▌Drug Design

Fragment-Based Drug Design Facilitated by Protein-Templated Click Chemistry: Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin

Milon Mondal+,[a] M. Yagiz Unver+,[a] Asish Pal,[b] Matthijs Bakker,[a] Stephan P. Berrier,[a] and Anna K. H. Hirsch*[a]

Abstract: There is an urgent need for the development of efficient methodologies that accelerate drug discovery. We demonstrate that the strategic combination of fragment linking/optimization and protein-templated click chemistry is an efficient and powerful method that accelerates the hit-identification process for the aspartic protease endothiapepsin. The best binder, which inhibits endothiapepsin with an IC_{50} value of 43 μM, represents the first example of triazole-based inhibitors of endothiapepsin. Our strategy could find application on a whole range of drug targets.

Despite recent developments in medicinal chemistry, there is a continuous need for the development of more efficient, rapid, and facile strategies to accelerate the drug-discovery process. In recent decades, fragment-based drug design (FBDD) has emerged as an effective and novel paradigm in drug discovery for numerous biological targets.[1–3] FBDD has higher hit rates and better coverage of the chemical space, enabling the use of smaller libraries than those used for high-throughput screening.[2] Since the first report of FBDD, it started to be more widely used in the mid-1990s[4] and has since expanded rapidly. Over the course of the past two decades, various pharmaceutical and biotechnology companies have used FBDD and developed more than 18 drugs that are currently in clinical trials.[5]

Upon identification of a fragment,[6] it has to be optimized to a hit/lead compound and eventually to a drug candidate by fragment growing, linking, merging, or optimization. On the one hand, fragment growing has become the optimization strategy of choice,[7–12] even though it is time consuming because it requires synthesis and validation of the binding mode of each derivative in the fragment–optimization cycle. To overcome this hurdle, we have previously developed strategies in which we combined fragment growing with dynamic combinatorial chemistry (DCC) to render the initial stage of the drug-discovery process more effective.[13] Fragment linking, on the other hand, is very attractive because of its potential for super-additivity (an improvement of ligand efficiency (LE) and not just maintenance of LE), but challenging as it requires the preservation of the binding modes of the individual fragments in adjacent pockets and identification of the best linker with an ideal fit.[14,15] It is presumably due to these challenges that there are only few reports of fragment linking,[4,16] demonstrating the efficiency of linking low-affinity fragments to higher-affinity binders.[17–24] We have recently reported a combination of DCC and fragment linking/optimization, which reduces the risks associated with fragment linking.[25]

In addition to DCC, protein-templated click chemistry (PTCC) has emerged as a powerful strategy to design/optimize a hit/lead for biological targets and holds the potential to reduce the risks associated with fragment-linking.[26,27] PTCC relies on the bio-orthogonal 1,3-dipolar cycloaddition of azide and alkyne building blocks facilitated by the protein target.[28] This highly exothermic reaction produces 1,4- and 1,5-triazoles, which are extremely stable under acidic/basic pH as well as in harsh oxidative/reductive conditions. Furthermore, triazoles can participate in H-bonding, π–π-stacking, and dipole–dipole interactions with the target protein and are a bioisostere of amide bonds. In PTCC, the individual azide and alkyne fragments bind to adjacent pockets of the protein and if the functional groups are oriented in a proper manner, the protein "clicks" them together to afford its own triazole inhibitor (Figure 1). We have therefore envisaged that the potentially synergistic combination of fragment linking and PTCC would represent an efficient hit/lead identification/optimization approach in medicinal chemistry. Here, we have combined fragment linking and PTCC by designing flexibility into the linker and letting the protein select the best combination of building blocks to identify a new class of hits for endothiapepsin, belonging to the pepsin-like aspartic proteases.

Aspartic proteases are a family of enzymes that are widely found in fungi, vertebrates, and plants, as well as in HIV retro-
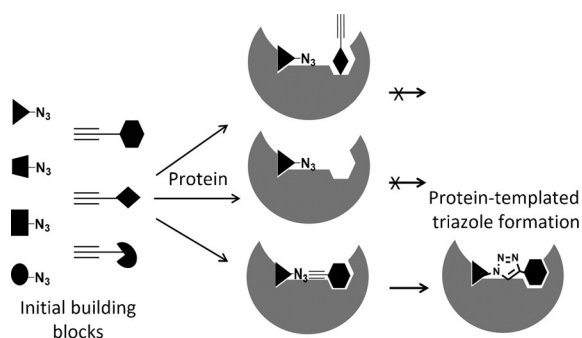
Figure 1. Schematic representation of protein-templated click chemistry leading to a triazole-based inhibitor starting from a library of azides and alkynes.

viruses. This class of enzymes plays a causative role in several important diseases such as malaria, Alzheimer's disease, hypertension, and AIDS.[29] Owing to its high degree of similarity with these drug targets, endothiapepsin has served as a model enzyme for mechanistic studies[30–32] as well as for the identification of inhibitors of renin[33] and β-secretase.[34] Endothiapepsin is a robust enzyme, is available in large quantities, crystallizes easily, and remains active at room temperature for more than three weeks, making this enzyme a convenient representative for aspartic proteases.[35] All aspartic proteases consist of two structurally similar domains, which contribute an aspartic acid residue to the catalytic dyad that is responsible for the water-mediated cleavage of the substrate's peptide bond.[31,32]

Although the linkage of two known inhibitors of acetylcholinesterase via a triazolyl linker using PTCC has been reported, the inhibitors that are linked do not qualify as fragments.[27] To the best of our knowledge, there is no report of fragment linking using PTCC. Herein, we describe how we combined fragment linking/optimization and PTCC for the efficient fragment-to-hit optimization of inhibitors of the aspartic protease endothiapepsin.

We used X-ray crystal structures of endothiapepsin in complex with fragments **1** and **2** (Protein Data Bank (PDB) codes:
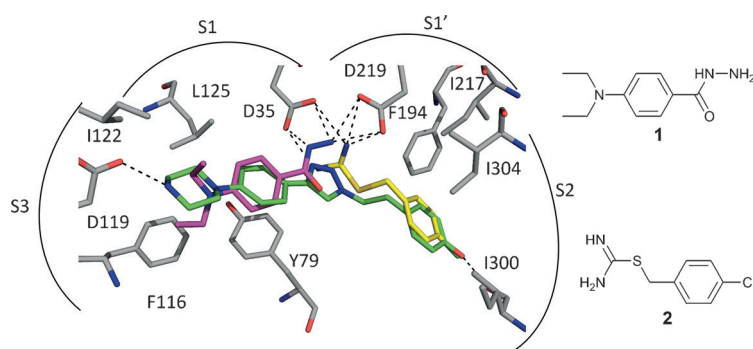

Figure 2. X-ray crystal structure of endothiapepsin in complex with fragments **1** and **2** (PDB code: 3PBZ and 3PLD, respectively) and a modeled potential triazole inhibitor in the active site.[36] Color code: protein skeleton: C: gray, O: red, and N: blue; fragment skeleton: C: purple, yellow and green, N: blue, O: red, Cl: green. Hydrogen bonds below 3.0 Å are shown as black, dashed lines.[38]

3PBZ and 3PLD, respectively, Figure 2), identified by Klebe and co-workers.[36] Both **1** and **2** are engaged in strong H-bonding interactions with the catalytic dyad consisting of amino acid residues D35 and D219, using their hydrazide and amidine groups, respectively (Figure 2). Except for the number of H-bond acceptors (four) for **1**, both fragments **1** and **2** obey Astex's "rule of three",[37] with a molecular weight (M_w) of 207 and 201 Da, three H-bond donors, four and two H-bond acceptors, two freely rotatable bonds and total polar surface areas (TPSAs) of 58.4 and 49.9 Å^2, respectively. At a concentration of 1 mM, fragments **1** and **2** display 89 and 84% inhibition of endothiapepsin, respectively. Considering their promising physicochemical properties, inhibitory potency, their small size (15 and 12 heavy atoms, respectively) and the fact that they bind to adjacent pockets of endothiapepsin, we chose them as a starting point for fragment linking/optimization into an inhibitor of endothiapepsin.

Fragments **1** and **2** occupy the S3 and S1 and the S2 and S1' pockets, respectively, and address the catalytic dyad using an H-bonding network (Figure 2). With the help of the molecular-modeling software Moloc[39] and the FlexX docking module in the LeadIT suite,[40] we linked these two fragments using a triazolyl linker. The newly introduced triazolyl moiety resides at the junction of the S1 and S1' pockets, where hydrazide and amidine groups of fragment **1** and **2**, respectively, were positioned. The triazolyl linker appeared to be ideally suited to address the catalytic dyad through a H-bonding network. Although the protonation of 1,2,3-triazole at pH 4.6, optimal for endothiapepsin, is unprecedented, given that its pK_a value in water is 1.2,[41] in the active site of endothiapepsin, the triazole is expected to bind in close proximity to the two Asp residues (D35 and D219), which will modulate the pK_a value, facilitating protonation. pK_a perturbation is a general phenomenon and has been observed, for instance, in several co-crystal structures of endothiapepsin in complex with heterocyclic fragments.[42] Hence, under acidic conditions, one of the N atoms of the triazole is likely protonated and engaged in a H-bonding interaction with residue D35. Careful analysis of known co-crystal structures of endothiapepsin[35,36] as well as hotspot analysis[43] of the active site of endothiapepsin suggested that the S2 pocket can host aromatic moieties, which can be involved in hydrophobic interactions with residues F194, I217, I304, and I300. The S3 pocket could accommodate a piperazine ring instead of the tertiary amine, which can be involved in an additional H-bonding interaction with residue D119. On the basis of molecular modeling and docking studies, we designed and optimized a series of triazole-based inhibitors. A superimposition of a designed potential triazole inhibitor and the two fragments is shown in Figure 2. All of the triazoles are engaged in H-bonding interactions with D35 and occupy the S3, S1, S1' and S2 pockets, the binding sites of fragments **1** and **2**.

Retrosynthesis of all designed triazole derivatives leads to nine azides (**3**–**11**) and the alkyne **12**, (Scheme 1). We also included alkynes **13**–**15** in our li-
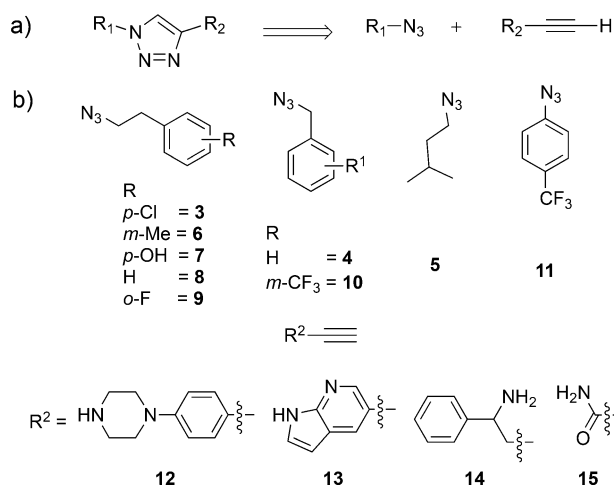
Scheme 1. a) Structures and retrosynthetic analysis of the designed triazole inhibitors starting from fragments 1 and 2; b) structures of the azides 3–11 and the alkynes 12–15.

Figure 3. Structure of the triazoles (17–20) identified using PTCC, inactive triazole 21.

brary, which were available from Syncom. While we obtained all azides from their corresponding bromides by treatment with sodium azide in 40–80% yield (Scheme S1 in the Supporting Information),[44,45] we synthesized alkyne 12 using a Sonogashira cross-coupling reaction, starting from the iodide 16 (Scheme S2 in the Supporting Information).

We set up a library, consisting of four alkynes 12–15 (100 μM each) and nine azides 3–11 (100 μM each), in the presence of a catalytic amount of protein (26 μM) to investigate whether the protein would select a pair of fragments from a library to template the formation of a triazole binder with high affinity (Scheme S8 in the Supporting Information). The advantage of PTCC is that it accelerates the screening time to two weeks by avoiding the synthesis of individual triazoles and reduces the amount of protein required in each individual analysis.

We used UPLC-TOF-SIM (selective ion monitoring) to analyze the formation of triazoles in the reaction mixture. SIM measurements are highly sensitive. We monitored for $[M+H]^+$ of all potential triazole products present in the library. After incubation of the protein at room temperature for two weeks (endothiapepsin is stable and active during this time period),[35] the library was analyzed using UPLC-TOF-SIM. To differentiate between the two regioisomers of triazoles (1,4- and 1,5-triazole), we set up two libraries using the same azide and alkyne building blocks, once in the presence of Cu^I-catalyst to selectively afford the 1,4-triazoles and once under Huisgen cycloaddition conditions to obtain both 1,4- and 1,5-triazoles. We also compared the PTCC reaction with the blank reaction (without protein) as well as the protein alone. We identified a total of four 1,4-triazoles (17–20), which are formed only in the presence of protein (Figure 3 and Figures S2–S9 in the Supporting Information). To establish that the active site of intact endothiapepsin is required for PTCC, we set up two control experiments. Repeating the reaction in the presence of saquinavir (100 μM, a strong inhibitor, $K_i = 48$ nM), or a catalytic amount of bovine serum albumin (BSA; 26 μM) did not lead to the formation of any triazoles.
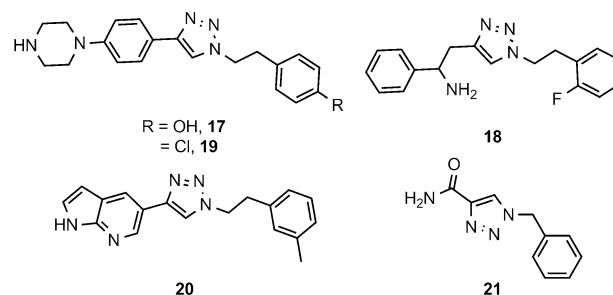
To investigate the biochemical activity of the binders identified by PTCC, we synthesized all four triazoles from their corresponding azide and alkyne precursors using the Cu^I-catalyzed 1,3-cycloaddition (Schemes S3–S6 in the Supporting Information). In addition, we synthesized an inactive triazole 21 to demonstrate the efficiency of PTCC (Figure 3 and Scheme S7 in the Supporting Information). We determined their inhibitory activity using a fluorescence-based assay adapted from the HIV-protease assay.[46]

The enzyme-activity assay confirmed the result of the PTCC experiment. Three out of the four triazoles indeed inhibit endothiapepsin with IC$_{50}$ values in the range of 43–121 μM (Figures S10–S12 in the Supporting Information). We were unable to determine the IC$_{50}$ value of 20 because of its poor solubility even at 250 μM using the maximum possible DMSO concentration for the assay. The inactive triazole 21, which was not observed in the PTCC but synthesized as a control, did not show any activity in the enzyme-activity assay. The most potent triazole inhibitor 17 displays an IC$_{50}$ value of 43 μM (Table 1). The experimental Gibbs free energies of binding (ΔG) and ligand efficiencies (LE), derived from the experimental IC$_{50}$ values using the Cheng–Prusoff equation,[47] correlate with the calculated values using the scoring function HYDE in the LeadIT

Table 1. The IC$_{50}$ values, ligand efficiency (LE), calculated and experimental Gibbs free energies of binding (ΔG) of triazole inhibitors.

Inhibitors	IC$_{50}$[a] [μM]	ΔG$_{EXPT}$[b] [kJ mol^{-1}]	LE[b]	ΔG$_{HYDE}$[c] [kJ mol^{-1}]
17	43±0	−27	0.25	−25
18	94±18	−25	0.26	−19
19	121±3	−24	0.22	−25
20	insoluble	–	–	−23
7	no inhibition	–	–	–
9	no inhibition	–	–	–
12	142±52	–	–	–
14	no inhibition	–	–	–

[a] 26 experiments were performed and only initial six experiments were considered to calculate the initial slope ($n = 6$), 11 different concentrations of inhibitor were used, starting at 1 mM; each experiment was carried out in duplicate and the errors are given in standard deviations (SD), [b] The Gibbs free energy of binding (ΔG$_{EXPT}$) and the ligand efficiencies (LEs) derived from the experimentally determined IC$_{50}$ values, [c] Values indicate the calculated Gibbs free energy of binding (ΔG$_{HYDE}$; calculated by the HYDE scoring function in the LeadIT suite).

suite ($\Delta G_{HYDE}(\mathbf{17}) = -25$ kJ mol^{-1}, Table 1).[48,49] This correlation is also valid for the other triazole inhibitors (Table 1).

To validate the predicted binding mode from fragment linking, we tried to soak crystals of endothiapepsin with the most potent triazole inhibitor **17**. Due to limited solubility, we were not able to obtain crystals of **17** with endothiapepsin. Based on the inhibitory potencies, replacement of −Cl in **19** with a −OH group in **17**, leads to a decrease in IC$_{50}$ value from 121 to 43 μM. This result indicates that the −OH group is involved in more favorable interactions than −Cl, which could be due to the H-bonding interaction with I300 in the S2 pocket, as illustrated by modeling studies (Figure 4a, and Figure S1 in the
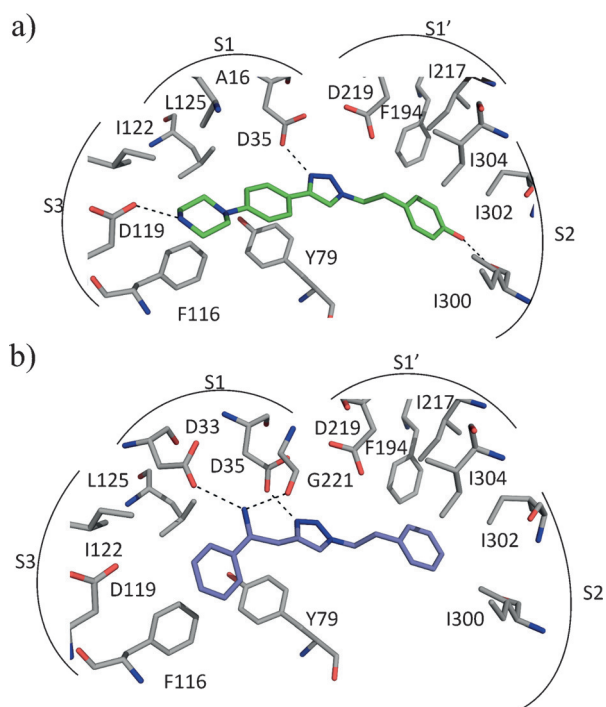
Figure 4. Moloc-generated modeled structures of: a) **17**, and b) (*S*)-**18** in the active site of endothiapepsin. Color code: inhibitor skeleton: C: green, violet, N: blue, O: red; enzyme skeleton: C: gray. H bonds below 3.2 Å are shown as black, dashed lines.

Supporting Information). Moreover, the alkyne **12** displays an IC$_{50}$ value of 142 μM (Figure S13 in the Supporting Information) and is present in both **17** and **19**, two identified triazoles. Fragment **12** is a privileged fragment for endothiapepsin and most probably the binding mode of **12** is retained in both **17** and **19**.

According to modeling and docking, as shown in Figure 4a, and Figure S1, respectively, both **17** and **19** address the catalytic dyad using their triazolyl linker to form direct H-bonds with D35. The NH group of both compounds is involved in a H-bonding interaction with D119 in the S3 pocket. The piperazinyl group of both triazoles occupies the S3 and part of the S1 pockets and is engaged in hydrophobic interactions with F116, I122, and L125, maintaining the binding mode of fragment **1**. The −Cl and −OH substituted phenyl groups of triazoles **17** and **19** occupy the S2 and part of S1′ pockets and are

involved in several hydrophobic contacts with I300, I302, I304, F194, and I217, maintaining the binding mode of fragment **2**.

Triazole **18** displays an IC$_{50}$ of 94 μM and (*S*)-**18** addresses the catalytic dyad using its triazolyl linker to form a direct H bond with D35, as indicated by modeling and docking studies (Figure 4b). The −NH$_2$ group of the triazole is engaged in H-bonding interactions with D33 and G221. Both phenyl substituents of the triazole (*S*)-**18** occupy the S3 and S2 pockets and are involved in hydrophobic interactions with F116, I122, and L125 in the S3 pocket, and I300, I302, I304, F194, and I217 in the S2 pocket, which preserve the binding mode of fragments **1** and **2**, respectively.

In conclusion, we have demonstrated for the first time that the strategic combination of fragment linking/optimization and PTCC is an efficient and powerful method that accelerates the hit-identification process for the aspartic protease endothiapepsin. We have exploited the sensitive UPLC-TOF-SIM method to identify the triazole binders templated by the protein. The best binder inhibits endothiapepsin with an IC$_{50}$ value of 43 μM. Due to the limited solubility of the triazoles identified, we were unable to obtain crystals of any triazole in complex with endothiapepsin. We have reported the first example of triazole-based inhibitors of endothiapepsin. The advantage of our approach is that, a catalytic amount of protein is sufficient to initiate and accelerate triazole formation from a sufficiently large library. Our strategic combination of methodologies proved to be very successful for the hit identification for the aspartic protease endothiapepsin and could be applied to a wide range of biological targets. It could be used in the early stages of drug development and holds the potential to greatly accelerate the drug-discovery process.

[1] D. A. Erlanson, R. S. Mcdowell, T. O. Brien, *J. Med. Chem.* **2004**, *47*, 3463–3482.
[2] P. J. Hajduk, J. Greer, *Nat. Rev. Drug Discovery* **2007**, *6*, 211–219.
[3] H. Chen, X. Zhou, A. Wang, Y. Zheng, Y. Gao, J. Zhou, *Drug Discovery Today* **2015**, *20*, 105–113.
[4] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531–1534.
[5] D. A. Erlanson, *Top. Curr. Chem.* **2011**, *317*, 1–32.
[6] E. H. Mashalidis, P. Śledź, S. Lang, C. Abell, *Nat. Protoc.* **2013**, *8*, 2309–2324.
[7] S. J. Taylor, *J. Med. Chem.* **2011**, *54*, 8174–8187.
[8] Y. Cheng, *J. Med. Chem.* **2011**, *54*, 5836–5857.

[9] M. Congreve, G. Chessari, D. Tisi, A. J. Woodhead, *J. Med. Chem.* **2008**, *51*, 3661 – 3680.

[10] G. Chessari, A. J. Woodhead, *Drug Discovery Today* **2009**, *14*, 668 – 675.

[11] E. Edink, *J. Am. Chem. Soc.* **2011**, *133*, 5363 – 5371.

[12] G. E. de Kloe, D. Bailey, R. Leurs, I. J. P. de Esch, *Drug Discovery Today* **2009**, *14*, 630 – 646.

[13] M. Mondal, D. E. Groothuisl, A. K. H. Hirsch, *Med. Chem. Commun.* **2015**, *6*, 1267 – 1271.

[14] D. C. Rees, M. Congreve, C. W. Murray, R. Carr, *Nat. Rev. Drug Discovery* **2004**, *3*, 660 – 672.

[15] S. Chung, J. B. Parker, M. Bianchet, L. M. Amzel, J. T. Stivers, *Nat. Chem. Biol.* **2009**, *5*, 407 – 413.

[16] P. J. Hajduk, *J. Am. Chem. Soc.* **1997**, *119*, 5818 – 5827.

[17] A. W. Hung, H. L. Silvestre, S. Wen, A. Ciulli, T. L. Blundell, C. Abell, *Angew. Chem. Int. Ed.* **2009**, *48*, 8452 – 8456; *Angew. Chem.* **2009**, *121*, 8604 – 8608.

[18] V. Borsi, V. Calderone, M. Fragai, C. Luchinat, N. Sarti, *J. Med. Chem.* **2010**, *53*, 4285 – 4289.

[19] J. J. Barker, *ChemMedChem* **2010**, *5*, 1697 – 1700.

[20] B. G. Szczepankiewicz, G. Liu, P. J. Hajduk, C. Abad-Zapatero, Z. Pei, Z. Xin, T. H. Lubben, J. M. Trevillyan, M. A. Stashko, S. J. Ballaron, H. Liang, F. Huang, C. W. Hutchins, S. W. Fesik, M. R. Jirousek, *J. Am. Chem. Soc.* **2003**, *125*, 4087 – 4096

[21] N. Howard, *J. Med. Chem.* **2006**, *49*, 1346 – 1355.

[22] D. J. Maly, I. C. Choong, J. A. Ellman, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2419 – 2424.

[23] E. E. Swayze, *J. Med. Chem.* **2002**, *45*, 3816 – 3819.

[24] A. M. Petros, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6587 – 6591.

[25] M. Mondal, N. Radeva, H. Fanlo-Virgós, S. Otto, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2016**, *55*, 9422 – 9426; *Angew. Chem.* **2016**, *128*, 9569 – 9574.

[26] S. K. Mamidyala, M. G. Finn, *Chem. Soc. Rev.* **2010**, *39*, 1252 – 1261.

[27] P. Thirumurugan, D. Matosiuk, K. Jozwiak, *Chem. Rev.* **2013**, *113*, 4905 – 4979.

[28] R. Huisgen, *1,3-Dipolar Cycloaddition Chemistry*, 1st ed., Wiley, New York, **1984**.

[29] J. B. Cooper, *Curr. Drug Targets* **2002**, *3*, 155 – 173.

[30] L. Coates, P. T. Erskine, S. Mall, R. Gill, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Eur. Biophys. J.* **2006**, *35*, 559 – 566.

[31] L. Coates, P. T. Erskine, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Biochemistry* **2001**, *40*, 13149 – 13157.

[32] L. Coates, H.-F. Tuan, S. Tomanicek, A. Kovalevsky, M. Mustyakimov, P. Erskine, J. Cooper, *J. Am. Chem. Soc.* **2008**, *130*, 7235 – 7237.

[33] J. Cooper, *Biochemistry* **1992**, *31*, 8142 – 8150.

[34] S. Geschwindner, L. L. Olsson, J. S. Albert, J. Deinum, P. D. Edwards, T. De Beer, R. H. A. Folmer, *J. Med. Chem.* **2007**, *50*, 5903 – 5911.

[35] M. Mondal, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259 – 3263; *Angew. Chem.* **2014**, *126*, 3324 – 3328.

[36] H. Köster, T. Craan, S. Brass, C. Herhaus, M. Zentgraf, L. Neumann, A. Heine, G. Klebe, *J. Med. Chem.* **2011**, *54*, 7784 – 7796.

[37] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discovery Today* **2003**, *8*, 876 – 877.

[38] Version 1.4.1, Pymol, L. Delano, http://www.pymol.org/.

[39] P. R. Gerber, K. Müller, *J. Comput. Aided. Mol. Des.* **1995**, *9*, 251 – 268.

[40] Version 2. 1. 3, BioSolveIT GmbH, S. Augustin: ‚LeadI Thttp://www.biosolveit.de, LeadIT.

[41] A. Albert, P. J. Taylor, *J. Chem. Soc. Perkin Trans. 2* **1989**, 1903 – 1905.

[42] N. Radeva, S. G. Krimmer, M. Stieler, K. Fu, X. Wang, F. R. Ehrmann, A. Metz, F. U. Huschmann, M. S. Weiss, U. Mueller, *J. Med. Chem.* **2016**, *59*, 7561 – 7575.

[43] H. Gohlke, M. Hendlich, G. Klebe, *Perspect. Drug Discovery Des.* **2000**, *20*, 115 – 144.

[44] S. S. Kulkarni, X. Hu, R. Manetsch, *Chem. Commun.* **2013**, *49*, 1193 – 1195.

[45] Y. Y. Yang, J. M. Ascano, H. C. Hang, *J. Am. Chem. Soc.* **2010**, *132*, 3640 – 3641.

[46] M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **2009**, *36*, 544 – 550.

[47] H. C. Cheng, *J. Pharmacol. Toxicol. Methods* **2001**, *46*, 61 – 71.

[48] N. Schneider, *J. Comput. Aided. Mol. Des.* **2012**, *26*, 701 – 723.

[49] I. Reulecke, G. Lange, J. Albrecht, R. Klein, M. Rarey, *ChemMedChem* **2008**, *3*, 885 – 897.