Rapid Discovery of Aspartyl Protease Inhibitors Using an Anchoring Approach

Markella Konstantinidou,[a] Francesca Magari,[b] Fandi Sutanto,[a] Jörg Haupenthal,[c] Varsha R. Jumde,[c, d] M. Yagiz Ünver,[c, d] Andreas Heine,[b] Carlos Jamie Camacho,[e] Anna K. H. Hirsch,[c, d, f] Gerhard Klebe,[b] and Alexander Dömling*[a]

Pharmacophore searches that include *anchors*, fragments contributing above average to receptor binding, combined with one-step syntheses are a powerful approach for the fast discovery of novel bioactive molecules. Here, we are presenting a pipeline for the rapid and efficient discovery of aspartyl protease inhibitors. First, we hypothesized that hydrazine could be a multi-valent warhead to interact with the active site Asp carboxylic acids. We incorporated the hydrazine anchor in a multicomponent reaction and created a large virtual library of hydrazine derivatives synthetically accessible in one-step. Next, we performed anchor-based pharmacophore screening of the libraries and resynthesized top-ranked compounds. The inhibitory potency of the molecules was finally assessed by an enzyme activity assay and the binding mode confirmed by several soaked crystal structures supporting the validity of the hypothesis and approach. The herein reported pipeline of tools will be of general value for the rapid generation of receptor binders beyond Asp proteases.

The discovery and development of novel drugs is a highly time, resource and investment-intensive undertaking with very low success rate if compared with other industrial development processes. Often it starts with a high throughput screening campaign, but the final discovery of a bioactive lead involves many different disciplines, including biochemistry, cell biology, pharmacology, structural biology and computational chemistry. Bottlenecks of early-stage discovery are often the time consuming and expensive high-throughput screening and the subsequent delineation and expansion of hits. We recently introduced a specialized pharmacophore search technology, AnchorQuery that brings interactive virtual screening of novel protein-protein interaction inhibitors to the desktop.[1,2] The technology is based upon a >30 million database of virtual compounds. Every library compound is accessible through one-step multi-component reaction (MCR) chemistry and contains an *anchor* motif that is bioisosteric to an amino-acid residue. An *anchor* is defined as an amino-acid side chain in the interface of a protein-protein interaction which is contributing above average to its energetics, for example a side chain that buries a large fraction of surface area at the core of the binding interface.[3] Anchors are usually part of energetic hot spots.[4] The value of AnchorQuery has been proven by the discovery of multiple novel and bioactive MCR scaffolds as direct or allosteric modulators of p53/MDM2[5] or PDK1.[6] The current limitation of AnchorQuery is that it was designed for small molecules mimicking amino acid side chains. However, the concept of an *anchor* combined with one-pot MCR chemistry could be useful not only in protein-protein interactions, but, as demonstrated in this report, it can be applied in other contexts such as fragment-based drug discovery. Thus, we provide here a generalized AnchorQuery pipeline of tools implemented for the discovery of novel Asp protease inhibitors (Figure 1).

[a] *Dr. M. Konstantinidou, F. Sutanto, Prof. Dr. A. Dömling*
Department of Pharmacy
Drug Design Group
University of Groningen
A. Deusinglaan 1
9713 AV Groningen (The Netherlands)
E-mail: a.s.s.domling@rug.nl
[b] *F. Magari, Prof. Dr. A. Heine, Prof. Dr. G. Klebe*
Institute of Pharmaceutical Chemistry
Philipps University Marburg
Marbacher Weg 6
35032 Marburg (Germany)
[c] *Dr. J. Haupenthal, Dr. V. R. Jumde, Dr. M. Y. Ünver, Prof. Dr. A. K. H. Hirsch*
Department of Drug Design and Optimization
Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)
Helmholtz Centre for Infection Research (HZI), Campus E8.1
66123 Saarbrücken (Germany)
[d] *Dr. V. R. Jumde, Dr. M. Y. Ünver, Prof. Dr. A. K. H. Hirsch*
Stratingh Institute for Chemistry
University of Groningen
Nijenborgh 7
9747 AG Groningen (The Netherlands)
[e] *Dr. C. J. Camacho*
Department of Computational and Systems Biology
University of Pittsburgh
3501 Fifth Avenue
Pittsburgh, PA 15260 (USA)
[f] *Prof. Dr. A. K. H. Hirsch*
Department of Pharmacy
Saarland University
Campus E8.1
66123 Saarbrücken (Germany)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cmdc.202000024

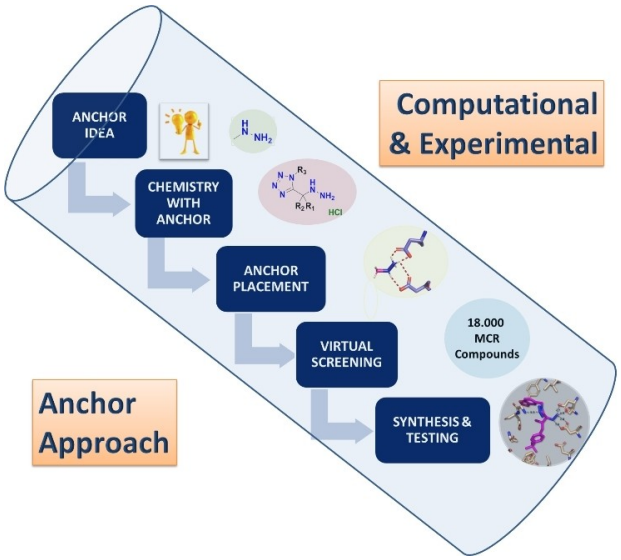

Figure 1. The overall workflow.

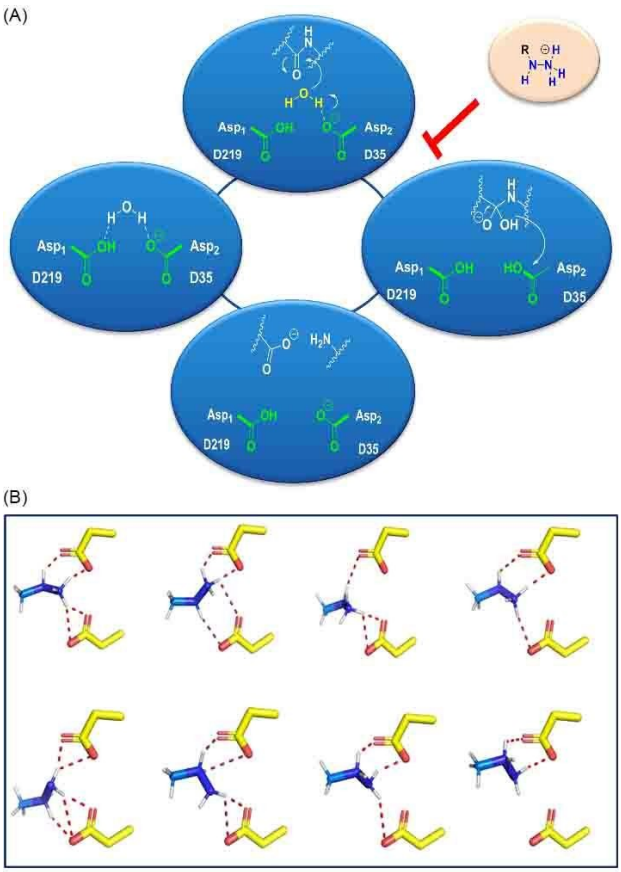

Figure 2. Hydrazine as a water mimicking warhead in Asp proteases. A) The generalized Asp protease mechanism and a hydrazine derivative as water mimic interacting with both Asp residues by hydrogen bonding and charge-charge interactions. B) Different possible binding poses of hydrazine between the two Asp of endothiapepsin.

We chose endothiapepsin as an archetypical Asp protease, which although is not a drug target per se, has received considerable attention as a relevant surrogate in drug discovery programs. Moreover, the enzyme can be easily obtained in large amounts and remains stable and active even after 20 days at room temperature.[7] The ease of crystallization, together with the considerable sequence similarity and folding architecture with related drug targets, explains its use in a hit-to-lead project for β-secretase inhibitors.[8] Interestingly, also renin inhibitors could be co-crystallized with endothiapespin, providing valuable information for the binding mode of the compounds.[9] Endothiapepsin is a monomer, with two structurally similar domains. Each domain contributes one aspartic acid to the catalytic dyad; D35 and D219 (Figure 2A). In the first step of the catalytic mechanism D35 is believed to be deprotonated, whereas D219 is protonated.[10]

Typical warheads for Asp proteases include primary and secondary amines, guanidines, amidines, hydrazides, carboxylic acids, alcohols, imidazoles and pyrazoles.[11] However, it is surprising the absence of a warhead with equal interaction to the two oxygens of an aspartic acid residue. The simplest structure in organic chemistry able to interact with two carboxylic acids bears two nitrogens, thus creating a hydrazine moiety (Figure 2B). While endothiapepsin is active in acidic pH, the hydrazine moiety has the advantage of being protonated under these conditions, thus forming ionic interactions with the carboxylic acids. NMR studies and quantum chemical calculations for alkyl- and arylhydrazines indicate that protonation is possible either with the exo- or the endo-nitrogen, providing a diverse arrangement of possible interactions (Figure 2B).[12] Hydrazine has unique attributes not present in common warheads for the potential of combined ionic and hydrogen bonds toward all four oxygen atoms of the catalytic dyad. Thus, we choose hydrazine as our warhead moiety.

We designed a scaffold that could be easily accessible with multi-component reaction chemistry (MCR) incorporating hydrazine as the warhead motif (Figure 3A).[13,14] Hydrazine is used as the amine component, in an Ugi-tetrazole reaction. The Ugi-tetrazole reaction was chosen due to shape complementarity of the scaffold with the target protein.[15] Synthetically, the scaffold is accessed in a two-step synthesis, starting from a 4-component Ugi-tetrazole reaction, followed by Boc-deprotection.[16] Diversity can be easily achieved through the oxo-component (aldehydes and ketones) and the isocyanides. The target compounds are isolated as HCl salts, due to the activity of the enzyme at acidic conditions.

Initially, we screened a small library of 17 derivatives of which five showed inhibitory activity (Figure 3B). For the biochemical evaluation, we employed a fluorescence-based assay adapted from an established HIV-protease assay.[17] Five compounds of the first set showed low to moderate inhibitory activity. In order to gain structural insights, a crystal structure for compound **3a** was obtained by soaking (Figure 5A, SI Figure S2). In this case, only the exo-nitrogen of the hydrazine

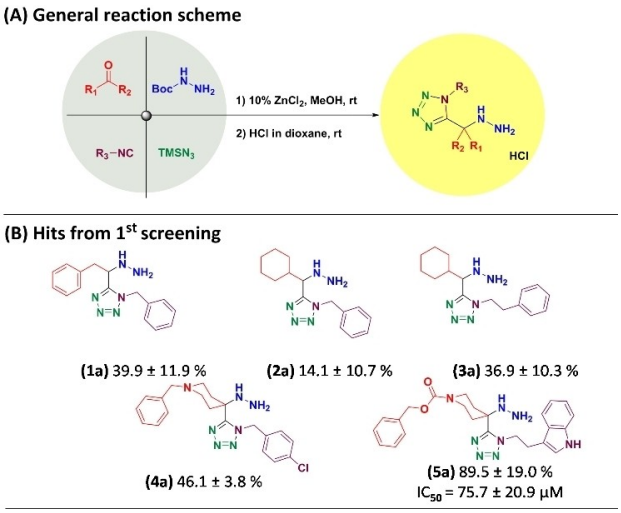

Figure 3. Chemistry and first hits. A) Design of an MCR incorporating hydrazine. B) Structures and % inhibition – IC_{50} values of initial hits (set 1).

warhead is interacting with the catalytic dyad. Interestingly, the tetrazole ring is forming a hydrogen bond with Gly80.

Next, we aimed to optimize the scaffold using the hydrazine moiety as an anchoring fragment. Thus, we developed a protocol for tailor-made virtual library screening. The workflow of this protocol has not been automated, but in contrast to AnchorQuery, there is no limitation to the design of the library, as long as the chemistry is deterministic (detailed protocol described in SI). Moreover, in contrast to public compound databases, a particular scaffold of interest can be optimized, by including commercially available starting materials.

The first step of the protocol is the enumeration of a virtual library, starting from commercially available starting materials (in this case: aldehydes and ketones). Isocyanides based on syntheses using primary amines or oxo components were included: starting from amines with the Ugi procedure[18] or from aldehydes/ketones with the Leuckart-Wallach procedure[19] or from the reaction of the glycine isocyanide (methyl 2-isocyanoacetate) with primary amines towards extended isocyanoacetamides.[20] The virtual libraries were created using Reactor[21] software including the post-modification of Boc-deprotection. In our library design, we included ~150 aldehydes/ketones and 120 isocyanides thus representing a chemical space of 18.000 possible combinations, not including stereoisomers. The Reactor-generated molecules were converted into 3D conformers using Moloc software. For the 3D anchoring of the hydrazine fragments, different protonation states and orientations between the catalytic aspartic acid residues (D35, D219) were considered and were used to position ("fix") the library against the fragments within the catalytic site. Pharmit software was used to remove clashes occurring during positioning of the library.[22] Moreover, at this stage geometrical cut-off criteria were applied, discarding molecules that clashed with the receptor. Lipinski's rule of five was applied to further filter putative candidates. A final energy minimization was performed with Moloc.[23]

Twelve optimized hits were selected, first by visually inspecting the poses and then by using the Scorpion software for quantitatively scoring the interactions.[24] In the end, the predicted compounds were synthesized and tested in the fluorescence-based assay and for the most active compounds, the IC_{50} values were determined (Figure 4).

3D structural geometries are key to understand the binding mode of the active compounds and to validate our approach regarding the docking workflow and the correlation between the docking poses and the crystal structures. We were able to obtain a crystal structure by soaking for the most active compound of the 2nd set, compound **8b** (Figure 5B, SI Figure S3). In this case, compound **8b** interacts with both the exo- and endo-nitrogens of the hydrazine warhead with the catalytic dyad. As in the case of compound **3a**, the tetrazole ring is involved in a hydrogen bond with the backbone NH of Gly80. Moreover, the benzodioxolic motif is involved in a hydrogen bond with the OH group of Tyr226. The molecule is also involved in multiple hydrophobic interactions. One more crystal structure was obtained for compound **3b** from the 2nd set (SI Figure S4). This smaller and more hydrophobic compound, although is still able to interact with the catalytic dyad, is lacking the formation of the hydrogen bond with Gly80. In the fluorescence-based assay, compound **3b** showed very low

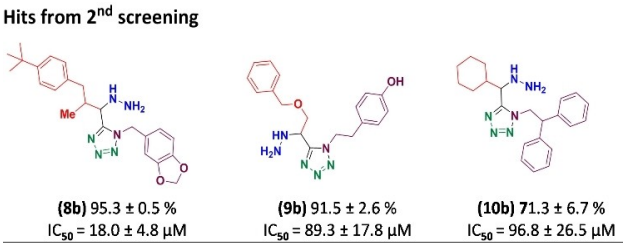

Figure 4. Structures and % inhibition – IC_{50} values optimized hits (set 2).

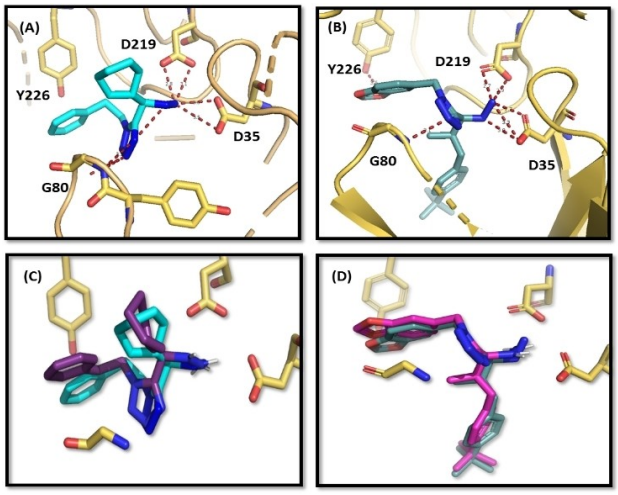

Figure 5. Structural analysis of inhibitors. A) Crystal Structures of **(3a) (PDB 6SCV)**,B) crystal structure of **(8b) (PDB 6RON)**. Hydrogen bonds are shown as red dashes. C–D). Overlap of crystal structures with predicted docking poses. For the docking with Moloc PDB 3PBZ was used as receptor.

inhibitory activity. The data from the crystal structures, together with the fluorescence-based assay results, gave valuable insight regarding the binding mode of the compounds and the structural features that are required for inhibition.

Since our aim is to evaluate the accuracy of the predictions regarding the docking workflow, we compared the obtained crystal structures with the docking poses of the compounds. In virtual screening, for each compound 10 conformers were generated (Figure 5C,D). A comparison of the crystal structures with the different docking poses showed that the overlap of the warhead was almost perfect and differences were mainly observed in the conformation of the terminal cyclohexyl ring. From the enumerated library, we immediately excluded compounds that were clashing with the receptor and we focused on compounds that had the right size and orientation to bind to the active site. Although, very weak binders, such as compound **3b** could not be excluded at this stage of the docking selection, they still provide interesting structural information for further optimization of the scaffold. It should be noted that accurate correlating of the binding poses with the biological activity is not possible and is beyond the aim of the developed workflow. However, this anchor-based approach shows how an anchor warhead can be incorporated in an MCR scaffold and be optimized without major synthetic effort.

In summary, we introduced a generalized protocol for the AnchorQuery approach which overcomes current limitations of amino-acidogenic *anchors*. *Anchors* are significantly affinity contributing fragments in protein binding and more general in receptor-ligand interactions. Thus, *anchor* fragments comprise valid starting points for growing leads that can be validated rapidly if combined with a high diversity convergent chemistry, such as MCR.

Thus, we designed an MCR scaffold with a novel warhead for aspartic proteases. In this approach, the scaffold could be accessed with a simple two-step methodology. The biological evaluation of the hits together with the determined crystal structures, indicate that the design and optimization of our libraries was successful. Although these are yet not highly potent inhibitors for this enzyme, we were able to analyze the interactions of our MCR scaffold and gained valuable insights regarding the adopted binding modes.

Moreover, the docking protocol for tailor-made virtual libraries can be applied to different chemical reactions and fragments, enabling computational evolution of libraries that are not part of public databases. The choice of the fragment-*anchor* is the determining step in this protocol and should include a sequence of atoms that are present as a common motif throughout the entire library. These atoms should significantly contribute to the binding interactions between the designed ligands and the protein. For instance, the *anchor* could be the motif binding in the enzyme's active site, whereas in protein-protein interactions, it could be a moiety deeply buried in the interface.

To the best of our knowledge, currently available docking software cannot optimize a specific scaffold/chemistry of interest by focusing on the possible combinations of commercially available starting materials. The libraries in this approach are not limited to multi-component reaction (MCR) scaffolds only but any sequence of organic reactions would work similarly. Broader chemistry schemes can be applied, including post-modifications. We envision future applications either for docking of novel scaffolds towards biological targets or for optimizing a scaffold of interest. As shown in this case study, departing from commercially available starting materials, thousands of compounds could potentially be accessed. Our protocol can significantly support the decision-making process of prioritizing docking hits as subsequent candidates for chemical synthesis and will lead to the requirement of fewer resources and in shorter times compared to strategies that still involve a significant serendipity and random trial component.

Experimental Section

See the Supporting Information for experimental details.

Acknowledgements

This research was supported (A.D., G.K., and J.C.C.) through ITN "Accelerated Early stage drug dIScovery" (AEGIS, grant agreement no. 675555). Moreover, funding was received from the US National Institutes of Health (NIH, grant 2R01GM097082-05), the European Lead Factory (IMI) (grant agreement no. 115489), the Qatar National Research Foundation (NPRP6-065-3-012), COFUNDs ALERT (grant agreement no. 665250), Prominent (grant agreement no. 754425), and KWF Kankerbestrijding grant (grant agreement no. 10504). Funding from the Helmholtz Association's Initiative and Networking Fund, COFUNDs ALERT (grant agreement No 665250) and the European Research Council (ERC starting grant 757913) is gratefully acknowledged (A.K.H.H.). Authors F.M., A.H., and G.K. thank the MX-team at BESSY II (Helmholtz-Zentrum Berlin, Germany) for their advice during data collection and particularly acknowledge the help and support of Dr. Manfred Weiss, Dr. Christian Feiler, Dr. Franziska Huschmann, and Dr. Jan Wollenhaupt. They also thank the Helmholtz-Zentrum Berlin for travel support.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: hydrazine-tetrazoles · MCR chemistry · docking protocol · aspartic protease · crystal structures

[1] D. R. Koes, A. Dömling, C. J. Camacho, *Protein Sci.* **2018**, *27(1)*,229–232.
[2] D. Koes, K. Khoury, Y. Huang, W. Wang, M. Bista, G. M. Popowicz, S. Wolf, T. A. Holak, A. Dömling, C. J. Camacho, *PLoS One* **2012**, *7(3)*, e32839.
[3] D. Rajamani, S. Thiel, S. Vajda, C. J. Camacho, *PNAS* **2004**, *101(31)*, 11287–11292.
[4] T. Clackson, J. A. Wells, *Science* **1995**, *267(5196)*, 383–386.

[5] E. Surmiak, C. G. Neochoritis, B. Musielak, A. Twarda-Clapa, K. Kurpiewska, G. Dubin, C. Camacho, T. A. Holak, A. Dömling, *Eur. J. Med. Chem.* **2017**, *126*, 384–407.
[6] E. Kroon, J. O. Schulze, E. Süß, C. J. Camacho, R. M. Biondi, A. Dömling, *Angew. Chem. Int. Ed. Engl.* **2015**,*54(47)*,13933–13936.
[7] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. Hirsch, *Angew. Chem. Int. Ed. Engl.* **2014**, *53(12)*, 3259–3263.
[8] S. Geschwindner, L. L. Olsson, J. S. Albert, J. Deinum, P. D. Edwards, T. de Beer, R. H. Folmer, *J. Med. Chem.* **2007**, *50(24)*, 5903–5911.
[9] B. Veerapandian, J. B. Cooper, A. Sali, T. L. Blundell, *J. Mol. Biol.* **1990**, *216(4)*, 1017–1029.
[10] a) L. Coates, H. F. Tuan, S. Tomanicek, A. Kovalevsky, M. Mustyakimov, P. Erskine, J. Cooper, *J. Am. Chem. Soc.* **2008**, *130(23)*, 7235–7237; b) L. Coates, P. T. Erskine, S. Mall, R. Gill, S. P. Wood, D. A. Myles, J. B. Cooper, *Eur. Biophys. J.* **2006**, *35(7)*, 559–566.
[11] N. Radeva, J. Schiebel, X. Wang, S. G. Krimmer, K. Fu, M. Stieler, F. R. Ehrmann, A. Metz, T. Rickmeyer, M. Betz, J. Winquist, A. Y. Park, F. U. Huschmann, M. S. Weiss, U. Mueller, A. Heine, G. Klebe, *J. Med. Chem.* **2016**, *59(21)*, 9743–9759 .
[12] A. Bagno, E. Menna, E. Mezzina, G. Scorrano, D. Spinelli, *J. Phys. Chem. A* **1998**, *102(17)*, 2888–2892.
[13] A. Dömling, I. Ugi, *Angew. Chem. Int. Ed. Engl.* **2000**, *39(18)*, 3168–3210.
[14] A. Dömling, W. Wang, K. Wang, *Chem. Rev.* **2012**, *112(6)*, 3083–3135.
[15] C. G. Neochoritis, T. Zhao, A. Dömling, *Chem. Rev.* **2019**, *119(3)*, 1970–2042.
[16] P. Patil, J. Zhang, K. Kurpiewska, J. Kalinowska-Tłuścik, A. Dömling, *Synthesis* **2016**, *48(8)*, 1122–1130.
[17] M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36(6)*, 544–550.
[18] I. Ugi, U. Fetzer, U. Eholzer, H. Knupfer, K. Offermann, *Angew. Chem. Int. Ed.* **1965**, *4(6)*, 472–484.
[19] a) C. G. Neochoritis, S. Stotani, B. Mishra, A. Dömling, *Org. Lett.*,**2015**, *17(8)*, 2002–2005; b) C. G. Neochoritis, T. Zarganes-Tzitzikas, S. Stotani, A. Dömling, E. Herdtweck, K. Khoury, A. Dömling, *ACS Comb. Sci.* **2015**, *17(9)*, 493–499.
[20] A. Dömling, B. Beck, T. Fuchs, A. Yazbak, *J. Comb. Chem.* **2006**, *8(6)*, 872–880.
[21] Reactor was used for enumeration and reaction modeling, J Chem 6.1 **2014** ChemAxon (http://www.chemaxon.com).
[22] a) J. Sunseri, D. R. Koes, *Nucleic Acids Res.* **2016**, *44*(Web Server issue): W442–W448; b) http://pharmit.csb.pitt.edu.
[23] http://www.moloc.ch/.
[24] a) B. Kuhn, J. E. Fuchs, M. Reutlinger, M. Stahl, N. R. Taylor, *J. Chem. Inf. Model.* **2011**, *51*, 3180–3198; b) http://saas1.desertsci.com/.

Manuscript received: January 15, 2020
Revised manuscript received: February 20, 2020
Version of record online: March 18, 2020