. López-Méndez, J. O. Sigurðsson, G. Montoya, J. V. Olsen, J. Nilsson, *A Conserved Motif Provides Binding Specificity to the PP2A-B56 Phosphatase*. *Mol. Cell* **2016**, *63*, 686–695.
[379] C. Ditchfield, V. L. Johnson, A. Tighe, R. Ellston, C. Haworth, T. Johnson, A. Mortlock, N. Keen, S. S. Taylor, *Aurora B couples chromosome alignment with anaphase by targeting BubR1, Mad2, and Cenp-E to kinetochores*. *J. Cell Biol.* **2003**, *161*, 267–280.
[380] C. Klebig, D. Korinth, P. Meraldi, *Bub1 regulates chromosome segregation in a kinetochore-independent manner*. *J. Cell Biol.* **2009**, *185*, 841–858.
[381] A. Musacchio, E. D. Salmon, *The spindle-assembly checkpoint in space and time*. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 379–393.
[382] M. A. Lampson, T. M. Kapoor, *The human mitotic checkpoint protein BubR1 regulates chromosome–spindle attachments*. *Nat. Cell Biol.* **2005**, *7*, 93–98.
[383] P. Meraldi, P. K. Sorger, *A dual role for Bub1 in the spindle checkpoint and chromosome congression*. *EMBO J.* **2005**, *24*, 1621–1633.
[384] J. G. DeLuca, W. E. Gall, C. Ciferri, D. Cimini, A. Musacchio, E. D. Salmon, *Kinetochore Microtubule Dynamics and Attachment Stability Are Regulated by Hec1*. *Cell* **2006**, *127*, 969–982.
[385] J. P. I. Welburn, M. Vleugel, D. Liu, J. R. Yates, M. A. Lampson, T. Fukagawa, I. M. Cheeseman, *Aurora B Phosphorylates Spatially Distinct Targets to Differentially Regulate the Kinetochore-Microtubule Interface*. *Mol. Cell* **2010**, *38*, 383–392.
[386] T. Kruse, G. Zhang, M. S. Y. Larsen, T. Lischetti, W. Streicher, T. Kragh Nielsen, S. P. Bjorn, J. Nilsson, *Direct binding between BubR1 and B56-PP2A phosphatase complexes regulate mitotic progression*. *J. Cell Sci.* **2013**, *126*, 1086–1092.
[387] L. T. Vassilev, C. Tovar, S. Chen, D. Knezevic, X. Zhao, H. Sun, D. C. Heimbrook, L. Chen, *Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1*. *Proc. Natl. Acad. Sci.* USA **2006**, *103*, 10660–10665.
[388] P. Zhou, A. A. Lugovskoy, G. Wagner, *A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins*. *J. Biomol. NMR* **2001**, *20*, 11–14.
[389] C.-R. Lee, Y.-H. Park, Y.-R. Kin, A. Peterkofsky, Y.-J. Seok, *Phosphorylation-Dependent Mobility Shift of Proteins on SDS-PAGE is Due to Decreased Binding of SDS*. *Bull. Korean Chem. Soc.* **2013**, *34*, 2063–2066.

[390] Z.-H. Peng, V. Sharma, S. F. Singleton, P. D. Gershon, *Synthesis and Application of a Chain-Terminating Dinucleotide mRNA Cap Analog*. *Org. Lett.* **2002**, *4*, 161–164.

[391] S. C. Devarkar, C. Wang, M. T. Miller, A. Ramanathan, F. Jiang, A. G. Khan, S. S. Patel, J. Marcotrigiano, *Structural basis for m7G recognition and 2′-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I*. *Proc. Natl. Acad. Sci.* USA **2016**, *113*, 596–601.

[392] R. Züst, L. Cervantes-Barragan, M. Habjan, R. Maier, B. W. Neuman, J. Ziebuhr, K. J. Szretter, S. C. Baker, W. Barchet, M. S. Diamond, et al., *Ribose 2′-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5*. *Nat. Immunol.* **2011**, *12*, 137–143.

[393] C. Schuberth-Wagner, J. Ludwig, A. K. Bruder, A.-M. Herzner, T. Zillinger, M. Goldeck, T. Schmidt, J. L. Schmid-Burgk, R. Kerber, S. Wolter, et al., *A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2′O-Methylated Self RNA*. *Immunity* **2015**, *43*, 41–51.

[394] S. Daffis, K. J. Szretter, J. Schriewer, J. Li, S. Youn, J. Errett, T.-Y. Lin, S. Schneller, R. Zust, H. Dong, et al., *2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members*. *Nature* **2010**, *468*, 452–456.

[395] F. Picard-Jean, C. Brand, M. Tremblay-Létourneau, A. Allaire, M. C. Beaudoin, S. Boudreault, C. Duval, J. Rainville-Sirois, F. Robert, J. Pelletier, et al., *2′-O-methylation of the mRNA cap protects RNAs from decapping and degradation by DXO*. *PLoS One* **2018**, *13*, e0193804.

[396] I. Shimada, in *Methods Enzymol.*, Elsevier, **2005**, pp. 483–506.

[397] H. Takahashi, T. Nakanishi, K. Kami, Y. Arata, I. Shimada, *A novel NMR method for determining the interfaces of large protein–protein complexes*. *Nat. Struct. Biol.* **2000**, *7*, 220–223.

[398] J. Marley, M. Lu, C. Bracken, *A method for efficient isotopic labeling of recombinant proteins*. *J. Biomol. NMR* **2001**, *20*, 71–75.

[399] X. Guo, J. G. Kempf, *Enhanced, simplified expression of perdeuterated hemoglobin for NMR structure and dynamics*. *Protein Expr. Purif.* **2010**, *72*, 8–18.

[400] O. Paliy, T. S. Gunasekera, *Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents*. *Appl. Microbiol. Biotechnol.* **2007**, *73*, 1169–1172.

[401] N. J. Greenfield, *Using circular dichroism spectra to estimate protein secondary structure*. *Nat. Protoc.* **2006**, *1*, 2876–2890.

[402] V. Tugarinov, P. M. Hwang, L. E. Kay, *Nuclear Magnetic Resonance Spectroscopy of High-Molecular-Weight Proteins*. *Annu. Rev. Biochem.* **2004**, *73*, 107–146.

[403] A. G. Tzakos, C. R. R. Grace, P. J. Lukavsky, R. Riek, *NMR techniques for very large proteins and rnas in solution*. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 319–342.

[404] H. P. Erickson, *Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy*. *Biol. Proced. Online* **2009**, *11*, 32–51.

[405] D. C. Desravines, I. Serna Martin, R. Schneider, P. J. Mas, N. Aleksandrova, M. R. Jensen, M. Blackledge, D. J. Hart, *Structural Characterization of the SMRT Corepressor Interacting with Histone Deacetylase 7*. *Sci. Rep.* **2017**, *7*, 3678.

[406] J. García-Nafría, J. F. Watson, I. H. Greger, *IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly*. *Sci. Rep.* **2016**, *6*, 27459.
[407] J. E. Cronan, J. Thomas, *Bacterial Fatty Acid Synthesis and its Relationships with Polyketide Synthetic Pathways*. *Methods Enzym.* **2009**, *459*, 395–433.
[408] S. Jackowski, *Regulation of Coenzyme A Biosynthesis*. *J. Bacteriol.* **1981**, *148*, 926–932.
[409] S. Wohnig, A. P. Spork, S. Koppermann, G. Mieskes, N. Gisch, R. Jahn, C. Ducho, *Total Synthesis of Dansylated Park's Nucleotide for High-Throughput MraY Assays*. *Chem. - Eur. J.* **2016**, *22*, 17813–17819.
[410] I. Edery, M. Altmann, N. Sonenberg, *High-level synthesis in Escherichia cofi of functional cap-binding eukaryotic initiation factor eIF-4E and affhity purification using a simplified cap-analog resin*. *Gene* **1988**, *74*, 517–525.
[411] N. J. Moerke, H. Aktas, H. Chen, S. Cantel, M. Y. Reibarkh, A. Fahmy, J. D. Gross, A. Degterev, J. Yuan, M. Chorev, et al., *Small-Molecule Inhibition of the Interaction between the Translation Initiation Factors eIF4E and eIF4G*. *Cell* **2007**, *128*, 257–267.
[412] S. G. Hyberts, K. Takeuchi, G. Wagner, *Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data*. *J. Am. Chem. Soc.* **2010**, *132*, 2145–2147.
[413] S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, G. Wagner, *Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling*. *J. Biomol. NMR* **2012**, *52*, 315–327.
[414] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *The CCPN data model for NMR spectroscopy: Development of a software pipeline*. *Proteins Struct. Funct. Bioinforma.* **2005**, *59*, 687–696.
[415] M. Bieri, E. J. d'Auvergne, P. R. Gooley, *relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and µs motion of proteins*. *J. Biomol. NMR* **2011**, *50*, 147–155.
[416] L. Whitmore, B. A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data*. *Nucleic Acids Res.* **2004**, *32*, W668–W673.