

University of Groningen

Expanding the toolbox of protein-templated reactions for early drug discovery

Unver, Muhammet

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Unver, M. (2017). *Expanding the toolbox of protein-templated reactions for early drug discovery*. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

**Expanding the toolbox of protein-templated
reactions for early drug discovery**

Muhammet Yağız Ünver

2017

The work described in this thesis was carried out at the Stratingh Institute for Chemistry, University of Groningen, The Netherlands.

This work was financially supported by the University of Groningen and the Ministry of Education, Culture and Science (Gravitation program 024.001.035, Research Center for Functional Molecular Systems).



Printed by: Proefschriftmaken

Cover design: Kaja Sitkowska

ISBN

978-94-034-0026-6 (printed version)

978-94-034-0029-7 (digital version)



university of
 groningen

Expanding the toolbox of protein- templated reactions for early drug discovery

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on
Monday 25 September 2017 at 16.15 hours

by

Muhammet Yağız Ünver

born on 30 November 1987

in Sarayönü, Turkey

Supervisors

Prof. A. K. H. Hirsch

Prof. B. L. Feringa

Assessment committee

Prof. A. J. Minnaard

Prof. G. Roelfes

Prof. J. Rademann

To my dear family

Table of Contents

Chapter 1. Protein-templated hit identification strategies in drug discovery

1.1	Drug discovery and development process	2
1.2	Fragment-based drug design	3
1.3	Target-guided synthesis	5
1.3.1	Dynamic combinatorial chemistry (DCC)	5
1.3.2	Kinetic target-guided synthesis (KTGS)	6
1.3.2.1	Reactions used	7
1.3.2.2	Practical aspects	9
1.3.2.3	Therapeutic scope	12
1.4	Aspartic proteases	22
1.5	Protein-protein interactions	23
1.6	Outline of this thesis	25
1.7	References	27

Chapter 2. Fragment-based drug design facilitated by protein-templated click chemistry: fragment-linking and -optimization of inhibitors of the aspartic protease endothiapepsin

2.1	Introduction	32
2.2	Results and discussion	34
2.2.1	Fragment-based drug design	34
2.2.2	Synthesis of building blocks (azides and alkynes)	35
2.2.3	Generation of the library	36
2.2.4	Synthesis of triazoles identified	39
2.2.5	Biochemical evaluation	41
2.2.6	Discussion	42
2.3	Conclusions	44
2.4	Experimental section	44
2.4.1	Fluorescence-based inhibition assay	44
2.4.2	Modeling and docking	45
2.4.3	PTCC experiments	45
2.4.4	General experimental details	47
2.4.5	Synthesis of azides, alkynes and triazoles	48
2.5	Contributions from co-authors	51
2.6	References	52

Chapter 3. In situ Ugi four-component reaction for the protein-templated identification of inhibitors of endothiapepsin

3.1	Introduction	56
3.2	Results and discussion	57
3.2.1	Design of the inhibitor with Ugi-4CR scaffold	57
3.2.2	Generation of the library	58
3.2.3	In situ Ugi reaction	59
3.2.4	Control experiments	61

3.2.5	Synthesis and biochemical evaluation of the inhibitors	65
3.2.6	Docking results	68
3.3	Conclusions	69
3.4	Experimental section	70
3.4.1	Fluorescence-based inhibition assay	70
3.4.2	Modeling and docking	70
3.4.3	Experimental procedures	70
3.4.4	General experimental details	73
3.4.5	Synthesis of Ugi products	73
3.5	Contributions from co-authors	78
3.6	References	79

Chapter 4. Protein-templated reductive amination for the identification of inhibitors of protein-protein interactions

4.1	Introduction	82
4.2	Results and discussion	84
4.2.1	Design of the inhibitor	84
4.2.2	Generation of the library	85
4.2.3	Synthesis of the core scaffold 3	85
4.2.4	Protein-templated reductive amination	86
4.2.5	Synthesis and biochemical evaluation of the inhibitors	88
4.3	Conclusions	91
4.4	Experimental section	92
4.4.1	Modeling	92
4.4.2	Protein purification procedure	92
4.4.3	FP assay	93
4.4.4	^1H - ^{15}N heteronuclear single quantum coherence (HSQC) NMR experiment	93
4.4.5	Experimental procedures	94
4.5	Contributions from co-authors	101
4.6	References	102

Chapter 5. Protein-templated esterification reaction for the inhibitors of aspartic protease endothiapepsin

5.1	Introduction	106
5.2	Results and discussion	107
5.2.1	Design of the ester inhibitor	107
5.2.2	Synthesis and biochemical evaluation of the designed inhibitor	108
5.2.3	Optimization studies	109
5.2.4	Generation of the library and the library reaction	110
5.2.5	Synthesis and biochemical evaluation of the library	111
5.2.6	Discussion	112
5.3	Conclusions	113
5.4	Experimental section	113
5.4.1	Fluorescence-based inhibition assay	113

5.4.2	Modeling and docking	113
5.4.3	Experimental procedures	113
5.4.4	General experimental details	114
5.4.5	General procedure for Steglich esterification /de-protection reaction	114
5.5	Contributions from co-authors	116
5.6	References	117

Chapter 6. Design and synthesis of bioisosteres of acylhydrazones as stable inhibitors of the aspartic protease endothiapepsin

6.1	Introduction	120
6.2	Results and discussion	121
6.2.1	Design of the bioisosteres	121
6.2.2	Synthesis of the bioisosteres	122
6.2.3	Biochemical evaluation	123
6.2.4	Crystallographic studies	124
6.3	Conclusions	126
6.4	Experimental section	127
6.4.1	Fluorescence-based inhibition assay	127
6.4.2	Modeling and docking	127
6.4.3	Crystallization, data collection and processing	127
6.4.4	General experimental details	127
6.4.5	Experimental procedures	127
6.5	Contributions from co-authors	131
6.6	References	132

Summary and outlook	133
----------------------------	-----

Samenvatting	137
---------------------	-----

Acknowledgment	143
-----------------------	-----

Chapter 1

Protein-templated hit-identification strategies in drug discovery

Today's drug discovery mainly relies on the synthesis of large numbers of compounds and testing with a variety of targets. Target-guided-synthesis (TGS) has emerged as a powerful alternative technique to current drug-discovery methods, which can accelerate the hit-identification process, and thus this long-term trajectory. In this chapter, we firstly discuss the drug-discovery process and fragment-based drug design (FBDD) briefly. Secondly, we introduce TGS in a broad context and discuss kinetic target-guided synthesis (KTGS) in more detail.

1.1 Drug discovery and development process

The drug-discovery process, from identification of a new active compound until gaining the regulatory approval, is a long and expensive path. This process takes approximately 10–15 years, sometimes even longer due to the complexity of drug development as well as the capital required to fund each phase of the process.^[1]

There are no simple solutions to shorten this lengthy process; however, methods used in each phase can improve the efficiency of the individual steps, and resulting in a significant acceleration of the overall process.

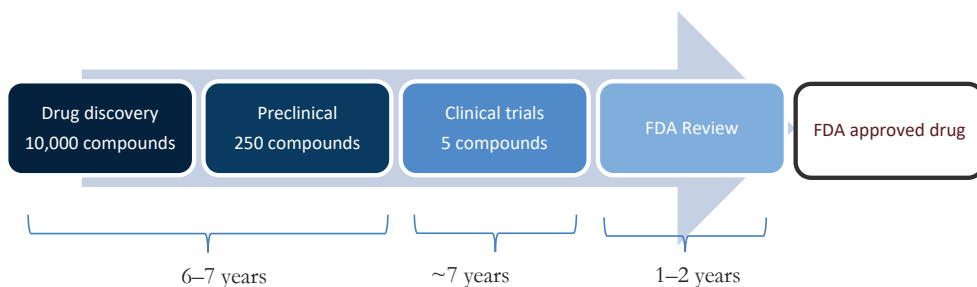


Figure 1. Drug discovery timeline.

A general drug-discovery process includes the following steps (Figure 1):

- Identification and validation of the target, assay development
- Hit identification
- Lead identification and –optimization
- Non-clinical safety
- Scale-up and production
- Clinical trials
- FDA approval

The first step, target identification, is the understanding of the disease and choosing a valid target molecule. Nowadays, in contrast to old drug discovery, it is very important to start with a clear understanding of the disease and the biological target itself.^[2] After selection of the biological target, which are mostly biomacromolecules such as DNA or a protein, hit compounds are identified by using various hit-identification techniques. There are several approaches at this stage such as high throughput screening (HTS), structure-based drug design (SBDD), fragment-based drug design and target-guided synthesis (TGS) that will be the main focus of this thesis.

Once the hit compound with desired inhibitory potency is identified, the next step is to explore the chemical space around the hit compound by making its analogues for structure–activity relationship (SAR) studies. Prior to human studies, lead compounds are tested on animal models. After successful pre-clinical studies and toxicity assessments, scientists should also consider scale-up and manufacturing issues such as ease of synthesis, costs, stability, shelf life etc. The last step are clinical studies on human for safety and efficacy of the drug and upon completion of clinical trials, the drug is ready for marketing.

1.2 Fragment-based drug design

Fragment-based drug design (FBDD) has become a very promising alternative method to current drug-discovery techniques.^[2–7] FBDD has some great benefits that makes it very successful compared to HTS, the most widely used method in pharmaceutical companies. Hit rates are higher than HTS as the chemical space is covered more efficiently by screening small molecules, in other words “fragments” instead of larger molecules.^[5] The other benefits are higher binding efficiency and more effective optimization capacity.

The first step in FBDD is the construction of fragment libraries. A suitable fragment library should have specific properties and some points should be critically considered; (i) the distinction between fragments and hits/leads. Congreve et al^[5] introduced ‘Astex’s rule of three’, a set of criteria for the design of fragment libraries. According to “rule of three”, a fragment should fulfill:

- molecular weight < 300 Da
- cLogP ≤ 3
- number of H-bond donors ≤ 3
- number of H-bond acceptors ≤ 3
- number of rotatable bonds ≤ 3

The rules are accepted and widely used by many medicinal chemists as efficient criteria for the selection of fragment libraries; (ii) the library should be structurally diverse enough to cover a large area of chemical space; (iii) as the fragments are weak binders, their biochemical assays are performed at high concentrations and therefore they should have good solubility.^[5,8]

Fragments are screened by using several techniques such as direct binding assays with higher concentrations,^[9–11] NMR-based screening,^[12,13] mass-spectrometry-based techniques^[14–18] or crystallographic techniques. Identified fragments can then be optimized to lead compounds.^[19]

Fragments are optimized to lead-like compounds via two main ways. These are fragment growing and fragment linking.^[20] In fragment growing, an initial fragment is optimized by introducing new functional groups to the fragment core to fill adjacent pockets of the active site, affording a lead-like compound (Figure 2).

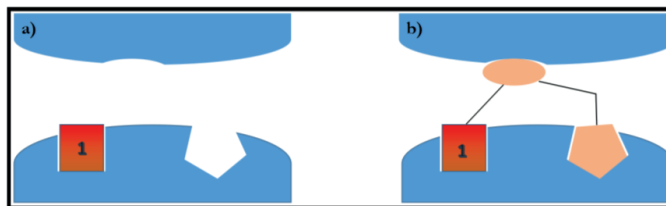


Figure 2. Fragment growing; a) fragment **1** binds to the target, b) fragment is grown to occupy adjacent pockets.

On the other hand, in fragment linking, two or more fragments binding to the pockets in close proximity are linked together via a linker with an optimal fit (Figure 3).

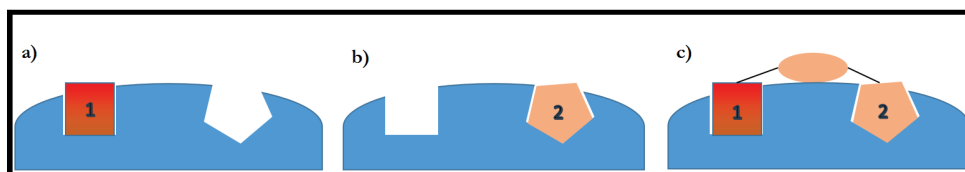


Figure 3. Fragment linking; a) fragment **1** binds to the one site of the target, b) fragment **2** binds to another site of the target, which is in close proximity, c) fragment **1** and **2** are linked together via a linker with optimal fit.

A theoretical benefit of fragment linking is the super additivity in ligand efficiency (LE) rather than preservation of LE.^[21] This advantage was also demonstrated experimentally,^[22] scientists were able to show the efficiency of fragment linking successfully for the first time. Although fragment linking is attractive, it is very challenging owing to the difficulty in finding a linker with optimal fit. On the other hand, fragment growing is time-consuming as it requires synthesis of each modified fragment and verification of the binding mode after each modification.^[4–8,19,20,23–26]

Therefore, to overcome the challenges in both linking and growing strategies in FBDD, we developed new techniques^[27] such as the use of protein-templated click reaction in fragment linking, which will be described in detail in Chapter 2.^[28]

1.3 Target- guided synthesis

Target-guided synthesis (TGS) is a powerful strategy in which the biological target selects its own inhibitors by assembling them from biocompatible reagents.^[29] There are two main strategies in this approach: dynamic combinatorial chemistry (DCC), in which the target selects and amplifies its own ligands from a library of products formed from reversibly connected building blocks and kinetic target-guided synthesis (KTGS) in which the assembly takes place in an irreversible manner. Both techniques hold great potential, yet they are still unconventional and remain relatively unexplored.^[29] Although the distinction between those two techniques is artificial in terms of application, in this thesis, we will categorize the methods according to the reversibility of the reaction and focus on irreversible reactions in the KTGS context.

1.3.1 Dynamic combinatorial chemistry

DCC has emerged as a powerful tool to identify binders for biological targets. It facilitates the reversible combination of building blocks by forming dynamic combinatorial libraries (DCLs) of potential binders in an efficient manner. Since the reaction facilitating DCLs is reversible, upon binding of the library members with the strongest affinity to the biological target, the product composition re-equilibrates, resulting in a shift in the equilibrium. Ultimately, the best binders are amplified, which circumvents the need for the synthesis and biochemical evaluation of each library member (Figure 4).^[30–34]

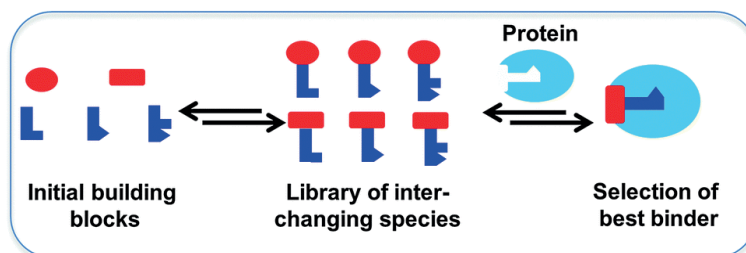
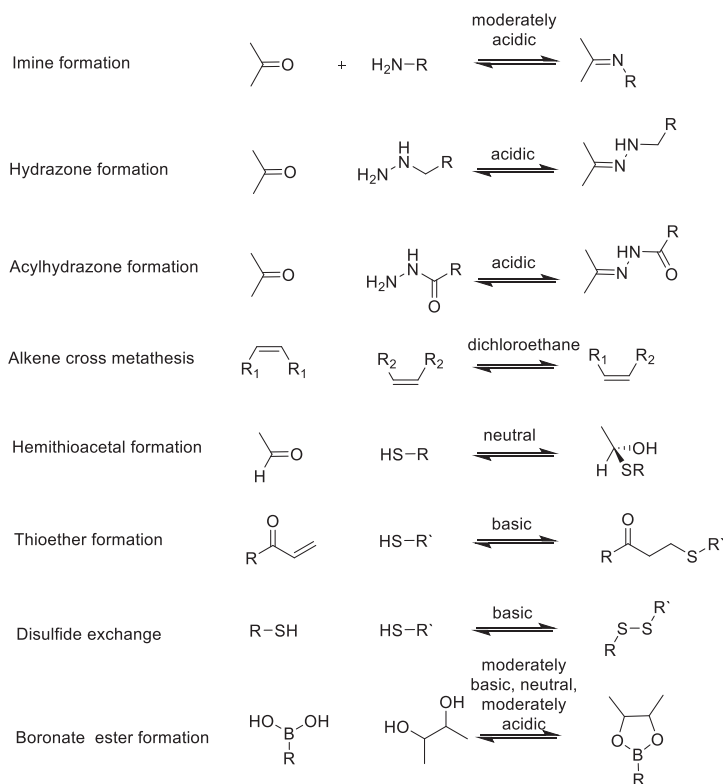


Figure 4. Schematic representation of the fundamental concept of DCC, adapted from Mondal et. al.^[34]

Several reversible reactions were applied to protein targets in the DCC context as can be seen in the Scheme 1.^[34–41] A number of issues should be taken into consideration in the selection of a compatible reaction for DCC; (i) the reaction must be carried out in aqueous media; (ii) equilibration of the DCL should be fast enough at the desired pH and the temperature where the protein is stable; (iii) the reaction should be chemoselective so that the cross-reactivity with functional groups in the library or with the target is avoided. The other critical points for a DCC

setup are: (i) selection of building blocks with comparable reactivity; (ii) in case mass spectroscopy is the analytical technique, avoiding the use of building blocks with identical molecular weights; (iii) high solubility of individual building blocks and the products to prevent precipitation in the reaction mixture; (iv) freezing the equilibrium prior to analysis and (v) selection of a proper analytical technique depending on the availability of the biological target.

In order to overcome the stringent requirements for DCC, dynamic ligation methods, in which the fragment combinations are screened one by one coupled with bioactivity-based detection, have been developed successfully.^[53]



Scheme 1. Reversible reactions used in DCC, adapted from Mondal et al.^[34]

1.3.2 Kinetic target-guided synthesis

KTGS represents another type of target-guided approach, which enables the assembly of the inhibitors in situ via an irreversible process. In this approach, the biological target accelerates the reaction in between complementary building blocks by bringing them in close proximity and proper orientation (Figure 5).^[42]

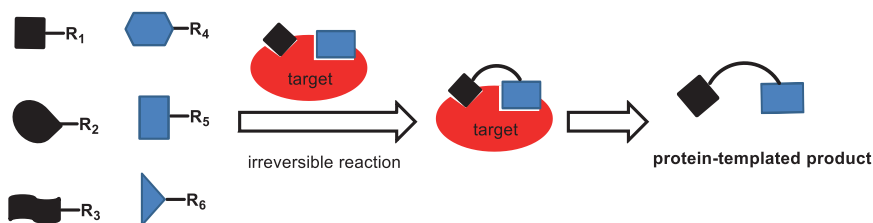


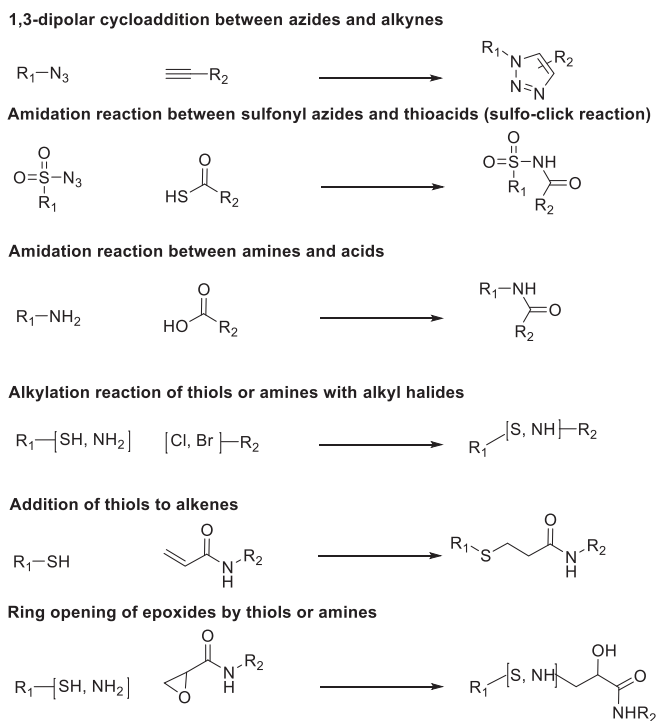
Figure 5. Schematic representation of the fundamental concept of KTGS.

The first step in KTGS is the selective binding of building blocks in the library mixture to the specific pockets, which is mostly the active site/targeted site of the protein. Once two building blocks or more (in Chapter 3, we will demonstrate the first example of using more than two building blocks) are bound to the adjacent pockets of the target in the proper orientation and in close proximity, an irreversible reaction takes place by assembling the best binder, leading to a stable complex with the protein. This technique does not require prior synthesis, purification and biochemical evaluation of the library members, enabling rapid and cheaper screening of large number of compounds. Therefore, it has a great potential to decrease the costs and the time needed to discover hits in the early phases of drug discovery.^[29]

The irreversible reactions used to connect fragments in KTGS should also be biocompatible; products formed as well as individual building blocks in the mixture should be stable and soluble under physiological conditions. In addition to this, a substantial rate difference between biomolecule-templated and blank reactions is required. Because the protein-templated reactions are only used for analytical purposes, the compounds formed in the reaction mixtures are mostly in trace amount, follow-up synthesis is required to confirm their activity. Therefore, the synthetic protocols to get the desired compounds should be readily available.^[42]

1.3.2.1 Reactions used in KTGS

As the criteria for the reactions in this field are very stringent, only a couple of reactions have been reported (Scheme 2).



Scheme 2. Chemical reactions reported for KTGS in the literature.

The most widely used reaction in KTGS is the 1,3-dipolar cycloaddition of azides and alkynes. Both azides and alkynes are biocompatible reagents, which makes this irreversible reaction suitable for KTGS. It is the first reaction ever reported in this context^[43,44] and allows for the formation of *syn*- and *anti*-triazoles, expanding the number of compounds screened.^[45–48] Sharpless^[49] and co-workers coined ‘in situ click reaction’ for the first time. The principle of this reaction is that the 1,3-dipolar cycloaddition can be catalyzed under metal-free conditions by the protein. The target binds to the initial blocks with the strongest affinity while giving them the proper orientation and close proximity and finally ‘clicks’ them together to form the corresponding triazoles. In Chapter 2, our recent contribution in this field will be discussed in detail.

The second reaction reported in this field is the sulfo-click reaction between electron-deficient sulfonyl azides or electron-rich azides such as alkyl or aryl azides and thioacids to form acyl sulfonamides.^[50,51]

Another reaction used in KTGS is the direct amidation reaction of a carboxylic acid with an amine. Gelin et al.^[52] observed the protein-templated amide formation without pre-activation of the carboxylic acid although acylation usually requires pre-activated carboxylic acids. Very

recently Rademann and co-workers^[53] reported the amidation reaction by using a set of activated carboxylic acid derivatives.

Protein-templated alkylation of thiols by halides was described by Chase et al.,^[54] demonstrating C-S bond formation between bromoacetylcarnitine and coenzyme A (CoA).

The last two reactions used in this field are the thio-Michael addition reaction, which is used for both DCC and KTGS^[50] and S_N2 ring opening of epoxides.^[55]

To date all the reports use two complementary building blocks and the variety of the reactions is limited. Therefore, expanding the toolbox of protein-templated reactions is timely and will be the main goal of this thesis.

1.3.2.2 Practical aspects of KTGS

Protein-templated reactions are usually performed at 37 °C. Reactions require a minimum of 24 h, and some reactions take even more than seven days. The building block concentration is usually high to overcome detection problems. All reported examples show that KTGS reactions necessitate 3–320 μM concentration and nearly stoichiometric amount of the biological target (0.37–320 μM).^[29]

Another important aspect in KTGS is the choice of analytical techniques as the products forming in the templated reactions are mostly in trace amounts. Therefore, selection of sensitive techniques is very important to detect the ligands. The first reported detection technique was MALDI-DIOS,^[49] subsequently LC-MS-SIM (SIM: selective ion monitoring) was used and improved the detection. Later, LC-HRMS-TOF was used to validate molecular formula of the identified hits.^[56] The last example of detection technique is X-ray crystallography reported by Gelin et al.^[52] to detect bound reagents.

The correlation between signal and affinity has been a debatable topic for a long time. Although some authors previously reported the existence of a direct correlation between signal intensity and affinity, recent reports indicate that inhibitory potency cannot directly be deduced from the KTGS experiment as described in the templated synthesis of hIDE inhibitors.^[56] The protein templates the formation of *syn*- and *anti*-triazoles with the same LC-MS signal, however 1,4-triazoles have the best activity.

Table 1. Experimental conditions for kinetic target-guided synthesis in the literature, adapted from Deprez and co-workers.^[29]

Target	Time	Temp.	Analytics	Target concentration	Reagent 1 concentration	Reagent 2 concentration
reagent 1: azides, reagent 2: alkynes (Huisgen 1,3 dipolar cycloaddition)						
^[49] <i>eel</i> AChE	1w	r.t	MALDI-DIOS LC-MS-ESI	1 μ M	3 μ M	66 μ M
^[58] <i>eel</i> or <i>m</i> AChE	6–24 h	r.t or 37 $^{\circ}$ C	LC-MS-SIM	1 μ M	1–4.6 μ M	6–24 μ M
^[59] <i>eel</i> or <i>m</i> AChE	6–24 h	37 $^{\circ}$ C	LC-MS-SIM	1 μ M	4.6 μ M	24 μ M
^[45] <i>m</i> AChE	2–7 d 9–13 d	37 $^{\circ}$ C	LC-MS-SIM	1 μ M	4.6 μ M	24 μ M
^[60] <i>b</i> CA-II	40 h	37 $^{\circ}$ C	LC-MS-SIM	30 μ M	400 μ M	60 μ M
^[61] <i>b</i> CA-II	36 h	37 $^{\circ}$ C	LC-MS-SIM	29 μ M	400 μ M	60 μ M
^[56] <i>HIV-1</i>	24 h	23 $^{\circ}$ C	LC-MS-SIM	15 μ M	100 μ M	500 μ M
^[62] <i>LsACbBP</i>	10 d	r.t.	LC-MS-SIM	1 mg/mL	100 μ M	100 μ M
^[63] <i>Sm</i> Chi A, B, C ¹	20h	37 $^{\circ}$ C	LC-MS-SIR	192 μ g/mL	100 μ M	300 μ M
^[64] <i>Tb</i> EthR	24–112h	37 $^{\circ}$ C	LC-MS-SIM	5 μ M	125 μ M	40 μ M
^[65] <i>h</i> IDE	72h	37 $^{\circ}$ C	LC-HRMS- TOF	4.7 μ M	100 μ M	100 μ M
^[57] <i>b</i> Bcr-Abl	4d	37 $^{\circ}$ C	LC-MS-ESI	0.37 μ M	370 μ M	370 μ M
^[66] <i>Sa</i> BPL	48 h	37 $^{\circ}$ C	LC-MS-SIM	2 μ M	370–500 μ M	370–500 μ M
^[67] <i>b</i> COX-2	18–20 h	37 $^{\circ}$ C	LC-MS-SIM	7 μ M	400 μ M	60 μ M

Target	Time	Temp.	Analytics	Target concentration	Reagent 1 concentration	Reagent 2 concentration
reagent 1: alkyl halides, reagent 2: α-mercaptotosylamine						
^[68] B _{CA-II}	48 h	25 °C	LC-MS-SIM	320 μ M	320 μ M–10 mM	320 μ M
reagent 1: acrylamides, reagent 2: thiols (acylated or not)						
^[69] mAChE	6 h	37 °C	LC-MS-SIM	1 μ M	4.6 μ M	24 μ M
reagent 1: NAD⁺, reagent 2: malonamide ester						
^[70] SIRT-1	ND	ND	MALDI-TOF	4 μ M	500 μ M	600 μ M
reagent 1: amine, reagent 2: carboxylic acid						
^[52] L _m NADK	ND	ND	X-RAY	9 mg/mL	ND	ND
reagent 1: sulfonyl azides, reagent 2: thioacids (protected; sulfo-click reaction)						
^[51] Bcl-XL	6 h	37 °C	LC-MS-SIM	2 μ M	20 μ M	20 μ M
reagent 1: amine, reagent 2: activated esters (amidation reaction)						
^[53] Factor Xa	2h	20 °C	HPLC/QTOF-MS	14.5 nM	0.285 mM	5 mM

Including the right control reactions is crucial in KTGS. The blank reaction in which the reaction is performed in the absence of the biological target serves as a negative control. The other control reaction to prove specificity of KTGS (positive control) is the use of another target such as bovine serum albumin (BSA) although the use of BSA is risky as it can catalyze many reactions itself.^[71] In addition, several protein mutants such as Bcl-XL^(F131A, D133A) for Bcl-XL, H223E for L_mNADK were used as a positive control for the corresponding templated reactions. Besides biological controls, synthetic controls are also important even though MS-TOF techniques can provide good information about the molecular formula of the compounds. In order to be sure that the peaks identified correspond to the compounds hypothesized, scientists compare their retention time by performing chemical synthesis in small quantities. Almost all published examples have those synthetic controls, especially in in situ click reaction to

differentiate the regioisomers. To differentiate between the *syn*- and *anti*-triazole, a copper catalyzed and a Huisgen cycloaddition or Ru-catalyzed click reaction were used, respectively.

1.3.2.3 Therapeutic scope of KTGS

Pioneering works in KTGS were published using acetylcholine esterase (AChE) as the biological target.^[49] Until recently, KTGS found applications mainly in the discovery of enzyme inhibitors. Nevertheless, recent studies revealed that KTGS is also applicable for the discovery of ligands for neurotransmitter ligand-gated ion channel (Ach-binding protein), EthR, a receptor of the TetR family as well as 14-3-3 protein-protein interactions (PPIs).^[29]

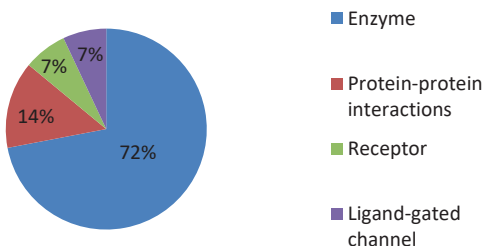


Figure 6. Classes of proteins used in KTGS.

Until now, 31 different KTGS experiments were performed by using 16 different targets, and 11 experiments were carried out using AChE (Table 2).

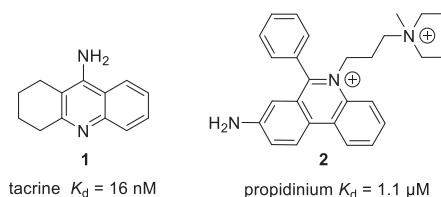
Table 2. Targets used in KTGS, adapted from Deprez and co-workers.^[29]

Species	Individual protein	Number of KTGS
<i>E. electricus</i> <i>T. californica</i> <i>M. musculus</i> <i>H. Sapiens</i>	Acetylcholine esterase	11
HIV-1 <i>Endothia parasitica</i> , <i>H. sapiens</i>	Aspartic protease Endothiapepsin Insulin-degrading enzyme	3 1 1
<i>S. mercescens</i>	Chitinase	2
<i>S. aureus</i>	Biotin-protein ligase	2
<i>P. falciparum</i>	Tryptophanyl tRNA synthetase	1
<i>L. monocytogenese</i>	NAD kinase	1
<i>H. sapiens</i>	Bcr-Abl	1
<i>B. taurus</i>	Carbonic anhydrase II	2
<i>H. sapiens</i>	Cyclooxygenase-2	1
<i>L. stagnalis</i> <i>A. californica</i>	Acetylcholine binding protein	1
<i>M. tuberculosis</i>	EthR	1
<i>H. sapiens</i>	Bcl-XL	2
<i>H. sapiens</i>	14-3-3 protein	1
<i>H. sapiens</i>	Factor Xa, Serine protease	1

Acetylcholine esterase (AChE)

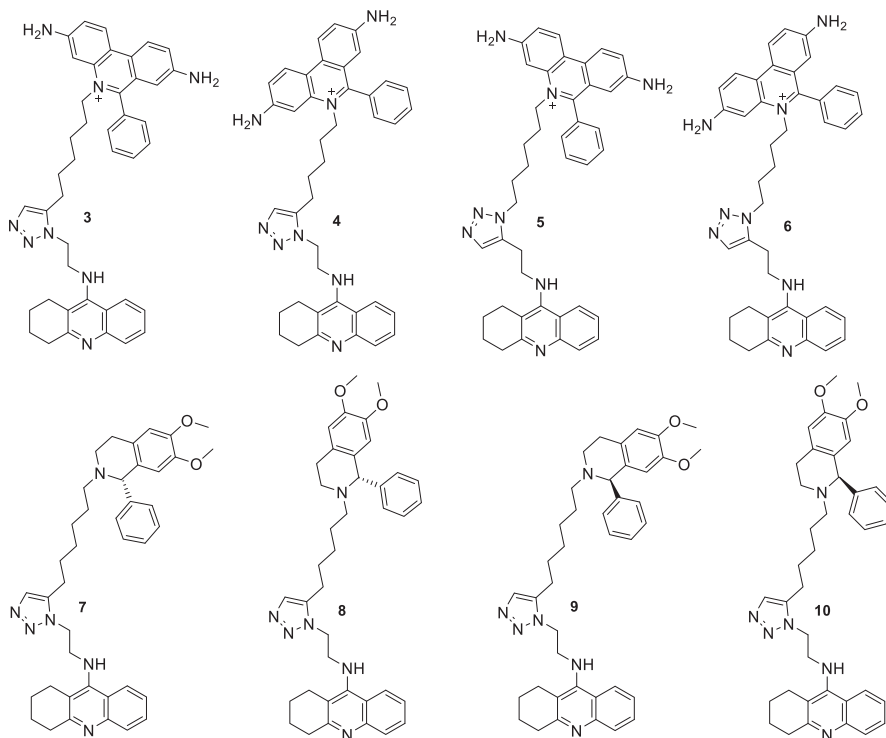
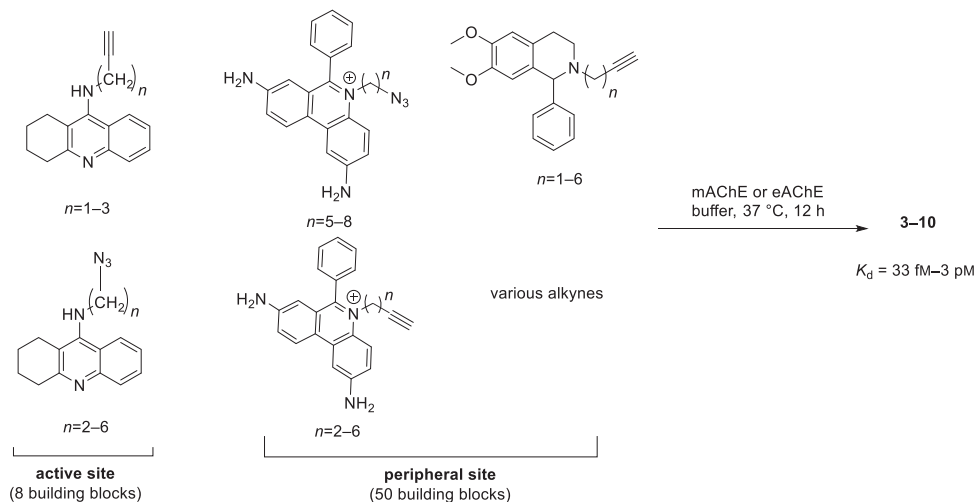
AChE, a protein which plays a key role in nerve impulse propagation, has become the most used protein in KTGS. This enzyme was validated as a drug target to cure Alzheimer's disease. It has a deep catalytic active site and large peripheral site on its surface, making it a convenient enzyme for KTGS.

The first series of inhibitors feature tacrine (**1**) (active site inhibitor) and propidium (**2**) (peripheral site binder) moieties in their scaffolds (Figure 7).

**Figure 7.** Structures of tacrine and propidium.

Tacrine strongly binds to the active site of AChE with a K_d of 16 nM and propidium is binding specifically to the peripheral site with moderate K_d value. Keeping those moieties

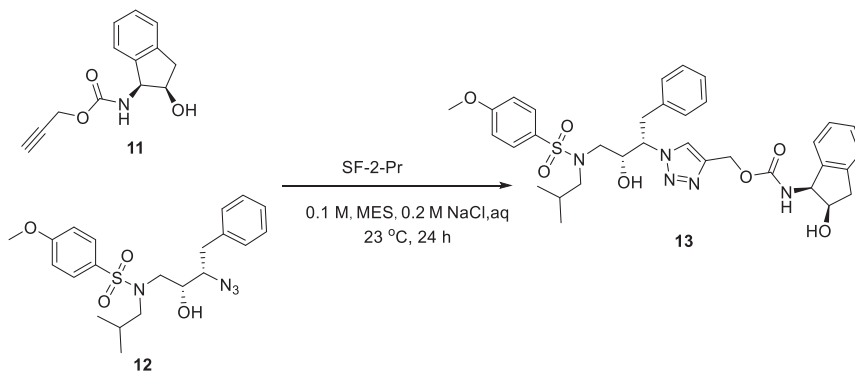
constant, the authors designed a library of compounds bearing azide and alkyne functional groups. By using in situ click chemistry approach they could screen more than 300 compounds and identified a series of triazoles **3–10** with femtomolar activities (Scheme 3).^[58]



Scheme 3. An example of KTGS using AChE.^[58]

HIV-I protease

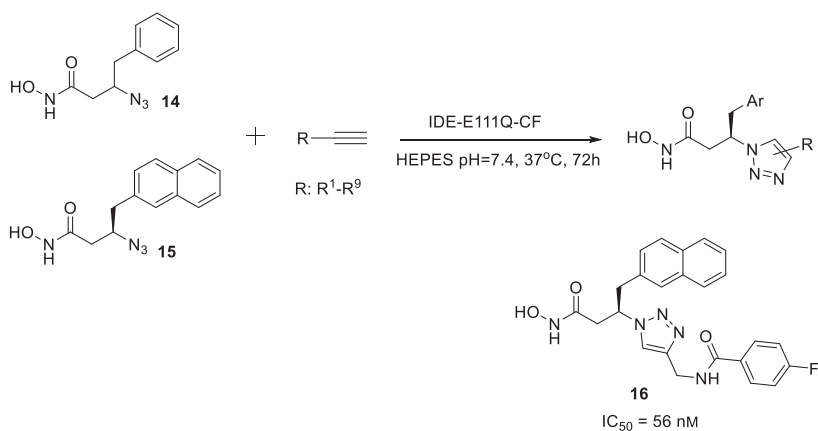
HIV-I protease is another target used in KTGS and is a retroviral aspartic protease, which has a crucial role in the life cycle of HIV.^[72] It is an important target for inhibition of viral replication. There are several inhibitors already available on the market such as indinavir. Being inspired by this inhibitor, Fokin and co-workers^[56] performed in situ click reaction using HIV-1-Pr SF-2-WTQ7K-Pr (SF-2-Pr) to illustrate the assembly of inhibitor **13** ($IC_{50} = 6$ nM) from alkyne **11** and azide **12** in the presence of the target.



Scheme 4. HIV-I-templated in situ click chemistry.^[56]

Insulin-degrading enzyme

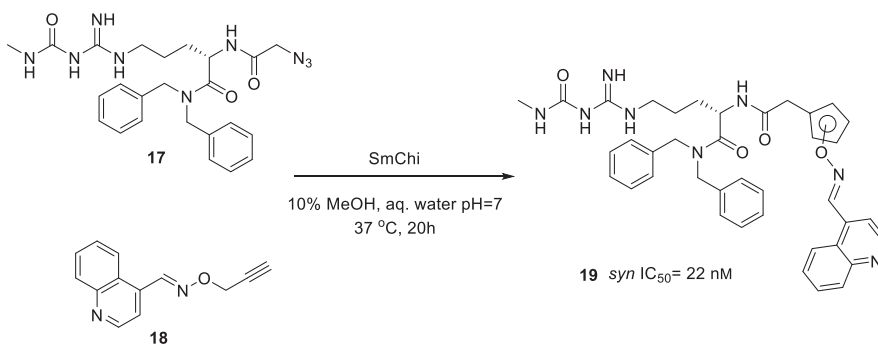
Insulin-degrading enzyme (IDE) is a protease, which plays a role in the cleavage of insulin or other bioactive peptides. It is in the M16 metalloenzyme family and has a large zinc binding site. Deprez et al.^[65] used KTGS to discover IDE inhibitors for the first time. They used in situ click chemistry approach and designed a library of compounds comprising azide-bearing hydroxamides **14** and **15** for the zinc binding site and various alkynes. They performed SAR studies and their best hit identified **16** displayed an IC_{50} value of 56 nM (Scheme 5).



Scheme 5. Insulin-degrading enzyme-templated in situ click chemistry.^[65]

Chitinase

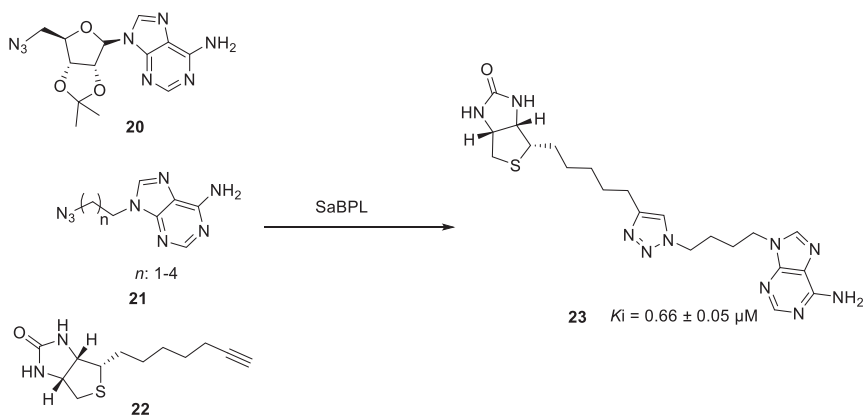
Chitinases are enzymes responsible for hydrolysis of chitin, the second most abundant polysaccharide in nature. They are biological targets for antifungals, antibacterial and antiparasitic agents. Omura et al. discovered an inhibitor with low-nanomolar activity by using *Serratia marcescens* (SmChi, a chitinase from *s. marcescens*)-templated click chemistry.^[63]



Scheme 6. *Serratia marcescens*-templated in situ click chemistry.^[63]

Biotin-protein ligase

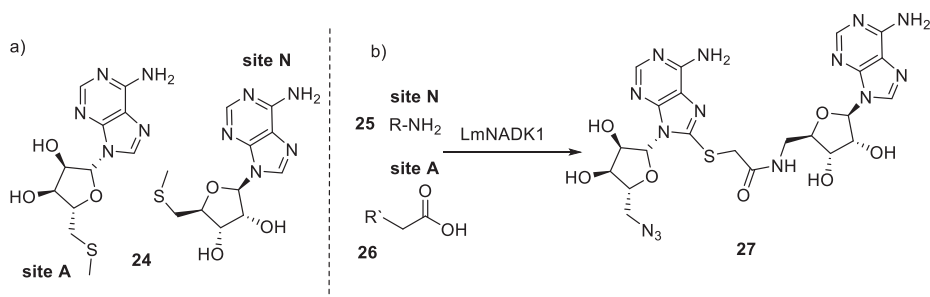
Since there is continuous need for novel antibiotics nowadays due to the increase in the drug-resistant pathogenic bacteria, biotin-protein ligase represents an important and promising drug target for new antibacterial research. Tieu et al. proved the applicability of KTGS by using a known triazole using *Streptococcus aureus* biotin-protein ligase (BPL).^[66]



Scheme 7. Assembly of **23** from **22** and **21** with *Streptococcus aureus* biotin-protein ligase-templated in situ click chemistry.^[66]

NAD kinase

NAD kinase from *Listeria monocytogenes* (LmNADK1) was another protein used in KTGS, representing an important target for antibiotic research. Gelin et al. reported that LmNADK1 templates the reaction between 5' amino-5'deoxyadenosine and carboxylic acids to afford amide bonds without prior activation of carboxylic acids.^[52] The active site of this enzyme has two binding sites (subsites A and N), during X-ray studies they observed that two molecules of adenosine derivatives are simultaneously binding to those two subpockets (Scheme 8a).

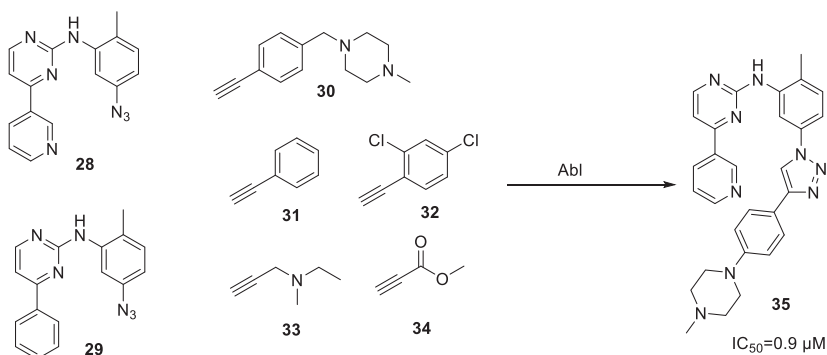


Scheme 8. a) Representative orientation of two bound ligands. b) *Listeria monocytogenes*-templated formation of **27**.^[52]

Upon those observations, the authors soaked two adenosine derivatives **25** and **26** in the presence of the enzyme. Subsequent X-ray studies proved the templated formation of an amide bond between those two derivatives to afford **27** (Scheme 8b).

Bcr-Abl

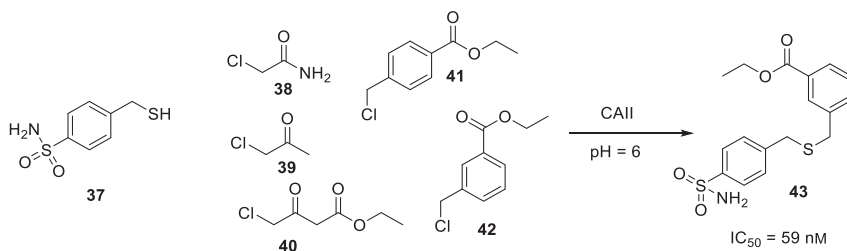
Bcr-Abl (tyrosine kinase) is an important target to cure leukemic cancer. Passerella et al.^[57] confirmed the capacity of this target in KTGS, assembling its own inhibitor from a pool of complementary building blocks derived from a known inhibitor reported with an IC_{50} of 0.9 μ M.



Scheme 9. Bcr-Abl-templated synthesis of inhibitor 35.^[57]

Carbonic anhydrase II

Carbonic anhydrase II (CA II) is a protein, which assists the interconversion between CO_2 and HCO_3^- , therefore it is responsible for the regulation of pH in the blood, renal reabsorption of NaCl and bicarbonate in the proximal tubule.^[29] Huc and co-workers disclosed the first example of KTGS using CA II, showing the rate enhancement of the reaction in the presence of the target^[68] while Kolb and co-workers published the same strategy to discover more potent inhibitors with improved affinity.^[60]

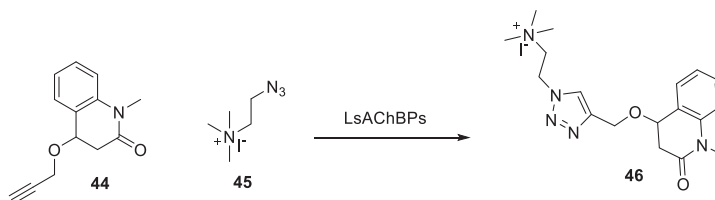


Scheme 10. First example of Carbonic anhydrase II in KTGS, templated synthesis of 43.^[68]

Acetylcholine binding protein

Acetylcholine binding protein (AChBP) is the homologous protein of nicotinic acetylcholine receptors (nAChR), a ligand-gated ion channel, which is an important target to cure Alzheimer's disease. It is homologous to the extracellular domain of nAChR.^[42] For the proof-of-

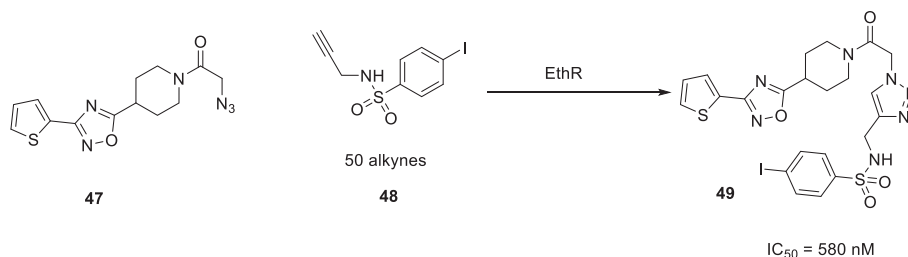
principle work, the authors firstly screened a library of triazoles by traditional synthesis against AChBP from *Lymnaea* (Ls), *Aplysia californica* (Ac) Y55W *Aplysia* mutant (AcY55W) and the best hit **46** was identified. In order to prove the methodology, the fragments of the best hit were incubated in the presence of Ls, Ac, AcY55W AChBP and they observed that Ls selectively templated the formation of **46** from the corresponding azide **44** and alkyne **45**.^[62]



Scheme 11. The best hit **46** is selectively templated by acetylcholine-binding protein.^[62]

EthR

EthR, a *Mycobacterium tuberculosis* gene (Rv3855), is responsible for the expression of EthA, which is a mycobacterial monooxygenase that has a causative role in the treatment of multidrug-resistant tuberculosis. Ethionamide is a prodrug activated by EthA and its active form inhibits InhA, which is actively involved in the biosynthesis of mycolic acid in the bacteria's wall. Accordingly, Deprez et al. proposed that inhibition of EthR would improve the efficacy of Ethionamide by increasing the transcription of EthA.^[64] Upon screening of the azide **47** with a library of alkynes, they observed the formation of hit **49** with the help of the protein's active site.

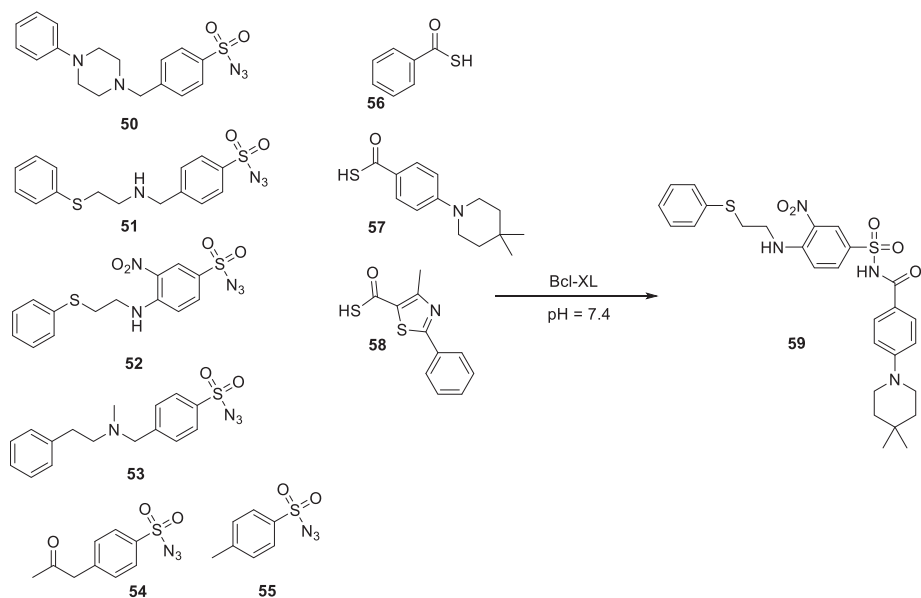


Scheme 12. *M. tuberculosis* gene-templated formation of **49**.^[64]

Bcl-XL

KTGS has also been applied to protein-protein interactions (PPIs). The first example is the Bcl-XL-templated sulfo-click reaction.^[73] PPIs are very important for many biological processes and represent an important target for therapeutics. Manetsch and co-workers performed a proof-of-principle work to demonstrate the compatibility of KTGS with PPIs. A

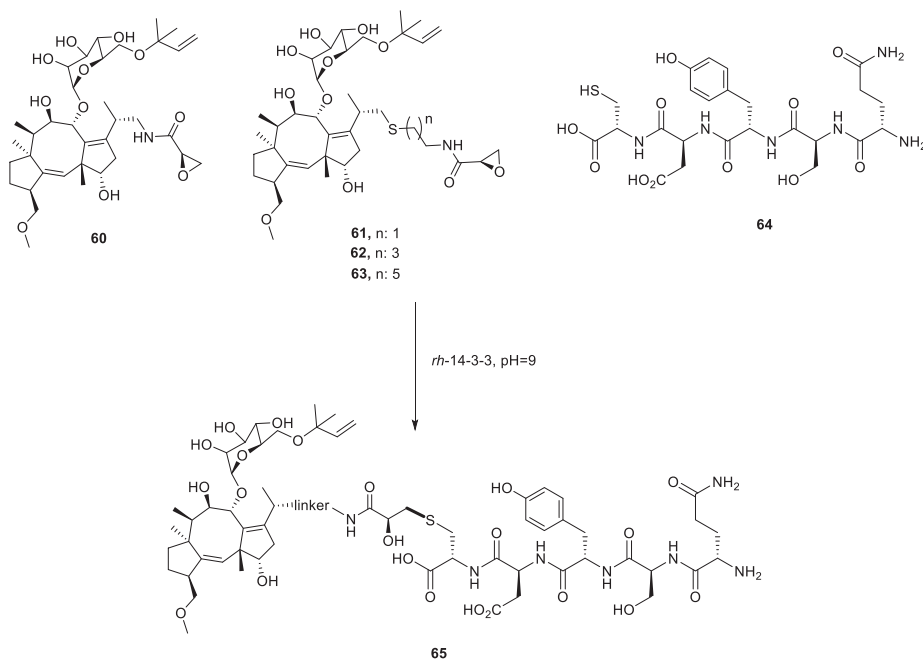
library of sulfonazides **50–55**, and thioacids **56–58**, mimics of Bcl-2 inhibitor, were designed and the library mixture was incubated with Bcl-XL. Subsequently, they observed the assembly of compound **59**, which was then proven to disturb Bcl-XL/Bak-BH3 protein-protein interactions.^[73]



Scheme 13. Bcl-XL-templated amide formation to afford compound **59**.^[73]

14-3-3

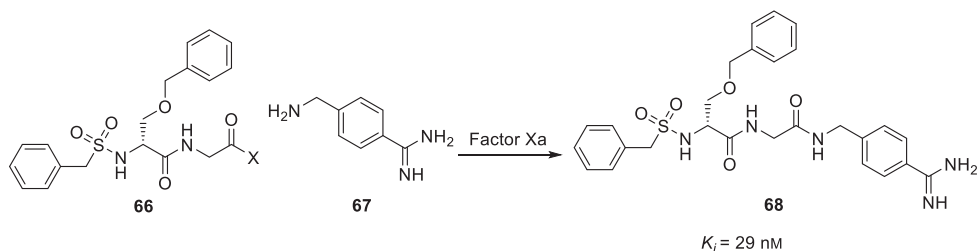
14-3-3 is the other target in PPI family, which found an application in the KTGS concept. It binds to key proteins involved in various roles such as apoptosis, transcription regulation and intracellular signaling. Okhanda et al. reported the template effect of recombinant human 14-3-3 protein for the in situ generation of inhibitors of PPIs.^[74] Four epoxide derivatives **60–63** and the peptide **64** were incubated with 14-3-3. Although the background reaction was still present, they could see a rate enhancement by almost 200% for compound **61**. When the linker is short (compound **60**) the authors still observed the simultaneous binding of the building blocks to the adjacent pockets (37% inhibition) but they were unable to react.^[74]



Scheme 14. 14-3-3-templated epoxide ring opening reaction for the formation of **65**.^[74]

Factor Xa

Factor Xa, a serine protease of the blood co-agulation cascade and target of antithrombotic drugs, has been used in the KTGS concept recently. Rademann and coworkers^[53] used factor Xa as a template for the formation of **68** from various activated carboxylic acid derivatives **66** and amine **67** by using fragment ligation screening.



Scheme 15. Factor Xa-templated amide formation of **68**.^[53]

1.4 Aspartic proteases

There are four main types of protease enzymes: aspartic, serine, cysteine and metallo proteases which catalyze the hydrolysis of polypeptide bonds selectively.^[75–77]

Aspartic proteases are mostly found in fungi, vertebrates, plants and viruses, and they are involved in several diseases such as hypertension, amyloid disease, malaria and AIDS.^[78] All aspartic proteases have a catalytic dyad comprised of two structurally similar domains. Each domain contributes an aspartic acid residue (D35 and D219) to form this catalytic dyad, which is responsible for the cleavage of the substrate's peptide bond via a bound water molecule (Figure 8).^[79,80]

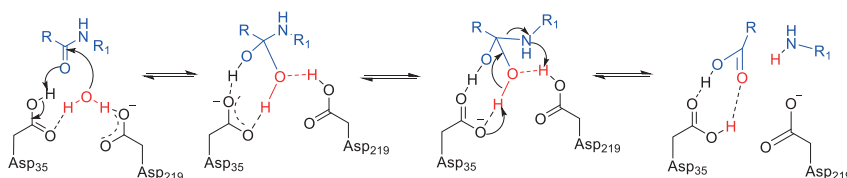


Figure 8. Mechanism of the hydrolysis of peptide bonds by aspartic proteases.^[79]

Most of the proteases are sequence-specific; they are capable of cleaving the peptide bonds selectively. This is commonly accomplished by having an enzyme site(s), which is complementary to substrate residue(s). The standard nomenclature is used to describe substrate/inhibitor residues (P3, P2, P1, P1', P2', P3'), which correspond to the enzyme binding site (S3, S2, S1, S1', S2', S3').^[77,81] Several experiments were performed using papain as a peptidic substrate to investigate the size of the active site of aspartic proteases, revealing that aspartic proteases have a large active site with several subpockets that accommodates one amino acid side chain.^[81] A representative example of subpockets in the aspartic protease, endothiapepsin in complex with pepstatin, which inhibits all members of the aspartic protease family, is depicted in Figure 9.

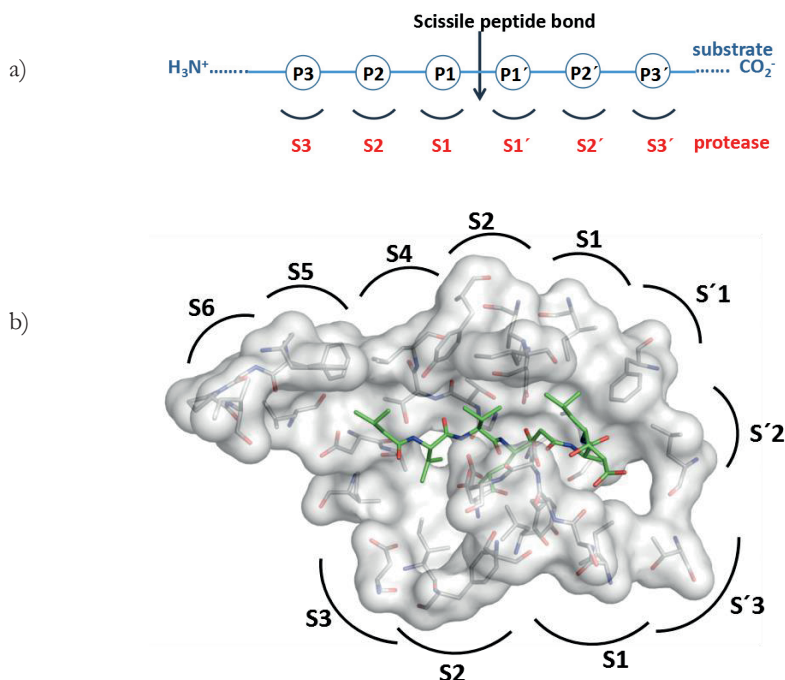


Figure 9. a) Nomenclature of the corresponding subpockets, X-ray crystal structure of the aspartic protease endothiapepsin cocrystallized with pepstatin, and the assigned subpockets. Color code: protein surface: gray; pepstatin skeleton: C: green, N: dark blue, O: red.

1.5 Protein-protein interactions

Protein-protein interactions (PPIs) play a central role in many biological functions such as intercellular communication and apoptosis, offering a new avenue for the discovery of new therapeutics.^[82] Targeting PPIs with small molecules is very challenging due to a number of factors. Those factors are mainly the lack of small-molecule starting points for future design, flatness of the interface, the problems in distinguishing real from artefactual bindings, and the general problems with the small-molecule library.

Natural products are often good starting points for medicinal chemists to develop new drugs. However, finding a small molecule bound to a protein-protein interface is very rare. Therefore, scientists are still seeking new techniques, as presented in section 1.3.2, to tackle this problem.

The shape of a typical protein-protein interface is the most challenging part in drug discovery. X-ray studies show that in most cases no deep cavities are available which can be

targeted by small molecules as in other targets.^[83] In addition, the protein-protein interface has a certain flexibility and adaptability which can cause overlooking the binding-site conformations that are well-fitted with small molecules but not visible in a single crystal structure.^[84-86]

Although artefacts are a general problem for all proteins, it is very difficult to exclude them in the case of PPIs. These systems are mostly screened by inhibition assays that require high compound/enzyme ratio. Another difficulty is the enzyme kinetic experiments to differentiate competitive from allosteric inhibition due to the challenge in determining which one of the two enzymes is binding to the substrate.^[82]

Despite the multiple challenges in the field, PPIs remain a popular target family. Developing technology as well as new screening techniques will allow scientists to have a deeper understanding of the field and more accurate selection of potential molecules.

1.6 Outline of this thesis

Despite advancements in technology in the medicinal chemistry field, drug discovery is still often a lengthy, expensive and difficult process. Therefore, shortening the duration, decreasing the costs and increasing the efficiency of the process are highly important goals.

The main aim of this thesis is to use modern and novel methodologies to access bio-active compounds in an efficient and fast way, and thus shortening the drug-discovery trajectory, in particular hit-identification. In order to fulfil our aim, we used the target-guided synthesis approach as well as combinations of methods.

In Chapter 2, we demonstrate the strategic combination of fragment-based drug design and protein-templated click chemistry (PTTC) to facilitate hit-identification. By using PTTC, we linked/optimized two fragments of endothiapepsin efficiently.

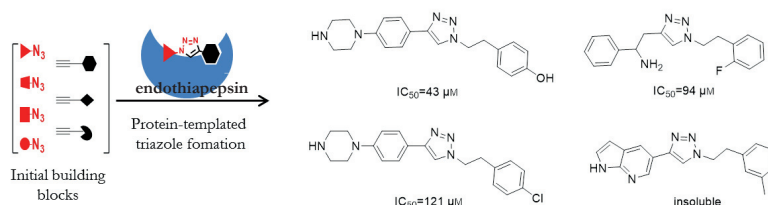


Figure 10. Protein-templated click reaction.

In Chapter 3, we introduce a novel protein-templated reaction to the toolbox of the protein-templated reactions, *in situ* Ugi 4-component reaction. Use of four building blocks, for the first time in this concept, enables rapid and efficient hit identification by opening up access to new regions of the chemical space.

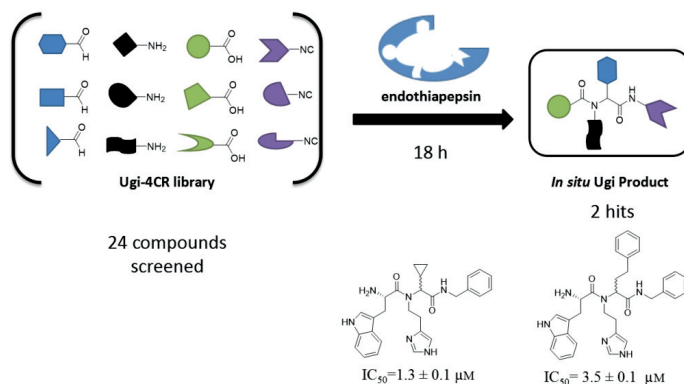


Figure 11. *In situ* Ugi 4-component reaction.

In Chapter 4, we disclose the applicability of target-guided synthesis to a challenging target, the Mdm2-p53 protein-protein interaction (PPI), by using a new protein-templated reaction. By using Mdm2 as a target, we screened a library of compounds via *in situ* reductive amination in one week and identified one hit, demonstrating the efficiency of the approach in targeting PPIs as well as flexible binding pockets.

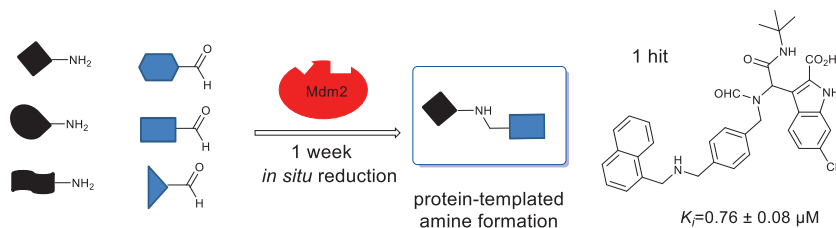


Figure 12. *In situ* reductive amination.

In Chapter 5, we describe our attempts to discover a new protein-templated reaction, namely *in situ* esterification. We performed optimization studies for the protein-templated esterification reaction and generated a library of ester inhibitors of endothiapepsin.

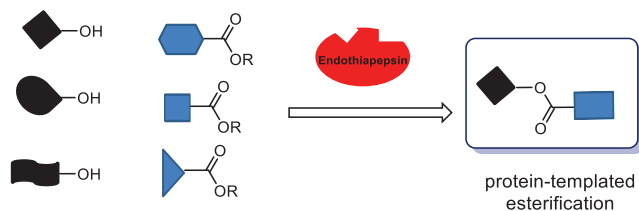


Figure 13. *In situ* esterification.

In Chapter 6, we demonstrate the design, synthesis and biochemical evaluation of bioisosteres of N-acylhydrazone which is published in our group previously. Replacement of the hydrazone moiety with an amide and ester linker afforded comparable activity against endothiapepsin.

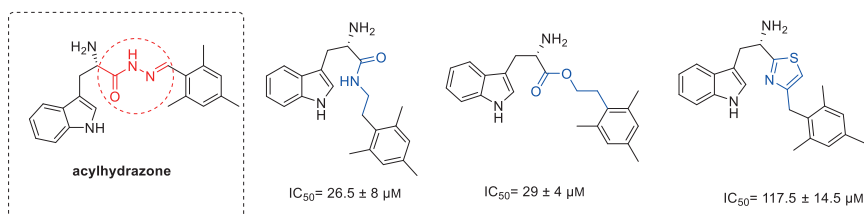


Figure 14. Bioisosteres of acylhydrazone inhibitors.

1.7 References

- [1] M. Everts, T. Cihlar, J. R. Bostwick, R. J. Whitley, *Annu. Rev. Pharmacol. Toxicol.* **2017**, *57*, annurev-pharmtox-010716-104533.
- [2] M. Ding, J. Eliashberg, S. Stremersch, *Innovation and Marketing in the Pharmaceutical Industry*, **2014**.
- [3] P. J. Hajduk, *J. Med. Chem.* **2006**, *49*, 6972–6976.
- [4] P. J. Hajduk, J. Greer, *Nat. Rev. Drug Discov.* **2007**, *6*, 211–219.
- [5] M. Congreve, G. Chessari, D. Tisi, A. J. Woodhead, *J. Med. Chem.* **2008**, *51*, 3661–3680.
- [6] G. Chessari, A. J. Woodhead, *Drug Discov. Today* **2009**, *14*, 668–675.
- [7] D. Joseph-McCarthy, A. J. Campbell, G. Kern, D. Moustakas, *J. Chem. Inf. Model.* **2014**, *54*, 693–704.
- [8] H. Chen, X. Zhou, A. Wang, Y. Zheng, Y. Gao, J. Zhou, *Drug Discov. Today*, **2015**, *20*, 105–113.
- [9] J. W. Kehoe, D. J. Maly, D. E. Verdugo, J. I. Armstrong, B. N. Cook, Y. Bin Ouyang, K. L. Moore, J. A. Ellman, C. R. Bertozzi, *Bioorganic Med. Chem. Lett.* **2002**, *12*, 329–332.
- [10] X. Wang, Y. Choe, C. S. Craik, J. A. Ellman, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2201–2204.
- [11] D. J. Maly, I. C. Choong, J. A. Ellman, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2419–2424.
- [12] S. Shuker, P. Hajduk, R. Meadows, S. W. Fesik, *Science*, **1996**, *274*, 1531–1534.
- [13] P. J. Hajduk, J. Dinges, J. M. Schkeryantz, D. Janowick, M. Kaminski, M. Tufano, D. J. Augeri, A. Petros, V. Nienaber, P. Zhong, et al., *J. Med. Chem.* **1999**, *42*, 3852–3859.
- [14] K. Breuker, *Angew. Chem. Int. Ed.* **2004**, *43*, 22–25.
- [15] D. A. Ockey, T. R. Gadek, *Bioorganic Med. Chem. Lett.* **2004**, *14*, 389–391.
- [16] D. A. Ockey, J. L. Dotson, M. E. Struble, J. T. Stults, J. H. Bourell, K. R. Clark, T. R. Gadek, *Bioorganic Med. Chem.* **2004**, *12*, 37–44.
- [17] Y. He, J. Yang, B. Wu, D. Robinson, K. Sprankle, P.-P. Kung, K. Lowery, V. Mohan, S. Hofstadler, E. E. Swayze, et al., *Bioorg. Med. Chem. Lett.* **2004**, *14*, 695–699.
- [18] S. A. Hofstadler, R. H. Griffey, *Chem. Rev.* **2001**, *101*, 377–390.
- [19] D. A. Erlanson, *Curr. Opin. Biotechnol.* **2006**, *17*, 643–652.
- [20] D. C. Rees, M. Congreve, C. W. Murray, R. Carr, *Nat. Rev. Drug Discov.* **2004**, *3*, 660–672.
- [21] W. P. Jencks, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4046–4050.
- [22] P. J. Hajduk, G. Sheppard, D. G. Nettlesheim, E. T. Olejniczak, S. B. Shuker, R. P. Meadows, D. H. Steinman, G. M. Carrera, P. A. Marcotte, J. Severin, et al., *J. Am. Chem. Soc.* **1997**, *119*, 5818–5827.
- [23] V. Borsi, V. Calderone, M. Fragai, C. Luchinat, N. Sarti, *J. Med. Chem.* **2010**, *53*, 4285–

- 4289.
- [24] G. E. de Kloe, D. Bailey, R. Leurs, I. J. P. de Esch, *Drug Discov. Today*, **2009**, *14*, 630–646.
- [25] D. A. Erlanson, R. S. McDowell, T. O. Brien, *J. Med. Chem.* **2004**, *47*, 3463–3483.
- [26] A. Kumar, A. Voet, K. Y. J. Zhang, *Curr. Med. Chem.* **2012**, *19*, 5128–5147.
- [27] M. Mondal, D. E. Groothuis, A. K. H. Hirsch, *Med. Chem. Comm.* **2015**, *6*, 1267–1271.
- [28] M. Mondal, M. Y. Unver, A. Pal, M. Bakker, S. P. Berrier, A. K. H. Hirsch, *Chem. - A Eur. J.* **2016**, *22*, 14826–14830.
- [29] D. Bosc, J. Jakhlal, B. Deprez, R. Deprez-Poulain, *Future Med. Chem.* **2016**, *8*, 381–404.
- [30] J. Li, P. Nowak, S. Otto, *J. Am. Chem. Soc.* **2013**, *135*, 9222–9239.
- [31] J.-M. Lehn, J.-M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151–60.
- [32] M. G. Simpson, M. Pittelkow, S. P. Watson, J. K. M. Sanders, *Chem. Rev.* **2006**, *106*, 3652–3711.
- [33] S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Dynamic Covalent Chemistry*, **2002**.
- [34] M. Mondal, A. K. H. Hirsch, *Chem. Soc. Rev.* **2015**, *44*, 2455–2488.
- [35] J. Tomasek, J. Schatz, *Green Chem.* **2013**, *15*, 2317.
- [36] B. Shi, M. F. Greaney, *Chem. Comm.* **2005**, *4*, 886–888.
- [37] K. C. Nicolaou, R. Hughes, S. Y. Cho, N. Winssinger, C. Smethurst, H. Labischinski, R. Endermann, *Angew. Chem. Int. Ed.* **2000**, *39*, 3823–3828.
- [38] Y. Patskovsky, L. Patskovska, S. C. Almo, I. Listowsky, *Biochemistry*, **2006**, *45*, 3852–3862.
- [39] A. J. Clipson, V. T. Bhat, I. McNae, A. M. Caniard, D. J. Campopiano, M. F. Greaney, *Chem. - A Eur. J.* **2012**, *18*, 10562–10570.
- [40] T. Bunyapaiboonsri, O. Ramström, S. Lohmann, J.-M. Lehn, L. Peng, M. Goeldner, *ChemBioChem*, **2001**, *2*, 438–444.
- [41] J. Liu, K. R. West, C. R. Bondy, J. K. Sanders, *Org. Biomol. Chem.* **2007**, *5*, 778–786.
- [42] E. Oueis, C. Sabot, P.-Y. Renard, *Chem. Comm.* **2015**, *51*, 12158–12169.
- [43] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
- [44] W. L. Mock, T. A. Irra, J. P. Wepsiec, T. L. Manimaran, *J. Org. Chem.* **1983**, *48*, 3619–3620.
- [45] R. Manetsch, A. Krasniński, Z. Radić, J. Rauschel, P. Taylor, K. B. Sharpless, H. C. Kolb, *J. Am. Chem. Soc.* **2004**, *126*, 12809–12818.
- [46] K. B. Sharpless, R. Manetsch, *Expert Opin. Drug Discov.* **2006**, *1*, 525–538.
- [47] H. C. Kolb, K. B. Sharpless, *Drug Discov. Today*, **2003**, *8*, 1128–1137.
- [48] S. W. Millward, H. D. Agnew, B. Lai, S. S. Lee, J. Lim, A. Nag, S. Pitram, R. Rohde, J. R. Heath, *Integr. Biol.* **2012**, *5*, 87–95.

- [49] W. G. Lewis, L. G. Green, F. Grynszpan, Z. Radić, P. R. Carlier, P. Taylor, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 1053–1057.
- [50] E. Oueis, C. Sabot, P.-Y. Renard, *Chem. Comm.* **2015**, *51*, 12158–12169.
- [51] X. Hu, J. Sun, H. G. Wang, R. Manetsch, *J. Am. Chem. Soc.* **2008**, *130*, 13820–13821.
- [52] M. Gelin, G. Poncet-Montange, L. Assairi, L. Morellato, V. Huteau, L. Dugue, O. Dussurget, S. Pochet, G. Labesse, *Structure*, **2012**, *20*, 1107–1117.
- [53] M. Jaegle, T. Steinmetzer, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, *56*, 3718–3722.
- [54] J. F. Chase, P. K. Tubbs, *Biochem. J.* **1969**, *111*, 225–235.
- [55] S.S. Kulkarni, X. Hu, K. Doi, H.-G. Wang, R. Manetsch, *ACS Chem. Biol.* **2011**, *6*, 724–732.
- [56] M. Whiting, J. Muldoon, Y.-C. Lin, S. M. Silverman, W. Lindstrom, A. J. Olson, H. C. Kolb, M. G. Finn, K. B. Sharpless, J. H. Elder, et al., *Angew. Chem. Int. Ed.* **2006**, *45*, 1435–1439.
- [57] C. Peruzzotti, S. Borrelli, M. Ventura, R. Pantano, G. Fumagalli, M. S. Christodoulou, D. Monticelli, M. Luzzani, A. L. Fallacara, C. Tintori, et al., *ACS Med. Chem. Lett.* **2013**, *4*, 274–277.
- [58] A. Krasinski, Z. Radić, R. Manetsch, J. Raushel, P. Taylor, K. B. Sharpless, H. C. Kolb, *J. Am. Chem. Soc.* **2005**, *127*, 6686–6692.
- [59] S. P. Nikas, R. Sharma, C. A. Paronis, S. Kulkarni, G. A. Thakur, D. Hurst, J. T. Wood, R. S. Gifford, G. Rajarshi, Y. Liu, et al., *J. Med. Chem.* **2015**, *58*, 665–681.
- [60] V. P. Mocharla, B. Colasson, L. V. Lee, S. Röper, K. B. Sharpless, C. H. Wong, H. C. Kolb, *Angew. Chem. Int. Ed.* **2004**, *44*, 116–120.
- [61] V. P. Mocharla, J. C. Walsh, H. C. Padgett, H. Su, B. Fueger, W. A. Weber, J. Czernin, H. C. Kolb, *ChemMedChem*, **2013**, *8*, 43–48.
- [62] N. P. Grimster, B. Stump, J. R. Fotsing, T. Weide, T. T. Talley, J. G. Yamauchi, Á. Nemez, C. Kim, K. Y. Ho, K. B. Sharpless, et al., *J. Am. Chem. Soc.* **2012**, *134*, 6732–6740.
- [63] T. Hirose, T. Sunazuka, A. Sugawara, A. Endo, K. Iguchi, T. Yamamoto, H. Ui, K. Shiomi, T. Watanabe, K. B. Sharpless, et al., *J. Antibiot.* **2009**, *62*, 277–282.
- [64] N. Willand, M. Desroses, P. Toto, B. Diricé, Z. Lens, V. Villeret, P. Rucktooa, C. Loch, A. Baulard, B. Deprez, *ACS Chem. Biol.* **2010**, *5*, 1007–1013.
- [65] R. Deprez-Poulain, N. Hennuyer, D. Bosc, W. G. Liang, E. Enée, X. Marechal, J. Charton, J. Totobenazara, G. Berte, J. Jahklal, et al., *Nat. Commun.* **2015**, *6*, 8250.
- [66] W. Tieu, T. P. Soares da Costa, M. Y. Yap, K. L. Keeling, M. C. J. Wilce, J. C. Wallace, G. W. Booker, S. W. Polyak, A. D. Abell, *Chem. Sci.* **2013**, *4*, 3533.
- [67] SIEMENS MEDICAL SOLUTIONS INC.: WO2006116736A2 (2006).

- [68] Á. Nguyen, I. Huc, *Angew. Chem. Int. Ed.* **2001**, *40*, 1774–1776.
- [69] C. K. Jones, N. Byun, M. Bubser, *Neuropsychopharmacol. Rev.* **2012**, *37*, 16–42.
- [70] T. Asaba, T. Suzuki, R. Ueda, H. Tsumoto, H. Nakagawa, N. Miyata, *J. Am. Chem. Soc.* **2009**, *131*, 6989–6996.
- [71] T. Hirose, N. Maita, H. Gouda, J. Koseki, T. Yamamoto, A. Sugawara, H. Nakano, S. Hirono, K. Shiomi, T. Watanabe, et al., *Proc. Natl. Acad. Sci.* **2013**, *110*, 15892–15897.
- [72] A. Brik, C. Wong, *Org. Biomol. Chem.* **2003**, *3*, 5–14.
- [73] X. Hu, R. Manetsch, *Chem. Soc. Rev.* **2010**, *39*, 1316–1324.
- [74] T. Maki, A. Kawamura, N. Kato, J. Ohkanda, *Mol. Biosyst.* **2013**, *9*, 940–943.
- [75] A. S. Ripka, D. H. Rich, *Curr. Opin. Chem. Biol.* **1998**, *2*, 441–452.
- [76] R. E. Babine, S. L. Bender, *Chem. Rev.* **1997**, *97*, 1359–1472.
- [77] D. Leung, G. Abbenante, D. P. Fairlie, *J. Med. Chem.* **2000**, *43*, 305–341.
- [78] J. Cooper, W. Quail, C. Frazao, S. I. Foundling, T. L. Blundell, C. Humblet, E. A. Lunney, W. T. Lowther, B. M. Dunn, *Biochemistry*, **1992**, *31*, 8142–8150.
- [79] L. Coates, P. T. Erskine, M. P. Crump, S. P. Wood, J. B. Cooper, *J. Mol. Biol.* **2002**, *318*, 1405–1415.
- [80] L. Coates, P. T. Erskine, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Biochemistry*, **2001**, *40*, 13149–13157.
- [81] I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *2*, 157–162.
- [82] M. R. Arkin, J. A. Wells, S. S. Francisco, *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317.
- [83] A. A. Bogan, K. S. Thorn, *J. Mol. Biol.* **1998**, *1*, 1–9.
- [84] B. Ma, M. Shatsky, H. J. Wolfson, *Protein science*, **2002**, *11*, 184–197.
- [85] I. Luque, E. Freire, *Proteins*, **2000**, *71*, 63–71.
- [86] S. J. Teague, *Nat. Rev. Drug Discov.* **2003**, *2*, 527–541.

Chapter 2

Fragment-based drug design facilitated by protein-templated click chemistry: fragment linking and -optimization of inhibitors of the aspartic protease endothiapepsin

Fragment-based drug design (FBDD) enables the efficient design of bioactive compounds. While there are numerous reports on FBDD using optimization of a hit by fragment growing/optimization, fragment linking has rarely been used. Protein-templated click chemistry is a hit-identification strategy, in which azides and alkynes are assembled irreversibly to afford the corresponding triazoles. In this chapter, we demonstrate that the strategic combination of fragment linking/optimization and protein-templated click chemistry (PTCC) 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 μ M, represents the first example of a triazole-based inhibitor of endothiapepsin.

M. Mondal*, **M. Y. Unver***, A. Pal, M. Bakker, S. Berrier and A. K. H. Hirsch, *Chem. Eur. J.* **2016**, 22, 14826–14830.

*(shared first author)

2.1 Introduction

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. Over the past 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 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 in clinical trials.^[5] Vemurafenib received FDA approval for the treatment of late-stage melanoma in 2011,^[6] representing the first approved drug designed by FBDD.^[7] In 2016, Venetoclax was approved for the treatment of chronic lymphocytic leukemia, being the second fragment-based designed drug in the market.^[8]

Upon identification of a fragment,^[9] 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,^[10-15] even though it is time-consuming as 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.^[16] 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.^[17,18] It is presumably due to these challenges that there are only few reports of fragment linking,^[4,19] demonstrating the efficiency of linking low-affinity fragments to higher-affinity binders.^[18,20-26] We have recently reported a combination of DCC and fragment linking/optimization, which reduces the risks associated with fragment linking.^[27]

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.^[28,29] PTCC relies on the bio-orthogonal 1,3-dipolar cycloaddition of azide and alkyne building blocks facilitated by the protein target.^[30] 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 the proper manner, the protein “clicks” them together to afford its own triazole inhibitor (Figure 1).

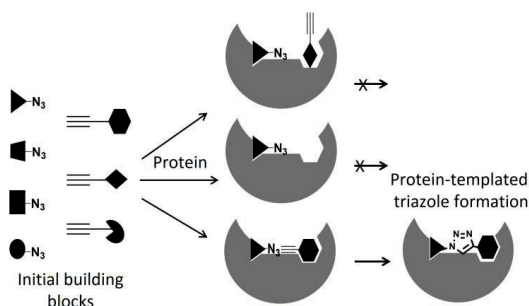


Figure 1. Schematic representation of protein-templated click chemistry leading to a triazole-based inhibitor starting from a library of azides and alkynes.

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, which are widely found in fungi, vertebrates, plants, as well as in HIV retro-viruses. This class of enzymes plays a causative role in several important diseases such as malaria, Alzheimer’s disease, hypertension and HIV.^[31] Owing to the high degree of similarity, endothiapepsin has served as a model enzyme for mechanistic studies^[32-34] as well as for the identification of inhibitors of renin^[35] and β -secretase.^[36] Endothiapepsin is a robust enzyme, is available in large quantity, crystallizes easily and remains active at room temperature for more than three weeks, making this enzyme a convenient representative for aspartic proteases.^[37] 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.^[33,34]

Although, the linkage of two known inhibitors of acetylcholinesterase via a triazole linker using PTCC has been pioneered by Sharpless and coworkers in 2005, the inhibitors that are linked do not qualify as fragments according to rule of three.^[29] To the best of our knowledge, there is no

report of fragment linking of actual fragments in a true FBDD context 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.

2.2 Results and discussion

2.2.1 Fragment-based drug design

We used X-ray crystal structures of endothiapepsin in complex with fragments **1** and **2** (Protein Data Bank (PDB) codes: 3PBZ and 3PLD, respectively, Figure 2), identified by Klebe and co-workers.^[38] 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",^[39] 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 Å² and 49.9 Å², 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.

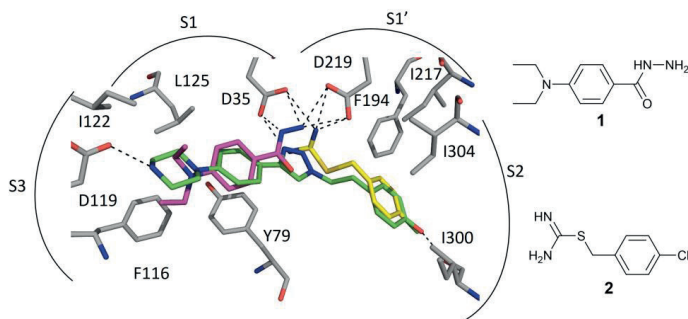


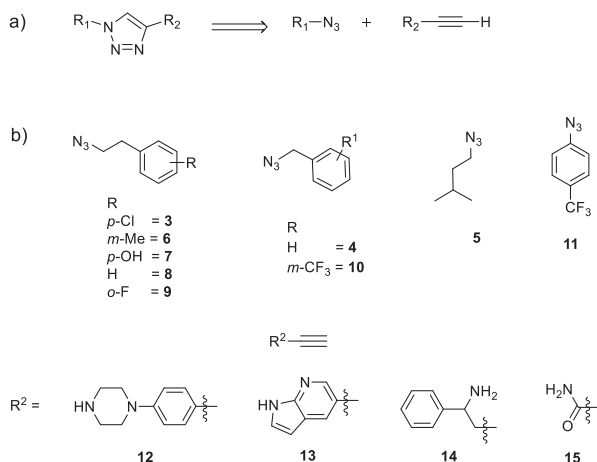
Figure 2. X-ray crystal structure of endothiapepsin in complex with fragments **1** and **2** (PDB code: 3PBZ and 3PLD, respectively) and a modeled potential triazole inhibitor in the active site.^[38] Color code: protein skeleton: C: gray, O: red, and N: blue; fragment skeleton: C: purple, yellow and green, N: blue, O: red, Cl: green. Hydrogen bonds below 3.0 Å are shown as black, dashed lines.^[40]

Fragments **1** and **2** occupy the S3 and S1 and the S2 and S1' pockets, respectively and address the catalytic dyad using a H-bonding network (Figure 2). With the help of the molecular-modeling software Moloc^[41] and the FlexX docking module in the LeadIT suite,^[42] we linked these two fragments using a triazole linker. The newly introduced triazole moiety resides at the

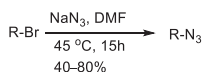
junction of the S1 and S1' pockets, where hydrazide and amidine groups of fragment **1** and **2**, respectively, were positioned. The triazole linker appeared to be ideally suited to address the catalytic dyad via a H-bonding network. Although, the protonation of 1,2,3-triazoles at pH 4.6, optimal for endothiapepsin, is unprecedented, given that their pK_a value in water is 1.2,^[43] 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.^[44] Hence, under the acidic condition, 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^[37,38] as well as hotspot analysis^[45] 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 piperazinyl 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**.

2.2.2 Synthesis of building blocks (azides and alkynes)

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 library, which were available from Syncom. All azides were obtained from their corresponding bromides by treatment with sodium azide in 40–80% yield (Scheme 2).^[46,47]

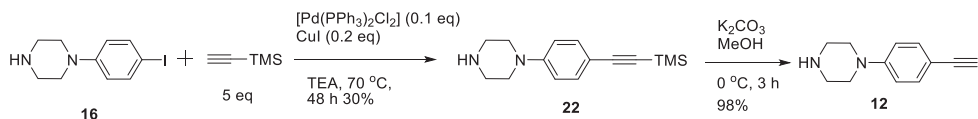


Scheme 1. a) Structures and retrosynthetic analysis of the designed triazole inhibitors starting from fragments **1** and **2**; b) structures of the azides **3–11** and the alkynes **12–15**.



Scheme 2. Synthesis of azides **3–11** starting from their corresponding bromides.

We synthesized alkyne **12** using a Sonogashira cross-coupling reaction, starting from the iodide **16** (Scheme 3).

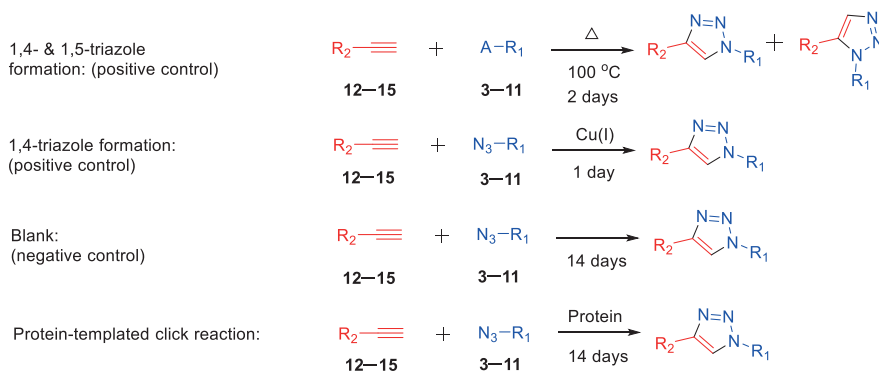


Scheme 3. Synthesis of alkyne **12**.

2.2.3 Generation of the library

We set up a library, consisting of four alkynes **12–15** (100 μM each) and nine azides **3–11** (100 μM each), in 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 4). 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),^[37] 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 (Scheme 4). We also compared the PTCC reaction with the blank reaction (without protein) as well as the protein alone.



Scheme 4. From bottom to top; Protein-templated triazole formation, blank reaction, Cu (I) catalyzed 1,4-triazole formation, Huisgen cycloaddition to afford 1,4- and 1,5-triazoles.

We identified a total of four 1,4-triazoles (**17-20**), which are formed only in the presence of protein (Figure 3). To establish that the active site of intact endothiapepsin is required for PTCC, we set up two control experiments. Repeating the reaction in presence of saquinavir (100 μM , a strong inhibitor, $K_i = 48\text{ nM}$), or a catalytic amount of BSA (26 μM) did not lead to the formation of any triazoles.

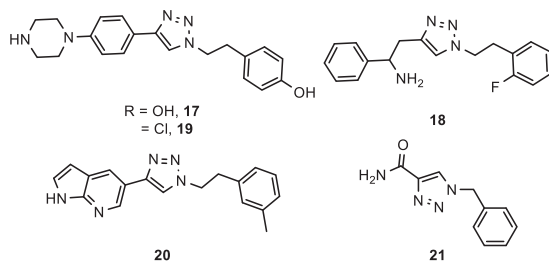


Figure 3. Structure of triazoles **17-20** identified using PTTC, inactive triazole **21**

A representative UPLC-TOF-SIM analysis of triazole **19** ($M+H^+ = 368$) is shown in Figure 4.

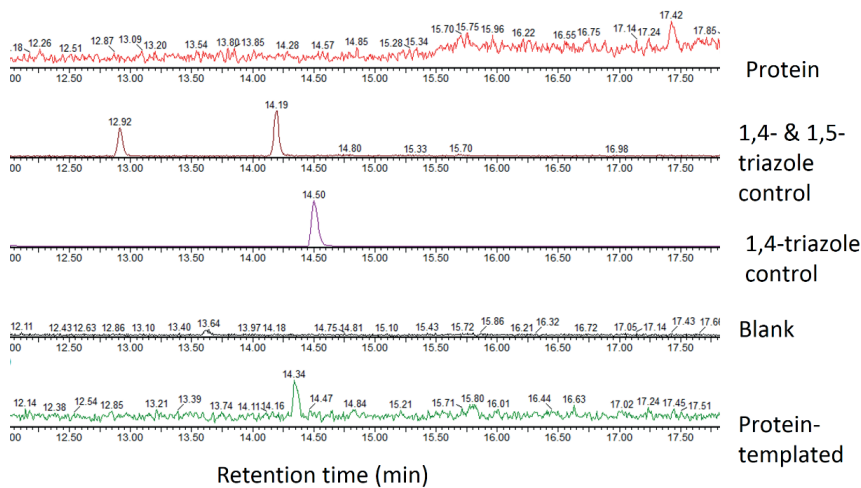
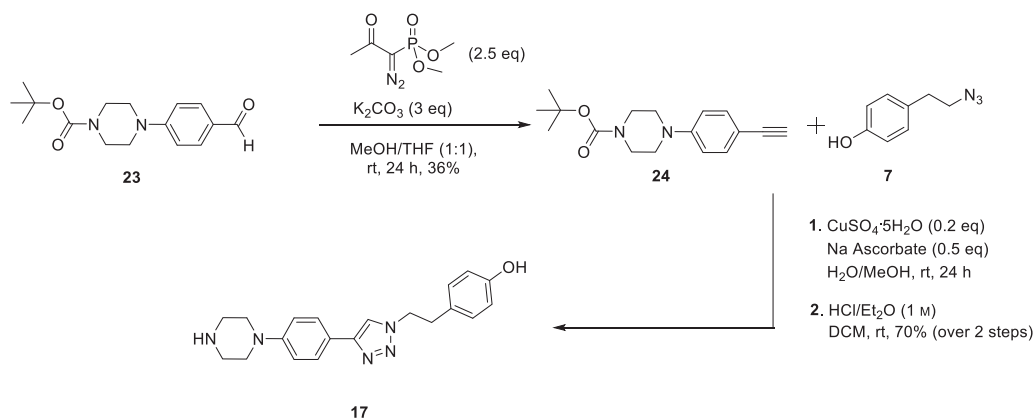


Figure 4. UPLC-TOF-SIM analysis of triazole **19** ($M+H^+ = 368$). Formation of **19** in PTCC was compared with the blank reaction, Cu (I) catalyzed 1,4-triazole formation, Huisgen cycloaddition for both 1,4- and 1,5-triazoles formation and protein fragmentation.

2.2.4 Synthesis of identified triazoles

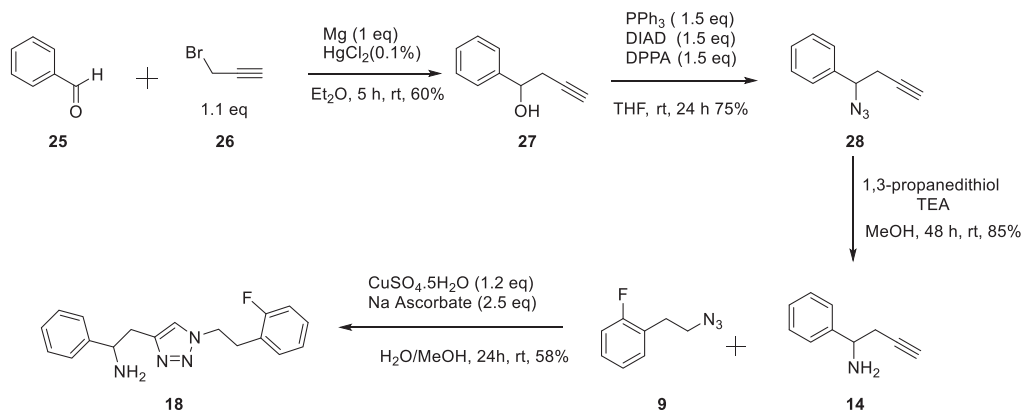
To investigate the biochemical activity of the binders identified by PTCC, we synthesized all four triazoles from their corresponding azide and alkyne precursors using the Cu(I)-catalyzed 1,3-cycloaddition. In addition, we synthesized an inactive triazole **21** to demonstrate the efficiency of PTCC. Their inhibitory activity was measured using a fluorescence-based assay adapted from the HIV-protease assay.^[48] We used Abz-Thr-Ile-Nle-*p*-nitro-Phe-Gln-Arg-NH₂ as a fluorogenic substrate, during measurement the fluorescence increased because of substrate hydrolysis by endothiapepsin. The initial slopes of the fluorescence in the inhibitor-containing wells were compared to the initial slope of the blanks for data analysis.

We synthesized triazole **17** starting from Boc-protected **23** in a Ohira-Bestman reaction to afford the corresponding alkyne **24** in 36% yield, followed by Cu(I)-catalyzed 1,3-cycloaddition with azide and deprotection of the Boc-protected piperazine (Scheme 5).



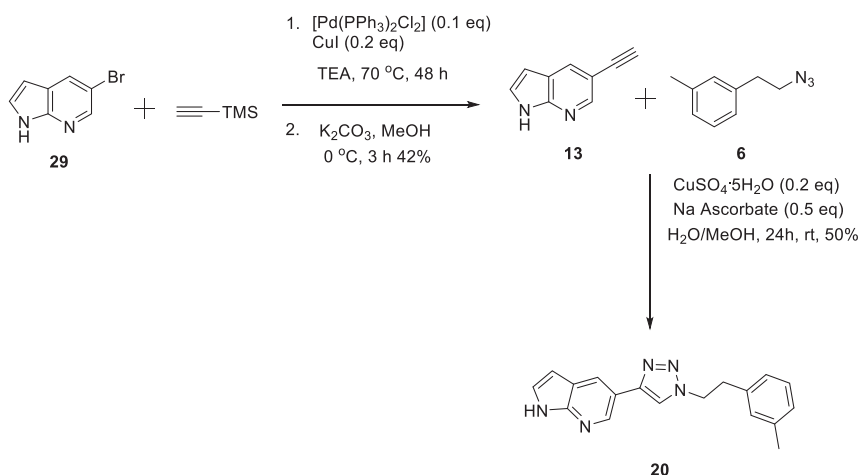
Scheme 5. Synthesis of triazole **17**, identified by PTCC.

Synthesis of triazole **18** was achieved starting from benzaldehyde (**25**) and propargyl bromide (**26**) in the presence of HgCl₂ to block the alkynyl proton (adapted protocol, S. Sezer et al.)^[49] to afford alcohol **27** in 60% yield, which was converted to the corresponding azide **28** in a Mitsunobu reaction in 75% yield. The azide was then reduced to amine **14** in 85% yield, followed by Cu(I)-catalyzed 1,3-cycloaddition with azide **9** (Scheme 6).



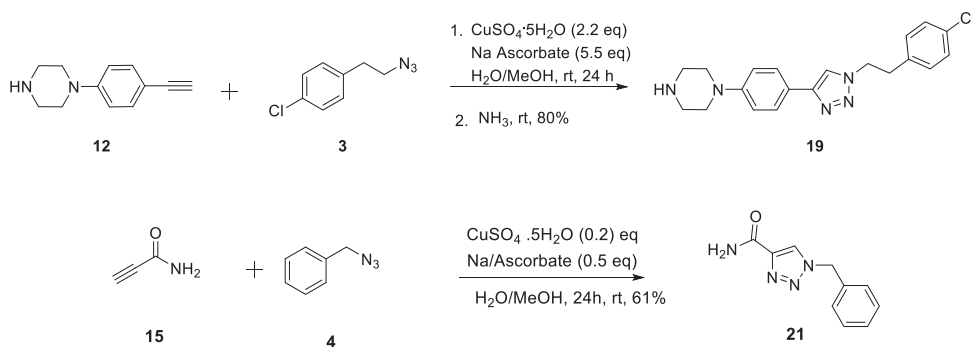
Scheme 6. Synthesis of triazole **18**, identified by PTCC.

Alkyne **13** was synthesized using Sonogashira cross-coupling reaction starting from the corresponding halide **29**, which was then used for the Cu(I)-catalyzed 1,3-cycloaddition reaction with azide **6** to afford triazole **20**.



Scheme 7. Synthesis of triazole **20**, identified by PTCC.

Finally, we synthesized triazole **19** starting from the corresponding azide and alkyne. In addition, we synthesized an inactive triazole **21** to demonstrate the efficiency of PTCC in selecting the most potent inhibitors (Scheme 8).



Scheme 8. Synthesis of triazole **19** (identified by PTCC) and **21** (inactive compound in PTTC).

2.2.5 Biochemical evaluation

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 (Figure 5 for **17**). 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 (Figure 5, Table 1).

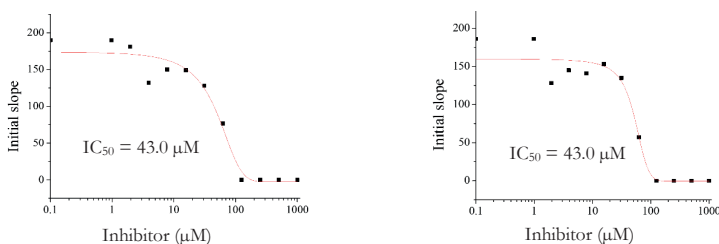


Figure 5. IC_{50} inhibition curves of **17** ($\text{IC}_{50} = 43 \pm 0 \mu\text{M}$). The inhibitor was measured in duplicate.

The experimental Gibbs free energies of binding (ΔG) and ligand efficiencies (LE), derived from the experimental IC_{50} values using the Cheng-Prusoff equation,^[50] correlate with the calculated values using the scoring function HYDE in the LeadIT suite ($\Delta G_{\text{HYDE}}(\mathbf{17}) = -25 \text{ kJ mol}^{-1}$, Table 1).^[51,52] This correlation is also valid for the other triazole inhibitors (Table 1).

Table 2. The IC₅₀ values, ligand efficiency (LE), calculated and experimental Gibbs free energy of binding (ΔG) of triazole inhibitors.

Inhibitors	IC ₅₀ ^[a] (μM)	$\Delta G_{\text{EXPT}}^{\text{[b]}}$ (kJmol^{-1})	LE ^[b]	$\Delta G_{\text{HYDE}}^{\text{[c]}}$ (kJmol^{-1})
17	43 \pm 0	-27	0.25	-25
18	94 \pm 18	-25	0.26	-19
19	121 \pm 3	-24	0.22	-25
20	Insoluble	-	-	-23
7	No inhibition	-	-	-
9	No inhibition	-	-	-
12	142 \pm 52	-	-	-
14	No inhibition	-	-	-

^a 26 experiments were performed and only 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 IC₅₀ values, ^c Values indicate the calculated Gibbs free energy of binding (ΔG_{HYDE} , calculated by the HYDE scoring function in the LeadIT suite).

2.2.6 Discussion

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 its limited solubility, we were unable to obtain crystals of **17** with endothiapepsin. Based on the inhibitory potencies, replacement of -Cl in **19** by a -OH group in **17**, leads to a decrease in IC₅₀ value from 121 μM 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 6a). Moreover, the alkyne **12** displays an IC₅₀ value of 142 μM (Table 1) 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**.

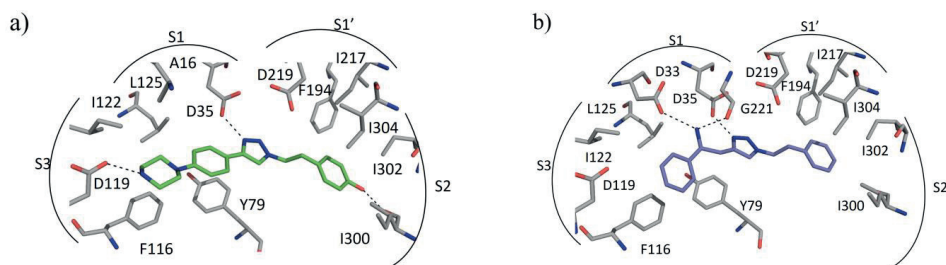


Figure 6. Moloc-generated modeled structures of: a) **17**, and b) (*S*)-**18** in the active site of endothiapepsin. Color code: inhibitor skeleton: C: green, violet, N: blue, O: red; enzyme skeleton: C: gray. H bonds below 3.2 Å are shown as black, dashed lines.

According to modeling and docking, as shown in Figures 6a and 7, respectively, both **17** and **19** address the catalytic dyad using their triazole 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**.

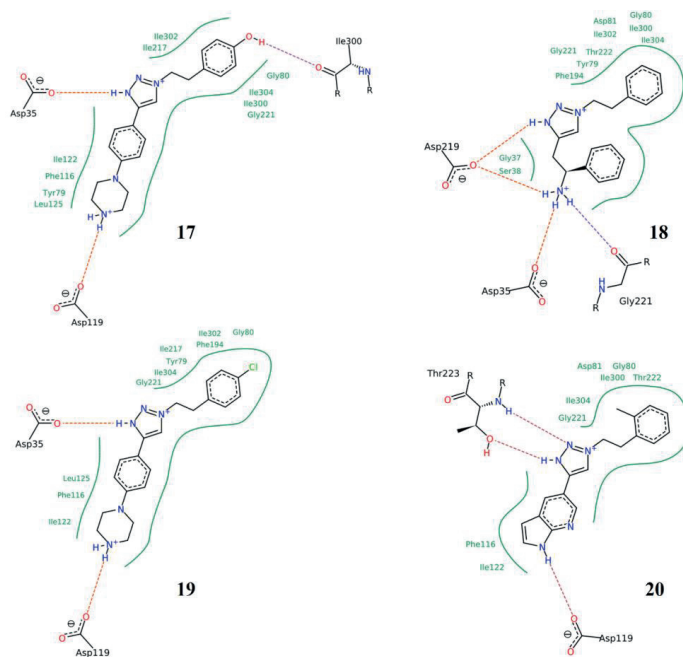


Figure 7. Schematic representation of the predicted binding modes of triazole-based inhibitors **17–20** in the active site of endothiapsin. These binding modes are the result of a docking run using the FlexX docking module with 30 poses and represent the top-scoring pose after HYDE scoring and careful visual inspection to exclude poses with significant inter- or intra-molecular clash terms or unfavorable conformations. The figures were generated with PoseView^[53] as implemented in the LeadIT[®] suite.^[42]

Triazole **18** displays an IC_{50} of 94 μM and (*S*)-**18** addresses the catalytic dyad using its triazole linker to form a direct H bond with D35 as indicated by modeling and docking studies (Figures 6b and 7). The –NH₂ 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, L125 in the S3 pocket, and I300, I302,

I304, F194, I217 in the S2 pocket, which preserve the binding mode of fragments **1** and **2**, respectively.

2.3 Conclusions

In conclusion, we 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 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 μ M, and is the first example of a triazole-based inhibitor of endothiapepsin. Due to the limited solubility of the triazoles identified; we were unable to obtain crystals of any triazole in complex with 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 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.

2.4 Experimental section

2.4.1 Fluorescence-based inhibition assay

Endothiapepsin was purified from Suparen[®] (kindly provided by DSM Food Specialties) by exchanging the buffer to sodium acetate buffer (0.1 M, pH 4.6) using a Vivaspin 500 with a molecular weight cutoff at 10,000 Da. Measurement of the absorption at 280 nm, assuming an extinction coefficient of 1.15 for 1 mg/mL solutions, afforded the protein concentration.^[54] Stock solutions (100 mM in DMSO) were prepared for all triazoles **17–21**. The final reaction volume was 200 μ L containing 0.4 nM endothiapepsin, 1.8 μ M substrate and 2.1% DMSO. The final concentration of inhibitors was 1000 μ M in the first well and subsequent half dilution in the next 10 wells. In the same way, blanks were prepared using DMSO instead of the inhibitor stock solutions. As substrate, Abz-Thr-Ile-Nle-*p*-nitro-Phe-Gln-Arg-NH₂ (purchased from Bachem) was used for the fluorescence screening assay. The assay was performed with flat bottom 96-well microplates (purchased from Greiner Bio-One) using a Synergy Mx microplate reader at an excitation wavelength of 337 nm and an emission wavelength of 414 nm. The K_m of the substrate toward endothiapepsin was known, 1.6 μ M.^[38] The assay buffer (0.1 M sodium acetate buffer, pH 4.6, containing 0.001% Tween 20) was premixed with the substrate and inhibitor; endothiapepsin was added directly before the measurement. As the substrate is a fluorogenic

substrate, during measurement the fluorescence increased because of substrate hydrolysis by endothiapepsin. The initial slopes of the fluorescence in the triazole-containing wells were compared to the initial slope of the blanks for data analysis. Each compound was measured in duplicate. The final result represents the average of both measurements.

2.4.2 Modeling and docking

Two X-ray crystal structures of complexes of endothiapepsin (PDB codes: 3PBZ and 3PLD) were used for our modeling.^[38] Several triazoles were designed using fragment linking. The energy of the system was minimized using the MAB force field as implemented in the computer program MOLOC,^[41] whilst keeping the protein coordinates fixed for the PDB code: 3PBZ. In all cases, the triazole addresses the catalytic dyad directly via hydrogen-bonding interactions. Taking inspiration from the co-crystal structures of endothiapepsin with eleven fragments,^[38] as well as from hot-spot analysis^[45] of the active site of endothiapepsin, several triazoles with different aromatic and aliphatic substituents were designed and subsequent energy minimization (MAB force field) was done using MOLOC. All types of interactions (hydrogen bonds and lipophilic interactions) between designed triazoles and protein were measured in MOLOC. All the designed triazoles were subsequently docked into the active site of endothiapepsin by using the FlexX docking module in the LeadIT suite.^[42] During the docking, the binding site in the protein was restricted to 9 Å around the cocrystallized ligands, and the 30 top (FlexX)-scored solutions were retained, and subsequently post-scored with the HYDE^[51] module in LeadIT v.2.1.3.^[42] After careful visualization to exclude poses with significant inter- or intramolecular clash terms or unfavorable conformations, the resulting solutions were subsequently ranked according to their binding energies. The top-ranked solutions identified in this way were chosen.

2.4.3 PTCC experiments

Protein-templated triazole formation using endothiapepsin: Endothiapepsin (25 µL, 1.052 mM in sodium acetate buffer 0.1 M, pH 4.6), the nine azides **3–11** (1 µL each, 100 mM in DMSO), alkynes **12–15** (1 µL each, 100 mM in DMSO) were added to a mixture of DMSO (62 µL) and sodium acetate buffer (900 µL, 0.1 M, pH 4.6). The reaction mixture was shortly vortexed and was allowed to stand at room temperature with occasional shaking. After 14 d, the library was analyzed by UPLC-TOF-SIM (electro-spray ionization, (ES+)) measurement because of its higher sensitivity and greater reliability for product identification.

Protein-templated triazole formation using endothiapepsin in presence of saquinavir:

Endothiapepsin (25 µL, 1.052 mM in sodium acetate buffer 0.1 M, pH 4.6), the nine azides **3–11** (1 µL each, 100 mM in DMSO), alkynes **12–15** (1 µL each, 100 mM in DMSO) and saquinavir

(1 μL , 100 mM in DMSO) were added to a mixture of DMSO (61 μL) and sodium acetate buffer (900 μL , 0.1 M, pH 4.6). The reaction mixture was shortly vortexed and was allowed to stand at room temperature with occasional shaking. After 14 d, the library was analyzed by UPLC-TOF-SIM (electro-spray ionization, (ES+)) measurement because of its higher sensitivity and greater reliability for product identification.

Blank reaction, negative control: The nine azides **3–11** (1 μL each, 100 mM in DMSO), alkynes **12–15** (1 μL each, 100 mM in DMSO) were added to a mixture of DMSO (62 μL) and sodium acetate buffer (925 μL , 0.1 M, pH 4.6). The reaction mixture was shortly vortexed and was allowed to stand at room temperature with occasional shaking. After 14 d, the library was analyzed by UPLC-TOF-SIM (ES+) measurement for the positive hits and compared it with protein-templated triazole formation.

Protein-templated triazole formation using BSA: BSA (2.7 μL , 10.0 mM in sodium acetate buffer 0.1 M, pH 4.6), the nine azides **3–11** (1 μL each, 100 mM in DMSO), alkynes **12–15** (1 μL each, 100 mM in DMSO) were added to a mixture of DMSO (62 μL) and sodium acetate buffer (922 μL , 0.1 M, pH 4.6). The reaction mixture was shortly vortexed and was allowed to stand at room temperature with occasional shaking. After 14 d, the library was analyzed by UPLC-TOF-SIM (electro-spray ionization, (ES+)) measurement because of its higher sensitivity and greater reliability for product identification.

Cu(I)-catalyzed 1,4-triazole formation, positive control: The nine azides **3–11** (4 μL each, 100 mM in DMSO), alkynes **12–15** (9 μL each, 100 mM in DMSO) were added to a mixture of DMSO (3 μL), aqueous copper sulfate (36 μL , stock solution 0.05 M), aqueous sodium ascorbate (360 μL , 100 mM) and sodium acetate buffer (529 μL , 0.1 M, pH 4.6). The reaction mixture was shortly vortexed and allowed to stand at room temperature with occasional shaking. After 1 d, the reaction was diluted with sodium acetate buffer (4x) and analyzed by UPLC-TOF (ES+) measurement and compared with the positive hits identified from protein-templated triazole formation.

Huisgen cycloaddition reaction, affording both 1,4- and 1,5-triazoles, positive control: The nine azides **3–11** (4 μL each, 100 mM in DMSO) and alkynes **12–15** (9 μL each, 100 mM in DMSO) were mixed. The reaction mixture was shortly vortexed and allowed to heat at 100 $^{\circ}\text{C}$ for 2 d with occasional shaking. After 2 d, the reaction was diluted with of sodium acetate buffer (5 mL, 0.1 M, pH 4.6) and analyzed by UPLC-TOF (ES+) measurement and compared with the positive hits identified from protein-templated triazole formation.

Protein: Endothiapepsin (25 μL , stock solution 1.052 mM in sodium acetate buffer 0.1 M, pH 4.6) was added to 75 μL of DMSO and 900 μL sodium acetate buffer 0.1 M, pH 4.6. After 14 days, the enzyme solution was analyzed by UPLC-TOF-SIM (ES+) measurements and compared with the positive hits identified from protein-templated triazole formation.

UPLC-TOF-SIM method

UPLC-TOF was performed using a Waters Acquity UPLC H-class system coupled to a Waters Xevo-G2 TOF. All analyses were performed using a reversed phase UPLC column (ACQUITY BEH C8 Column, 130 \AA , 1.7 μm , 2.1 mm x 150 mm). Positive-ion mass spectra were acquired using ES ionization, injecting 3 μL of sample; column temperature 35 $^{\circ}\text{C}$; flow rate 0.3 mL/min. The eluents, acetonitrile and water contained 0.1% of formic acid. The library components were eluted with a gradient from 95% \rightarrow 59% over 20 min, then at 5% over 1 min, followed by 5% for 2 min.

The UPLC-TOF-SIM method was used to analyze the formation of triazoles in PTCC and blank reactions. SIM measurements are highly sensitive, where a minute amount of compound can be detected by the mass spectrometer. $[M+H]^+$ were monitored using the ($M \rightarrow M+4$) mass range to ensure correct isotope patterns for all possible potential triazole products both in PTCC and blank reactions. The 1,3-cycloaddition products in the protein-templated reaction were identified by comparison of their retention time with those formed by the Cu(I)-catalyzed reaction, Huisgen cycloaddition reaction and by their molecular weight.

2.4.4 General experimental details

Starting materials and reagents were purchased from Aldrich or Acros. Yields refer to analytically pure compounds and have not been optimized. All solvents were reagent-grade and if necessary, SPS-grade. Column chromatography was performed on silica gel (Silicycle[®] SiliaSepTM 40–63 μm 60 \AA). TLC was performed with silica gel 60/Kieselguhr F254. Solvents used for the column chromatography were pentane, ethyl acetate, dichloromethane and methanol. ^1H and ^{13}C ^{19}F spectra were recorded at 400 MHz on a Varian AMX400 spectrometer (400 MHz for ^1H , 101 MHz for ^{13}C and 376 MHz for ^{19}F) at 25 $^{\circ}\text{C}$.

Chemical shifts (δ) are reported relative to the residual solvent peak. Splitting patterns are indicated as (s) singlet, (d) doublet, (t) triplet, (q) quartet, (m) multiplet, (br) broad. The coupling constants (J) are given in Hz. High-resolution mass spectra were recorded with an FTMS orbitrap (Thermo Fisher Scientific) mass spectrometer. FT-IR were measured on a PerkinElmer FT-IR spectrometer. Melting points were measured on a Stuart[®] SMP11 50 W melting point apparatus.

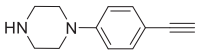
HPLC conditions: column, Symmetry® C8 3.5µm, 4.6 x 150 mm; flow rate 0.5 mL min⁻¹; wavelength, 254 nm; temperature, 23 °C; gradient, H₂O/MeCN (0.1% TFA) from 95% → 5% over 15 min, then at 5% for 1 min.

2.4.5 Synthesis of azides, alkynes and triazoles

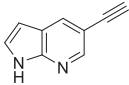
General procedure for azide synthesis

To a 5-mL flask under a nitrogen atmosphere, the corresponding bromide (0.7 mmol, 1 eq) was added, followed by addition of dry DMF (2 mL) and NaN₃ (228 mg, 3.5 mmol). The reaction was stirred at 45 °C overnight. The reaction mixture was dissolved in water (10 mL) and extracted with dichloromethane (3 x 5 mL). The organic layers were combined, dried over Na₂SO₄, filtered and evaporated. The remaining liquid was dissolved in diethyl ether and extracted with aqueous HCl solution (2–3 x 0.5 M) to remove residual dimethylformamide. This provided the pure products in 40–80% yield. All azides are commercially available or known in the literature.^[46,47]

1-(4-Ethynylphenyl)piperazine (**12**)

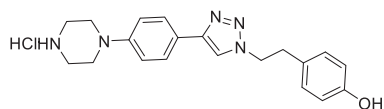
 To a 100 mL Schlenk flask charged with 1-(4-iodophenyl)piperazine (100 mg, 0.35 mmol), [Pd(PPh₃)₂Cl₂] (0.1 eq), CuI (0.2 eq), distilled and degassed Et₃N (30 mL) and trimethylsilylacetylene (0.25 mL, 1.75 mmol) were added. The resulting reaction mixture was heated for 24 h at 70 °C. The resulting mixture was concentrated in vacuo. A quick purification with silica gel using DCM/MeOH (85:15) as an eluent, afforded the crude product **22**, which was used for the following deprotection step. A 50 mL flask charged with the crude product **22** and a saturated solution of potassium carbonate in MeOH (10 mL) was left to stir at 0 °C for 3 h. To the reaction mixture, water (20 mL) was added and extracted with DCM (2 x 20 mL). The combined organic phases were dried over Mg₂SO₄, filtered and concentrated in vacuo to afford the desired compound **12** as a brown solid in 30% yield.^[55,56] The spectral data correspond to those reported in the literature.^[57]

5-Ethynyl-1*H*-pyrrolo[2,3-*b*]pyridine (**13**)

 To a 25-mL Schlenk flask charged with 5-bromo-1*H*-pyrrolo[2,3-*b*]pyridine (**29**, 250 mg, 1.26 mmol), [Pd(PPh₃)₂Cl₂] (0.1 eq), CuI (0.2 eq), distilled and degassed Et₃N (10 mL) and trimethylsilylacetylene (1.3 mL, 8.90 mmol) were added. The resulting reaction mixture was heated to 80 °C for 48 h, and concentrated in vacuo. Purification with silica gel using mixture of ethyl acetate/pentane (1:4) as an eluent afforded the crude product, which was directly used in the following step without any further purification. A flask charged with a crude and a saturated solution of potassium carbonate (20 mL) was left to stir at 0 °C for 3 h. To the reaction mixture, water (20 mL) was added and then extracted with DCM (2 x 20 mL). The

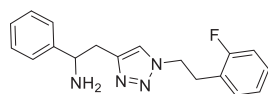
combined organic phases were dried over Mg_2SO_4 , filtered and concentrated in vacuo to afford desired compound **13** as brown solid in 42% yield.^[55,56] The spectral data correspond to those reported in the literature.^[58]

4-(2-(4-(4-(Piperazin-1-yl)phenyl)-1H-1,2,3-triazol-1-yl)ethyl)phenol (**17**)



To a solution of *tert*-butyl 4-(4-ethynylphenyl)piperazine-1-carboxylate **24** (60 mg, 0.21 mmol) and 4-(2-azidoethyl)phenol **9** (29 mg, 0.17 mmol) in 3 mL $\text{H}_2\text{O}/\text{MeOH}$ (1:1) mixture, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (0.2 eq) and Na/Ascorbate (0.5 eq) were added. After stirring for 24 h at room temperature, the reaction mixture was diluted with water (10 mL) and extracted with DCM (3 x 10 mL). The separated organic layer was dried over MgSO_4 , filtered and concentrated in vacuo. A quick purification with column chromatography using mixture of MeOH/DCM (1:9) as an eluent, afforded *N*-*boc* protected triazole, which was directly dissolved in dichloromethane (2 mL) and $\text{HCl}/\text{diethyl ether}$ (1 M, 4 mL) was added under an argon atmosphere. The mixture was stirred for 15 h at 20 °C. The resulting light-brown precipitate was collected and washed with Et_2O . Complete removal of the solvent afforded the HCl salt of the desired compound **17** as a pale brown solid in 70% yield. mp: > 230 °C (decomposition). ^1H NMR (400 MHz, CD_3OD) δ 8.52 (s, 1H), 7.69 (d, $J = 8.4$ Hz, 2H), 7.17 (d, $J = 8.4$ Hz, 2H), 6.99 (d, $J = 8.2$ Hz, 2H), 6.69 (d, $J = 8.2$ Hz, 2H), 4.76 (t, $J = 7.1$ Hz, 2H), 3.56 (t, $J = 5.2$ Hz, 4H), 3.39 (t, $J = 5.2$ Hz, 4H), 3.21 (t, $J = 7.1$ Hz, 2H). ^{13}C NMR (101 MHz, CD_3OD) δ 157.7, 152.6, 145.5, 130.9, 128.7, 128.3, 124.5, 118.5, 117.8, 116.6, 55.3, 46.8, 44.6, 36.1. IR (cm^{-1}): 2929, 1743, 1459, 1240, 1164, 1031. HRMS (ESI) calcd for $\text{C}_{20}\text{H}_{24}\text{N}_5\text{O}$ $[M+\text{H}]^+$: 350.1975, found: 350.1977.

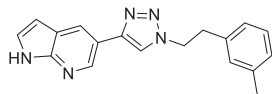
2-(1-(2-Fluorophenethyl)-1H-1,2,3-triazol-4-yl)-1-phenylethan-1-amine (**18**)



To a solution of 1-phenylbut-3-yn-1-amine (**14**, 50 mg, 0.34 mmol) and 1-(2-azidoethyl)-2-fluorobenzene (**7**, 47 mg, 0.28 mmol) in water/methanol (4 mL, 1:1) mixture, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (1.2 eq) and Sodium ascorbate (2.5 eq) were added. After stirring for 24 h at room temperature, the reaction mixture was diluted with water (10 mL) and extracted with DCM (3 x 10 mL). The separated organic layer was dried over MgSO_4 , filtered and concentrated in vacuo. After purification with column chromatography using mixture of MeOH/DCM (1:9) as an eluent, the desired compound **18** was obtained as an off-white solid in 58% yield. mp: 56–57 °C. ^1H NMR (400 MHz, CDCl_3) δ 7.37–7.18 (m, 6H), 7.06–6.90 (m, 4H), 4.52 (t, $J = 7.2$, 2H), 4.32 (t, $J = 6.7$, 1H), 3.19 (t, $J = 7.1$, 2H), 3.10–2.92 (m, 2H), 2.63 (br s, 3H). ^{13}C NMR (101 MHz, CDCl_3) δ 161.1 (d,

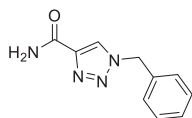
$^1J(\text{C},\text{F}) = 245.6$), 131.1 (d, $^3J(\text{C}-\text{F}) = 4.6$), 129.0 (d, $^3J(\text{C}-\text{F}) =$), 128.5 (2C), 127.3, 126.5, 124.3 (d, $^4J(\text{C}-\text{F}) = 3.6$), 123.9 (d, $^2J(\text{C}-\text{F}) = 15.6$), 122.2, 115.4 (d, $^2J(\text{C}-\text{F}) = 21.7$), 113.7, 55.5, 49.8, 35.3, 30.5. ^{19}F NMR (376 MHz, CDCl_3) δ -119.06. IR (cm^{-1}) 3365, 1933, 1589, 1493, 1451, 1363, 1229, 1052, 759, 702. HRMS (ESI) calcd for $\text{C}_{18}\text{H}_{19}\text{FN}_4$ $[M+\text{H}]^+$: 311.1666, found: 311.1668.

5-(1-(3-Methylphenethyl)-1H-1,2,3-triazol-4-yl)-1H-pyrrolo[2,3-b]pyridine (20)

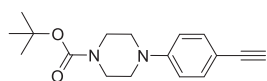


To a solution of *tert*-butyl 5-ethynyl-1H-pyrrolo[2,3-b]pyridine (**13**, 28 mg, 0.24 mmol) and 1-(2-azidoethyl)-3-methylbenzene (**6**, 35 mg, 0.2 mmol) in a water/methanol (4mL, 1:1) mixture, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (0.2 eq) and sodium ascorbate (0.5 eq) were added. After stirring for 24 h at room temperature, the reaction mixture was diluted with water (10 mL) and extracted with DCM (3 x 10 mL). The separated organic layer was dried over MgSO_4 , filtered and concentrated in vacuo. After purification with column chromatography using mixture of MeOH/DCM (1:9) as an eluent, the desired compound **20** was obtained as a brown solid in 50% yield. mp: 70–72 °C. ^1H NMR (400 MHz, $(\text{CD}_3)_2\text{SO}$) δ 11.73 (s, 1H), 8.68 (s, 1H), 8.54 (s, 1H), 8.34 (d, $J = 1.9$ Hz, 1H), 7.57 – 7.48 (m, 1H), 7.18 (t, $J = 7.5$ Hz, 1H), 7.09 (s, 1H), 7.06–7.01 (m, 2H), 6.52 (dd, $J = 3.5, 1.6$ Hz, 1H), 4.66 (t, $J = 7.4$ Hz, 2H), 3.20 (t, $J = 7.4$ Hz, 2H), 2.27 (s, 3H). ^{13}C NMR (101 MHz, $(\text{CD}_3)_2\text{SO}$) δ 148.1, 145.1, 140.2, 137.4, 137.4, 129.3 (2C), 128.3, 127.2, 127.0, 125.7 (2C), 124.4, 120.5, 100.1, 50.6, 35.5, 20.9. IR (cm^{-1}) 3661, 2980, 2886, 1383, 1256, 1158, 1072, 957. HRMS (ESI) calcd for $\text{C}_{18}\text{H}_{17}\text{N}_5$ $[M+\text{H}]^+$: 304.1556, found: 304.1558.

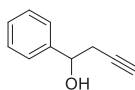
1-Benzyl-1H-1,2,3-triazole-4-carboxamide (21)



To a solution of propiolamide (**15**, 40 mg, 0.58 mmol) and (azidomethyl)benzene (**4**, 64 mg, 0.48) in a water/methanol (4 mL, 1:1) mixture, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (0.2 eq) and sodium ascorbate (0.5 eq) were added. After stirring for 24 h at room temperature, the reaction mixture was diluted with water (10 mL) and extracted with DCM (3 x 10mL). The separated organic layer was dried over MgSO_4 , filtered and concentrated in vacuo. After purification with column chromatography using mixture of MeOH/DCM (1:9) as an eluent, the desired compound **21** was obtained as yellow solid in 61% yield. mp > 225 °C (decomposition). ^1H NMR (400 MHz, $(\text{CD}_3)_2\text{SO}$) δ 8.59 (s, 1H), 7.85 (s, 1H), 7.46 (s, 1H), 7.42 – 7.28 (m, 5H), 5.64 (s, 2H). ^{13}C NMR (101 MHz, $(\text{CD}_3)_2\text{SO}$) δ 161.4, 143.1, 135.6, 128.8 (2C), 128.2, 127.9(2C), 126.6, 53.0. IR(cm^{-1}): 3406, 3093, 1649, 1404, 723. HRMS (ESI) calcd for $\text{C}_{10}\text{H}_{10}\text{N}_4\text{O}$ $[M+\text{H}]^+$: 203.0927, found: 203.0928.

***Tert*-butyl 4-(4-ethynylphenyl)piperazine-1-carboxylate (**24**)**

To a solution of 1-boc-4-(4-formylphenyl)piperazine **23** (200 mg, 0.69 mmol) in MeOH (10 mL), K_2CO_3 (285 mg, 2.07 mmol) and dimethyl(1-diazo-2-oxopropyl)phosphonate (185 μ L, 1.72 mmol) were added. After stirring for 24 h at room temperature, reaction mixture was diluted with water (30 mL) and extracted was done with DCM (3 x 30 mL). The separated organic layer was dried over $MgSO_4$, filtered and concentrated in vacuo. After purification with column chromatography using ethyl acetate/pentane (1:4) mixture as an eluent, the desired compound **24** was obtained as a white solid in 35% yield.^[59] The spectral data correspond to those reported in the literature.^[60]

1-Phenylbut-3-yn-1-ol (27**)**

To a stirred suspension of Mg turnings (720 mg, 30 mmol), iodine (two crystals) and $HgCl_2$ (10 mg) in anhydrous Et_2O (10 mL) at 25 °C, equipped with a reflux condenser, was added dropwise a mixture of propargyl bromide (**26**, 1.7 mL, 22 mmol) in anhydrous Et_2O (5 mL). The mixture was heated to reflux for 30 min. The mixture was cooled down to 0 °C and then benzaldehyde (**25**, 2 mL, 20 mmol) in anhydrous diethylether (3 mL) was added dropwise. The resultant mixture was stirred for 5 h at room temperature. The reaction mixture was hydrolyzed with saturated ammonium chloride solution (10 mL) and with 1 N HCl (2 mL). The resultant mixture was extracted with diethyl ether (3 x 30 mL). The combined organic phase was washed with brine (20 mL), dried over $MgSO_4$, and evaporated in vacuo. The crude product was purified by flash column chromatography using mixture of ethyl acetate/pentane (1:5) as an eluent, the desired compound was obtained as a light yellow oil in 60% yield.^[49] The spectral data correspond to those reported in the literature.^[61]

2.5 Contributions from co-authors

PTCC experiments, modelling/docking and biochemical evaluation were performed by M. Mondal, UPLC-TOF-SIM measurements were done by A. Pal and part of the synthesis was done by M. Bakker and S. Berrier during their Master's and Bachelor's projects, respectively.

2.6 References

- [1] D. A. Erlanson, R. S. McDowell and T. O. Brien, *J. Med. Chem.* **2004**, *47*, 3463–3482.
- [2] P. J. Hajduk, J. Greer, *Nat. Rev. Drug Discov.* **2007**, *6*, 211–219.
- [3] H. Chen, X. Zhou, A. Wang, Y. Zheng, Y. Gao, J. Zhou, *Drug Discov. Today* **2015**, *20*, 105–113.
- [4] S. B. Shuker, P. J. Hajduk, R. P. Meadows and S. W. Fesik, *Science* **2013**, *274*, 1531–1534.
- [5] D. A. Erlanson, *Top. Curr. Chem.* **2012**, *317*, 1–32.
- [6] G. Bollag, P. Hirth, J. Tsai, J. Zhang, P. N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, et al., *Nature* **2010**, *467*, 596–9.
- [7] G. Bollag, J. Tsai, J. Zhang, C. Zhang, P. Ibrahim, K. Nolop, P. Hirth, *Nat. Rev. Drug Discov.* **2012**, *11*, 873–886.
- [8] A. J. Souers, J. D. Levenson, E. R. Boghaert, S. L. Ackler, N. D. Catron, J. Chen, B. D. Dayton, H. Ding, S. H. Enschede, W. J. Fairbrother, et al., *Nat. Med.* **2013**, *19*, 202–208.
- [9] E. H. Mashalidis, P. Ślędź, S. Lang, C. Abell, *Nat. Protoc.* **2013**, *8*, 2309–24.
- [10] S. J. Taylor, A. Abeywardane, S. Liang, I. Muegge, A. K. Padyana, Z. Xiong, M. Hill-Drzewi, B. Farmer, X. Li, B. Collins, et al., *J. Med. Chem.* **2011**, *54*, 8174–8187.
- [11] Y. Cheng, T. C. Judd, M. D. Bartberger, J. Brown, K. Chen, R. T. Freneau, D. Hickman, S. A. Hitchcock, B. Jordan, V. Li, et al., *J. Med. Chem.* **2011**, *54*, 5836–5857.
- [12] M. Congreve, G. Chessari, D. Tisi, A. J. Woodhead, *J. Med. Chem.* **2008**, *51*, 3661–3680.
- [13] G. Chessari, A. J. Woodhead, *Drug Discov. Today* **2009**, *14*, 668–675.
- [14] E. Edink, P. Rucktooa, K. Retra, A. Akdemir, T. Nahar, O. Zuiderveld, R. Van Elk, E. Janssen, P. Van Nierop, J. Van Muijlwijk-Koezen, et al., *J. Am. Chem. Soc.* **2011**, *133*, 5363–5371.
- [15] G. E. de Kloe, D. Bailey, R. Leurs, I. J. P. de Esch, *Drug Discov. Today* **2009**, *14*, 630–646.
- [16] M. Mondal, D. E. Groothuis, A. K. H. Hirsch, *Med. Chem. Comm.* **2015**, *6*, 1267–1271.
- [17] D. C. Rees, M. Congreve, C. W. Murray, R. Carr, *Nat. Rev. Drug Discov.* **2004**, *3*, 660–672.
- [18] S. Chung, J. B. Parker, M. Bianchet, L. M. Amzel, J. T. Stivers, *Nat. Chem. Biol.* **2009**, *5*, 407–413.
- [19] P. J. Hajduk, G. Sheppard, D. G. Nettlesheim, E. T. Olejniczak, S. B. Shuker, R. P. Meadows, D. H. Steinman, G. M. Carrera, P. a. Marcotte, J. Severin, et al., *J. Am. Chem. Soc.* **1997**, *119*, 5818–5827.
- [20] A. W. Hung, H. L. Silvestre, S. Wen, A. Ciulli, T. L. Blundell, C. Abell, *Angew. Chem. Int. Ed.* **2009**, *48*, 8452–8456.
- [21] V. Borsi, V. Calderone, M. Fragai, C. Luchinat, N. Sarti, *J. Med. Chem.* **2010**, *53*, 4285–

- 4289.
- [22] J. J. Barker, O. Barker, S. M. Courtney, M. Gardiner, T. Hestekamp, O. Ichihara, O. Mather, C. a G. N. Montalbetti, A. Müller, M. Varasi, et al., *ChemMedChem* **2010**, *5*, 1697–1700.
- [23] N. Howard, C. Abell, W. Blakemore, G. Chessari, M. Congreve, S. Howard, H. Jhoti, C. W. Murray, L. C. a Seavers, R. L. M. Van Montfort, *J. Med. Chem.* **2006**, *49*, 1346–1355.
- [24] D. J. Maly, I. C. Choong, J. a Ellman, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2419–2424.
- [25] E. E. Swayze, E. a. Jefferson, K. A. Sannes-Lowery, L. B. Blyn, L. M. Risen, S. Arakawa, S. A. Osgood, S. A. Hofstadler, R. H. Griffey, *J. Med. Chem.* **2002**, *45*, 3816–3819.
- [26] A. M. Petros, J. R. Huth, T. Oost, C. M. Park, H. Ding, X. Wang, H. Zhang, P. Nimmer, R. Mendoza, C. Sun, et al., *Bioorganic Med. Chem. Lett.* **2010**, *20*, 6587–6591.
- [27] M. Mondal, M. Y. Unver, A. Pal, M. Bakker, S. P. Berrier, A. K. H. Hirsch, *Chem. - A Eur. J.* **2016**, *22*, 14826–14830.
- [28] S. K. Mamidyala, M. G. Finn, *Chem. Soc. Rev.* **2010**, *39*, 1252–1261.
- [29] P. Thirumurugan, D. Matosiuk and K. Jozwiak, *Chem. Rev.* **2013**, *113*, 4905–4979.
- [30] R. Huisgen, *1,3-Dipolar Cycloaddition Chemistry*, Wiley, New York, 1st edn., 1984.
- [31] J. B. Cooper, *Curr. Drug Targets* **2002**, *3*, 155–173. **n.d.**
- [32] 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.
- [33] L. Coates, P. T. Erskine, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Biochemistry* **2001**, *40*, 13149–13157.
- [34] L. Coates, H.-F. Tuan, S. Tomanicek, A. Kovalevsky, M. Mustyakimov, P. Erskine and J. Cooper, *J. Am. Chem. Soc.* **2008**, *130*, 7235–7237.
- [35] J. Cooper, W. Quail, C. Frazao, S. I. Foundling, T. L. Blundell, C. Humblet, E. A. Lunney, W. T. Lowther, B. M. Dunn, *Biochemistry* **1992**, *31*, 8142–8150.
- [36] 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.
- [37] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259–3263.
- [38] H. Köster, T. Craan, S. Brass, C. Herhaus, M. Zentgraf, L. Neumann, A. Heine, G. Klebe, *J. Med. Chem.* **2011**, *54*, 7784–7796.
- [39] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discov. Today* **2003**, *8*, 876–877.
- [40] Version 1.4.1, Pymol, L. Delano, <http://www.pymol.org/>
- [41] P. R. Gerber, K. Müller, *J. Comput. Aided. Mol. Des.* **1995**, *9*, 251–268.

- [42] Version 2. 1. 3, BioSolveIT GmbH, S. Augustin. <http://www.biosolveit.de>, LeadIT.
- [43] A. Albert, P. J. Taylor, *J. Chem. Soc. Perkin Trans.* **1989**, 1903–1905.
- [44] N. Radeva, S. G. Krimmer, M. Stieler, K. Fu, X. Wang, F. R. Ehrmann, A. Metz, F. U. Huschmann, M. S. Weiss, U. Mueller, et al., *J. Med. Chem.* **2016**, *59*, 7561–7575.
- [45] H. Gohlke, M. Hendlich, G. Klebe, *Perspect. Drug Discov. Des.* **2000**, *20*, 115–144.
- [46] S. S. Kulkarni, X. Hu, R. Manetsch, *Chem. Comm.* **2013**, *49*, 1193.
- [47] Y. Y. Yang, J. M. Ascano, H. C. Hang, *J. Am. Chem. Soc.* **2010**, *132*, 3640–3641.
- [48] M. V Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544–550.
- [49] S. Sezer, Y. Gümrukçü, E. Şahin, C. Tanyeli, *Tetrahedron: Asymmetry* **2008**, *19*, 2705–2710.
- [50] H. C. Cheng, *J. Pharmacol. Toxicol. Methods* **2001**, *46*, 61–71.
- [51] N. Schneider, S. Hindle, G. Lange, R. Klein, J. Albrecht, H. Briem, K. Beyer, H. Claussen, M. Gastreich, C. Lemmen, et al., *J. Comput.-Aided Mol. Des.* **2012**, *26*, 701–723.
- [52] I. Reulecke, G. Lange, J. Albrecht, R. Klein, M. Rarey, *ChemMedChem* **2008**, *3*, 885–897.
- [53] K. Stierand, M. Rarey, *ACS Med. Chem. Lett.* **2010**, *1*, 540–545.
- [54] M. K. Larson and J. R. Whitaker, *J. Dairy Sci.* **1970**, *53*, 262–269.
- [55] A. Elangovan, Y. H. Wang, T. I. Ho, *Org. Lett.* **2003**, *5*, 1841–1844.
- [56] T. Oka, K. Fujiwara, A. Murai, *Tetrahedron* **1998**, *54*, 21–44.
- [57] *WO, Pat.*, WO2013014039A1, **2013**.
- [58] Y. Li, M. Shen, Z. Zhang, J. Luo, X. Pan, X. Lu, H. Long, D. Wen, F. Zhang, F. Leng, et al., *J. Med. Chem.* **2012**, *55*, 10033–10046.
- [59] S. Müller, B. Liepold, G. J. Roth, H. J. Bestmann, *Synlett* **1996**, *6*, 521–522.
- [60] *WO, Pat.*, WO2014048878A1, **2014**.
- [61] V. Sandgren, T. Agback, P. O. Johansson, J. Lindberg, I. Kvarnström, B. Samuelsson, O. Belda, A. Dahlgren, *Bioorganic Med. Chem.* **2012**, *20*, 4377–4389.

Chapter 3

In situ Ugi four-component reaction for the protein-templated identification of inhibitors of endothiapepsin

The discovery of new bioactive compounds is a long and expensive process, which calls for the development of new techniques that can speed up hit identification and render it more efficient. In this context, a new protein-templated reaction for in situ selection of binders of protein targets would represent an invaluable addition to the portfolio available to medicinal chemists. In this work, we report the first example of an in situ Ugi reaction for the discovery of novel inhibitors of the aspartic protease endothiapepsin. Use of a protein-templated four-component reaction enables screening of a library of potential inhibitors in an efficient and rapid manner. This new protein-templated multicomponent reaction goes beyond known protein-templated reactions such as the in situ click reaction and opens up access to new regions of the chemical space. The Ugi products we identified are low-micromolar inhibitors, demonstrating the efficiency of this approach.

3.1 Introduction

Despite recent developments in drug discovery in terms of technology and techniques, there is a pressing need for the discovery of fast and efficient hit-identification strategies. Target guided synthesis (TGS) is a powerful approach to discover hit compounds by using the biological target itself in ligand selection. Two main methods, dynamic combinatorial chemistry (DCC) and kinetic target-guided synthesis (KTGS) have emerged.^[1] In KTGS, the target is actively involved in ligand selection by assembling its own inhibitors via an irreversible reaction from a library of complementary building blocks, whereas DCC assembles ligands via reversible process.^[2] Only a few protein-templated reactions for KTGS^[3–7] were reported so far because of the stringent requirements: The irreversible reaction needs to be compatible with physiological conditions, the building blocks need to be inert towards biomolecules and a substantial difference in reaction rate between the blank and biomolecule-templated reaction is required.

Multicomponent reactions (MCRs) are one-pot reactions in which more than two starting materials are incorporated to form a new adduct comprising most of the structural motifs of the starting building blocks.^[18] The Ugi four-component reaction (Ugi-4CR), discovered in 1959,^[19] is one of the most important MCRs, which affords dipeptide-like structures from isocyanides, carboxylic acids, aldehydes and amines.^[20] This elegant reaction has found numerous applications in drug discovery, including hit- and lead-identification as well as the generation of large libraries of analogues.^[18,20] Owing to the high exploratory power with regard to chemical space and biocompatibility, it represents an attractive reaction for KTGS. Herein, we describe the first use of the Ugi-4CR in KTGS for the identification of inhibitors.

The Ugi-4CR reaction represents a novel reaction for KTGS, in which the enzyme templates the synthesis of its own binders from a pool of four types of building blocks. It follows the same concept as other protein-templated reactions in terms of simultaneous binding of building blocks to adjacent pockets of the protein target, enabling the assembly of the corresponding binders. The great advantage of this method over other reported protein-templated reactions is the simultaneous screening of four subpockets and relative ease of accessing structurally complex binders and their derivatives for further optimization via a one-pot reaction, starting from simple and commercially available building blocks (Figure 1).

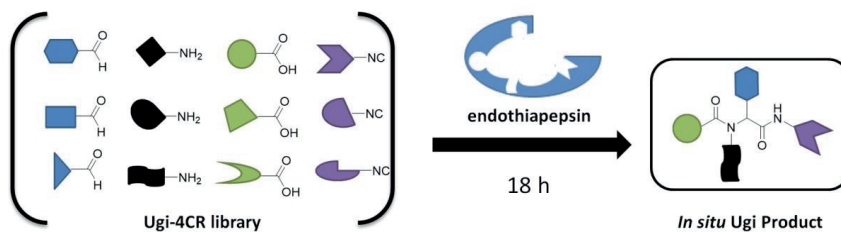


Figure 1. Schematic representation of in situ Ugi 4-component reaction leading to a dipeptide-like Ugi product from a library of aldehydes, amines, carboxylic acids and isocyanides.

Endothiapepsin is a member of the class of pepsin-like aspartic proteases, which play a causative role in several diseases such as malaria, Alzheimer's disease, hypertension and HIV. The high degree of similarity makes endothiapepsin a convenient model enzyme for mechanistic studies,^[21–23] and the identification of inhibitors of renin^[24] and β -secretase.^[25] For more detailed information about endothiapepsin, see Chapter 1.^[22,23,26]

To increase the power and scope of KTGS, it is necessary to expand the number of biocompatible reactions and building-block diversity. To the best of our knowledge, all the reports of KTGS represent examples of in situ reactions of two types of building blocks. In this work, for the first time, we described a new protein-templated reaction with four sets of fragments, which allows fast and efficient screening of a large portion of the chemical space.

3.2 Results and discussion

3.2.1 Design of the inhibitor with Ugi-4CR scaffold

We selected endothiapepsin as a model enzyme for our proof-of-principle work. After careful screening of the chemical structures of known inhibitors in terms of ease of modification to dipeptide-like compounds and affinity, we designed a novel, potential dipeptide inhibitor **2** based on inhibitor **1**, which we had reported previously by using DCC and *de novo* structure-based drug design ($IC_{50} = 12.8 \pm 0.4 \mu\text{M}$, PDB: 4KUP).^[26] Starting from the X-ray crystal structure of inhibitor **1** in complex with endothiapepsin, we used the molecular-modeling software Moloc^[27] and the FlexX docking module in the LeadIT suite for the structure-based design of **2**.^[28]

The indole moiety of inhibitor **1** functions as an anchor in the active site of the protein. Inhibitor **1** occupies the S1 and S2 pockets of the protein and is involved in three H bonding interactions with the catalytic dyad (D35: 2.8 Å and 3.2 Å, D219: 2.9 Å) via its α -amino group. Additionally, while the NH of the acylhydrazone forms a H-bond with T222 (2.8 Å), the indolic

NH group engages in a H-bond with the carboxylate of D81 (3.0 Å).^[26] We designed inhibitor **2** as an Ugi product that should maintain the same interactions as inhibitor **1**: H-bonding interactions with D35 (3.3 Å, 2.8 Å) and D219 (2.8 Å). The indolic NH group donates one H-bond to D81 (3.0 Å) and the carbonyl group of the amide bond accepts a H-bond from T222 (2.7 Å). An additional H-bonding interaction can be observed with G80 (2.7 Å) with the other carbonyl group, which should be hosted in the S2' pocket. By converting our parent inhibitor into a 4-CR scaffold, the imidazolyl group (amine component) can occupy the S2' pocket where it engages in hydrophobic interactions with T79, G37, G221, while the phenyl group (isocyanide component) should be bound in between the S1' and S2 pockets, benefiting from hydrophobic interactions with I302, I300 and I217 (Figure 2).

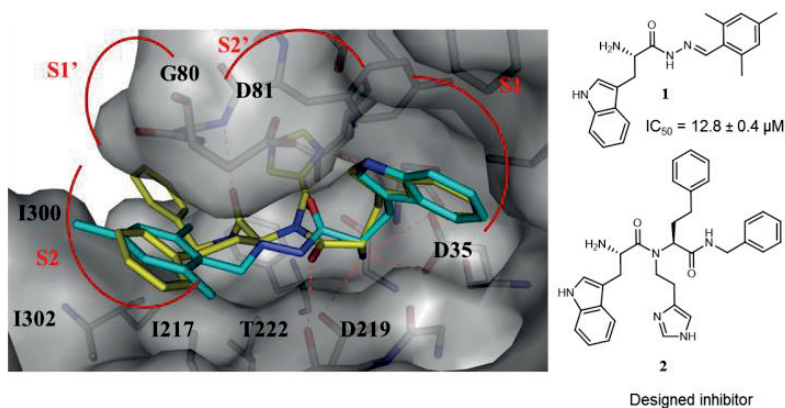
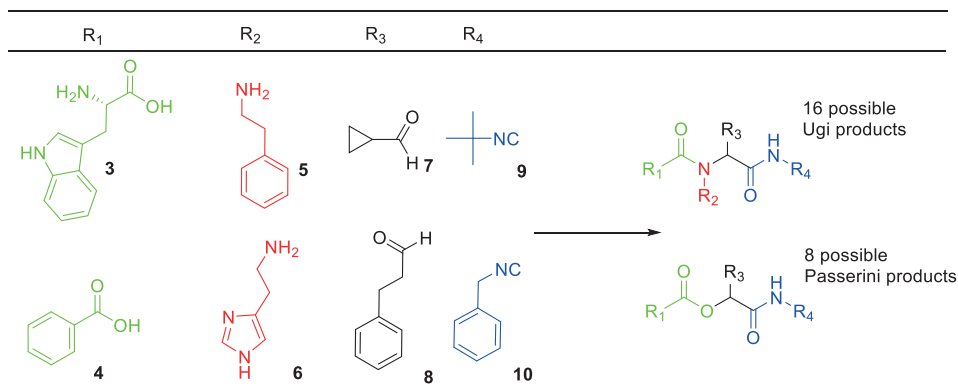


Figure 2. X-ray crystal structure of endothiapepsin in complex with inhibitor **1** (PDB: 4KUP)^[26] superimposed with designed inhibitor **2**. Color code: inhibitor **1** skeleton: C: cyan, N: blue, O: red; inhibitor **2** skeleton: C: yellow, N: blue, O: red; protein backbone: gray; dashed lines: H-bonding interactions below 3.3 Å.

3.2.2 Generation of the library

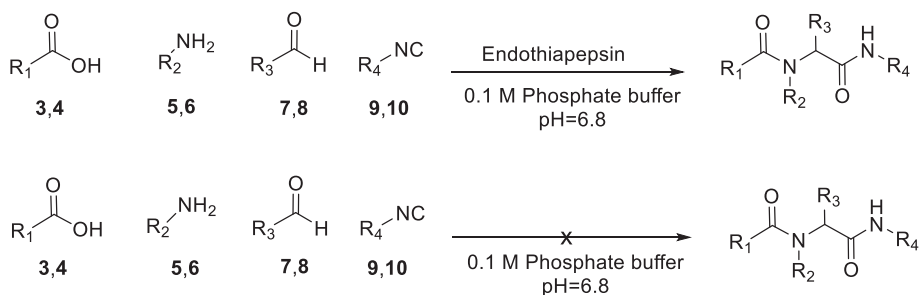
To generate the combinatorial library, we selected two different building blocks for each of the four components (compounds **3–10**, Scheme 1) with comparable reactivity and solubility. We used benzoic acid as one of the acid components, a negative control for our docking studies, to demonstrate that we maintain the same interactions with the catalytic dyad: H-bonding interactions with D35 and D219 as in inhibitor **1**. The Ugi-4CR requires high concentrations (0.5–2.0 M) to afford the corresponding products. The Passerini reaction is a 3-CR in which an aldehyde, carboxylic acid and isocyanide react to form α -acyloxyamides.^[29] This reaction is the main side reaction of the Ugi-4CR. The mechanism and the products formed are similar for both reactions, which should enable the simultaneous screening of a higher number of products.



Scheme 1. Selection of building blocks for the in situ Ugi 4-CR, which affords sixteen possible Ugi and eight possible Passerini reaction products with their diastereomeric pairs.

3.2.3 In situ Ugi reaction

Having selected the building blocks, which are commercially available, we set up two reactions in parallel, a protein-templated reaction and a blank reaction, at pH=6.8 (phosphate buffer 0.1 M, 10% DMSO) by mixing two carboxylic acids **3** and **4** two amines **5** and **6**, two aldehydes **7** and **8** and two isocyanides **9** and **10** (100 μ M each) (Scheme 2).



Scheme 2. From top to bottom; in situ Ugi reaction, blank reaction.

The optimized reaction conditions have a building-block concentration of 100 μM to prevent product formation in the reference reaction and to reduce the amount of protein used. To the protein-templated reaction, we added a catalytic amount of endothiapepsin (25 μM) and analyzed both reaction mixtures after each day by using UPLC-TQD-SIR (SIR: selective-ion recording) for each of the possible 16 Ugi and 8 Passerini products. Although there are other possible side products derived from this library such as imine formation, Pictet-Spengler reaction, side products arising from the dual functionality of tryptophan (**3**) or cyclization products (each possibility was analyzed), we only focus on those products, which are most likely to bind and formed in the library mixture, for clarity. The SIR technique enables fast and sensitive screening of specific molecular weights (Mws, up to 8 Mws per injection) regardless of very low concentrations. By screening eight Mws per injection using the same concentrations for protein-templated and reference reactions, we analyzed each reaction in 3 SIR measurements (8-fold faster than commonly used SIM (selective ion monitoring) technique) and detected the formation of two Ugi products **2** and **11** only in the presence of endothiapepsin after 18 h (Figures 3-4).

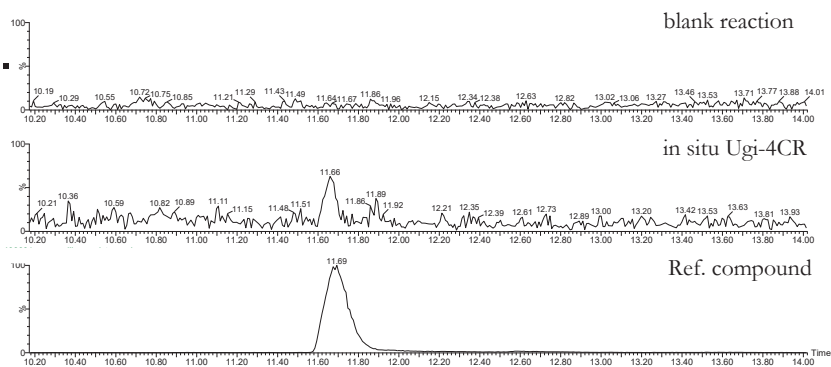


Figure 3. UPLC-TQD-SIR analysis of compound **2** ($[M+H]^+ = 549$). Formation of **2** by in situ Ugi-4CR was compared with the blank reaction and synthesized compound **2**.

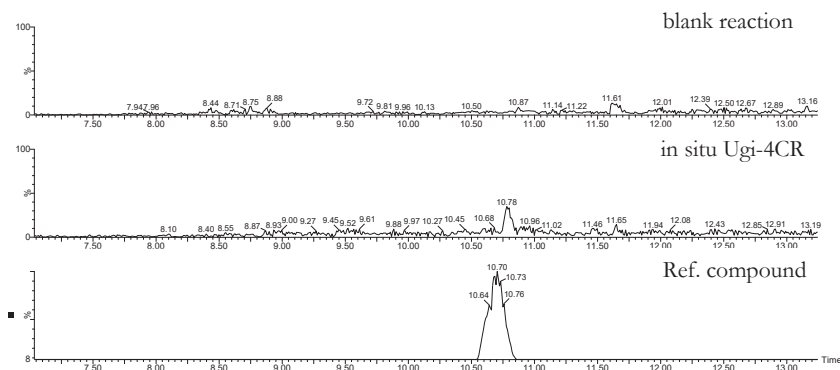


Figure 4. UPLC-TQD-SIR analysis of compound **11** ($[M+H]^+ = 485$). Formation of **11** by in situ Ugi-4CR was compared with the blank reaction and synthesized compound **11**.

3.2.4 Control experiments

To demonstrate that the active site of endothiapepsin is required for product formation, we repeated the reaction in the presence of 25 μM bovine serum albumin (BSA) instead of endothiapepsin and in the presence of a strong inhibitor of endothiapepsin, saquinavir (100 μM , $K_i = 48$ nM). No product formation was observed in these control experiments, indicating that binding of the components to the active site of endothiapepsin is essential for the formation (Figures 5 and 6).

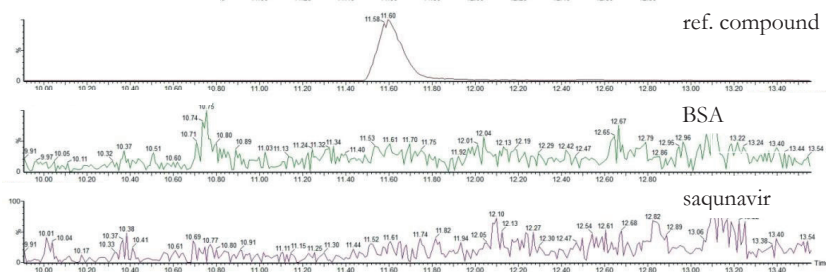


Figure 5. Control reactions with BSA and saquinavir, compared with the synthesized compound **2**.

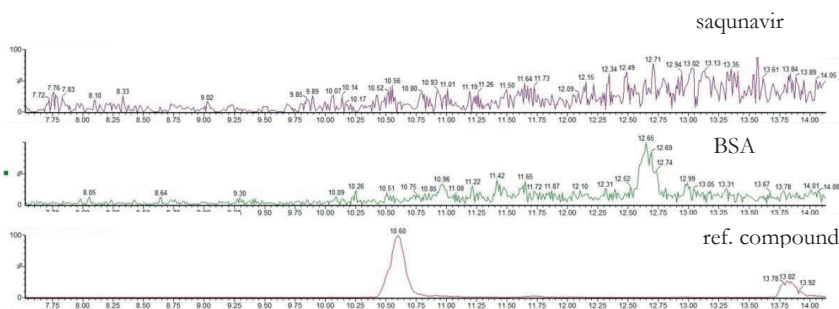


Figure 6. Control reactions with BSA and saquinavir, compared with the synthesized compound **11**.

To check for possible protein modification in the active site of endothiapepsin, we run several experiments in parallel under identical conditions: the whole library mixture **3–10** incubated with endothiapepsin, individual building blocks **3**, **5–8**, **10** incubated with endothiapepsin and as a reference only endothiapepsin in the reaction buffer. After 18 h, we evaluated the activity of the enzyme in each reaction using an adaptation from the fluorescence-based assay for HIV protease.^[30] The activity was not affected, demonstrating that no modification occurred in the active site (Figure 7).

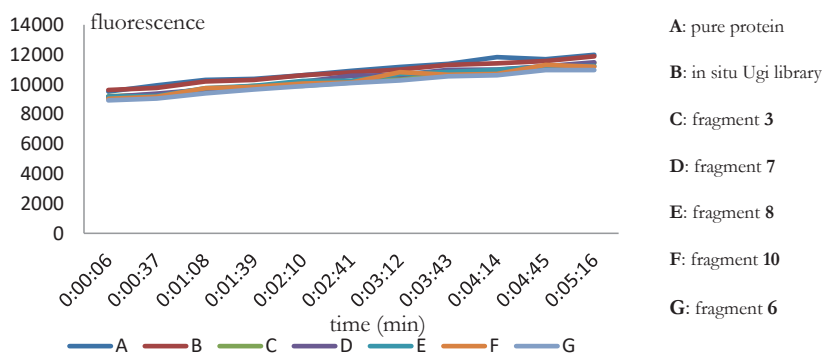


Figure 7. Protein modification control experiments.

To investigate whether the individual building blocks bind to endothiapepsin, we used saturation-transfer difference (STD) NMR spectroscopy.^[31] STD-NMR spectroscopy enables the characterization of target–ligand interactions in solution. The basic principle of the technique is that the protons, which are in close contact with the target protein receive a higher degree of saturation, resulting in stronger STD-NMR signals. Protons, which are not involved in the interaction or having weaker interaction with the target reveal no STD NMR signals. This technique is an excellent tool to investigate how a ligand interacts with the target. It is easy to implement and it requires small amount of the protein. STD-NMR experiments of building

blocks **3**, **6–8** and **10**, which are comprised in the hit compound's skeleton formed in the protein-templated reaction, showed that all building blocks except for aldehyde **7** have interactions with the target, which may be ascribed to its small size.

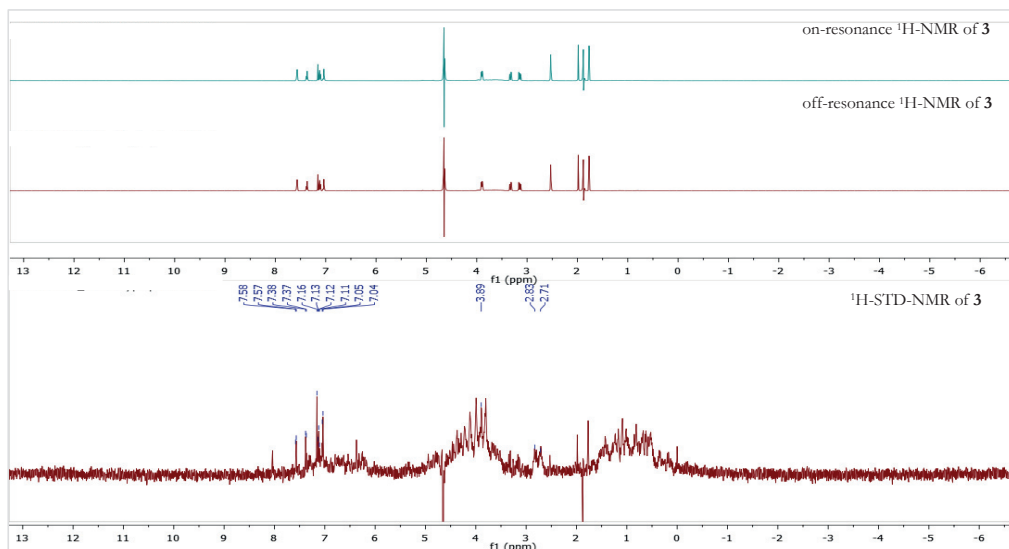


Figure 8. A representative example of STD-NMR studies for fragment **3**.

In addition, we performed competition ^1H -STD-NMR experiments between each of the above fragments and a known inhibitor of endothiapepsin (bisacylhydrazone **20**, $\text{IC}_{50} = 54$ nM, Figure 9).^[31] The results of the competition ^1H -STD-NMR experiments revealed that the bisacylhydrazone displaces **3** and **6** from the binding site of the enzyme. Given that bisacylhydrazone **20** (PDB: 5HCT) is a strong inhibitor and binds in the active site of the enzyme, this experiment demonstrates that both **3** and **6** bind to the same pocket of the enzyme as the bisacylhydrazone inhibitor **20**. On the other hand, the signals of fragments **8** and **10** did not disappear in the STD-NMR competition experiment with bisacylhydrazone **20**, indicating that they occupy different subpockets of the enzyme's active site. As a result, we conclude that the two fragments **3** and **6** bind in the core of the active site of endothiapepsin just like the bisacylhydrazone **20**, whereas fragments **8** and **10** occupy alternative subpockets in proximity to enable the protein-templated formation of the products. A representative example of the ^1H -STD-NMR competition experiments can be seen in Figures 9 and 10.

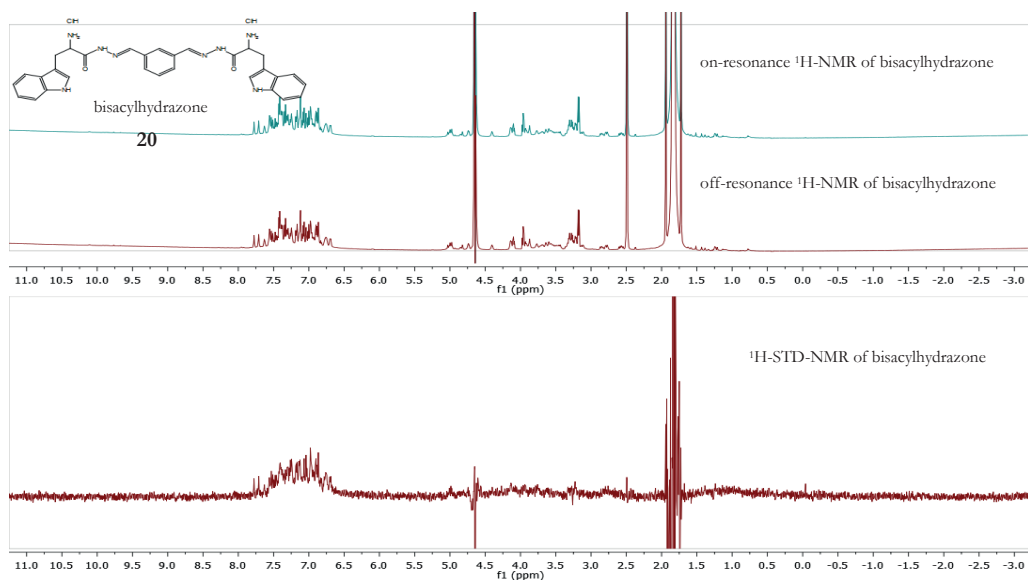


Figure 9. STD-NMR studies for the bisacylhydrazone **20**.

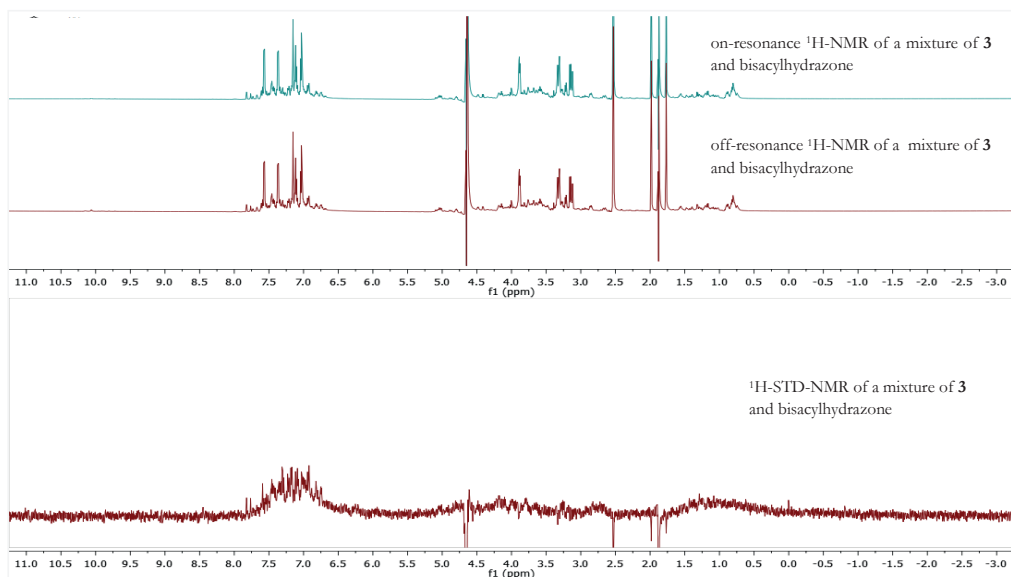


Figure 10. A representative example for the ^1H -STD-NMR competition study for fragment **3** and bisacylhydrazone **20**.

3.2.6 Synthesis and biochemical evaluation of the inhibitors

In order to confirm that the Ugi products formed are indeed inhibitors of endothiapepsin, we synthesized both compounds. The Ugi-4CR reaction with Boc-protected tryptophan (**12**), amine **6**, aldehydes **7** and **8** and isocyanide **10** followed by HCl-mediated deprotection of the Boc-group of the corresponding Ugi products afforded compounds **2** and **11** in 40% and 35% yield, respectively (Scheme 3). We isolated the final compounds as diastereomeric mixtures and used them without further separation.

Evaluation of the inhibitory potency of compounds **2** and **11** showed that compounds **2** and **11**, which were formed in the presence of protein, are potent inhibitors with IC_{50} values of $1.3 \pm 0.1 \mu\text{M}$ and $3.5 \pm 0.1 \mu\text{M}$, respectively (Figures 11).

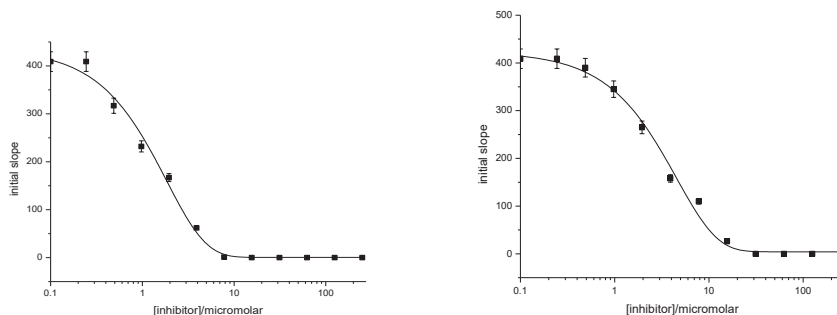
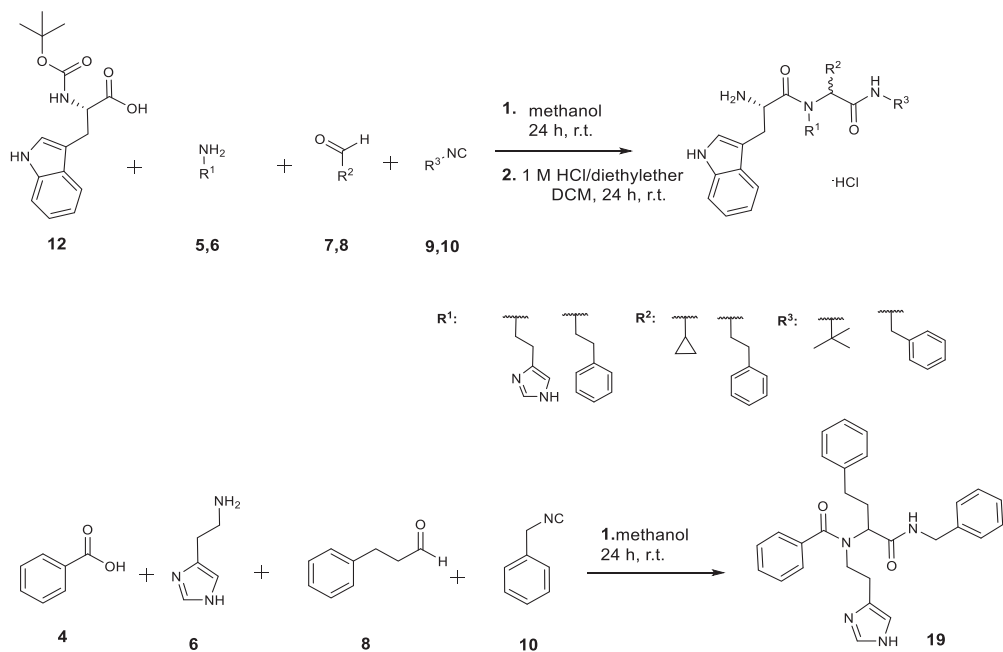


Figure 11. From left to right: IC_{50} inhibition curves of **2** ($IC_{50} = 1.3 \pm 0.1 \mu\text{M}$) and **11** ($IC_{50} = 3.5 \pm 0.1 \mu\text{M}$). The inhibitors were measured in duplicate.

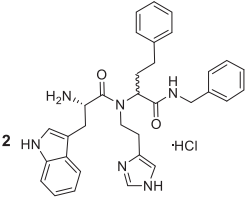
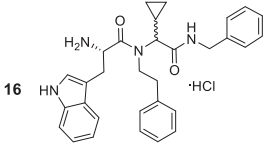
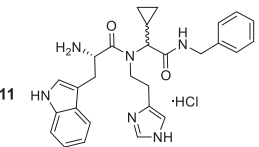
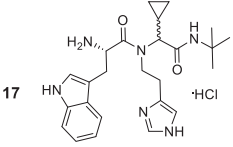
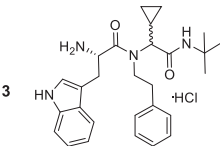
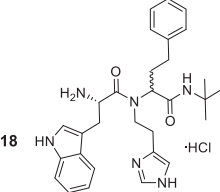
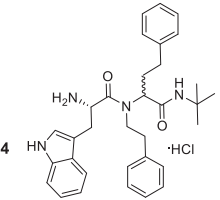
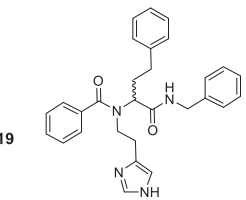
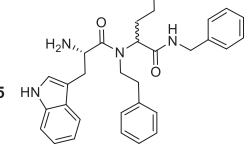
We could separate only one diastereomer of inhibitor **2**, which showed a slight improvement in activity ($IC_{50} = 0.89 \pm 0.9 \mu\text{M}$). Further attempts to separate diastereomers of hit compounds by using various separation techniques failed, presumably owing to the polarity of the compounds.



Scheme 3. Synthetic strategy towards compounds **2**, **11** and **13–19**.

As the in situ Ugi-4CR reaction represents the first example in the field, we synthesized a library of possible compounds including the eight Ugi-4CR products **2**, **11**, **13–18** using *L*-Trp as the acid component and one Ugi-4CR product using benzoic acid **19** (Table 1).

Table 1. Synthesis and biochemical evaluation of selected compounds derived from the combinatorial library.

Compound	Yield (%)	IC ₅₀ (μM)	Compound	Yield (%)	IC ₅₀ (μM)
	40	1.3 ± 0.1		40	10 ± 5
	35	3.5 ± 0.1		61	16 ± 3
	80	6 ± 0.2		37	3.3 ± 0.5
	60	129 ± 35		32	No inhibition
	75	19 ± 3			

Synthesis and subsequent biochemical evaluation of the library demonstrated that all combinations with *L*-Trp as the acid component show activity against endothiapepsin in the range of 1.3–129 μM. Compounds **11** and **18** show activities in the same range as the hits and the analogue **19** with the benzoic acid component did not show any inhibition at the starting concentration of 250 μM as we expected. As can be concluded from Table 1, in situ Ugi-4CR is a powerful method for the selection of the best binders from this library. The slight preference of compound **11** over **18** could be due to the content ratio of the diastomeric mixture.

3.2.7 Docking results

We tried to soak crystals of endothiapepsin with inhibitors **2** and **11** to confirm the binding mode but were unsuccessful due to the limited solubility of the compounds. Our docking studies for both diastereomers of **2** and **11** using the FlexX docking module in the LeadIT^[28] suite followed by evaluation using the scoring function HYDE in SEESAR^[32] show that the inhibitor **11** can occupy the same pockets as inhibitor **2**. As can be seen in Figure 12, the inhibitor is involved in a H-bonding network with the catalytic dyad via its α -amino group (D35: 3.3 Å, 2.8 Å, D219: 2.8 Å). The H-bonding between the NH of the indolyl moiety with D81 (3.1 Å), the carbonyl group in the S2' pocket with G80 (2.8 Å) and the interaction between the other carbonyl group with T222 (3 Å) should be retained (Figures 12 and 13).

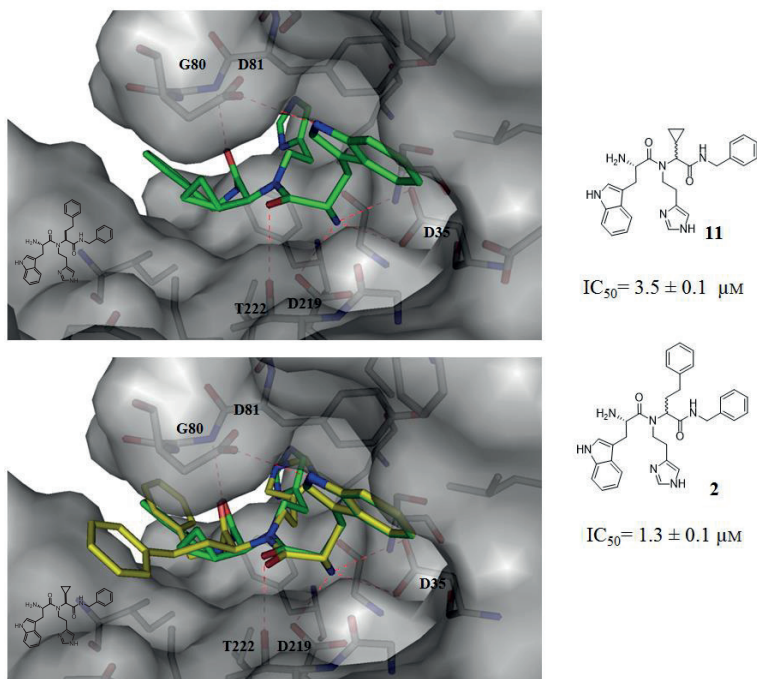


Figure 12. Compounds formed in presence of endothiapepsin. Top: top-ranked pose of inhibitor **11** generated by docking using the FlexX docking module in the LeadIT^[28] suite followed by evaluation using the scoring function HYDE in SEESAR.^[32] Bottom: superimposition of top-ranked poses of inhibitors generated by docking using the FlexX docking module in the LeadIT^[28] suite followed by hydrescoring in SEESAR.^[32] Color code: inhibitor **2** skeleton: C: yellow, N: blue, O: red; inhibitor **11** skeleton: C: green, N: blue, O: red; protein backbone: gray; dashed lines: H-bonding interactions below 3.3 Å.

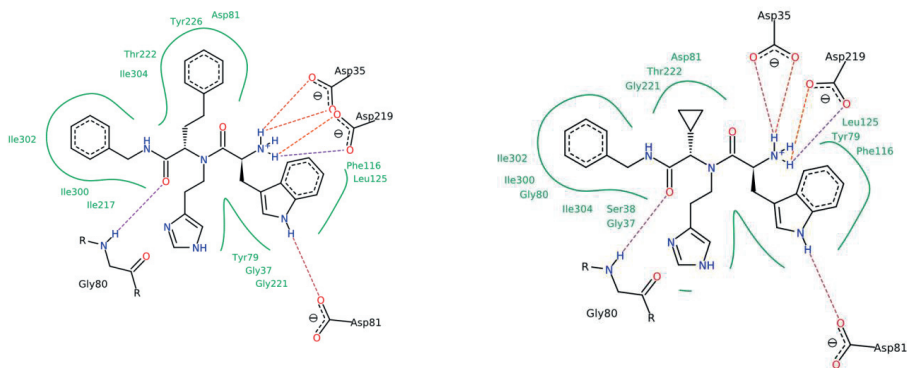


Figure 13. Predicted binding mode of compounds (*S,S*)-**2** and (*S,S*)-**11**. These binding modes are the result of a docking run using the FlexX docking module with 30 poses and represent the top-scoring pose after HYDE scoring with SEESAR^[32] and careful visual inspection to exclude poses with significant inter- or intra-molecular clash terms or unfavorable conformations. The figures were generated with PoseView^[33] as implemented in the LeadIT suite.^[28] Docking studies were performed for both diastereomers of the inhibitors **2** and **11**.

3.3 Conclusions

In conclusion, herein, we describe the in situ Ugi-4CR, which represents the first example in drug discovery. As a result, the mechanism of the in situ reaction needs to be studied in detail. We screened 24 compounds (16 Ugi+8 Passerini) products by using commercially available building blocks and a catalytic amount of endo-thiapepsin in 18 h, which is comparatively fast for KTGS (6 h–2 weeks). Our efficient strategy circumvents the need for synthesis, purification and evaluation of each individual compound. Both inhibitors identified were synthesized in two synthetic steps. Thanks to the use of the Ugi-4CR, the inhibitors identified can be optimized in a straightforward manner. Our novel protein-templated 4-CR strategy could find application in the early stages of drug discovery, namely hit identification, on a range of drug targets.

3.4 Experimental section

3.4.1 Fluorescence-based inhibition assay

For fluorescence-based inhibition assay, see Chapter 2, Section 2.4.1.

3.4.2 Modeling and docking

The X-ray crystal structure of the complex of endothiapepsin (PDB code: 4KUP) with compound **1** was used for our modeling studies. The energy of the system was minimized using the MAB force field as implemented in the computer program MOLOC.^[27] Taking inspiration from the cocrystal structure of endothiapepsin with compound **1**, as well as from hot-spot analysis^[34] of the active site of endothiapepsin, a new Ugi scaffold was designed and subsequent energy minimization (MAB force field) was done using MOLOC. All types of interactions (H-bonding, lipophilic and repulsive interactions) between designed Ugi products and protein were analyzed in MOLOC. Designed compounds were subsequently docked into the active site of endothiapepsin by using the FlexX docking module in the LeadIT suite.^[28] During the docking, the binding site in the protein was restricted to 6.5 Å around the cocrystallized ligand **1**, and the 30 top (FlexX)-scored solutions were retained, subsequently post-scored with SEESAR,^[32] and the best scored pose was selected.

3.4.3 Experimental procedures

In situ Ugi reaction

Endothiapepsin (25 µL, 1.0 mM in phosphate buffer 0.1 M, pH 6.8), the seven building blocks **4–10** (1 µL each, 100 mM in DMSO) and *L*-tryptophan (**3**) (2.5 µL, 25 mM in DMSO) were added to a mixture of DMSO (91.5 µL) and phosphate buffer (900 µL, 0.1 M, pH 6.8). The reaction mixture was allowed to rotate at room temperature with 10 rpm. After 18 h, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

In situ Ugi-4CR experiments in presence of saquinavir

Endothiapepsin (25 µL, 1.0 mM in phosphate buffer 0.1 M, pH 6.8), the 7 building blocks **4–10** (1 µL each, 100 mM in DMSO), *L*-tryptophan (**3**) (2.5 µL, 25 mM in DMSO) and saquinavir (1 µL, 100 mM in DMSO) were added to a mixture of DMSO (91.5 µL) and phosphate buffer (900 µL, 0.1 M, pH 6.8). The reaction mixture was allowed to rotate at room temperature with 10 rpm. After 18 h, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

Blank reaction, negative control:

The 7 building blocks **4–10** (1 μL each, 100 mM in DMSO) and *L*-tryptophan (**3**) (2.5 μL , 25 mM in DMSO) were added to a mixture of DMSO (91.5 μL) and phosphate buffer (900 μL , 0.1 M, pH 6.8). The reaction mixture was allowed to rotate at room temperature with 10 rpm. After 18 h, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

In situ Ugi experiments using BSA

BSA (25 μL , 1.0 mM in phosphate buffer 0.1 M, pH 6.8), the 7 building blocks **4–10** (1 μL each, 100 mM in DMSO) and *L*-tryptophan (**3**) (2.5 μL , 25 mM in DMSO) were added to a mixture of DMSO (91.5 μL) and phosphate buffer (900 μL , 0.1 M, pH 6.8). The reaction mixture was allowed to rotate at room temperature with 10 rpm. After 18 h, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

Protein

Endothiapepsin (25 μL , stock solution 1.0 mM in phosphate buffer 0.1 M, pH 6.8) was added to 75 μL of DMSO and 900 μL phosphate buffer 0.1 M, pH 6.8. After 18 h, the enzyme solution was analyzed by UPLC-TQD-SIR (ESI+) measurements and compared with the positive hits identified from in situ Ugi product formation.

Adaptation protocol for protein-modification test experiment

Seven reactions were started in parallel, an in situ Ugi reaction with all library members **3–10** and individual fragments **3, 6–8** (100 μM) were incubated with endothiapepsin (25 μM) in phosphate buffer (0.1 M, pH 6.8). Finally, pure protein was incubated in the same reaction buffer as a reference. After 18 h, reaction mixtures were diluted to 4 nM and were directly used as protein stock solutions. Endothiapepsin was purified from Suparen® (kindly provided by DSM Food Specialties) by exchanging the buffer to potassium phosphate buffer (0.1 M, pH 6.8) using a Vivaspin 500 with a molecular weight cut-off at 10,000 Da. Measurement of the absorption at 280 nm, assuming an extinction coefficient of 1.15 for 1 mg/mL solutions, afforded the protein concentration. The final reaction volume was 200 μL containing 0.4 nM endothiapepsin (from each test reaction), 1.8 μM substrate and 2.1% DMSO. As substrate, Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH₂ (purchased from Bachem) was used for the fluorescence screening assay. The assay was performed with flat bottom 96-well microplates (purchased from Greiner Bio-One) using a Synergy Mx microplate reader at an excitation wavelength of 337 nm and an emission wavelength of 414 nm. The K_m of the substrate toward endothiapepsin was known, 1.6 μM .^[35]

The assay buffer (phosphate buffer 0.1 M, pH 6.8), containing 0.001% Tween 20) was premixed with the substrate, endothiapepsin was added directly before the measurement. As the substrate is a fluorogenic substrate, during measurement the fluorescence increased because of substrate hydrolysis by endothiapepsin. The initial slopes of the fluorescence in the wells containing different protein solutions were compared for data analysis. Each compound was measured in duplicate.

UPLC-TQD-SIR method

UPLC-TQD was performed using a Waters Acquity UPLC H-class system coupled to a Waters TQD. All analyses were performed using a reversed-phase UPLC column (ACQUITY HSS T3 Column, 130 Å, 1.8 µm, 2.1 mm x 150 mm). Positive-ion mass spectra were acquired using ES ionization, injecting 10 µL of sample; column temperature 35 °C; flow rate 0.3 mL/min. The eluents, acetonitrile and water contained 0.1% of formic acid. The library components were eluted with a gradient from 95% → 30% over 20 min, then at 5% over 1 min, followed by 5% for 2 min.

The UPLC-TQD-SIR method was used to analyze the formation of Ugi products in in situ and blank reactions. SIR measurements are highly sensitive, where a minute amount of compound can be detected by the mass spectrometer. $[M+H]^+$ were monitored using the full mass range to ensure correct isotope patterns for all possible potential Ugi products both for in situ Ugi and blank reactions. The Ugi products in the protein-templated reaction were identified by comparison of their retention time with those synthesized using conventional methods.

¹H-STD-NMR binding experiments

All ¹H-STD-NMR experiments were performed at 25 °C on a Varian 600 MHz spectrometer. The on-resonance irradiation on endothiapepsin was set to 0 ppm. In each experiment, one ligand (3 mM) and endothiapepsin (10 µM) were present.

All fragments of two Ugi hit compounds **2** and **11**, identified in the protein-templated Ugi reaction, were tested to verify their interaction with the protein by STD-NMR technique.

First, the ¹H-NMR spectra of the fragments Trp.HCl (**3**), histamine (**6**), 3-phenylpropanal (**8**), cyclopropanecarboxaldehyde (**7**), benzylisocyanide (**10**) in NaOAc buffer (100 mM, NaOAc of pD=4.7 in deuterated water (pH = 4.3)) and 5% DMSO-*d*₆, except for 3-phenylpropanal (**8**) (10% DMSO-*d*₆ was used for the solubility reason) in the presence of protein (off-resonance ¹H NMR spectra) were recorded.

Subsequently, on-resonance $^1\text{H-NMR}$ spectra were recorded with minimum 512 scans. $^1\text{H-STD-NMR}$ spectra were obtained by subtracting the on-resonance spectrum from the corresponding off-resonance spectrum. $^1\text{H-STD-NMR}$ analysis revealed that all five fragments except for the cyclopropanecarboxaldehyde (**7**) bind to the protein with good to moderate interaction realized by the intensity of the peaks in the STD NMR.

$^1\text{H-STD-NMR}$ competition experiments

First, the $^1\text{H-STD-NMR}$ of each fragment (final concentration 3 mM), was recorded as mentioned above. Subsequently, bisacylhydrazone **20** (reported inhibitor, $\text{IC}_{50} = 54 \text{ nM}$)^[31] was added to the same NMR tube in order to have a final concentration of 3 mM, subsequently a second $^1\text{H-STD NMR}$ spectrum was recorded showing the appearance of bisacylhydrazone peaks and disappearance (or reduced intensity) of peaks of some fragments such as TrpHCl and histamine, indicating that they are displaced by the bisacylhydrazone **20** and bind in the active site of endothiapsin where bisacylhydrazone is hosted (PDB: 5HCT). This experiment confirms that fragments TrpHCl and histamine bind to the same binding pocket of the enzyme.

3.4.4 General experimental details

For general experimental conditions, see Chapter 2, Section 2.4.4

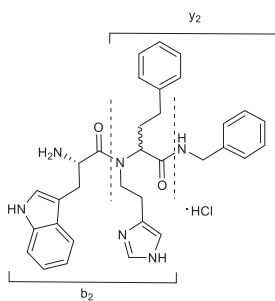
Chemical shifts (δ) are reported relative to the residual solvent peak. The $^1\text{H-NMR}$ spectra except for compound **19** (mixture of diastereomers and rotamers, see NMR spectra) were not interpreted due to their complexity. High-resolution mass spectra were recorded with an FTMS orbitrap (Thermo Fisher Scientific) mass spectrometer.

3.4.5 Synthesis of the Ugi products

General procedure for Ugi/Deprotection reaction

To a 10-mL round-bottomed flask charged with MeOH (4 mL), the corresponding aldehyde **7** or **8** (1 mmol, 1 eq.), amine **5** or **6** (1 mmol, 1 eq.), carboxylic acid **12** or **4** (1 mmol, 1 eq.) and isocyanide **9** or **10** (1 mmol, 1 eq.) were added. The reaction mixture was stirred at room temperature for 24 h. The reaction mixture was concentrated in vacuo and a quick purification over silica gel using EtOAc/pentane (1:1) afforded the corresponding *N*-Boc-protected Ugi product, which was directly used for the following deprotection step. The crude product was dissolved in DCM (4 ml), and HCl/diethyl ether (1 M, 10 mL) was added. After 24 h stirring at r.t., the resulting white precipitate was collected and washed with Et_2O . Complete removal of the solvent afforded the HCl salt of the desired compounds.

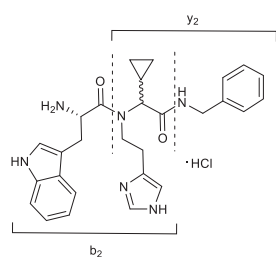
2-((S)-N-(2-(1H-Imidazol-4-yl)ethyl)-2-amino-3-(1H-indol-3-yl)propanamido)-N-benzyl-4-phenylbutanamide hydrochloride (2)



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), histamine (**6**, 101mg, 1 mmol), 3-phenylpropanal (**8**, 134 mg, 1 mmol) and benzyl isocyanide (**10**, 117mg, 1 mmol) in MeOH (4 mL) afforded the desired product **2** as a white solid as a diastereomeric mixture in 40% yield.

^{13}C NMR (101 MHz, CD_3OD) δ 172.2, 171.8, 171.6, 171.4, 171.2, 171.1 170.9, 170.7, 142.1, 142.0, 141.6, 141.5, 139.9, 139.6, 139.5, 139.4, 138.1, 138.0, 137.9, 134.8, 132.6, 132.5, 130.2, 129.9, 129.8, 129.6, 129.5, 129.5, 129.5, 129.4, 129.3, 129.2, 128.9, 128.8, 128.7, 128.5, 128.4, 128.4, 128.4, 128.3, 128.3, 128.2, 128.2, 127.6, 127.2, 127.1, 126.1, 126.0, 125.9, 125.6, 123.1, 123.0, 120.7, 120.6, 120.5, 120.3, 119.2, 119.1, 119.0, 117.4, 117.1, 113.0, 112.9, 112.8, 112.8, 112.6, 107.9, 107.8, 107.6, 107.2, 61.5, 61.0, 60.7, 59.2, 52.5, 52.4, 52.1, 51.8, 49.9, 49.7, 49.5, 49.3, 49.1, 48.9, 44.5, 44.5, 44.3, 43.9, 43.8, 33.7, 33.5, 33.3, 32.9, 32.5, 32.2, 31.9, 29.4, 29.1, 28.5, 27.6, 25.6, 24.3 HRMS (ESI) calcd for $\text{C}_{33}\text{H}_{36}\text{N}_6\text{O}_2$ $[\text{M}+\text{H}]^+$: 549.2972, found: 549.2940. calcd for y_2 fragment (m/z): 363.2179, found 363.2161, calcd for b_2 fragment (m/z): 442.2243, found 442.2213.

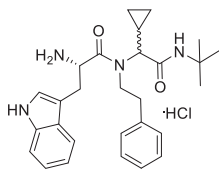
(2S)-N-(2-(1H-Imidazol-4-yl)ethyl)-2-amino-N-(2-(benzylamino)-1-cyclopropyl-2-oxoethyl)-3-(1H-indol-3-yl)propanamide hydrochloride (11)



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), histamine (**6**, 101mg, 1 mmol), cyclopropanecarboxyaldehyde (**7**, 70 mg, 1 mmol) and benzyl isocyanide (**10**, 117mg, 1 mmol) in MeOH (4 mL) afforded the desired product **11** as a white solid as a diastereomeric mixture in 35% yield.

^{13}C NMR (101 MHz, CD_3OD) δ 173.7, 172.7, 172.5, 172.0, 171.7, 171.1, 170.5, 139.9, 139.8, 139.7, 138.2, 138.1, 138.0, 134.8, 129.7, 129.6, 129.5, 129.4, 128.6, 128.5, 128.4, 128.3, 128.2, 128.1, 126.2, 125.9, 125.6, 123.3, 123.1, 123.0, 122.9, 120.6, 120.5, 120.4, 120.3, 118.9, 119.8, 117.6, 117.3, 117.1, 113.81, 112.9, 112.6, 107.9, 107.7, 107.6, 67.4, 66.8, 66.7, 52.6, 52.4, 52.0, 51.9, 49.9, 49.7, 49.5, 49.3, 49.1, 48.6, 46.9, 44.7, 44.2, 44.1, 43.7, 29.2, 29.1, 25.9, 24.5, 14.2, 11.7, 11.3, 10.9, 7.7, 7.3, 6.2, 5.9, 5.2, 4.8, 4.1 HRMS (ESI) calcd for $\text{C}_{28}\text{H}_{32}\text{N}_6\text{O}_2$ $[\text{M}+\text{H}]^+$: 485.2659, found: 485.2640, calcd for y_2 fragment (m/z): 299.1866, found 299.1859, calcd for b_2 fragment (m/z): 378.1925, found 378.1912.

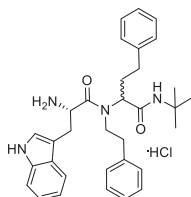
(2*S*)-2-Amino-*N*-(2-(*tert*-butylamino)-1-cyclopropyl-2-oxoethyl)-3-(1*H*-indol-3-yl)-*N*-phenethylpropanamide hydrochloride (13**)**



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), phenylethylamine (**5**, 121 mg, 1 mmol), cyclopropanecarboxaldehyde (**7**, 70 mg, 1 mmol) and *tert*-Butyl isocyanide (**9**, 83mg, 1 mmol) in MeOH (4 mL) afforded the desired product **13** as a white solid as a diastereomeric mixture in 80% yield.

^{13}C NMR (101 MHz, CD_3OD) δ 171.2, 170.3, 170.2, 170.0, 169.7, 169.6, 169.6, 169.3, 169.2, 169.1, 139.3, 139.0, 138.0, 137.7, 137.7, 136.8, 136.7, 136.6, 136.6, 128.3, 128.3, 128.7, 128.2, 128.2, 128.1, 128.0, 127.4, 127.0, 126.9, 126.6, 126.3, 126.2, 126.1, 125.9, 125.9, 124.6, 124.5, 124.4, 124.3, 124.1, 121.9, 121.7, 121.6, 121.6, 121.3, 119.1, 119.1, 118.7, 117.9, 117.5, 117.5, 117.5, 111.6, 111.5, 111.4, 111.1, 106.7, 106.7, 106.3, 106.3, 106.3, 105.7, 65.5, 65.2, 65.1, 65.0, 51.5, 51.4, 51.3, 51.2, 51.1, 51.1, 50.9, 50.6, 50.5, 48.3, 48.1, 45.6, 45.5, 35.7, 35.6, 34.3, 34.1, 27.8, 27.8, 27.7, 27.6, 27.5, 27.5, 27.5, 26.5, 26.3, 13.3, 10.6, 10.5, 10.1, 5.3, 5.2, 4.3, 4.3, 4.2, 4.1, 3.9, 3.2. HRMS (ESI-) calcd for $\text{C}_{28}\text{H}_{35}\text{N}_4\text{O}_2$ [$M - \text{H}$] $^-$: 459.2754, found: 459.2764.

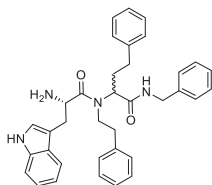
2-((*S*)-2-Amino-3-(1*H*-indol-3-yl)-*N*-phenethylpropanamido)-*N*-(*tert*-butyl)-4-phenylbutanamide hydrochloride (14**)**



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), phenylethylamine (**5**, 121 mg, 1 mmol), 3-phenylpropanal (**8**, 134 mg, 1 mmol) and *tert*-Butyl isocyanide (**9**, 83mg, 1 mmol) in MeOH (4 mL) afforded the desired product **14** as a white solid as a diastereomeric mixture in 60 % yield.

^{13}C NMR (101 MHz, CD_3OD) δ 171.9, 171.2, 171.2, 171.0, 171.0, 170.9, 170.9, 170.7, 170.7, 169.9, 142.2, 142.1, 141.8, 141.4, 140.4, 139.0, 138.8, 138.5, 138.3, 138.2, 138.0, 137.9, 137.8, 137.3, 130.3, 129.9, 129.9, 129.9, 129.8, 129.7, 129.7, 129.7, 129.6, 129.6, 129.6, 129.5, 129.4, 129.4, 129.4, 129.4, 129.4, 129.3, 129.3, 129.1, 128.5, 128.4, 128.2, 128.2, 128.1, 128.1, 128.0, 127.9, 127.6, 127.5, 127.3, 127.3, 127.2, 127.1, 127.1, 127.0, 126.6, 125.8, 125.8, 125.6, 125.5, 125.4, 123.1, 122.9, 122.9, 120.5, 120.5, 120.3, 119.2, 119.2, 119.0, 118.9, 112.8, 112.8, 112.8, 112.6, 107.9, 107.9, 107.9, 107.7, 107.1, 83.7, 81.2, 62.2, 61.3, 61.0, 60.8, 59.6, 59.0, 52.9, 52.8, 52.7, 52.7, 52.7, 52.6, 52.5, 52.4, 52.4, 51.6, 49.8, 49.5, 49.2, 49.1, 48.4, 46.9, 46.49, 43.28, 37.12, 36.78, 35.41, 35.32, 34.92, 34.76, 33.45, 33.36, 33.25, 33.19, 33.03, 32.9, 32.6, 32.4, 32.4, 32.2, 32.1, 31.8, 29.3, 29.2, 29.1, 28.8, 28.7, 28.7, 28.1, 27.7, 27.66. HRMS (ESI-) calcd for $\text{C}_{33}\text{H}_{39}\text{N}_4\text{O}_2$ [$M - \text{H}$] $^-$: 523.3067, found: 523.3079.

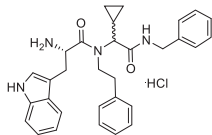
2-((S)-2-Amino-3-(1*H*-indol-3-yl)-*N*-phenethylpropanamido)-*N*-benzyl-4-phenylbutanamide hydrochloride (15)



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), phenylethylamine (**5**, 121 mg, 1 mmol), 3-phenylpropanal (**8**, 134 mg, 1 mmol) and benzyl isocyanide (**10**, 117mg, 1 mmol) in MeOH (4 mL) afforded the desired product **15** as a white solid as a diastereomeric mixture in 75% yield.

^{13}C NMR (101 MHz, CD_3OD) δ 172.0, 171.7, 171.3, 171.0, 170.9, 170.8, 164.3, 142.1, 142.1, 141.6, 141.4, 140.3, 140.2, 139.9, 139.6, 139.6, 139.4, 139.0, 138.8, 138.2, 138.1, 138.0, 137.8, 130.2, 130.0, 129.9, 129.7, 129.6, 129.6, 129.6, 129.5, 129.5, 129.5, 129.4, 129.4, 129.4, 129.3, 129.1, 129.0, 128.8, 128.8, 128.6, 128.4, 128.4, 128.3, 128.3, 128.2, 128.1, 128.0, 127.6, 127.6, 127.6, 127.4, 127.4, 127.2, 127.1, 127.1, 126.1, 125.8, 125.6, 125.5, 123.2, 123.2, 123.0, 120.8, 120.6, 120.6, 120.5, 119.2, 119.1, 118.8, 118.8, 112.9, 112.8, 112.8, 107.9, 107.8, 107.7, 107.5, 61.4, 60.8, 59.1, 52.8, 52.5, 51.8, 49.5, 49.4, 49.2, 49.0, 47.1, 47.0, 44.5, 44.4, 44.3, 44.1, 36.9, 36.7, 35.4, 34.7, 33.5, 33.4, 32.8, 32.5, 32.1, 32.1, 29.3, 29.2, 29.1, 28.6. HRMS (ESI-) calcd for $\text{C}_{36}\text{H}_{37}\text{N}_4\text{O}_2$ [$M\text{-H}$]: 557.2911, found: 557.2923.

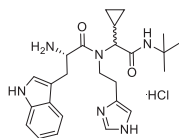
2-((S)-*N*-(2-(1*H*-Imidazol-4-yl)ethyl)-2-amino-3-(1*H*-indol-3-yl)propanamido)-*N*-(*tert*-butyl)-4-phenylbutanamide hydrochloride (16)



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), histamine (**5**, 101mg, 1 mmol), 3-phenylpropanal (**8**, 134 mg, 1 mmol) and *tert*-Butyl isocyanide (**9**, 83mg, 1 mmol) in MeOH (4 mL) afforded the desired product **16** as a white solid as a diastereomeric mixture in 37% yield.

^{13}C NMR (101 MHz, CD_3OD) δ 172.3, 171.5, 171.4, 171.3, 171.2, 171.1, 171.1, 170.8, 170.1, 142.4, 142.2, 142.1, 142.0, 141.8, 141.7, 138.0, 137.9, 137.9, 137.8, 137.7, 134.7, 132.7, 132.5, 131.1, 131.0, 130.1, 129.8, 129.6, 129.5, 129.5, 129.4, 129.4, 129.3, 129.3, 129.2, 129.1, 128.5, 128.3, 128.2, 128.1, 127.6, 127.3, 127.2, 127.1, 127.0, 126.1, 126.0, 125.9, 125.8, 125.6, 123.0, 122.9, 122.3, 122.8, 120.5, 120.4, 120.3, 120.2, 119.2, 119.1, 119.0, 117.4, 117.3, 117.2, 117.19, 117.05, 116.8, 112.9, 112.8, 112.8, 112.6, 108.0, 107.8, 107.7, 107.7, 107.5, 106.7, 61.9, 61.4, 61.0, 60.9, 60.5, 59.2, 54.8, 52.9, 52.9, 52.8, 52.7, 52.7, 52.5, 52.6, 52.4, 52.4, 52.1, 51.8, 51.6, 49.8, 47.4, 45.1, 44.9, 43.8, 43.2, 34.8, 33.7, 33.4, 33.4, 33.2, 32.9, 32.7, 32.6, 32.1, 31.9, 30.8, 30.7, 30.6, 30.3, 29.3, 29.1, 28.8, 28.7, 28.7, 28.7, 28.7, 27.9, 27.7, 27.6, 25.7, 25.5, 24.9, 24.4, 24.1, 23.6. HRMS (ESI-) calcd for $\text{C}_{30}\text{H}_{37}\text{N}_6\text{O}_2$ [$M\text{-H}$]: 513.2972 found: 513.2985.

(2S)-2-Amino-N-(2-(benzylamino)-1-cyclopropyl-2-oxoethyl)-3-(1*H*-indol-3-yl)-N-phenethylpropanamide hydrochloride (17)

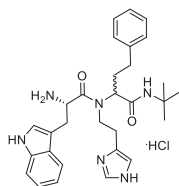


General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), phenylethylamine (**5**, 121 mg, 1 mmol), cyclopropanecarboxaldehyde (**7**, 70 mg, 1 mmol) and benzyl isocyanide (**10**, 117mg, 1 mmol) in MeOH (4 mL) afforded the desired product **16** as a white solid as a diastereomeric mixture in 40% yield.

¹³C NMR (101 MHz, CD₃OD) δ 173.4, 172.5, 171.9, 171.6, 171.0, 170.6, 170.5, 140.5, 140.3,

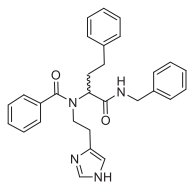
139.9, 139.8, 139.7, 139.5, 139.4, 139.1, 138.1, 138.1, 138.0, 130.1, 129.9, 129.8, 129.7, 129.7, 129.6, 129.6, 129.6, 129.5, 129.5, 129.5, 129.5, 129.4, 129.4, 129.4, 128.6, 128.5, 128.3, 128.3, 128.3, 128.2, 128.2, 128.1, 128.0, 128.0, 127.5, 127.5, 127.3, 125.9, 125.8, 125.5, 125.5, 125.4, 123.1, 123.0, 122.7, 120.5, 120.5, 120.4, 120.2, 119.1, 118.9, 118.8, 118.7, 113.0, 112.8, 112.8, 112.5, 108.1, 107.8, 107.6, 107.2, 67.2, 67.0, 66.6, 66.5, 65.9, 52.8, 52.7, 52.6, 52.5, 52.4, 51.9, 49.7, 47.1, 46.7, 44.5, 44.3, 44.2, 44.1, 36.9, 36.9, 35.5, 35.3, 30.8, 30.7, 29.1, 29.1, 29.0, 28.1, 26.9, 23.6, 14.4, 14.2, 11.9, 11.6, 10.6, 7.4, 7.1, 6.2, 6.0, 5.3, 5.2, 4.2. HRMS (ESI-) calcd for C₃₁H₃₃N₄O₂ [M-H]⁻: 493.2598, found: 493.2610.

(2S)-N-(2-(1*H*-Imidazol-4-yl)ethyl)-2-amino-N-(2-(*tert*-butylamino)-1-cyclopropyl-2-oxoethyl)-3-(1*H*-indol-3-yl)propanamide hydrochloride (18)



General procedure starting from commercially available Boc-*L*-Trp (**12**, 304 mg, 1 mmol), histamine (**6**, 101mg, 1 mmol), cyclopropanecarboxaldehyde (**7**, 70 mg, 1 mmol) and *tert*-Butyl isocyanide (**9**, 83mg, 1 mmol) in MeOH (4 mL) afforded the desired product **17** as a white solid as a diastereomeric mixture in 61% yield.

¹³C NMR (101 MHz, CD₃OD) δ 172.9, 172.8, 172.1, 171.6, 171.5, 171.1, 171.0, 170.5, 138.0, 137.9, 134.7, 134.6, 134.1, 132.8, 132.5, 131.8, 131.5, 131.3, 130.3, 128.6, 128.4, 128.3, 128.2, 126.0, 125.9, 125.8, 123.1, 123.0, 122.9, 122.7, 120.5, 120.4, 120.1, 119.2, 119.0, 119.0, 118.8, 117.5, 117.3, 117.1, 113.1, 112.8, 112.4, 107.9, 107.5, 107.5, 106.8, 67.2, 66.6, 66.6, 66.2, 52.9, 52.8, 52.4, 52.4, 52.3, 52.0, 51.8, 46.5, 46.0, 44.0, 33.0, 30.7, 30.6, 30.5, 30.5, 30.4, 30.2, 29.1, 29.0, 28.9, 28.9, 28.8, 27.8, 27.6, 25.8, 25., 24.8, 24.5, 23.6, 14.5, 14.4, 11.7, 11.6, 6.8, 6.5, 5.1, 5.4, 5.3, 5.2, 5.1, 4.5. HRMS (ESI) calcd for C₂₅H₃₃N₆O₂ [M-H]⁻: 449.2659, found: 449.2670.

***N*-(2-(1*H*-Imidazol-4-yl)ethyl)-*N*-(1-(benzylamino)-1-oxo-4-phenylbutan-2-yl)benzamide (19)**

General procedure starting from commercially available benzoic acid (**4**, 122mg, 1 mmol), histamine (**5**, 101mg, 1 mmol), 3-phenylpropanal (**8**, 134 mg, 1 mmol) and benzyl isocyanide (**10**, 117mg, 1 mmol) in MeOH (4 mL) afforded the desired product **19** as a white solid in 80% yield.

^1H NMR (400 MHz, CDCl_3) δ 8.58 (brs, 1H), 8.17 (brs, 1H), 7.43 – 6.11 (m, 17H), 4.74 – 4.17 (m, 3H), 3.87 – 3.30 (m, 2H), 3.18 – 1.86 (m, 6H). ^{13}C NMR (101 MHz, CDCl_3) δ 173.2, 171.2, 140.8, 138.3, 135.8, 134.4, 133.3, 129.6, 128.5, 128.4, 127.8, 127.3, 126.2, 117.4, 58.9, 49.0, 43.6, 32.9, 30.2, 25.9. HRMS (ESI) calcd for $\text{C}_{29}\text{H}_{30}\text{N}_4\text{O}_2$ $[\text{M}+\text{H}]^+$: 467.2441 found: 467.2443.

3.5 Contributions from co-authors

STD-NMR experiments and part of the synthesis were performed by Dr. V. R. Jumde. M. Witte participated in the discussion of the results and commented on the manuscript.

3.6 References

- [1] M. Jaegle, E. L. Wong, C. Tauber, E. Nawrotsky, C. Arkona, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, *56*, 7358–7378.
- [2] E. Oueis, C. Sabot, P.-Y. Renard, *Chem. Commun.* **2015**, *51*, 12158–12169.
- [3] J. Inglese, S. J. Benkovic, *Tetrahedron* **1991**, *47*, 2351–2364.
- [4] S. E. Greasley, T. H. Marsilje, H. Cai, S. Baker, S. J. Benkovic, D. L. Boger, I. A. Wilson, *Biochemistry* **2001**, *40*, 13538–13547.
- [5] W. G. Lewis, L. G. Green, F. Grynszpan, Z. Radić, P. R. Carlier, P. Taylor, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 1053–1057.
- [6] R. Manetsch, A. Krasinski, Z. Radić, J. Raushel, P. Taylor, K. B. Sharpless, H. C. Kolb, *J. Am. Chem. Soc.* **2004**, *126*, 12809–12818.
- [7] A. Krasinski, Z. Radić, R. Manetsch, J. Raushel, P. Taylor, K. B. Sharpless, H. C. Kolb, *J. Am. Chem. Soc.* **2005**, *127*, 6686–6692.
- [8] X. Hu, J. Sun, H. G. Wang, R. Manetsch, *J. Am. Chem. Soc.* **2008**, *130*, 13820–13821.
- [9] M. Gelin, G. Poncet-Montange, L. Assairi, L. Morellato, V. Huteau, L. Dugue, O. Dussurget, S. Pochet, G. Labesse, *Structure* **2012**, *20*, 1107–1117.
- [10] J. F. Chase, P. K. Tubbs, *Biochem. J.* **1969**, *111*, 225–235.
- [11] R. Nguyen, I. Huc, *Angew. Chem. Int. Ed.* **2001**, *40*, 1774–1776.
- [12] E. Oueis, F. Nachon, C. Sabot, P.-Y. Renard, *Chem. Commun.* **2014**, *50*, 2043–5.
- [13] T. Maki, A. Kawamura, N. Kato, J. Ohkanda, *Mol. Biosyst.* **2013**, *9*, 940–943.
- [14] T. Asaba, T. Suzuki, R. Ueda, H. Tsumoto, H. Nakagawa, N. Miyata, *J. Am. Chem. Soc.* **2009**, *131*, 6989–6996.
- [15] L. Weber, *Drug Discov. Today Technol.* **2004**, *1*, 261–267.
- [16] M. Jaegle, T. Steinmetzer, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, *56*, 3718–3722.
- [17] D. Bosc, J. Jakhlal, B. Deprez, R. Deprez-Poulain, *Future Med. Chem.* **2016**, *8*, 381–404.
- [18] A. Dömling, W. Wang, K. Wang, *Chem. Rev.* **2012**, *112*, 3083–3135.
- [19] I. Ugi, C. Steinbrückner, *Angew. Chem.* **1960**, *71*, 267–268.
- [20] P. Slobbe, E. Ruijter, R. V. A. Orru, *Med.Chem.Comm.* **2012**, *3*, 1189–1218.
- [21] 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.
- [22] L. Coates, H. H. Tuan, S. Tomanicek, A. Kovalevsky, M. Mustyakimov, P. Erskine, J. Cooper, *Eur. Biophys. J.* **2008**, *130*, 7235–7237.
- [23] L. Coates, P. T. Erskine, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Biochemistry* **2001**, *40*, 13149–13157.

- [24] J. Cooper, W. Quail, C. Frazao, S. I. Foundling, T. L. Blundell, C. Humblet, E. a Lunney, W. T. Lowther, B. M. Dunn, *Biochemistry* **1992**, *31*, 8142–8150.
- [25] 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.
- [26] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259–3263.
- [27] P. R. Gerber, K. Müller, *J. Comput. Aided. Mol. Des.* **1995**, *9*, 251–268.
- [28] BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, LeadIT, version 2. 1. 3.
- [29] L. Passerini, M.; Simone, *Gazz. Chim. Ital.* **1921**, *51*, 126–129.
- [30] M. V Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544–550.
- [31] M. Mondal, Nedyalka Radeva, Hugo Fanlo-Virgos, S. Otto, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2016**, *55*, 9422–9426.
- [32] BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, SeeSar, version 5.3.
- [33] K. M. R. Stierand, *ACS Med. Chem. Lett.* **2010**, *1*, 540.
- [34] H. Gohlke, M. Hendlich, G. Klebe, *Perspect. Drug Discov. Des.* **2000**, *20*, 115–144.
- [35] H. Köster, T. Craan, S. Brass, C. Herhaus, M. Zentgraf, L. Neumann, A. Heine, G. Klebe, *J. Med. Chem.* **2011**, *54*, 7784–7796.

Chapter 4

Protein-templated reductive amination for the identification of inhibitors of protein-protein interactions

In this chapter, we present another protein-templated reaction in the kinetic target-guided synthesis context, which was used to target protein-protein interactions, a challenging field in drug discovery. By using a protein-templated reductive amination, we screened a library of potential inhibitors of p53-Mdm2 protein-protein interaction and identified one hit showing a K_i value of 0.76 μM , demonstrating the efficiency of the approach regardless of the characteristics of flexible binding pockets.

M. Y. Unver, T. Felicetti, A. Twarda-Clapa, R. van der Vlag, F. Kassim, C. Ermis, C. G. Neochoritis, B. Musielak, A. Dömling, T. A. Holak, A. K. H. Hirsch, *under revision*.

4.1 Introduction

Discovery of fast and efficient techniques to identify bioactive compounds constitutes an important part in today's drug discovery. Target-guided synthesis (TGS) is a powerful approach in which the target selects its own inhibitors by assembling the corresponding binders from a library of complementary building blocks or by binding and amplifying them from a library of compounds formed in a reversible reaction.^[1] The two main methods in TGS are kinetic target-guided synthesis (KTGS) and dynamic combinatorial chemistry (DCC).^[2] In KTGS, the biological target accelerates the irreversible reaction between complementary building blocks upon binding, whereas in DCC a reversible reaction between building blocks affords a dynamic combinatorial library (DCL) from which the biological target selects and amplifies the best binder.^[3] Both techniques hold the potential to accelerate drug discovery and are still relatively underexplored, especially in terms of target scope and availability of the biocompatible reactions.

KTGS is a promising hit-identification strategy but only a few reactions with a limited number of targets have been reported so far.^[1,2,4-16] Most of them focus on acetylcholine esterase (AChE) from various animal species, as well as a few other targets.^[1] We will address both bottlenecks and report a novel protein-templated reaction using a new target for KTGS.

Protein-protein interactions (PPIs) are involved role in many biological functions such as intercellular communication and apoptosis. Targeting PPIs using small molecules is considered challenging given the flatness of the interface, a lack of small molecule starting points for the future design and the difficulties in distinguishing real from artefactual binding.^[17] We discussed PPIs in Chapter 1 in more detail.

p53 is a tumor-suppressor protein that is activated by cellular stress or damage and leads to cell-cycle arrest, apoptosis and DNA repair. Mdm2 is the negative regulator of the p53 protein and its overexpression leads to loss of p53 function.^[18] Mdm2 has well-defined and deep pocket, unusual for PPIs, accommodating a hotspot triad consisting of Trp23, Leu26 and Phe19 from p53. Therefore, design of high-affinity ligands to inhibit Mdm2 should focus on these hotspot amino acids of p53. Recently we found the Leu26 pocket to be a flexible pocket, which is enlarged upon ligand binding, making it very difficult to target by using structure-based drug design (SBDD) or other computational techniques such as virtual screening. Therefore, KTGS holds the potential to explore this flexible binding pocket by letting the protein sample which combination of building blocks ideally suited to fill it represent a valuable and efficient approach.^[19]

Use of protein-templated reactions to interrupt PPIs has only been shown for the Bcl-XL/BAX interaction^[20] and the 14-3-3 protein as discussed in Chapter 1.^[13] These targets feature

deep cavities in their binding pocket which makes them suitable for KTGS just like Mdm2. The reactions used for these two targets are the click/sulfo-click and S_{N2} thiol ring opening of epoxides, respectively.

Reductive amination represents another protein-templated reaction in which the biological target templates the synthesis of its own inhibitors from a pool of different aldehydes and amines by forming amines upon in situ reduction of the corresponding imines. Assembly of the inhibitors is achieved by reversible imine formation between aldehydes and amines after simultaneous binding to the adjacent pockets followed by reduction, affording the corresponding products featuring an irreversible amine bond (Figure 1).

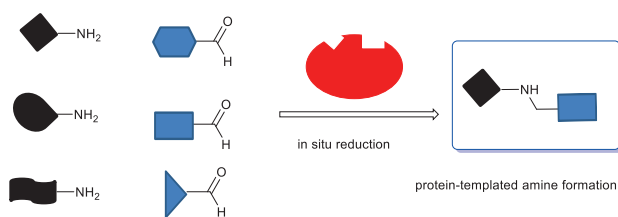


Figure 1. In situ protein-templated reductive amination.

The reversible reaction between an aldehyde and an amine to afford imines has been used by several groups for DCC using various biological targets.^[21–27] As the imines formed are unstable compounds, reduction was necessary; and the compounds were synthesized and tested as amine analogues with the risk that activity might change upon reduction.

Herein, we represent the first example of protein-templated reductive amination in the KTGS context for the identification of the inhibitors of PPIs which has never been applied to Mdm2-p53 interaction, an important interaction for cancer research, and the first application of KTGS for a flexible binding pocket (Leu26 pocket).

4.2 Results and discussion

4.2.1 Design of the inhibitor

After the discovery of the extended Leu26 pocket in Mdm2 shown by us.^[19] we set out to explore this flexible pocket. Having selected Mdm2 as a target, we designed our potential inhibitor scaffold starting from the X-ray crystal structure of inhibitor **1** in complex with Mdm2 (Figure 2, $K_i = 0.6 \mu\text{M}$, PDB: 4MDN).^[19]

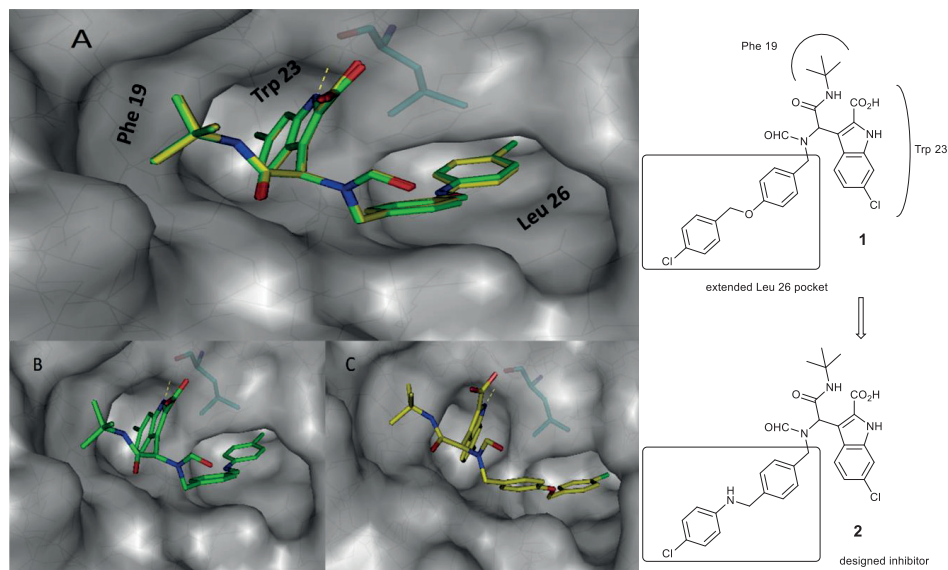
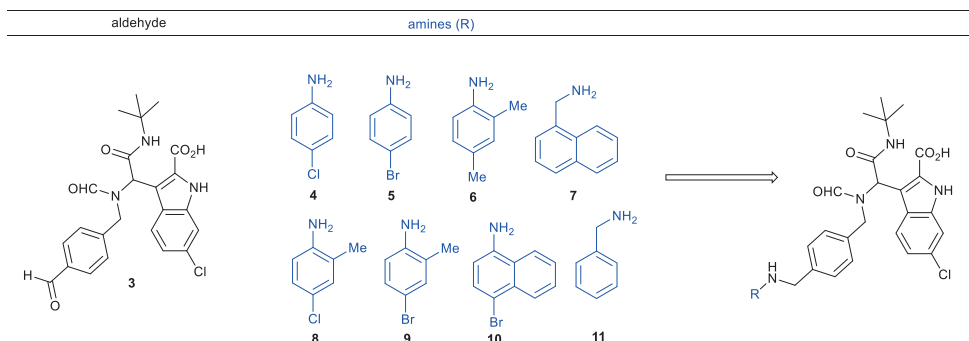


Figure 2. A) X-ray crystal structure of Mdm2 in complex with inhibitor **1** superimposed with designed inhibitor **2** (PDB: 4MDN). B) Designed inhibitor. Color code: inhibitor **2** skeleton: C: green, N: blue, O: red; C) inhibitor skeleton **1**: C: yellow, N: blue, O: red; protein backbone: gray; dashed lines: H-bonding interactions below 3.3 Å.

We designed and optimized a new inhibitor by using the molecular modeling program MOLOC.^[28] Inhibitor **1** occupies the three subpockets of Mdm2: the 6-chloroindole-2-hydroxamic acid moiety is hosted by the Trp23 pocket, the isobutyl group in the Phe19 pocket and the large 4-chlorobenzyl phenyl ether was found to occupy the enlarged Leu26 subpocket. In order to occupy this extended subpocket in an optimal way by using the reductive amination reaction, we designed the new scaffold by converting the 4-chlorobenzyl phenyl ether moiety into an amine, which can be assembled from the corresponding aldehyde **3** and amine **4** followed by in situ reduction (Scheme 1).

4.2.2 Generation of the library

Following the design of the initial inhibitor, we generated a combinatorial library by using aldehyde **3** as a core scaffold and eight different amines **4–11** to explore and fill the Leu26 pocket in the best manner.

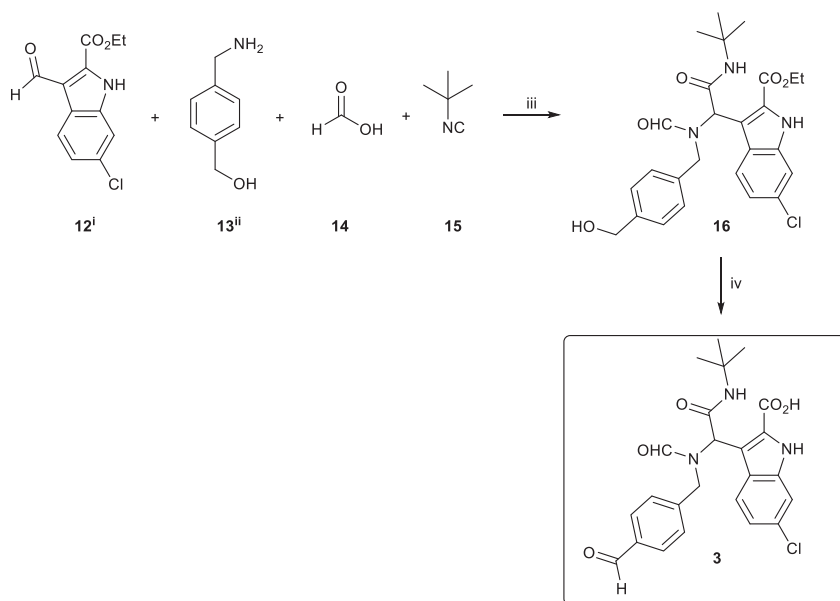


Scheme 1. Selection of building blocks for the protein-templated reductive amination, which affords eight possible amine products.

The use of KTGS to fill/explore flexible pockets is unprecedented. To explore this binding pocket further and possibly open-up the Leu26 pocket, we used benzyl- and naphthylamines in the library in addition to the aromatic amines to extend the length of the linker between the core scaffold and the amines. A potential extension in the pocket would be an important finding for future drug development.

4.2.3. Synthesis of the core scaffold **3**

Having selected the building blocks from commercially available amines, we synthesized the core scaffold **3** as shown in the Scheme 2.



Conditions and reagents: i) one-step synthesis^[29] ii) one-step synthesis^[30] iii) MeOH, 30 °C, 6 d, 52%; iv) a) DMP, CH₂Cl₂, rt, 3 h, b) 1 M LiOH, H₂O, EtOH, rt, 18 h, 85%

Scheme 2. Synthesis of the aldehyde building block **3**.

Literature protocols afforded compounds **12** and **13** each in one step.^[29,30] A four-component Ugi reaction of the aldehyde **12**, the amine **13**, formic acid (**14**) and *tert*-butyl-isocyanide (**15**) afforded the corresponding Ugi product **16** in 52% yield. Oxidation of the alcohol and subsequent hydrolysis of the ester led to the final aldehyde **3** in 85% yield over two steps.

4.2.4 Protein-templated reductive amination

Using the synthesized aldehyde building block and eight different commercially available amines **4–11**, we set up two experiments in parallel, a protein-templated reaction and a blank reaction at pH=6.8 (0.1 M phosphate buffer, 10% DMSO). In both reactions, we optimized the reaction conditions to a concentration of building blocks and reducing agent of 100 and 200 μM, respectively. One of the stringent requirements for KTGS is a substantial difference in reaction rate between the blank and the protein-templated reaction. As the imine formation between an aldehyde and an amine is a fast reaction, we used very dilute conditions to prevent product formation in the reference reaction for a certain time. As a result, less protein is required, an important consideration especially for precious proteins. By using a reducing agent in the reaction mixture from the beginning, the imines in the reaction mixture are stabilized and the imine formation is frozen by forming amines, which increases the amount of products formed to a

detectable level. Therefore, the two reactions were started by mixing all amines **4–11** (100 μM), the aldehyde **3** (100 μM) and NaCNBH_3 (200 μM). To the protein-templated reaction, we added Mdm2 (100 μM) (Figure 3). After careful screening of the reaction mixtures by using the UPLC-TQD-SIR (SIR: selective-ion recording) technique for seven days, we observed the formation of one hit compound only in the protein-templated reaction (Figure 4).

Protein- templated reductive amination

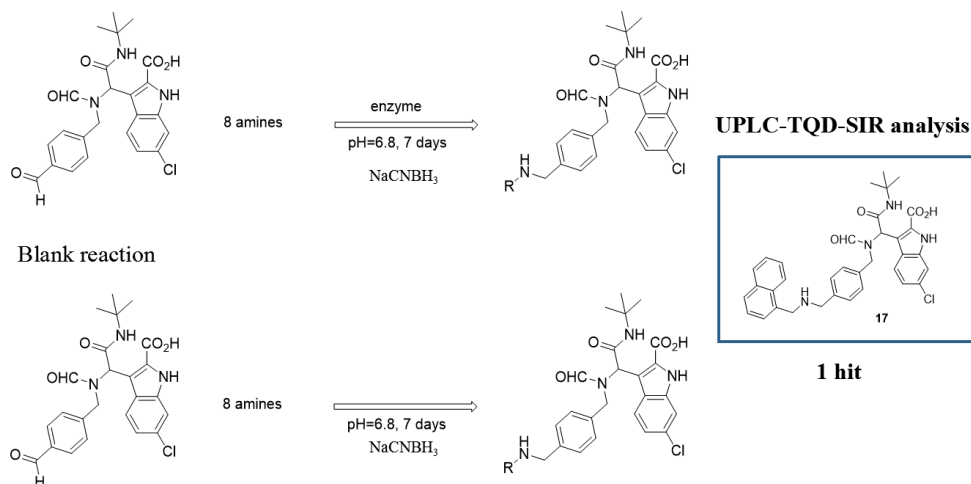


Figure 3. Protein-templated and blank reactions, formation of one hit after seven days.

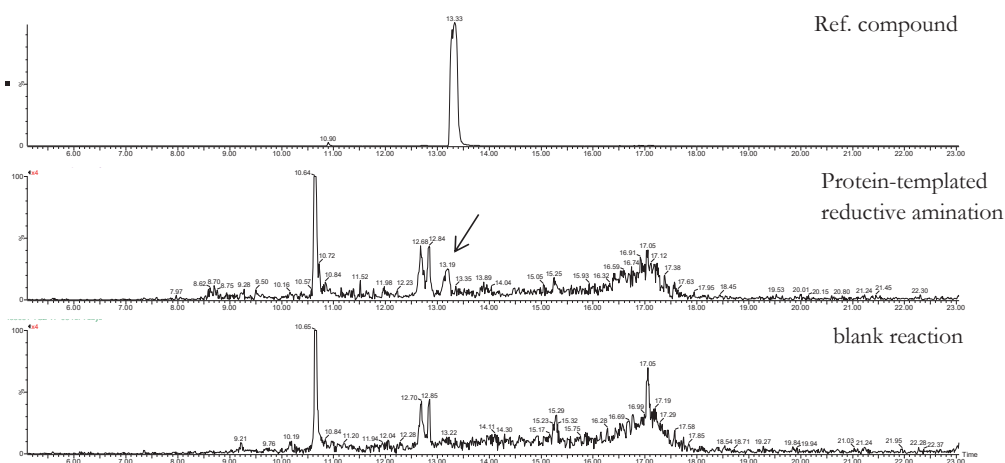
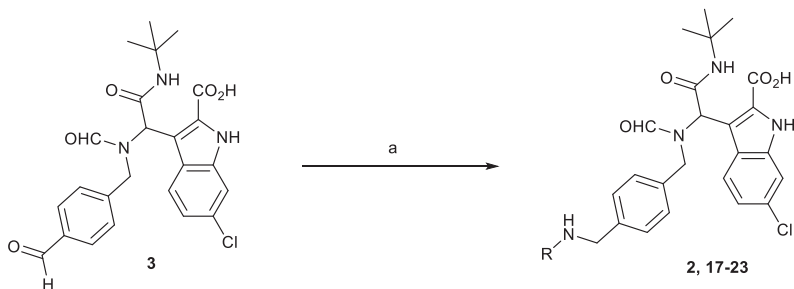


Figure 4. UPLC-TQD-SIR analysis of compound **17** ($[M+H]^+ = 611$). Formation of **17** by protein-templated reductive amination was compared with the blank reaction and synthesized compound **17**.

The SIR technique enables fast and sensitive screening of specific molecular weights (eight molecular weights per injection) regardless of very low concentrations. UPLC-TQD samples were prepared with the same concentration for both reactions, enabling us to detect the formation of compound **17** only in the presence of Mdm2. To demonstrate that the targeted site of Mdm2 is required for product formation, we repeated the same reaction in the presence of 100 μM bovine serum albumin (BSA) instead of Mdm2 and in the presence of a strong inhibitor of Mdm2, nutlin-3a (100 μM , IC_{50} = 90 nM). No product formation was observed in these control experiments. To exclude that the selection of the best combination is due to a difference in reaction rate, we started parallel reactions by using individual building blocks under the same conditions and the products appeared in each reaction after 10 days.

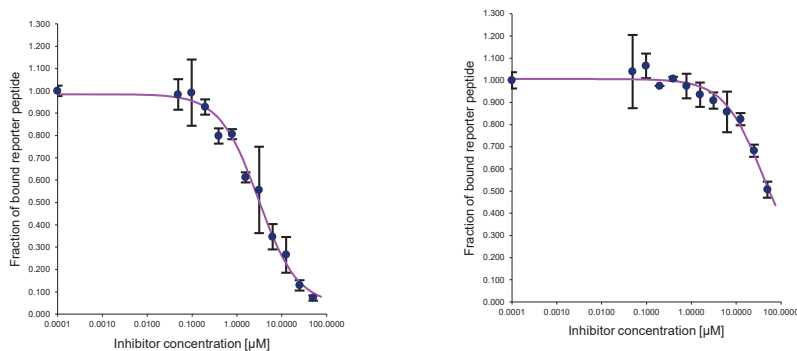
4.2.5. Synthesis and biochemical evaluation of the inhibitors

In order to confirm that the product formed in the presence of protein is indeed an inhibitor of the p53-Mdm2 interaction, we synthesized compound **17**, by adapting the reductive amination protocol and isolated the final compound in 52% yield as racemic mixture and used it for bioassays without further chiral separation (Scheme 3). Evaluation of the inhibitory potency using a fluorescence polarization assay confirmed that the compound **17** is a potent inhibitor ($K_i = 0.76 \pm 0.08 \mu\text{M}$, Figure 5).



^aConditions and reagents: amines **4-11**, pyrrolidine, 4 Å MS, Na(CH₃CO₂)₃BH, dry CH₂Cl₂, rt, 18 h

Scheme 3. Synthesis of the inhibitors **2**, **17-23**.



Mdm2 ($K_i=0.76 \pm 0.08 \mu\text{M}$)

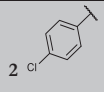
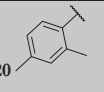
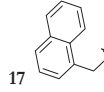
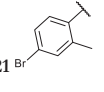
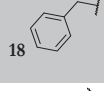
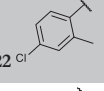
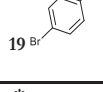
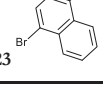
Mdmx ($K_i=12.16 \pm 2.15 \mu\text{M}$)

Figure 5. Inhibitory activity of compound **17** towards Mdm2/MdmX.

As the concept represents the first example for this target, we synthesized all possible Ugi products **2**, **17–23** (Table 1), starting from the core scaffold **3** by using the same adapted reductive amination protocol in 19–52% yields (Scheme 3).

After the synthesis of all the library members, we evaluated the inhibitory potency of compounds **2**, **17–23** by using the same fluorescence polarization assay. Mdmx is another p53 binding protein and shows significant homology with Mdm2. Therefore, we also tested all compounds against Mdmx (Table 1) by aiming to inhibit both targets.

Table 1. Biochemical evaluation of all possible products.

R:	K_i (μM) Mdm2	K_i (μM) Mdmx	R:	K_i (μM) Mdm2	K_i (μM) Mdmx
 2 Cl 34%	0.40 ± 0.05	4.18 ± 0.27	 20 33%	0.75 ± 0.08	7.08 ± 0.52
 17 52%	0.76 ± 0.08	12.2 ± 2.1	 21 Br 33%	0.49 ± 0.04	4.61 ± 0.39
 18 31%	3.18 ± 0.18	Not active	 22 Cl 22%	0.47 ± 0.04	3.73 ± 0.30
 19 Br 19%	0.63 ± 0.07	4.56 ± 0.49	 23 Br 12%	$0.25 \pm 0.02^*$	$3.28 \pm 0.43^*$

* 66% pure

As can be seen from Table 1, the majority of the compounds except for inhibitor **18** showed an activity in the same range and the hit **17** identified in the protein-templated reaction has lower activity according to biochemical evaluation of the corresponding amines. Due to the difficulties in the purification of compound **23**, we could not calculate the exact K_i for this compound. One of the main problems in using imine chemistry is the instability of the compounds, and therefore the requirement for in situ reduction. In this work, there are two main reasons why the activities of compounds are not correlating with the protein-templated reaction. The first well known reason is the risk of losing activity upon imine reduction. All the literature reports on imine-based DCC were conducted with the assumption of preservation of activity upon reduction.^[10] However, in situ reduction plays a vital role in our case because the amines formed in compounds **17** and **18** are aliphatic and the rest are aromatic amines. This clearly affects the protonation state of the amines, and thus their binding modes. Compound **17** is the only compound formed in this library in the presence of Mdm2 in a background-free reaction based on control experiments with BSA and nutlin-3a, strongly supporting the templated formation of inhibitor **17**. In addition to this, the imines formed in the reaction mixture are more rigid scaffolds, which can enable the further opening of the Leu26 pocket. In the case of compound **17**, the substituent is relatively longer chain than in compounds **19–23** which will potentially be hosted in the extended pocket in the imine form, and upon reduction of the corresponding imine the more flexible amine can easily change its binding mode, resulting in loss of activity. Nevertheless, we demonstrated the Mdm2-templated formation of **17**, which showed the applicability of KTGS in PPIs to explore flexible pockets, also demonstrated the risk of reduction in imine chemistry to determine the activities of the compounds. We performed crystallization studies to confirm the binding mode, however due to solubility problems of compound **17**, we could not obtain any crystals.

To confirm the inhibitory activity of compound **17**, we performed the ^1H - ^{15}N Heteronuclear Single Quantum Coherence (HSQC) NMR experiment. This method is based on monitoring of chemical shift changes in protein amide backbone resonances upon its interaction with a small molecule. For this experiment, the uniformly ^{15}N -labeled Mdm2 was titrated with an increasing concentration of compound **17** and ^1H - ^{15}N HSQC spectra were recorded after each new portion of the inhibitor has been added. In the course of titration, the shifts of the cross-peaks assigned to the amino acids of Mdm2 affected by binding of **17** were observed, which corroborates the binding (Figure 6).

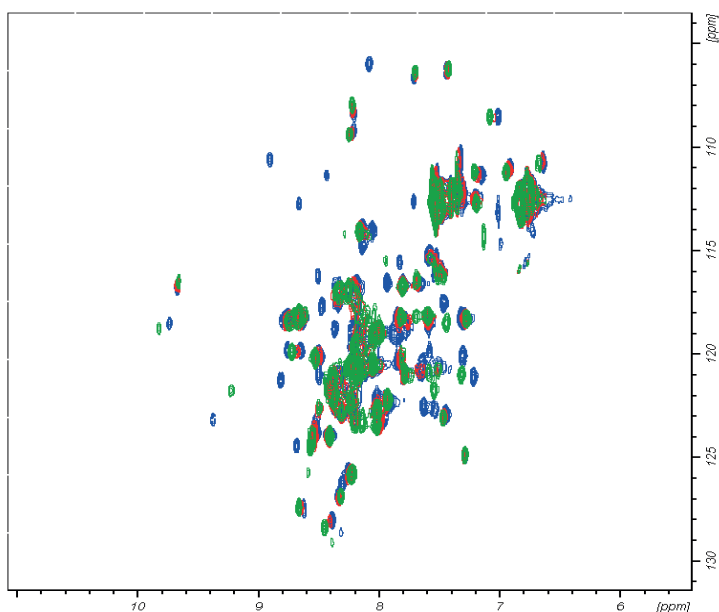


Figure 6. Superimposed ^1H - ^{15}N HSQC NMR spectra: blue-reference Mdm2 spectrum, red-4:1 (Mdm2: inhibitor **17**) titration step, green-1:5 (Mdm2: inhibitor **17**; over titration).

4.3 Conclusions

In conclusion, herein, we reported the first example of a reductive amination in KTGS. We expand the number of bio-compatible reactions available for medicinal chemists and show that KTGS can be applied when targeting PPIs. By using our target Mdm2 in situ, we screened a library of compounds in one-pot, revealing one hit after seven days. The synthesized compound emerged as a good lead compound showing a $K_i=0.76 \pm 0.08 \mu\text{M}$ activity. Our novel protein-templated reductive amination strategy could find applications in the early stages of drug discovery, namely hit identification/optimization, on a range of drug targets.

4.4 Experimental section

4.4.1 Modeling

The X-ray crystal structure of the complex of Mdm2 (PDB code: 4MDN) with compound **1** was used for our modeling studies.

For modeling studies, see Chapter 2, Section 2.4.2

4.4.2 Protein purification procedures

Mdm2

Fragments of human Mdm2 (residues 1–118 in pET46) were expressed in the *E. coli* BL21-CodonPlus (DE3) RIL strain. Cells were cultured in a total volume of 5 L of LB or minimal medium at 37 °C, and induced with 1 mM IPTG at OD_{600nm} of 0.8. Protein was expressed for 5 h at 37 °C. Cells were collected by centrifugation, re-suspended in 120 mL PBS with protease inhibitor cocktail and lysed by sonication. Inclusion bodies that were collected by centrifugation, washed twice with 120 mL PBS containing 0.05% Triton-X100 and once with 120 mL PBS and centrifuged after each wash. Purified inclusion bodies were solubilized in 20 mL of 6 M guanidine hydrochloride in 100 mM Tris-HCl, pH 8.0, containing 1 mM EDTA and 10 mM β-mercaptoethanol. The protein was dialyzed against 1 L of 4 M guanidine hydrochloride, pH 3.5 supplemented with 10 mM β-mercaptoethanol. Following, the protein was refolded by dropwise addition into 1 l of 10 mM Tris-HCl, pH 7.0, containing 1 mM EDTA and 10 mM β-mercaptoethanol and slow mixing overnight at 4 °C. Ammonium sulfate was added to the final concentration of 1.5 M, mixed for 2 h and centrifuged. The refolded protein was recovered on Butyl Sepharose 4 Fast Flow (GE Healthcare) previously equilibrated with refolding buffer containing 1.5 M (NH₄)₂SO₄. Mdm2 was eluted using 100 mM Tris-HCl, pH 7.2, containing 5 mM β-mercaptoethanol. Fractions containing the protein were pooled, concentrated to <10 mL and further purified by gel filtration on S75 16/600 column (GE Healthcare) in 50 mM phosphate buffer pH 7.4 containing 150 mM NaCl and 5 mM DTT (FP/NMR buffer).

MdmX

Fragments of human MdmX (residues 1–134 in pET46) were expressed in *E. coli* BL21-CodonPlus (DE3) RIL strain. Cells were cultured in a total volume of 5 L of LB or minimal medium at 37 °C and induced with 0.5 mM IPTG at OD_{600nm} of 0.6. Protein was expressed for 12 h at 20 °C. Cells were collected by centrifugation, resuspended in 120 mL 50 mM NaH₂PO₄ pH 8.0 with 300 mM NaCl, 10 mM imidazole and protease inhibitor cocktail and lysed by sonication. The protein was purified under native conditions. The lysate was cleared by centrifugation and loaded on Ni-chelating sepharose (GE Healthcare) previously equilibrated with lysis buffer (50

mM NaH₂PO₄ pH 8.0 with 300 mM NaCl and 10 mM imidazole). MdmX was eluted with 50 mM NaH₂PO₄ pH 8.0 containing 300 mM NaCl and 300 mM imidazole. Fractions containing the protein were pooled, concentrated and further purified by gel filtration on S75 16/600 column (GE Healthcare) in the FP/NMR buffer.

4.4.3 FP Assay

Fluorescence Polarization (FP) assay was used to monitor interactions between Mdm2 and MdmX proteins and their inhibitors.

For each assay, fresh protein stocks of Mdm2 (1–118) and MdmX (1–134) were thawed, and the protein concentrations were determined using Bradford method. Assay buffer contained 50 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA and 5% DMSO. Corning black 96-well NBS assay plates were used.

The binding affinity of P2 peptide (sequence: LTFEHYWAQLTS, labeled with carboxyfluorescein) towards Mdm2 and MdmX was first determined. For this purpose, 10 nM of the fluorescent P2 peptide was contacted with serial dilutions of tested protein (range from 750 to 0.012 nM for Mdm2 and from 3750 to 0.10 nM for MdmX) in a final volume of 100 μL and fluorescence polarization was determined. K_d was determined by fitting the curve described by below equation to experimental data:

$$FP = FP_{min} \frac{(FP_{max} - FP_{min}) \cdot c}{K_d + c}$$

where FP is the determined value of fluorescence polarization, FP_{min} - fluorescence polarization for ligand only, FP_{max} - fluorescence polarization at protein concentration saturating the ligand, and c – protein concentration.

Competition binding assay was performed using 10 nM fluorescent P2 peptide and optimal protein concentration for the measurement calculated based on determined K_d according to Huang, 2003 ($f_0 = 0.8$).^[33] Tested compounds were dissolved in DMSO at 50 μM. Serial dilutions (50 μM to 0.05 μM) were prepared in DMSO.

All the experiments were prepared in duplicates and plates were read 15 min after mixing of all assay components. Fluorescence polarization was determined using Tecan InfinitePro F200 plate reader with the 485 nm excitation and 535 nm emission filters. Fluorescence polarization values were expressed in millipolarization units (mP).

4.4.4 ¹H-¹⁵N Heteronuclear Single Quantum Coherence (HSQC) NMR experiment

Uniform ¹⁵N isotope labeling was obtained by expression of the protein in the M9 minimal media containing ¹⁵NH₄Cl as the sole nitrogen source. Final step of purification of Mdm2 for NMR consisted of gel filtration into the NMR buffer (50 mM phosphate buffer pH 7.4 containing 150

mM NaCl, 5 mM DTT). 10% (v/v) of D₂O was added to the samples to provide lock signal. Stock solutions of inhibitors of Mdm2/MdmX used for titration were prepared in d₆-DMSO. The samples were prepared by adding 50 mM ligand stock solution to the protein solution containing the ¹⁵N-labeled Mdm2 fragment at a concentration of 0.3 mM. 2D ¹H-¹⁵N correlated heteronuclear single quantum coherence (HSQC) NMR spectrum was recorded at 2–3 different ligand/protein ratios. All the spectra were recorded at 300 K using a Bruker Avance 600 MHz spectrometer. ¹H-¹⁵N heteronuclear correlations were obtained using the fast HSQC pulse sequence.^[31] Assignment of the amide groups of Mdm2 was obtained according to Stoll *et al.*, 2001.^[32] The spectra were processed with TopSpin 3.2 software.

4.4.5 Experimental Procedures

Protein-templated reductive amination

Mdm2 (162 μL, 0.308 mM in phosphate buffer 0.1 M, pH 6.8), the eight building blocks **4–11** (0.5 μL each, 100 mM in DMSO) and the aldehyde **3** (0.5 μL, 100 mM in DMSO) were added to a mixture of DMSO (45.5 μL) and phosphate buffer (288 μL, 0.1M, pH 6.8). Finally, NaCNBH₃ (1 μL, 100 mM in CH₃CN) was added. The reaction mixture was allowed to rotate at room temperature with 10 rpm with Falc rotary shaker. After 7 days, the library was analyzed by UPLC-TQD-SIR (electrospray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

Blank reaction, negative control

The eight building blocks **4–11** (0.5 μL each, 100 mM in DMSO) and the aldehyde **3** (0.5 μL, 100 mM in DMSO) were added to a mixture of DMSO (45.5 μL) and phosphate buffer (450 μL, 0.1 M, pH 6.8). Finally, NaCNBH₃ (1 μL, 100 mM in CH₃CN) was added. The reaction mixture was allowed to rotate at room temperature with 10 rpm with Falc rotary shaker. After 7 days, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

Protein-templated reductive amination experiments using BSA

The eight building blocks **4–11** (0.5 μL each, 100 mM in DMSO) and the aldehyde **3** (0.5 μL, 100 mM in DMSO) were added to a mixture of DMSO (45.5 μL) and BSA (450 μL, 0.110 mM in phosphate buffer 0.1 M, pH 6.8). Finally, NaCNBH₃ (1 μL, 100 mM in CH₃CN) was added. The reaction mixture was allowed to rotate at room temperature with 10 rpm with Falc rotary shaker. After 7 days, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

Protein-templated reductive amination experiments in the presence of Nutlin-3a

Mdm2 (162 μL , 0.308 mM in phosphate buffer 0.1 M, pH 6.8), nutlin-3a (0.5 μL , 100 mM in DMSO), the eight building blocks **4–11** (0.5 μL each, 100 mM in DMSO) and the aldehyde **3** (0.5 μL , 100 mM in DMSO) were added to a mixture of DMSO (45.0 μL) and phosphate buffer (288 μL , 0.1M, pH 6.8). Finally, NaCNBH_3 (1.0 μL , 100 mM in CH_3CN) was added. The reaction mixture was allowed to rotate at room temperature with 10 rpm with Falc rotary shaker. After 7 days, the library was analyzed by UPLC-TQD-SIR (ESI+) measurement because of its higher sensitivity and greater reliability for product identification.

Protein: Mdm2 (162 μL , 0.308 mM in phosphate buffer 0.1 M, pH 6.8) was added to 50 μL of DMSO and phosphate buffer (288 μL 0.1 M, pH 6.8). After 7 days, the enzyme solution was analyzed by UPLC-TQD-SIR (ESI+) measurements and compared with the positive hit identified from protein-templated reductive amination reaction.

UPLC-TQD-SIR method

UPLC-TQD was performed using a Waters Acquity UPLC H-class system coupled to a Waters TQD. All analyses were performed using a reversed-phase UPLC column (ACQUITY BEH C8 Column, 130 \AA , 1.7 μm , 2.1 mm x 150 mm). Positive-ion mass spectra were acquired using ES ionization, injecting 10 μL of sample; column temperature 35 $^\circ\text{C}$; flow rate 0.3 mL/min. The eluents, acetonitrile and water contained 0.1% of formic acid. The library components were eluted with a gradient from 95% \rightarrow 30% over 20 min, then at 5% over 1 min, followed by 5% for 2 min.

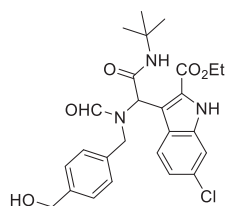
The UPLC-TQD-SIR method was used to analyze the formation of products in protein-templated and blank reactions. SIR measurements are highly sensitive, where a minute amount of compound can be detected by the mass spectrometer. $[M+H]^+$ were monitored using the full mass range to ensure correct isotope patterns for all possible potential products both for protein-templated and blank reactions. The product in the protein-templated reaction was identified by comparison of its retention time with that of synthesized compound using conventional methods on small scale.

General Experimental Details

For general experimental conditions, see Chapter 2, Section 2.4.4

Ethyl 3-(2-(*tert*-butylamino)-1-{formyl[4-(hydroxymethyl)benzyl]amino}-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylate (16**)**

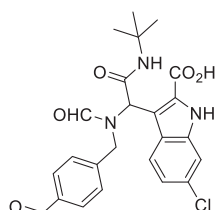
To a 50-mL round-bottomed flask charged with MeOH (10 mL), the corresponding amine **13** (0.40 g, 2.92 mmol), aldehyde **12** (0.63 g, 2.92 mmol), formic acid (**14**) (0.11 mL, 2.92 mmol) and *tert*-butyl isocyanide (**15**) (0.33 mL, 2.92 mmol) were added. The reaction mixture was stirred at room temperature for six days. Then, the reaction mixture was evaporated under vacuum, and the crude was purified by flash column chromatography eluting with CH₂Cl₂/MeOH (98:2→95:5). Compound **16** was obtained as a pale, yellow solid (772 mg, 52 % yield). m. p. = 113–115 °C; HRMS (ESI) calcd for C₂₆H₃₀ClN₅O₅ [M+H]⁺: 500.1947, found: 500.1946.



¹H NMR (400 MHz, CD₃OD) (major rotamer) δ ppm 8.37 (s, 1H, CHO), 7.78 (d, 1H, *J* = 8.8 Hz, H-4), 7.53 (br s, 1H, indole-NH), 7.38 (d, 1H, *J* = 1.9 Hz, H-7), 7.11 (dd, 1H, *J* = 1.9 and 6.8 Hz, H-5), 6.99 (d, 2H, *J* = 8.0 Hz, H-2' and H-6'), 6.75 (d, 2H, *J* = 8.0 Hz, H-3' and H-5'), 6.18 (s, 1H, CH), 5.10 (d, 1H, *J* = 15.3 Hz, ½ x CH₂), 4.68 (d, 1H, *J* = 16.2 Hz, ½ x CH₂), 4.46 (s, 2H, CH₂), 4.36 (q, 2H, *J* = 7.1 Hz, OCH₂), 1.40 (t, 3H, *J* = 7.2 Hz, CH₃), 1.22 (s, 9H, (CH₃)₃); ¹³C NMR (101 MHz, CD₃OD) (mixture of rotamers) δ ppm 169.8, 169.5, 165.1, 164.5, 160.7, 160.6, 139.9, 139.8, 136.6, 136.4, 136.4, 135.9, 130.6, 130.4, 128.0, 127.0, 126.9, 125.9, 125.7, 125.4, 125.3, 124.6, 122.0, 121.8, 121.1, 120.9, 114.4, 112.5, 111.8, 111.7, 63.4, 63.3, 60.8, 60.6, 56.9, 52.7, 51.3, 51.2, 49.7, 46.2, 27.4, 27.4, 13.3, 13.3.

3-{2-(*tert*-Butylamino)-1-[formyl(4-formylbenzyl)amino]-2-oxoethyl}-6-chloro-1*H*-indole-2-carboxylic acid (3**).**

To a solution of **16** (0.40 g, 0.80 mmol) in CH₂Cl₂ (20 mL), Dess-Martin periodinane (0.34 g, 0.80 mmol) was added portionwise. The reaction was stirred at room temperature for 3 h. Then, the mixture was quenched with a sat. aq. solution of NaHCO₃ (7 mL) and an aq. solution of Na₂S₂O₃ (10%, 3 mL) and extracted with CH₂Cl₂. The organic layer was washed with a sat. aq. solution of NaHCO₃ and a sat. aq. solution of NaCl, dried over MgSO₄, filtered and evaporated to dryness to give the oxidized intermediate as pale yellow solid that was quickly dissolved in EtOH (10 mL). Then, an aq. solution of LiOH (1 M, 0.19 g, 8.00 mmol, 10 mL) was added dropwise, and the reaction mixture was stirred at room temperature for 18 h. Then, the mixture was acidified with 1 M HCl to pH=6 and the precipitate was filtered. Compound **3** was obtained as a pale yellow solid (325 mg, 85 % yield). m. p. = 211–213 °C; HRMS (ESI) calcd for C₂₄H₂₄ClN₅O₅ [M+H]⁺: 470.1477, found 470.1482.



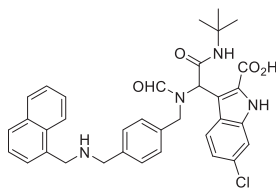
^1H NMR (400 MHz, CD_3OD) (major rotamer) δ ppm 9.80 (s, 1H, $\text{CH}'\text{O}$), 8.42 (s, 1H, CHO), 7.78 (d, 1H, $J = 8.2$ Hz, H-4), 7.56 (br s, 1H, indole-NH), 7.49 (d, 2H, $J = 7.4$ Hz, H-3' and H-5'), 7.32 (s, 1H, H-7), 7.10 (d, 1H, $J = 8.2$ Hz, H-5), 6.91 (d, 2H, $J = 7.4$ Hz, H-2' and H-6'), 6.29 (s, 1H, CH), 5.19 (d, 1H, $J = 15.9$ Hz, $\frac{1}{2}$ CH_2), 4.35 (d, 1H, $J = 14.8$ Hz, $\frac{1}{2}$ CH_2), 1.25 (s, 9H, $(\text{CH}_3)_3$); ^{13}C NMR (101 MHz, CD_3OD) (mixture of rotamers) δ ppm 192.2, 192.1, 165.2, 165.2, 144.4, 135.0, 128.6, 128.5, 127.5, 127.1, 126.6, 125.7, 125.7, 125.6, 124.6, 124.6, 121.7, 121.6, 121.0, 120.9, 120.8, 120.8, 111.7, 111.6, 111.6, 56.8, 52.5, 51.1, 46.4, 46.4, 27.4, 27.3.

General procedure for compounds 2, 18–23.

To a solution of compound **3** (1 eq), amine (1 eq), and pyrrolidine (0.1 eq) in dry CH_2Cl_2 , 4 Å MS (100 mg per mmol) and $\text{Na}(\text{CH}_3\text{CO}_2)_3\text{BH}$ (2 eq) were added. The reaction mixture was stirred at room temperature for 18 h under a nitrogen atmosphere. Then, the mixture was filtered over Celite[®], and the filtrate was evaporated under vacuum. The crude product was purified by flash chromatography column, eluting with ammonia infused (1/1 (v/v) ammonia/ CH_2Cl_2 mixed and CH_2Cl_2 phase was extracted) $\text{CH}_2\text{Cl}_2/\text{MeOH}$ (90:10).

3-{2-(*tert*-Butylamino)-1-[formyl(4-[(1-naphthylmethyl)amino]methyl)benzyl]amino]-2-oxoethyl}-6-chloro-1H-indole-2-carboxylic acid (**17**).

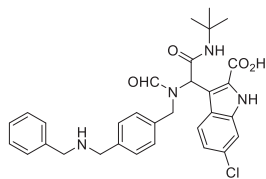
General procedure, starting from 1-naphthylmethylamine (73.3 mg, 0.12 mmol) as amine, compound **17** was obtained as a white solid (38 mg, 52% yield). m.p. 214–217 °C; HRMS (ESI) calcd for $\text{C}_{35}\text{H}_{35}\text{ClN}_4\text{O}_4$ [$M+\text{H}$]⁺: 611.2419, found 611.2425.



^1H NMR (400 MHz, $\text{DMSO}-d_6$) (major rotamer) δ ppm 11.48 (br s, 1H), 8.26 (s, 1H), 8.08 (d, 1H, $J = 9.3$ Hz), 7.91 (d, 1H, $J = 9.3$ Hz), 7.84 (d, 1H, $J = 8.1$ Hz), 7.65 (d, 1H, $J = 8.8$ Hz), 7.58 (br s, 1H), 7.52–7.43 (m, 4H), 7.32 (d, 1H, $J = 1.8$ Hz), 7.15 (d, 2H, $J = 8.1$ Hz), 7.03–6.99 (m, 1H), 6.93 (d, 2H, $J = 8.0$ Hz), 6.27 (s, 1H, CH), 4.84 (d, 1H, $J = 15.5$ Hz), 4.34 (d, 1H, 15.5 Hz), 4.32 (s, 2H), 3.85 (s, 2H), 1.09 (s, 9H); ^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) (mixture of rotamers) δ ppm 169.6, 164.2, 163.9, 137.4, 135.7, 133.7, 131.8, 131.7, 128.8, 128.4, 128.3, 127.6, 127.2, 126.5, 126.2, 125.7, 125.4, 124.3, 122.2, 120.4, 112.2, 57.0, 51.9, 50.8, 48.9, 47.1, 28.9, 28.7.

3-[1-[[4-[(Benzylamino)methyl]benzyl](formyl)amino]-2-(*tert*-butylamino)-2-oxoethyl]-6-chloro-1*H*-indole-2-carboxylic acid (18).

General procedure starting from benzylamine (10.7 mg, 0.1 mmol) as amine, compound **18** was obtained as a white solid (17 mg, 31% yield). m.p. 241–243 °C; HRMS (ESI) calcd for C₃₁H₃₄ClN₄O₄ [*M*+*H*]⁺: 561.2263, found 561.2255.

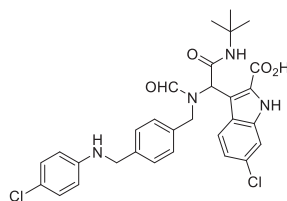


¹H NMR (400 MHz, DMSO-*d*₆) (major rotamer) δ ppm 11.48 (br s, 1H), 8.24 (s, 1H), 7.63 (d, 1H, *J* = 8.7 Hz), 7.56 (br s, 1H), 7.39–7.26 (m, 6H), 7.11 (d, 2H, *J* = 8.0 Hz), 6.98 (dd, 1H, *J* = 1.9 and 8.7 Hz), 6.91 (d, 2H, *J* = 8.0 Hz), 6.48 (s, 1H), 4.79 (d, 1H, *J* = 15.6 Hz), 4.31 (d, 1H, *J* = 15.6 Hz), 3.83 (s, 2H), 3.78 (s, 2H), 1.11 (s, 9H); ¹³C

NMR (101 MHz, DMSO-*d*₆) (mixture of rotamers) δ ppm 169.7, 164.2, 137.7, 135.9, 135.5, 129.4, 129.2, 128.7, 128.7, 128.6, 128.0, 127.6, 127.5, 126.2, 125.5, 122.1, 120.24, 112.1, 57.1, 51.1, 50.9, 50.8, 47.0, 28.9, 28.7.

3-(2-(*tert*-Butylamino)-1-(*N*-(4-(((4-chlorophenyl)amino)methyl)benzyl)formamido)-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylic acid (2).

General procedure starting from 4-chloroaniline (12.7 mg, 0.1 mmol) as amine, compound **19** was obtained as a white solid (20 mg, 34% yield). m. p. >119 °C (degradation); HRMS C₃₀H₂₉Cl₂N₃O₄ [*M*+*H*]⁺: 580.1644 (calc.), 580.1636 (found).

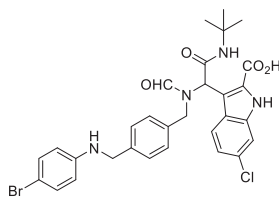


¹H NMR (400 MHz, CD₃OD) (major rotamer) δ ppm 8.35 (s, 1H), 7.71 (d, 1H, *J* = 8.8 Hz), 7.45 (s, 1H, indole-NH), 7.35 (d, 1H, *J* = 1.8 Hz), 7.11 (d, 2H, *J* = 7.9 Hz), 7.04–6.97 (m, 5H), 6.53 (d, 2H, *J* = 8.8 Hz), 6.41 (s, 1H), 5.02 (d, 2H, *J* = 15.2 Hz), 4.35 (d, 1H, *J* = 15.1), 4.17 (s, 2H), 1.14 (s, 9H); ¹³C NMR (101 MHz, CD₃OD)

δ ppm 169.7, 169.5, 165.2, 164.6, 147.4, 147.4, 138.3, 138.2, 136.2, 136.1, 136.0, 135.3, 130.1, 129.2, 128.2, 127.4, 127.3, 126.4, 126.3, 125.5, 124.7, 122.0, 121.7, 120.9, 120.7, 120.6, 113.6, 113.5, 111.7, 111.6, 56.8, 52.8, 51.0, 50.9, 49.7, 46.8, 46.7, 27.4, 27.3.

3-(1-(*N*-(4-(((4-Bromophenyl)amino)methyl)benzyl)formamido)-2-(*tert*-butylamino)-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylic acid (19)

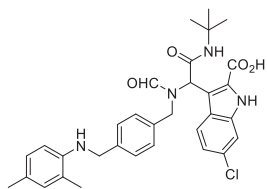
General procedure starting from 4-bromoaniline (17.2 mg, 0.1 mmol) as amine, compound **20** was obtained as a white solid (12 mg, 19% yield). m. p. >129 °C (degradation); HRMS C₃₀H₃₀BrClN₄O₄ [*M*+*H*]⁺: 624.1139 (calc.), 624.1136 (found).



^1H NMR (400 MHz, CD_3OD) (major rotamer) δ ppm 8.35 (s, 1H, CHO), 7.69 (d, 1H, $J = 8.8$ Hz), 7.44 (br s, 1H, indole-NH), 7.34 (d, 1H, $J = 1.9$ Hz), 7.14-7.11 (m, 4H), 7.02-6.97 (m, 3H), 6.49 (d, 2H, $J = 8.7$ Hz), 6.45 (s, 1H, CH), 5.00 (d, 1H, $J = 14.9$ Hz, $\frac{1}{2} \times \text{CH}_2$), 4.37 (d, 1H, $J = 15.0$ Hz, $\frac{1}{2} \times \text{CH}_2$), 4.18 (s, 2H, CH_2), 1.14 (s, 9H, $(\text{CH}_3)_3$); ^{13}C NMR (101 MHz, CD_3OD) (mixture of rotamers) δ ppm 170.2, 169.7, 166.5, 165.2, 164.6, 147.8, 147.8, 138.3, 137.9, 136.4, 135.5, 135.4, 135.2, 131.1, 129.0, 128.9, 127.7, 126.6, 126.5, 125.9, 125.6, 124.9, 121.8, 121.4, 120.33, 120.1, 114.1, 114.1, 111.5, 111.3, 110.8, 107.5, 107.4, 56.8, 53.2, 50.8, 50.7, 49.6, 46.7, 27.4, 27.3.

3-(2-(*tert*-Butylamino)-1-(*N*-(4-(((2,4-dimethylphenyl)amino)methyl)benzyl)formamido)-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylic acid (20)

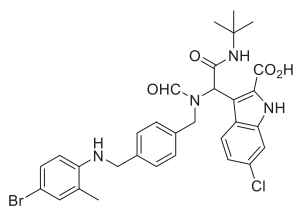
General procedure starting from 2,4-dimethylaniline (12.1 mg, 0.1 mmol) as amine, compound **21** was obtained as a white solid (19mg, 33% yield). m. p. >132 °C (degradation); HRMS: $\text{C}_{32}\text{H}_{35}\text{ClN}_4\text{O}_4$, $[M+H]^+$ 575.2420 (calc.), 575.2412 (found)



^1H NMR (400 MHz, CD_3OD) (major rotamer) δ ppm 8.36 (s, 1H), 7.70 (d, 1H, $J = 8.8$ Hz), 7.45 (s, 1H, indole-NH), 7.29 (d, 1H, $J = 1.8$ Hz), 7.12 (d, 2H, $J = 8.0$ Hz), 7.02-6.92 (m, 3H), 6.84-6.77 (m, 2H), 6.40-6.37 (m, 2H), 5.04 (d, 1H, $J = 15.0$ Hz), 4.33 (d, 1H, $J = 15.1$), 4.24 (s, 2H), 2.15 (s, 3H), 2.13 (s, 3H), 1.14 (s, 9H); ^{13}C NMR (101 MHz, CD_3OD) (mixture of rotamers) δ ppm 169.7, 165.2, 143.7, 138.9, 138.7, 135.3, 135.1, 130.2, 129.2, 127.6, 126.6, 126.6, 126.5, 126.3, 125.5, 125.5, 124.9, 122.2, 121.9, 121.4, 120.4, 120.4, 120.3, 111.5, 110.6, 110.4, 56.8, 5.0, 50.9, 50.8, 49.7, 27.4, 27.3, 19.0, 16.3.

3-(1-(*N*-(4-(((4-Bromo-2-methylphenyl)amino)methyl)benzyl)formamido)-2-(*tert*-butylamino)-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylic acid (21)

General procedure starting from 4-bromo-2-methylamine (19 mg, 0.1 mmol) as amine, compound **22** was obtained as a white solid (21 mg, 33% yield). m. p. >141 °C (degradation); HRMS: $\text{C}_{31}\text{H}_{32}\text{BrClN}_4\text{O}_4$ $[M+H]^+$ 639.1368 (calc.), 639.1371 (found).

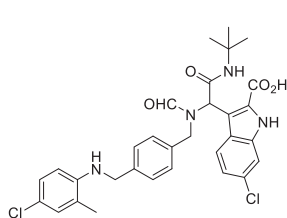


^1H NMR (400 MHz, CD_3OD) (major rotamer) δ ppm 8.36 (s, 1H, CHO), 7.71 (d, 1H, $J = 8.8$ Hz), 7.33 (d, 1H, $J = 1.8$ Hz), 7.12-6.99 (m, 5H), 6.92 (d, $\frac{1}{2}\text{H}$, $J = 8.0$ Hz), 6.51 (d, $\frac{1}{2}\text{H}$, $J = 8.0$ Hz), 6.40 (s, 1H), 6.32 (d, 1H, $J = 8.6$ Hz), 5.04 (d, 1H, $J = 15.1$ Hz), 4.35 (d, 1H, $J = 15.1$), 4.26 (s, 2H), 2.14 (s, 3H), 1.13 (s, 9H); ^{13}C NMR (101 MHz, CD_3OD) (mixture of rotamers) δ 170.0, 169.7,

165.7, 165.3, 164.6, 145.4, 145.3, 138.5, 138.3, 136.3, 135.7, 135.6, 135.2, 131.7, 129.5, 129.3, 128.9, 127.7, 126.4, 126.2, 125.8, 125.7, 124.9, 124.4, 124.3, 121.9, 121.5, 120.5, 120.4, 111.6, 111.6, 111.5, 111.4, 111.3, 110.7, 107.5, 107.4, 56.9, 53.2, 51.0, 50.8, 49.7, 47.0, 46.6, 46.6, 27.5, 27.3, 16.2.

3-(2-(*tert*-Butylamino)-1-(*N*-(4-(((4-chloro-2-methylphenyl)amino)methyl)benzyl)formamido)-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylic acid (22)

General procedure starting from 4-chloro-2-methylamine (14 mg, 0.1 mmol) as amine, compound **23** was obtained as a white solid (13.9 mg 22% yield). m. p. >123 °C (degradation); HRMS: C₃₁H₃₂Cl₂N₄O₄ [*M*+*H*]⁺: 595.1873 (calc.), 595.1874 (found).



¹H NMR (400 MHz, CD₃OD) (major rotamer) δ ppm 8.35 (s, 1H), 7.72 (d, 1H, *J* = 8.8 Hz), 7.34 (d, 1H, *J* = 1.8 Hz), 7.09 (d, 1H, *J* = 8.0 Hz), 7.06–7.00 (m, 1H), 6.98–6.87 (m, 4H), 6.48 (d, ½H, *J* = 8.0 Hz), 6.36 (d, ½H, *J* = 8.6 Hz), 6.33 (s, 1H), 5.06 (d, 1H, *J* = 15.1 Hz), 4.33 (d, 1H, *J* = 15.3), 4.26 (s, 2H), 2.15 (s, 3H), 1.15 (s, 9H); ¹³C NMR (101 MHz, CD₃OD) (mixture of rotamers) δ ppm 169.9, 169.6, 165.3, 164.6, 144.9, 144.9, 138.6, 138.4, 136.2, 136.0, 135.8, 135.2, 129.9, 129.7, 128.9, 127.6, 126.4, 126.1, 125.9, 125.7, 125.6, 124.8, 124.0, 123.9, 122.0, 121.6, 120.7, 120.6, 120.5, 120.5, 112.6, 111.7, 111.6, 110.9, 110.8, 56.9, 53.0, 51.0, 50.9, 49.7, 46.9, 46.8, 46.7, 27.5, 27.3, 16.2.

3-(1-(*N*-(4-(((4-Bromonaphthalen-1-yl)amino)methyl)benzyl)formamido)-2-(*tert*-butylamino)-2-oxoethyl)-6-chloro-1*H*-indole-2-carboxylic acid (23)

Following the general procedure and using 1-amino-4-chloronaphthalene as amine, compound **24** was obtained as a yellow solid in 12% corrected yield. The reaction was repeated several times, including longer reaction times of up to 70 h, but the product could not be obtained in >95% purity. Due to the small scale, low yields, costly starting material, as well as the issue that the *R_f* values of the product is the same as the starting material and an unknown side-product, **24** was used for biologically testing with a purity of 66% based on UPLC-MS (uv 256 nm). m. p. 165–170 °C (degradation); HRMS: C₃₄H₃₂BrClN₄O₄, 675.1368 (calc.), 675.1357 (found).

4.5 Contributions from co-authors

Synthesis of the compounds was partially done by T. Felicetti, F. Kassim during her Master's project, R. van der Vlag and C. Ermis during his internship. FP assays were performed by A. Twarda-Clapa and NMR binding experiment was performed by B. Musielak in the laboratory of T.D. Holak. C.G. Neochoritis and A. Dömling supplied some building blocks and exchanged knowledge.

4.6 References

- [1] D. Bosc, J. Jakhlal, B. Deprez, R. Deprez-Poulain, *Future Med. Chem.* **2016**, *8*, 381–404.
- [2] M. Jaegle, E. L. Wong, C. Tauber, E. Nawrotsky, C. Arkona, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, *56*, 7358–7378.
- [3] M. Mondal, A. K. H. Hirsch, *Chem. Soc. Rev.* **2015**, *44*, 2455–2488.
- [4] S. E. Greasley, T. H. Marsilje, H. Cai, S. Baker, S. J. Benkovic, D. L. Boger, I. A. Wilson, *Biochemistry* **2001**, *40*, 13538–13547.
- [5] J. Inglese, S. J. Benkovic, *Tetrahedron* **1991**, *47*, 2351–2364.
- [6] W. G. Lewis, L. G. Green, F. Grynszpan, Z. Radić, P. R. Carlier, P. Taylor, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 1053–1057.
- [7] R. Manetsch, A. Krasinski, Z. Radić, J. Raushel, P. Taylor, K. B. Sharpless, H. C. Kolb, *J. Am. Chem. Soc.* **2004**, *126*, 12809–12818.
- [8] A. Krasinski, Z. Radić, R. Manetsch, J. Raushel, P. Taylor, K. B. Sharpless, H. C. Kolb, *J. Am. Chem. Soc.* **2005**, *127*, 6686–6692.
- [9] X. Hu, J. Sun, H. G. Wang, R. Manetsch, *J. Am. Chem. Soc.* **2008**, *130*, 13820–13821.
- [10] M. Gelin, G. Poncet-Montange, L. Assairi, L. Morellato, V. Huteau, L. Dugue, O. Dussurget, S. Pochet, G. Labesse, *Structure* **2012**, *20*, 1107–1117.
- [11] J. F. Chase, P. K. Tubbs, *Biochem. J.* **1969**, *111*, 225–235.
- [12] R. Nguyen, I. Huc, *Angew. Chem. Int. Ed.* **2001**, *40*, 1774–1776.
- [13] T. Maki, A. Kawamura, N. Kato, J. Ohkanda, *Mol. Biosyst.* **2013**, *9*, 940–943.
- [14] T. Asaba, T. Suzuki, R. Ueda, H. Tsumoto, H. Nakagawa, N. Miyata, *J. Am. Chem. Soc.* **2009**, *131*, 6989–6996.
- [15] L. Weber, *Drug Discov. Today Technol.* **2004**, *1*, 261–267.
- [16] E. Oueis, F. Nachon, C. Sabot, P.-Y. Renard, *Chem. Comm.* **2014**, *50*, 2043–5.
- [17] A. Dömling, *Curr. Opin. Chem. Biol.* **2008**, *12*, 281–291.
- [18] K. Khoury, G. M. Popowicz, T. A. Holak, A. Dömling, *MedChemComm* **2011**, *2*, 246–260.
- [19] M. Bista, S. Wolf, K. Khoury, K. Kowalska, Y. Huang, E. Wrona, M. Arciniega, G. M. Popowicz, T. A. Holak, A. Dömling, *Structure* **2013**, *21*, 2143–2151.
- [20] S. K. T. Synthesis, *ACS Chem. Biol.* **2011**, 724–732.
- [21] M. F. Schmidt, A. Isidro-Llobet, M. Lisurek, A. El-Dahshan, J. Tan, R. Hilgenfeld, J. Rademann, *Angew. Chem. Int. Ed.* **2008**, *47*, 3275–3278.
- [22] A. Valade, D. Urban, J. M. Beau, *ChemBioChem* **2006**, *7*, 1023–1027.
- [23] S. Zameo, B. Vauzeilles, J. M. Beau, *Angew. Chem. Int. Ed.* **2005**, *44*, 965–969.
- [24] S. Zameo, B. Vauzeilles, J. M. Beau, *European J. Org. Chem.* **2006**, 5441–5444.

-
- [25] G. Nasr, E. Petit, C. T. Supuran, J. Y. Winum, M. Barboiu, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6014–6017.
- [26] G. Nasr, E. Petit, D. Vullo, J. Y. Winum, C. T. Supuran, M. Barboiu, *J. Med. Chem.* **2009**, *52*, 4853–4859.
- [27] Z. Fang, W. He, X. Li, Z. Li, B. Chen, P. Ouyang, K. Guo, *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5174–5177.
- [28] P. R. Gerber, K. Müller, *J. Comp.-Aided Mol. Des.* **1995**, *9*, 251–268.
- [29] G. M. Popowicz, A. Czarna, S. Wolf, K. Wang, W. Wang, A. Dömling, T. A. Holak, *Cell Cycle* **2010**, *9*, 1104–1111.
- [30] J. Zheng, Y. Li, Y. Sun, Y. Yang, Y. Ding, Y. Lin, W. Yang, *ACS Appl. Mater. Interfaces* **2015**, *7*, 7241–7250.
- [31] S. Mori, C. Abeygunawardana, M. O. van Zijl, Johnson, *J. Magn. Reson. B*, **1995**, *108*, 94–98.
- [32] R. Stoll, C. Renner, S. Hansen, S. Palme, C. Klein, A. Belling, W. Zeslawski, M. Kamionka, T. Rehm, P. Mühlhahn, et al., *Biochemistry* **2001**, *40*, 336–344.
- [33] X. Huang, *J. Biomol. Screen.* **2003**, *8*, 34–38.

Chapter 5

Protein-templated esterification reaction for the inhibitors of aspartic protease endothiapepsin

In this chapter, we discuss our attempts to use protein-templated esterification for the kinetic target-guided synthesis of the inhibitors of endothiapepsin. We describe the design of the ester inhibitor, its protein-templated formation, our attempts to optimize the library reaction and the synthesis of a library of ester inhibitors.

5.1 Introduction

Kinetic target-guided synthesis (KTGS) is a powerful strategy to identify binders of protein targets in an efficient and faster way.^[1,2] Our attempts to extend the scope of KTGS were demonstrated in the previous chapters. In this work, we aimed to add another reaction to the toolbox of protein-templated reactions, namely protein-templated esterification reaction (PTER).

PTER has the same principle with the other protein-templated reactions, enabling assembly of the inhibitors from two types building blocks bearing an alcohol and acid/ester functionality (Figure 1).

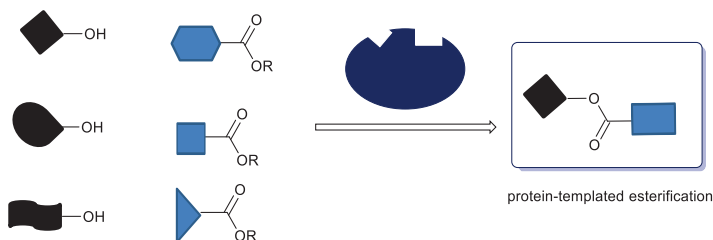


Figure 1. Protein-templated esterification reaction.

The target is introduced to the pool of building blocks having alcohol and acid/ester functionalities to select high affinity binders for each of the subpockets. Upon selection of the best combination in the adjacent pockets of the target, assembly of the inhibitors occurs by forming an ester linker between these two complementary building blocks. In other words, the protein holds these fragments in the proper orientation whilst bringing them in close proximity to facilitate the formation of the corresponding esters. As in the other protein-templated reactions, it facilitates the screening of a library of compounds without individual synthesis, characterization and biochemical evaluation of all library members.

Esterification is a reversible reaction under very acidic/basic conditions or in the presence of an enzyme and found application in dynamic combinatorial chemistry (DCC).^[3,4] Herein, we will summarize our attempts to use this reaction in the KTGS context for the first time by using the model enzyme endothiapepsin.

5.2 Results and discussion

5.2.1 Design of the ester inhibitor

We selected endothiapepsin as the target due to the advantages mentioned in the Chapters 2 and 3. Starting from the acylhydrazone inhibitor **1**, published in our group,^[5] we designed a new inhibitor featuring an ester functionally.

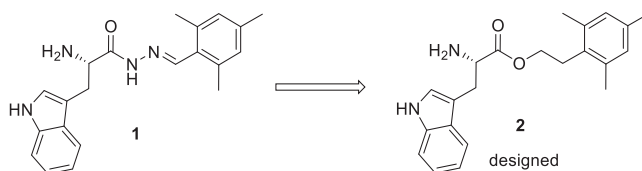


Figure 2. Design of the initial ester inhibitor **2**.

During our design, we used the X-ray crystal structure of inhibitor **1** in complex with endothiapepsin (PDB: 4KUP) and aimed to keep the same interactions in the active site by filling the same pockets with the parent inhibitor **1**. By using the proposed bioisostere **2**, we presumed that we would not lose any key interaction with the catalytic dyad (D35 and D210) and have acceptable affinity for the target, which is important to enable protein-templated reaction optimization.

We performed modeling and docking studies by using the molecular modeling program SEESAR^[6] and the FlexX docking module in the LeadIT suite for the structure-based design of **2**.^[7] According to our modeling studies, the designed inhibitor **2** can be hosted in the same pockets as the parent inhibitor **1** to engage in the same interactions in the active site (Figure 3).

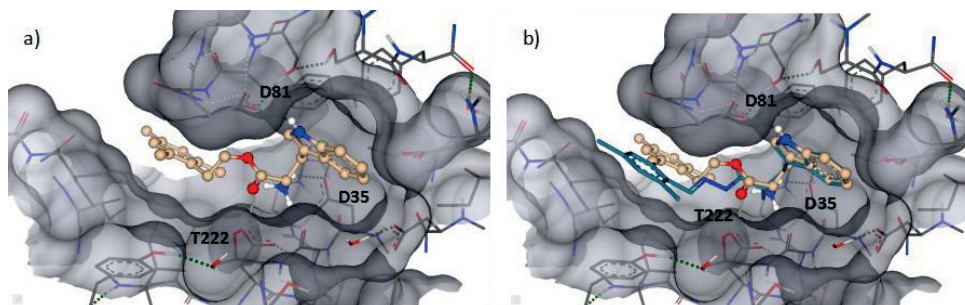
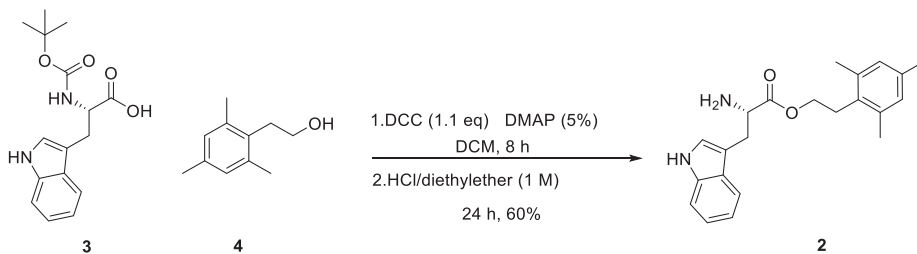


Figure 3. a) Top-ranked pose of inhibitor **2** generated by docking using the FlexX docking module in the LeadIT^[7] suite followed by evaluation using the scoring function HYDE in SEESAR.^[6] b) X-ray crystal structure of endothiapepsin in complex with inhibitor **1** superimposed with the designed inhibitor **2** (PDB: 4KUP). Color code: inhibitor **2** skeleton: C: cyan, N: blue, O: red; inhibitor skeleton **1**: C: light blue, N: blue, O: red; protein backbone: gray; dashed lines: H-bonding interactions below 3.3 Å.

5.2.2 Synthesis and biochemical evaluation of the designed inhibitor

Prior to optimization studies, the designed inhibitor was synthesized in two combined synthetic steps to validate its affinity with the target (Scheme 1).



Scheme 1. Synthetic pathway towards compound **2**.

Starting from Boc-protected tryptophan (**3**) and the amine, **4** a Steglich^[8] esterification using dicyclohexylcarbodiimide (DCC) in the presence of the catalytic amount of DMAP afforded the corresponding ester. Deprotection of the Boc-group on the tryptophan moiety afforded the desired compound **2** in 60% overall yield.

To evaluate the inhibitory potency of compound **2** we used an adaptation from the fluorescence-based assay for HIV protease,^[9] showing an IC_{50} of $29 \pm 4 \mu\text{M}$ (Figure 4).

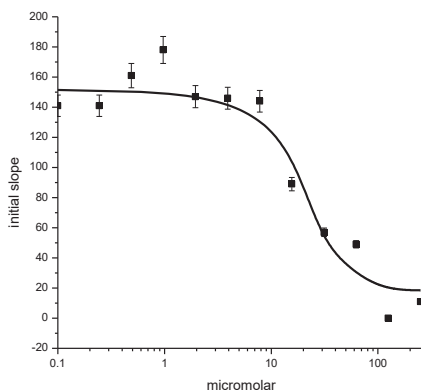
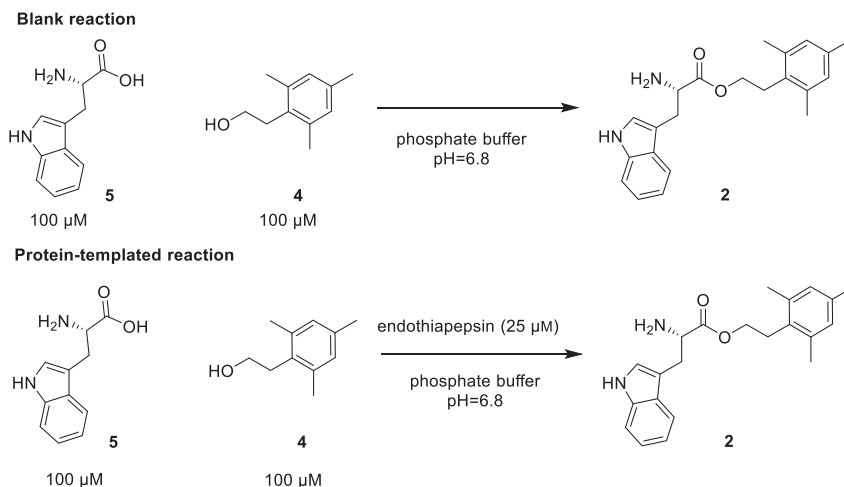


Figure 4. IC_{50} inhibition curve of **2** ($IC_{50} = 29 \pm 4 \mu\text{M}$). The inhibitor was measured in duplicate.

5.2.3 Optimization studies

Having selected a starting point with a desired affinity, we started two test reactions under the same conditions optimized for protein-templated click reaction (Chapter 2) and in situ Ugi reaction (Chapter 3). The first reaction is a blank reaction in which carboxylic acid **5** and alcohol **4** (100 μ M each) were mixed in the reaction buffer. The second reaction, protein-templated reaction, was started under identical conditions with a catalytic amount of endothiapepsin (25 μ M, Scheme 2).



Scheme 2. Test reactions for the protein-templated esterification reaction.

Due to the stringent requirements of KTGS, we aimed for a substantial rate difference between the blank and the protein-templated reactions. Therefore, we used 100 μ M building block concentration and an unactivated acid **5**. As the esterification reaction is reversible under acidic and basic conditions, we started both reactions under neutral conditions where the reaction is not reversible (pH=6.8). We analyzed the reaction mixture every day by using UPLC-TQD-SIR (SIR: selective-ion recording) technique, which enables fast and sensitive screening of specific molecular weights (Mws, up to 8 Mws per injection) regardless of very low concentrations, and observed the formation of compound **2** only in the presence of endothiapepsin after one week. No product formation was observed in the blank reaction (Figure 3).

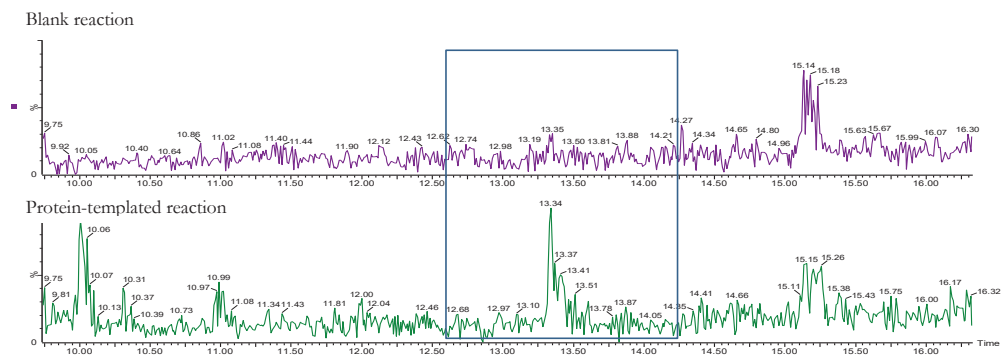
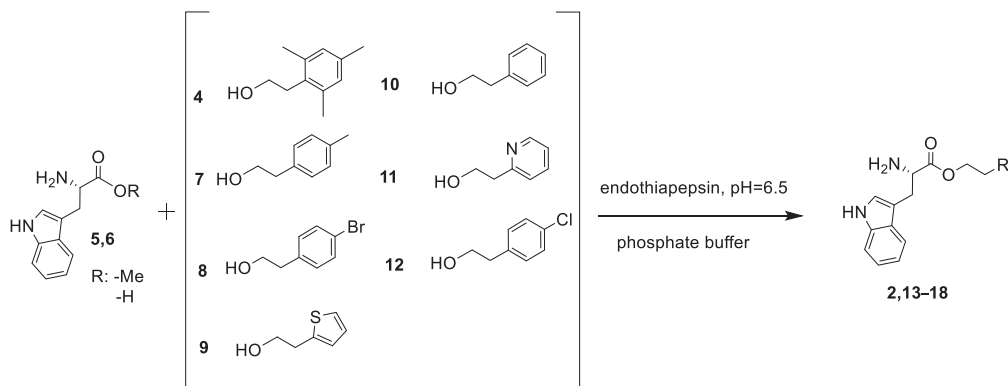


Figure 5. UPLC-TQD-SIR chromatograms of the test reactions.

5.2.4 Generation of the library and the library reaction

After the successful test reactions, we generated a library of compounds **2**, **13–18** which can be assembled from the carboxylic acid **5**/methyl ester **6** and seven different alcohols **4**, **7–12** (Scheme 3).

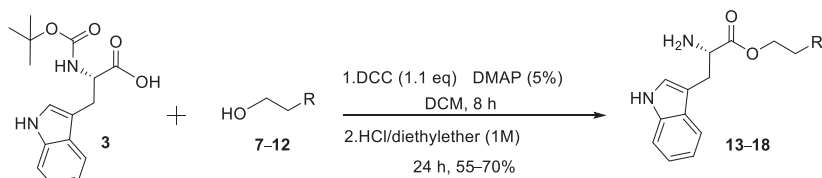


Scheme 3. Generation of the library and the library reaction.

Using the optimized conditions for the test reaction (100 μM building blocks, 25 μM endo-thiapsin), did not afford any product in the protein-templated reaction; therefore we performed optimization studies for the library reaction under various conditions. Starting from carboxylic acid **5**, we repeated the same reaction at higher concentrations 100–500 μM building blocks and 25–125 μM endo-thiapsin (Table 2). None of these conditions enabled the protein-templated esterification. We synthesized whole library prior to further optimization and tested against our target to make sure that some of the library compounds have sufficient affinity for the target as well as to use them as a reference for UPLC-TQD-SIR measurements.

5.2.5 Synthesis and biochemical evaluation of the library

Starting from Boc-protected tryptophan (**3**), we synthesized the whole library **13–18** in 45–88% yield by using the same protocol as for inhibitor **2** and evaluated all library members against endo-thiapsin by using the same fluorescence-type bioassay as described in the previous chapters (Table 1).



Scheme 4. Synthesis and biochemical evaluation of the library.

Table 1. Synthesis and biochemical evaluation results of the compounds **2**, **13–18**.

Compound	IC ₅₀ (μM)	Compound	IC ₅₀ (μM)
2 	29 ± 4	16 	36 ± 9
13 	64 ± 1	17 	No inhibition
14 	118 ± 22	18 	No inhibition
15 	11 ± 1		

As can be seen in the Table 1, compounds **2**, **13–19** show activities in the range of 11–118 μM , whereas the compounds **17** and **18** did not show any inhibition. The best hit in this library inhibits endothiapepsin with an IC_{50} of $11 \pm 1 \mu\text{M}$.

5.2.6 Discussion

Starting from carboxylic acid **5** and the alcohols **4**, **7–12** did not afford the protein-templated selection of the binders although we could see the template effect by using individual building blocks **4** and **5**.

Further attempts using the ester **6** (Scheme 3) under various conditions such as acidic water ($\text{pH}=4.6$) as the reaction medium, various protein concentration 10%–100% and longer reaction duration (up to 16 days) did not template the formation of any ester products (Table 2). This could be due to the lack of sufficient reactivity of carboxylic acid **5** or ester **6**, therefore requiring pre-activation of the acid component. This can be achieved by using several activated acids as described in the recent study by Rademann and coworkers for the protein-templated amidation reaction.^[10] In addition, of the seven compounds, five show a reasonable activity, which may lead to reduced amounts of products formed in the reaction mixture; making their detection particularly challenging.

Table 2. Conditions screened for the library reaction.

Building block concentration (μM)	Protein amount (μM)	Remarks*
100	25	acid 5 , $\text{pH}=6.8$ (phosphate buffer)
100	50	ester 6 , $\text{pH}=6.8$ (phosphate buffer)
100	100	ester 6 , $\text{pH}=6.8$ (phosphate buffer)
100	25	acid 5 , $\text{pH}=4.8$ (acidic water)
300	30	ester 6 , $\text{pH}=6.8$ (phosphate buffer)
300	75	ester 6 , $\text{pH}=6.8$ (phosphate buffer)
300	150	ester 6 , $\text{pH}=6.8$ (phosphate buffer)
300	75	acid 5 , $\text{pH}=6.8$ (phosphate buffer)
500	125	acid 5 , $\text{pH}=6.8$ (phosphate buffer)
500	125	ester 6 , $\text{pH}=6.8$ (phosphate buffer)
500	250	ester 6 , $\text{pH}=6.8$ (phosphate buffer)

*Reaction duration up to 16 days.

5.3 Conclusions

In this work, we set out to establish the protein-templated esterification reaction in the KTGS context. We demonstrated the protein-templated esterification of the inhibitor **2** with by using single complementary building blocks **4** and **5**. However, our attempts using multiple building blocks under various conditions did not afford any binder by assembly from this library. Therefore, we synthesized a library of esters in 45–88% yield and evaluated them against endothiapsin, showing activities in the range of 11–118 μM .

5.4 Experimental section

5.4.1 Fluorescence-based inhibition assay

For fluorescence-based inhibition assay, see Chapter 2, Section 2.4.1.

5.4.2 Modeling and docking

Taking inspiration from the cocrystal structure of endothiapsin with compound **1** (PDB code: 4KUP), a new ester scaffold was optimized by using SEESAR.^[6]

For docking studies, see Chapter 2, Section 2.4.2.

5.4.3 Experimental procedures

Protein-templated esterification

Endothiapsin (25–100 μL , 1.0 mM in phosphate buffer 0.1 M, pH 6.8) and the eight building blocks **4**, **5/6–12** (1–5 μL each, 100 mM in DMSO) were added to a mixture of DMSO (10%) and phosphate buffer (900 μL , 0.1 M, pH 6.8). The reaction mixture was allowed to rotate at room temperature with 10 rpm with Falc rotary shaker. After 1–16 days, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

Blank reaction, negative control:

The eight building blocks **4**, **5/6–12** (1–5 μL each, 100 mM in DMSO) were added to a mixture of DMSO (10%) and phosphate buffer (900 μL , 0.1 M, pH 6.8). The reaction mixture was allowed to rotate at room temperature with 10 rpm with Falc rotary shaker. After 1–16 days, the library was analyzed by UPLC-TQD-SIR (electro-spray ionization, (ESI+)) measurement because of its higher sensitivity and greater reliability for product identification.

UPLC-TQD-SIR method

UPLC-TQD was performed using a Waters Acquity UPLC H-class system coupled to a Waters TQD. All analyses were performed using a reversed-phase UPLC column (ACQUITY HSS T3 Column, 130 Å, 1.8 µm, 2.1 mm x 150 mm). Positive-ion mass spectra were acquired using ES ionization, injecting 10 µL of sample; column temperature 35 °C; flow rate 0.3 mL/min. The eluents, acetonitrile and water contained 0.1% of formic acid. The library components were eluted with a gradient from 95% → 50% over 15 min, then at 50% over 5 min.

The UPLC-TQD-SIR method was used to analyze the formation of ester products in protein-templated and blank reactions. SIR measurements are highly sensitive, where a minute amount of compound can be detected by the mass spectrometer. $[M+H]^+$ were monitored using the full mass range to ensure correct isotope patterns for all possible potential ester products both for protein-templated and blank reactions.

5.4.4 General Experimental Details

For general experimental conditions, see Chapter 2, Section 2.4.4

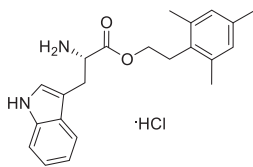
UPLC was performed using a Waters Acquity UPLC H-class system coupled to a Waters TQD. All analyses were performed using a reversed-phase UPLC column (ACQUITY C8 Column, 130 Å, 1.8 µm, 2.1 mm x 150 mm). The eluents, acetonitrile and water contained 0.1% of formic acid. The compounds were eluted with a gradient from 95% → 30% over 20 min, then at 5% over 1 min, followed by 5% for 2 min.

5.4.5 General procedure for Steglich^[8] esterification /de-protection reaction

N-Boc-Tryptophan (**3**) (1.2 mmol, 1 eq) was dissolved in anhydrous DCM. Alcohols **4**, **7–12** (1.2 mmol, 1 eq) were added to the reaction, followed by a catalytic amount of DMAP (5 mol%). The reaction was cooled to 0 °C, and a solution of DCC (1.32 mmol, 1.1 eq) in anhydrous dichloromethane was added dropwise over 10 mins. The reaction was allowed to warm to room temperature, and a precipitate was observed after 20 min. The reaction was stirred for a further 6–12 h. The reaction was filtered to remove the DCC urea by-product and the solvent removed in-vacuo. A quick purification with column chromatography using mixture of DCM/EtOAc (1:0 to 9:1) as an eluent afforded the *N*-boc protected product which was directly dissolved in DCM (3 mL) and HCl/diethyl ether (1 M, 15 mL) was added under nitrogen atmosphere. After stirring at r.t. for 24 h, the resulting white precipitate was collected and washed with Et₂O

Compounds **13**, **15**, and **16** were prepared according to procedures reported in the literature.^[10]

2,4,6-Trimethylphenethyl *L*-tryptophanate hydrochloride (**2**)

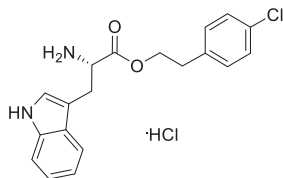


General procedure starting from commercially available Boc-*L*-Trp (**3**, 365 mg, 1.2 mmol) and 2-mesitylethan-1-ol (**4**, 197 mg, 1.2 mmol) afforded the desired product **2** as a white solid (278 mg, 60% yield).

m. p. >170 °C (degradation) $[\alpha]_{\text{D}}^{19} = -21^{\circ}$ (c=0.02, MeOH).

¹H NMR (400 MHz, Methanol-*d*₄) δ 7.47 (d, *J* = 7.9 Hz, 1H), 7.40 (d, *J* = 8.2 Hz, 1H), 7.17 – 7.11 (m, 2H), 7.05 (t, *J* = 7.5 Hz, 1H), 6.82 (s, 2H), 4.32 (t, *J* = 6.5 Hz, 1H), 4.20 (t, *J* = 7.7 Hz, 2H), 3.45 – 3.33 (m, 2H), 2.97–2.82 (m, 2H), 2.28 (s, 6H), 2.20 (s, 3H). ¹³C NMR (101 MHz, Methanol-*d*₄) δ 170.5, 138.4, 137.7, 137.1, 131.3, 130.0, 128.3, 125.8, 122.9, 120.3, 118.8, 112.7, 107.5, 65.9, 54.7, 28.9, 27.6, 20.9, 20.1. HRMS (ESI) calcd for C₂₂H₂₇N₂O₂ $[M+H]^+$: 351.2067, found: 351.2070.

4-Chlorophenethyl *L*-tryptophanate hydrochloride (**14**)

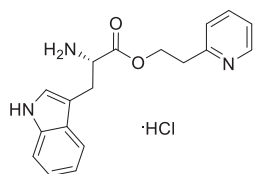


General procedure starting from commercially available Boc-*L*-Trp (**3**, 365 mg, 1.2 mmol) and 2-(4-chlorophenyl)ethan-1-ol (**12**, 187 mg, 1.2 mmol) afforded the desired product **14** as a white solid (219 mg, 58% yield).

m. p. >158 °C (degradation), $[\alpha]_{\text{D}}^{20} = +28.2^{\circ}$ (c=0.015, MeOH)

¹H NMR (400 MHz, Methanol-*d*₄) δ 7.45 (d, *J* = 7.9 Hz, 1H), 7.39 (d, *J* = 8.1 Hz, 1H), 7.29 – 6.99 (m, 7H), 4.31 (t, *J* = 6.7, 2.4 Hz, 2H), 4.24 (t, *J* = 6.7 Hz, 1H), 3.40 – 3.23 (m, 2H), 2.73 – 2.83 (m, 2H). ¹³C NMR (101 MHz, Methanol-*d*₄) δ 169.0, 136.8, 136.3, 132.1, 130.2, 128.2, 126.8, 124.1, 121.6, 118.9, 117.4, 111.3, 106.2, 66.3, 53.3, 33.5, 26.4. HRMS (ESI) calcd for C₁₉H₂₀ClN₂O₂ $[M+H]^+$: 343.1208, found: 343.1211.

2-(Pyridin-2-yl)ethyl *L*-tryptophanate hydrochloride (**17**)



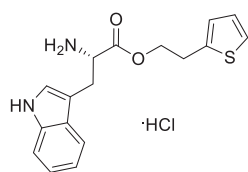
General procedure starting from commercially available Boc-*L*-Trp (**3**, 365 mg, 1.2 mmol) and 2-(pyridin-2-yl)ethan-1-ol (**11**, 148 mg, 1.2 mmol) afforded the desired product **17** as a white solid in (259 mg, 75% yield).

m. p. >195 °C (degradation) $[\alpha]_{\text{D}}^{20} = +66.0^{\circ}$ (c=0.01, MeOH)

¹H NMR (400 MHz, Methanol-*d*₄) δ 8.64 (d, *J* = 6.0 Hz, 1H), 8.40 (t, *J* = 7.9 Hz, 1H), 7.89 (t, *J* = 6.8 Hz, 1H), 7.62 (d, *J* = 7.9 Hz, 1H), 7.44 (d, *J* = 7.9 Hz, 1H), 7.34 (d, *J* = 8.1 Hz, 1H), 7.19 – 7.09 (m, 2H), 7.05 (t, *J* = 7.5 Hz, 1H), 4.61– 4.56 (m, 1H), 4.45 – 4.32 (m, 1H), 4.34 (t, *J* = 7.0 Hz, 1H), 3.39 – 3.31 (m, 3H), 3.24 – 3.17(m, 1H). ¹³C NMR (101 MHz, Methanol-*d*₄) δ 168.9,

152.9, 146.5, 141.0, 136.5, 127.3, 126.7, 125.2, 123.9, 121.5, 118.9, 117.4, 111.4, 106.3, 63.2, 53.18, 31.9, 26.4. HRMS (ESI) calcd for $C_{18}H_{20}N_3O_2$ $[M+H]^+$: 310.1550, found: 310.1552.

2-(Thiophen-2-yl)ethyl *L*-tryptophanate hydrochloride (**18**)



General procedure starting from commercially available Boc-*L*-Trp (**3**, 365 mg, 1.2 mmol) and 2-(thiophen-2-yl)ethan-1-ol (**9**, 154 mg, 1.2 mmol) afforded the desired product **18** as a white solid (221 mg, 63% yield).

m. p. >186 °C (degradation); $[\alpha]_D^{20} = +21.3^\circ$ (c=0.015, MeOH)

^1H NMR (400 MHz, Methanol- d_4) δ 7.48 (d, $J = 7.9$ Hz, 1H), 7.39 (d, $J = 8.1$ Hz, 1H), 7.23 (d, $J = 5.1$ Hz, 1H), 7.18 – 7.09 (m, 2H), 7.06 (t, $J = 7.5$ Hz, 1H), 6.93 – 6.90 (m, 1H), 6.84 (d, $J = 3.4$ Hz, 1H), 4.35 (t, $J = 6.6$ Hz, 2H), 4.33 – 4.25 (m, 1H), 3.41 (dd, $J = 15.0, 5.8$ Hz, 1H), 3.36 – 3.26 (m, 1H), 3.15 – 3.01 (m, 2H). ^{13}C NMR (101 MHz, Methanol- d_4) δ 169.0, 139.1, 136.8, 126.8, 126.5, 125.5, 124.0, 123.7, 121.5, 118.9, 117.3, 111.2, 106.1, 66.3, 53.3, 28.2, 26.2. HRMS (ESI) calcd for $C_{17}H_{19}N_2O_2S_1$ $[M+H]^+$: 315.1162, found: 315.1163.

5.5 Contributions from co-authors

A part of the protein-templated library reaction optimization studies were performed by L. Monjas.

5.6 References

- [1] X. Hu, R. Manetsch, *Chem. Soc. Rev.* **2010**, *39*, 1316–1324.
- [2] D. Bosc, J. Jakhlal, B. Deprez, R. Deprez-Poulain, *Future Med. Chem.* **2016**, *8*, 381–404.
- [3] W. He, Z. Fang, Z. Yang, D. Ji, K. Chen, K. Guo, *RSC Adv.* **2015**, *5*, 23224–23228.
- [4] R. C. Brachvogel, M. von Delius, *Eur. J. Org. Chem.* **2016**, *22*, 3662–3670.
- [5] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259–3263.
- [6] BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, SeeSar, version 5.3
- [7] BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, LeadIT, version 2. 1. 3.
- [8] B. Neises, W. Steglich, *Angew. Chem. Int. Ed. English* **1978**, *17*, 522–524.
- [9] M. V Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544–550.
- [10] M. Jaegle, T. Steinmetzer, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, 3718–3722.

Chapter 6

Design and synthesis of bioisosteres of acylhydrazones as stable inhibitors of the aspartic protease endothiapepsin

In this chapter, we describe the design, synthesis and biochemical evaluation of bioisosteres of acylhydrazones. We applied bioisosterism strategies to a previously reported acylhydrazone derivative, which is selected as hit-compound using a combination of dynamic combinatorial chemistry and *de novo* structure based drug design due to its promising inhibitory profile against endothiapepsin ($IC_{50} = 12 \pm 0.4 \mu\text{M}$). Among the series of three bioisosteres, two compounds are as potent as the hit. Unlike the labile acylhydrazones, these new bioisosteres do not liberate toxic hydrazides upon hydrolysis.

6.1 Introduction

We previously discovered acylhydrazone-based inhibitors of the aspartic protease endothiapepsin using dynamic combinatorial chemistry (DCC) in combination with *de novo* structure-based drug design.^[1-2] The most potent lead compound inhibits endothiapepsin, previously introduced in Chapters 3, with an IC_{50} value of 54 nM. However, it is important to focus on the behavior of the acylhydrazones *in vivo*. The major setback of acylhydrazones is their lack of stability under physiological conditions as they hydrolyze into an aldehyde and a hydrazide at acidic pH. Notably, formation of hydrazides can lead to potential toxicity issues. Therefore, it is highly desirable to identify stable substitute maintaining the key interactions in the active site of the protein without making significant changes in chemical structure. Acylhydrazones are considered to be a privileged structure in medicinal chemistry as they have the potential to interact with various biological targets.^[3] Introducing bioisosteres of the acylhydrazone moiety is therefore of paramount importance to obtain potential drug candidates.

Bioisosteres have been introduced as a fundamental strategy to improve the compatibility of the parent hit/lead compounds. As such bioisosteres contribute to the field of medicinal chemistry in terms of improving the potency, enhancing the selectivity, altering the physical properties, reducing/redirecting the metabolism, eliminating/modifying toxicophores and acquiring novel intellectual property.^[4]

To the best of our knowledge, there is no report so far on bioisosteres of acylhydrazones. Herein, we describe the design, synthesis and biochemical evaluation of the three bioisosteres of the acylhydrazone **1** (Figure 1), the first acylhydrazone inhibitor of endothiapepsin. Importantly, unlike the labile acylhydrazones, these new bioisosteres are not prone to hydrolysis, liberating toxic hydrazides.

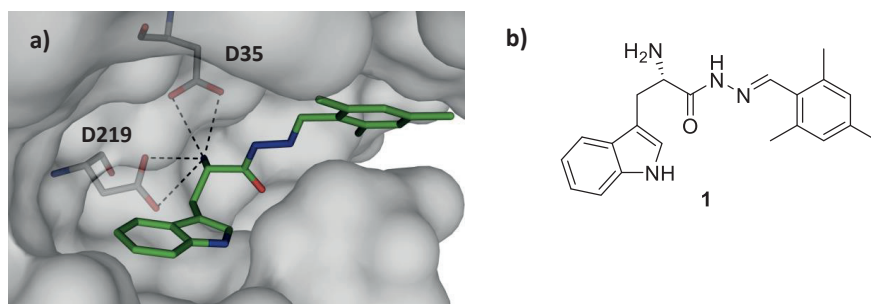


Figure 1: a) Overview of acylhydrazone **1** in complex with endothiapepsin (Color code: protein backbone: gray, inhibitor **1**: C: green, N: blue, O: red; PDB code: 4KUP).^[1] b) Structure of acylhydrazone inhibitor **1**.

6.2 Results and discussion

6.2.1 Design of the bioisosteres

We chose the X-ray crystal structure of endothiapepsin in complex with acylhydrazones **1** (PDB: 4KUP, Figure 1) as starting point for the design of the stable bioisosteres of the labile acylhydrazone moiety. The hit **1** displays an IC_{50} value of 12.8 μ M and a ligand efficiency (LE) of 0.27. It interacts with the catalytic dyad via H-bonding interactions (Asp35 (2.8 Å, 3.2 Å) and Asp219 (2.9 Å)) through its α -amino group.

The bioisosteres were designed using two ligand-based drug design approaches, namely Recore in the LeadIT suite^[5] and the molecular-modeling software MOLOC,^[6] which afforded various compounds displaying heterocyclic, ester or amide linkages (Figures 2 and 3).

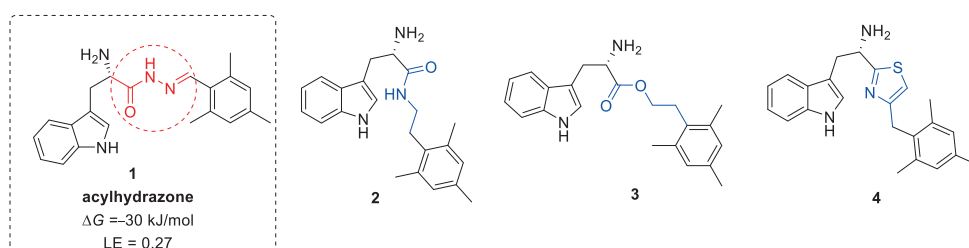


Figure 2. Proposed bioisosteres (**2–4**) of the acylhydrazone **1**, as stable inhibitors of endothiapepsin.

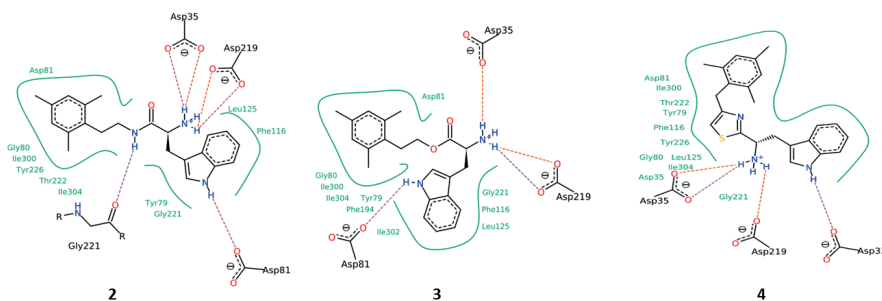
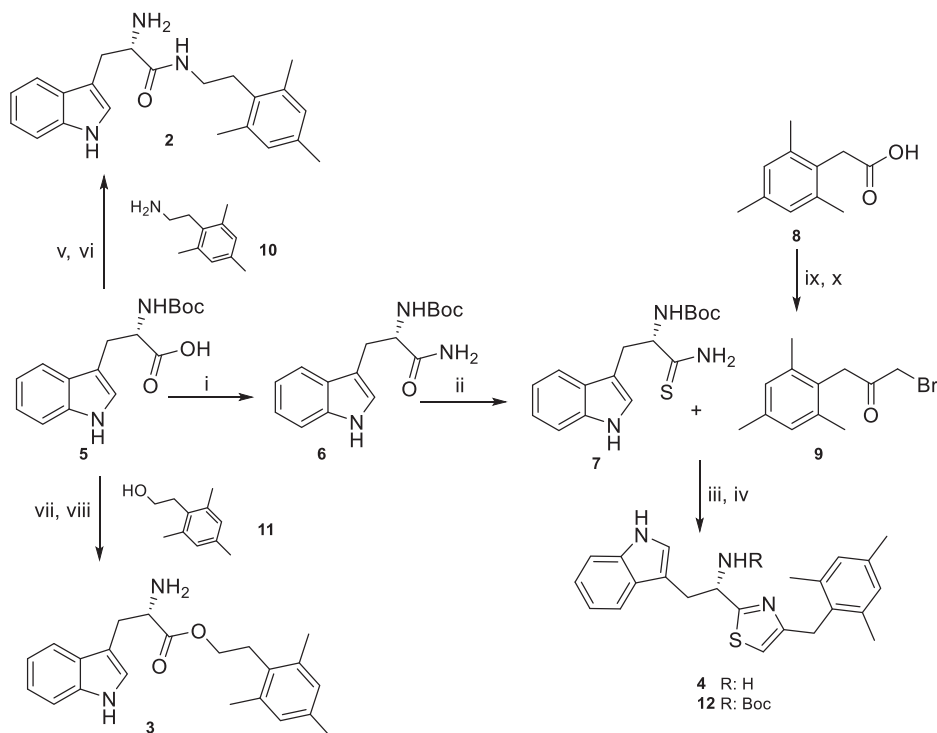


Figure 3. Predicted binding modes of acylhydrazone-derived bioisosteres **2–4** in the active site of endothiapepsin. These binding modes are the result of a docking run using the FlexX docking module with 30 poses and represent the top-scoring pose after HYDE scoring with SEESAR and careful visual inspection to exclude poses with significant inter- or intra-molecular clash terms or unfavorable conformations. The figures were generated with PoseView as implemented in the LeadIT suite.

Figure 3 shows the predicted binding modes of the three bioisosteres in the active site of endothiapepsin. These binding modes are the result of a docking run using the FlexX docking module with 30 poses and represent the top-scoring pose after HYDE scoring. Inspection of the co-crystal structure of endothiapepsin with acylhydrazone **1** in the active site of endothiapepsin shows that the aromatic part of the compound such as indolyl and/or mesityl groups are able to form π -stacking interactions with the amino acid residues of the protein backbone. In all of the structures (Figure 3), the binding mode of the indolyl- and mesityl- moieties is preserved. It was observed that the α -amino groups of all bioisosteres **2–4** form charge-assisted hydrogen bonds to the catalytic dyad (Asp35 and Asp 219) as well as additional interactions with the Asp81, Gly221. The NH group of the amide donates an H-bond to Gly221 in **2**. It was observed that in all of the bioisostere compounds the indolyl–NH form H bonds with Asp 81 and Asp33. In addition to these, the thiazol ring of **4** is involved in several hydrophobic interactions with the protein backbone.

6.2.2 Synthesis of the bioisosteres

We synthesized the bioisostere **2** with the amide linker using very mild peptide coupling reaction conditions followed by deprotection of the Boc-group in the subsequent step. Starting from *N*-Boc-tryptophan (**5**) and 2-mesitylethanamine hydrochloride (**10**) in the presence of a weak base carbonyldiimidazole, we obtained the corresponding amide in 80% yield and subsequently deprotected in presence of TFA to afford **2** in quantitative yield.^[7] To synthesize the ester **3**, a Steglich esterification was used as described in Chapter 5.^[8] We synthesized the bioisostere **4** from the building blocks thioamide **7** and ketobromide **9**, which can be accessed from *N* _{α} -Boc-L-tryptophan (**5**) and mesitylacetic acid (**8**), respectively, in two steps.^[8] Subsequent deprotection of compound **12** afforded the bioisostere **4** in quantitative yield. The first step to obtain thioamide **7**, consists of the formation of amide **6** followed by thionation using Lawesson's reagent. On the other hand, compound **9** was synthesized applying modified Arndt-Eistert reaction conditions starting from mesitylacetic acid (**8**).



i) ClCOOEt, Et₃N, dry THF, aq. NH₃, ii) Lawesson's reagent, dry DCM, iii) EtOH, reflux, 4 h, iv) TFA, DCM, v) 2-mesitylethanamine hydrochloride, 1,1'-Carbonyldiimidazole, THF, rt, overnight, vi) TFA, DCM, 0 °C to rt, 1.5 h, vii) 2-Mesitylethanol, DCC, DMAP (5%) DCM, 8 h, viii) HCl/Et₂O 1M, 24 h, ix) SOCl₂, dry toluene, reflux, 3 h, x) a. TMS-diazomethane, ether, b. 47.5% aq. HBr.

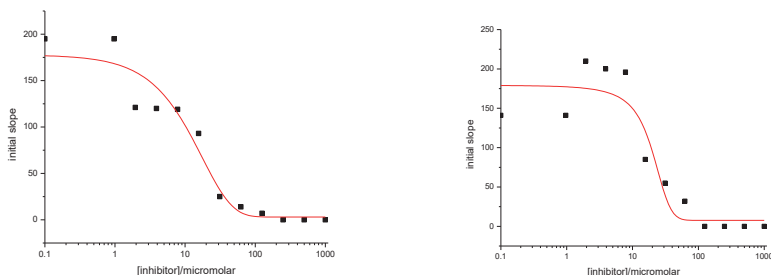
Scheme 1. Synthesis of bisosteres.

6.2.3 Biochemical evaluation

To investigate the biochemical activity of the designed bisosteres, we performed fluorescence-based inhibition assays adapted from the HIV-protease^[10] as described in the previous Chapters 2, 3 and 5. All of the three designed bisosteres inhibit the activity of endothiapepsin to a different extent. The most potent inhibitor, the amide bisostere **2**, displays an IC₅₀ value of 17.6 μM, comparable with the parent acylhydrazone **1** (IC₅₀ = 12.8 μM, Table 1, Figure 4).

Table1. Biochemical evaluation of bioisosteres 2–4.

Compound	IC ₅₀ (μM)
2	17.6 ± 6.8
3	28.7 ± 4.1
4	193.7 ± 11.4

**Figure 4.** A representative example of the IC₅₀ inhibition curves of amide bioisostere **2** (IC₅₀ = 17.6 ± 6.8 μM). The inhibitors were measured in duplicate.

6.2.4 Crystallographic studies

To verify the predicted binding mode of the bioisosteres, we soaked crystals of endothiapepsin with the most potent bioisostere **2** and determined the crystal structure of **2** in complex with endothiapepsin at 1.58 Å resolution (PDB code: 5OJE). The structure features clear electron density for the ligand, as shown in Figure 5.

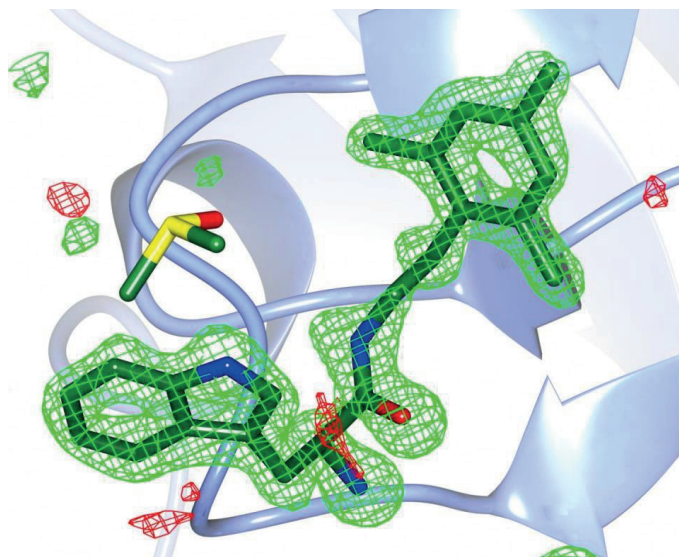


Figure 5. Electron density omit-map of the crystal structure of endothiapepsin in complex with compound 2 and a coordinated DMSO molecule. Fo-Fc Map contoured at 3.3σ (color code: protein cartoon: light blue; carbon: green; oxygen: red; nitrogen: blue; sulfur: yellow).

Upon closer examination, the binding mode of the ligand is similar to the docked pose shown in Figure 3. The amino group of the ligand forms two H bonds with the Asp35 (2.8 Å) and Asp219 (2.8 Å). The secondary amine of the indolyl ring forms a H bond with Asp81 with a distance of 3.0 Å. The hydrophobic part of the indolyl is engaged in hydrophobic interactions with Phe116, Leu125, Tyr79 and Gly221. The mesityl moiety is involved in hydrophobic interactions with Ile300 and 304, Tyr226, Gly80 and C β of Asp81. The only difference between the predicted and the real interactions is observed at the amide linkage. In contradiction to the computation, the nitrogen atom does not form a H bond with the oxygen atom of Gly221, the distance is 4.3 Å. Instead, the hydroxyl group of Thr222 acts as a H-bond acceptor and forms a H-bond (3.0 Å) with the nitrogen atom of the ligand, which is also shown in Figure 6.

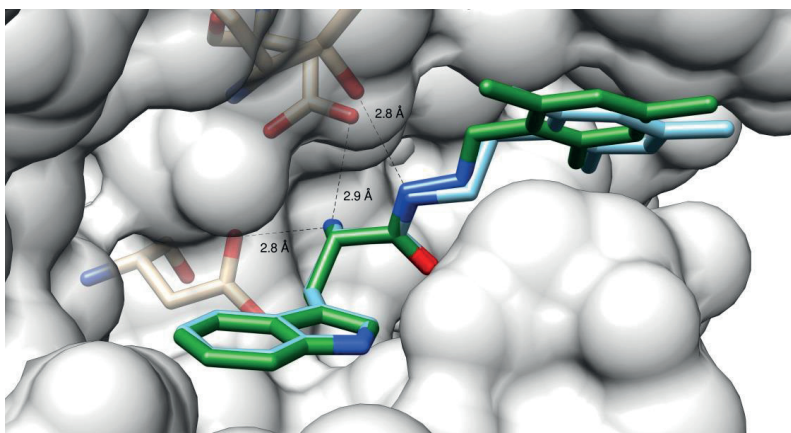


Figure 6. Superimposition of the acylhydrazone inhibitor **1** and the peptidic bioisostere **2**. (color code: protein backbone: C: gray, O: red, N: blue, **1**: C: green, **2**: C: light blue, N: blue, O: red, dashed lines: H-bonding interactions below 3.0 Å).

Due to the slightly bent shape of the coordinated ligand both aromatic groups are able to form hydrophobic interactions (distance of 4.2 Å) with one DMSO molecule, shown in Figure 7. This DMSO molecule is well coordinated and seems to displace several water molecules. This may be important for the stabilization of the ligand bound to the protein. A similar DMSO molecule can be observed in the previous structure (PDB code: 4KUP).

6.3 Conclusions

We report the successful replacement of the acid-sensitive and hydrolyzable acylhydrazone linker of parent hit **1**, as stable and equipotent inhibitors of endothiapepsin. We designed and synthesized three bioisosteres and evaluated them for their inhibitory activity against endothiapepsin. Compounds **2** and **3**, possessing amide and ester linkers respectively, display comparable IC_{50} values as the parent hit **1**, while compound **4** is an order of magnitude weaker than the parent hit.

We solved the crystal structure of amide **2** ($IC_{50} = 17.6 \pm 6.8 \mu\text{M}$) in complex with endothiapepsin, verifying predicted binding mode. Taken together, we demonstrate that acylhydrazones can be replaced without changing the binding whilst preserving the activity, demonstrating that acylhydrazone-based DCC is a powerful tool to identify hits, which can then be optimized to stable lead compounds in a straightforward manner.

6.4 Experimental section

6.4.1 Fluorescence-based inhibition assay

For fluorescence-based inhibition assay, see Chapter 2, Section 2.4.1.

6.4.2 Modeling and docking

For modeling and docking studies, see Chapter 2, Section 2.4.2.

6.4.3 Crystallization, data collection and processing

The protein, endothiapepsin, was purchased commercially as Suparen 600. It was purified by ultrafiltration using 10 kDa cut-off and washing with 0.1 mol/L Na/Acetate buffer, pH= 4.6. After washing, the protein concentration was adjusted to 5 mg/mL.

Crystals were obtained using sitting drop vapor-diffusion method with a drop volume of 4 μ L, using streak-seeding with microcrystals. The conditions were 0.1 mol/L NaAc, 0.1 mol/L NH₄Ac, 50% PEG4000. Crystallization occurred within two days. Soaking of the ligand was achieved within 24 h in mother liquor containing additional 0.1 mol/L ligand and 25% glycerol. Crystals were fished and flash-frozen in liquid nitrogen. Data collection was done using an Incoatec 1 μ m X-ray source equipped with an Oxford Cryosystems and a Mar 345dtb image plate detector.

The structure was solved using molecular replacement and refined using ccp4i2.^[11]

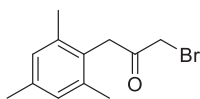
6.4.4 General experimental details

For general experimental conditions, see Chapter 2, Section 2.4.4

Optical rotations were measured in MeOH on a Schmidt & Haensch polarimeter (Polartronic MH8) with a 10 cm cell (*c* given in g/100 mL).

6.4.5 Experimental procedures

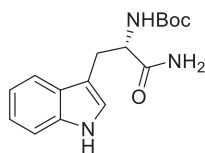
1-Bromo-3-mesitylpropan-2-one (**9**)



In a dry 100-mL three-necked flask, mesitylacetic acid (**8**) (1 eq., 1.5 g, 8.4 mmol) was dissolved in dry toluene (10 mL). Thionyl chloride (13.8 eq., 8.5 mL, 116 mmol) was added to this reaction mixture, and was refluxed for 3 h. The colorless solution was concentrated in vacuo to obtain a yellow crude oil, which was dissolved in ether (8 mL). This solution was added dropwise to trimethylsilyldiazomethane in ether (2.6 eq., 10.8 mL, 2 M, 21.7 mmol) at 0–5 °C. The yellow solution was left to stir at room temperature for 1 h followed by the dropwise addition of aqueous HBr (10.8 mL, 47.5%) and the orange emulsion was refluxed for 2.5 h. It was cooled down to room temperature and the ether layer was separated and washed with water. The aqueous layer was extracted with

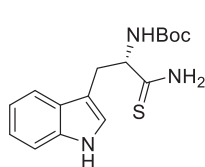
ether (3 x 35 mL). The combined organic layers were washed with saturated aqueous NaHCO₃ solution and water sequentially, dried over MgSO₄, filtered and concentrated in vacuo. The white solid was purified by column chromatography (SiO₂, DCM/pentane, 2:3, stained in Phosphomolybdic acid (PMA)) to yield **9** (2.56 mmol, 655 mg, 31%) as a white solid. ¹H-NMR (400 MHz, chloroform-*d*): δ 6.88 (d, *J* = 6.6 Hz, 2H), 3.98 (d, *J* = 1.8 Hz, 1H), 3.91 (d, *J* = 1.9 Hz, 1H), 3.67 (dd, *J* = 5.7, 1.9 Hz, 2H), 2.49 – 1.89 (m, 9H). ¹³C-NMR (101 MHz, chloroform-*d*): (equilibrium of keto-enol form) δ 199.8, 172.1, 137.1, 137.1, 137.0, 136.7, 129.3, 129.0, 128.6, 127.9, 52.1, 41.2, 35.0, 33.7, 21.0, 21.0, 20.4, 20.3. m. p. = 52–55 °C;

N-Boc-L-tryptophan amide (**6**)



In a 50-mL three-necked flask was dissolved *N*_α-Boc-L-tryptophan (**5**) (1 eq., 2 g, 6.57 mmol) in dry THF (14 mL). To the reaction mixture was added triethylamine (2.6 eq., 2.38 mL, 17.09 mmol) to obtain a colorless solution. The reaction mixture was cooled in an ice bath, and ethyl chloroformate (2 eq., 1.26 mL, 13.14 mmol) was added to obtain a white thick suspension. The reaction mixture was stirred in an ice bath for 30 min followed by addition of aqueous ammonia solution (1.58 mL, 25%). The reaction mixture was left to stir at room temperature for 30 min and the solvent was removed in vacuo. The reaction mixture was dissolved in EtOAc, and the organic phase was washed with saturated aqueous NaHCO₃ solution, dried over MgSO₄, filtered, concentrated in vacuo and purified by column chromatography (SiO₂, MeOH/DCM, 1:9) to yield **2** (3.80 mmol, 1.15 g, 58%) as a white solid. ¹H-NMR (400 MHz, dimethyl sulfoxide-*d*₆): δ 10.79 (s, 1H), 7.60 (d, *J* = 7.9 Hz, 1H), 7.40 – 7.25 (m, 2H), 7.12 (s, 1H), 7.08 – 6.89 (m, 4H), 6.64 (d, *J* = 8.3 Hz, 1H), 4.14 (td, *J* = 8.8, 4.7 Hz, 1H), 3.07 (dd, *J* = 14.6, 4.6 Hz, 1H), 2.89 (dd, *J* = 14.6, 9.2 Hz, 1H), 1.31 (s, 9H). ¹³C-NMR (101 MHz, dimethyl sulfoxide-*d*₆): δ 174.0, 155.2, 136.0, 127.4, 123.5, 120.8, 118.5, 118.1, 111.2, 110.4, 77.9, 54.9, 28.2, 27.8. [α]_D¹⁹ = + 0.135 (*c* = 0.68 in MeOH). m. p. = 110–115 °C; HRMS (ESI) calcd for C₁₆H₂₁N₃O₃Na [*M*+Na]⁺: 326.1475, found: 326.1476.

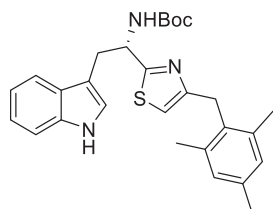
N-Boc-L-tryptophan thioamide (**7**)



To a 50-mL three-necked flask was added *N*-Boc-L-tryptophan amide (**6**) (1 eq., 790 mg, 2.60 mmol) and Lawesson's reagent (0.55 eq., 580 mg, 1.44 mmol) in dry DCM (10 mL) to obtain a white suspension. The reaction mixture was heated to reflux for 3 h to obtain a clear yellow-orange solution, which was cooled down to room temperature. To the reaction mixture, saturated aqueous NaHCO₃ solution (20 mL) was added, and this suspension was left to stir overnight.

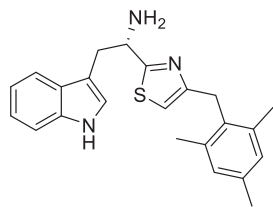
The reaction mixture was extracted with DCM (3 x 30 mL) and the combined organic layers were washed with saturated aqueous NaHCO₃ solution and dried over MgSO₄, filtered and concentrated in vacuo to yield a yellow solid. The crude was purified by column chromatography (SiO₂, EtOAc/pentane, 1:1) to yield **3** (1.42 mmol, 454 mg, 55%) as an off-white solid. ¹H-NMR (400 MHz, chloroform-*d*): δ 8.09 (s, 1H, -NH), 7.74 (d, *J* = 7.8 Hz, 1H), 7.37 (d, *J* = 8.1 Hz, 1H), 7.24 – 7.04 (m, 3H), 5.40 (br s, 1H), 4.73 (q, *J* = 7.2 Hz, 1H), 3.56 – 3.05 (m, 2H), 1.61 (br s, 2H), 1.42 (s, 9H). ¹³C-NMR (101 MHz, chloroform-*d*): δ 208.8, 155.5, 136.2, 127.3, 123.5, 122.3, 119.8, 118.9, 111.4, 110.4, 80.6, 60.7, 31.8, 28.4. [α]_D¹⁹ = + 0.22 (*c* = 0.56 in MeOH). m. p. = 73–75 °C HRMS (ESI) calcd for C₁₆H₂₁N₃O₂SNa [*M*+Na]⁺: 342.1247, found: 342.1248.

tert-Butyl (S)-(2-(1*H*-indol-3-yl)-1-(4-mesitylthiazol-2-yl)ethyl carbamate (12)



In a 25-mL round-bottomed flask was dissolved 1-bromo-3-mesitylpropan-2-one (**9**) (1 eq., 51.4 mg, 0.201 mmol) and *N*-Boc-L-tryptophan thioamide (**7**) (1.5 eq., 96.3 mg, 0.302 mmol) in absolute ethanol (6 mL). The yellow reaction mixture was heated to reflux for 4 h, and conversion was monitored by TLC (EtOAc/pentane, 3:7). The reaction mixture was concentrated in vacuo, and the orange liquid crude material was dissolved in a mixture of water and DCM. The organic layer was separated and the aqueous layer was extracted with DCM (3 x 15 mL). The combined organic layers were collected, dried over MgSO₄, filtered and concentrated in vacuo to afford the crude, which was purified by column chromatography (SiO₂, EtOAc/pentane, 3:7) to yield **6** (0.072 mmol, 34.1 mg, 36%) as a white solid, which was used in next step without further purification. ¹H-NMR (400 MHz, chloroform-*d*): δ 8.14 (s, 1H), 7.37 (d, *J* = 8.0 Hz, 1H), 7.32 (d, *J* = 8.1 Hz, 1H), 7.16 (t, *J* = 7.5 Hz, 1H), 7.03 (t, *J* = 7.5 Hz, 1H), 6.90 (s, 2H), 6.84 (s, 1H), 6.26 (s, 1H), 5.35 (br s, 1H), 4.13 (m, 2H), 3.57 – 3.33 (m, 2H), 2.30 (s, 3H), 2.28 (s, 6H), 1.43 (s, 9H). ¹³C-NMR (101 MHz, chloroform-*d*): δ 172.1, 155.7, 137.1, 136.9, 136.6, 136.0, 132.6, 129.0, 128.6, 123.4, 122.0, 119.5, 118.8, 113.0, 111.1, 80.0, 52.0, 35.0, 28.4, 21.0, 20.3, 20.0.

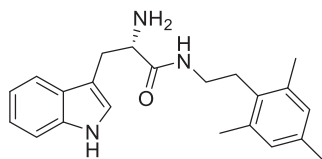
(S)-2-(1*H*-Indol-3-yl)-1-(4-mesitylthiazol-2-yl)ethan-1-amine (4)



In a 25-mL round-bottomed flask was dissolved **4** (1 eq., 26.1 mg, 0.055 mmol) in dry DCM (1 mL). The reaction mixture was cooled to 0 °C, and trifluoroacetic acid (TFA) (16 eq., 0.067 mL, 0.878 mmol) was added. The solution was stirred in an ice bath for 5 min and at room temperature for 1 h. The reaction mixture was concentrated in vacuo, and the crude was dissolved in chloroform.

The organic layer was washed with saturated aqueous NaHCO_3 solution and saturated aqueous NaCl solution, dried over MgSO_4 , filtered and concentrated in vacuo. The crude was purified by column chromatography (SiO_2 , EtOAc /pentane, 1:1 \rightarrow 1:0) to yield **7** (0.031 mmol, 11.5 mg, 56%) as a yellow-orange semi-solid. $^1\text{H-NMR}$ (400 MHz, chloroform-*d*): δ 8.12 (s, 1H), 7.62 (d, $J = 7.8$ Hz, 1H), 7.36 (d, $J = 7.8$ Hz, 1H), 7.20 (t, $J = 7.8$ Hz, 1H), 7.11 (t, $J = 7.8$ Hz), 7.05 (s, 1H), 6.90 (s, 2H), 6.32 (s, 1H), 4.61 (dd, $J = 12.1, 5.3$ Hz, 1H), 4.14 (s, 2H), 3.50 (dd, $J = 12.1, 5.3$ Hz, 1H), 3.11 (dd, $J = 12.1, 5.3$ Hz, 1H), 2.29 (s, 3H), 2.27 (s, 6H), 1.83 (s, 2H). $^{13}\text{C-NMR}$ (101 MHz, methanol-*d*₄): δ 178.1, 156.4, 138.1, 137.7, 137.1, 133.4, 129.9, 128.8, 124.8, 122.4, 119.8, 119.2, 114.0, 112.3, 111.1, 55.6, 35.7, 32.2, 21.0, 20.1. $[\alpha]_{\text{D}}^{19} = +0.073$ ($c = 0.15$ in MeOH). HRMS (ESI): calcd for $\text{C}_{23}\text{H}_{26}\text{N}_3\text{S}$ $[M+\text{H}]^+$: 376.1842, found 376.1837; calcd for $\text{C}_{23}\text{H}_{25}\text{N}_3\text{SNa}$ $[M+\text{Na}]^+$: 398.1661, found 398.1656.

(S)-2-Amino-3-(1*H*-indol-3-yl)-*N*-(2,4,6-trimethylphenethyl)propanamide (**2**)



Under nitrogen atmosphere with ice cooling, *N*-Boc-*L*-tryptophan (**5**) (1 eq., 610 mg, 2 mmol) was dissolved in dry THF (5 mL), 1,1'-carbonyl bis-1*H*-imidazole (1 eq., 324 mg, 2 mmol) was added. After stirring the solution for 2 h at room temperature, under ice cooling, mesitylethylamine (**8**) (1 eq.,

400 mg, 2 mmol) was added and left to stir overnight at room temperature. This solution was concentrated in vacuo, the residue obtained was dissolved in ethyl acetate (50 mL), washed with water, 0.1 M HCl, saturated NaHCO_3 aqueous solution and with saturated NaCl sequentially. The organic layer was dried over anhydrous MgSO_4 , filtered, and concentrated in vacuo. It was monitored by TLC (DCM/ MeOH 98:2, stained with Ninhydrin), showing complete conversion. The crude was used in the next step without any further purification.

Trifluoroacetic acid (16mmol, 21.5 mmol, 1.65 mL) was added to a solution of the protected amide compound from the previous step (1 eq., 1.334 mmol, 600 mg) in dry DCM (6mL) under nitrogen gas atmosphere with ice cooling and left to stir for 5 minutes, followed by 1 hour at room temperature. After concentrating the reaction mixture to dryness under reduced pressure, chloroform was added to the residue. The organic layer was washed with saturated aqueous NaHCO_3 solution (3 x 15 mL) and saturated aqueous NaCl (15 mL), dried over anhydrous MgSO_4 , filtered and concentrated in vacuo to obtain the final product **2**, (400 mg, quantitative). $^1\text{H NMR}$ (400 MHz, Methanol-*d*₄) δ 7.65 (dt, $J = 7.8, 1.0$ Hz, 1H), 7.37 (dt, $J = 8.2, 0.9$ Hz, 1H), 7.15 – 6.93 (m, 3H), 6.78 (s, 2H), 3.61 (t, $J = 6.7$ Hz, 1H), 3.27 – 2.93 (m, 4H), 2.60 (t, $J = 8.3$ Hz, 2H), 2.25 (s, 6H), 2.21 (s, 3H). $^{13}\text{C NMR}$ (101 MHz, Methanol-*d*₄) δ 178.6, 139.7, 138.9, 138.0, 134.9, 131.4, 130.4, 126.2, 124.0, 121.4, 121.1, 113.9, 112.8, 58.6, 40.9, 33.9, 31.7, 22.5, 21.5.

$[\alpha]_D^{19} = + 0.387$ ($c = 0.39$ in MeOH). m. p. = 56–60 °C HRMS (ESI) calcd for $C_{22}H_{28}N_3O$
 $[M+H]^+$: 350.2226, found 350.2232

6.5 Contributions from co-authors

R.M. Gierse, F. Magari, A. Heine, G. Klebe performed the protein crystallography studies. M. Mondal performed a part of the synthesis and modeling/docking, V. R. Jumde synthesized a part of the inhibitors and R. van Lier synthesized some building blocks during her Bachelor's research project.

6.6 References

- [1] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259 – 3263.
- [2] M. Mondal, N. Radeva, H. Fanlo-Virgýs, S. Otto, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2016**, *55*, 9422 –9426.
- [3] R.A. Massarico Serafim et al. *Eur. J. of Med. Chem.* **2014**, *82*, 418-425.
- [4] N. A. Meanwell, *J. Med. Chem.* **2011**, *54*, 2529–2591.
- [5] BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, LeadIT, version 2.1.3
- [6] P. R. Gerber, K. Muller, *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251 – 268.
- [7] V.V. Sureshbabu, S.A.Naik, G. Nagendra, *Synth Commun.* **2009**, *39*, 395-406.
- [8] N.D. Koduri, H. Scott, B. Hileman, J. D. Cox, M. Coffin, L. Glicksberg, S. R. Hussaini, *Org. Lett.* **2012**, *14*, 440-443.
- [9] A. Gangjee, J. Yang, M. A. Ihnat, S. Kamat, *Bioorg. Med. Chem.* **2003**, *11*, 5155-5170.
- [10] M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544 – 550.
- [11] <http://www.ccp4.ac.uk/>

Summary and Outlook

Despite recent developments in technology and techniques in today's drug discovery, accessing new bioactive compounds is still challenging. It is of utmost importance both for medicinal chemists as well as pharmaceutical companies to accelerate this lengthy trajectory. In this thesis, we aimed to establish novel strategies to facilitate faster and more efficient hit identification, the early stage of the drug-discovery process. In order to achieve this goal, we used target-guided synthesis (TGS) in which the biological target, in our case a protein, templates the formation of its own inhibitors from a pool of complementary building blocks. This efficient technique accelerates the drug-discovery process, namely hit-identification/optimization, enabling the selection of the best binders without prior synthesis, purification and biochemical evaluation of each individual compound.

In Chapter 1, we discuss the difficulties in today's drug discovery and how this trajectory could be improved and accelerated by using the TGS approach. Firstly, we give an overall introduction to the drug-discovery process and fragment-based drug design (FBDD). Secondly, we focus on target-guided synthesis in a broad perspective, subsequently kinetic-target guided synthesis (KTGS) in detail. KTGS is a type of TGS approach in which the assembly of the inhibitors takes place in an irreversible manner. This elegant approach has found numerous applications in drug discovery, especially in hit-identification. However, it is still premature in terms of the target scope and the number of compatible reactions. In this chapter, we present a detailed literature overview, the reaction and the target scope as well as some practical aspects in conducting KTGS experiments.

In Chapter 2, we demonstrate a strategic combination of two known techniques, protein-templated click chemistry (PTTC) and FBDD, to facilitate hit-identification. FBDD starts with the identification of the fragments, which are optimized in two ways: fragment growing and fragment linking. Fragment-growing is studied in more detail in the literature, whereas fragment-linking is rarely reported owing to the challenge in identifying a linker with optimal fit. Therefore, we aim to tackle this challenge as it is very attractive because of the potential for super-additivity. To overcome this hurdle, we used KTGS, namely PTCC, to link/optimize two fragments to afford inhibitors of endothiapepsin. We generated a library of azides and alkynes derived from fragments identified as inhibitors of endothiapepsin and selected the best combination by using the target itself in catalytic amount. We used the UPLC-TOF-SIM method to identify four of the 36 possible triazole combinations as the best binders. The hits identified have IC_{50} values of 43–121 μ M. This method represents a successful example of fragment-linking facilitated by PTCC, affording the first triazole inhibitor of endothiapepsin. As it is a proof-of-principle work and the target is a model enzyme, we did not perform further rounds of optimization. A next

development of this method should be the use of real drug targets with iterative rounds of optimization.

In Chapter 3, we present a novel protein-templated reaction, the in situ Ugi-four component reaction (Ugi-4CR), to facilitate hit identification by opening up access to a larger part of chemical space. By using a protein-templated multicomponent reaction for the first time, we screened a library of complex molecules derived from two multicomponent reactions (Ugi-4CR and Passerini) by using a catalytic amount of protein in a very short time. This new protein-templated reaction goes beyond PTCC given that it enables simultaneous screening of four subpockets and provides access to the hits identified in one synthetic step. To demonstrate this concept, we generated a library of building blocks comprised of aldehydes, amines, carboxylic acids and isocyanides and used endothiapsin in situ, enabling the selection of two hits with low micromolar activity (IC_{50} = 1.3 and 3.5 μ M) in 18 h. Thanks to the use of the Ugi-4CR, the inhibitors identified can be optimized in a straightforward manner. As it is the first example in the field, the mechanism should be studied in detail and the target scope and the number of multicomponent reactions in the KTGS toolbox should be expanded.

In Chapter 4, we demonstrate the applicability of KTGS in targeting protein-protein interactions (PPIs) as well as flexible binding pockets by using a new reaction in KTGS. By using a protein-templated reductive amination, we screened a library of potential inhibitors of the p53-Mdm2 PPI and identified one hit showing a K_i value of 0.76 μ M. Mdm2 has a flexible pocket, the Leu26 pocket, which is enlarged upon ligand binding, making it difficult to target by using structure-based drug design (SBDD) or other computational techniques such as virtual screening. Imine chemistry has found applications in the dynamic combinatorial chemistry (DCC) context and freezing the equilibrium was mostly achieved by in situ reduction with the risk of loss in affinity. In our work, we used reductive amination reaction in a KTGS setting by using a reducing agent from the beginning, making the reaction irreversible. The disadvantage of using this technique is the difficulty to compare activities of the amines after reduction since the selection takes place at the imine stage. Therefore, future focus in this reaction should be on the synthesis of the mimic of the imine scaffold formed in the protein-templated reaction.

In Chapter 5, we disclose our attempts to use an esterification reaction for the first time in KTGS. In this work, we used the model enzyme endothiapsin and generated a library of building blocks bearing alcohol and acid/ester functional groups. Our first attempts afforded the templated formation of an ester inhibitor from a carboxylic acid and alcohol building blocks. Starting from the optimized conditions for the library reaction did not afford the assembly of any products under various conditions. There are many drawbacks in using esterification reactions in

this context. Firstly, hydrolysis of the esters may occur in the reaction buffer, which should be monitored carefully for the time duration needed by the templated-reaction. Another important point is the use of a protease as a model enzyme. As this family of enzymes is responsible for hydrolysis, another target family may be a better option for further optimization studies. In addition, fragment ligation screening methods could be more suitable method to identify the best binders.

In Chapter 6, we describe the design, synthesis and biochemical evaluation of bioisosteres of the acylhydrazone inhibitor ($IC_{50}=12 \pm 0.4 \mu M$), which was developed in our group previously and used as a parent inhibitor for chapters 3 and 5. We synthesized three new compounds and evaluated their inhibitory activity against endothiapepsin, showing that the best bioisostere inhibits endothiapepsin with an IC_{50} value of $27 \pm 8 \mu M$. We resolved the cocrystal structure of one bioisostere in complex with endothiapepsin. Therefore, we could demonstrate replacements for the sensitive and hydrolyzable acylhydrazone linker without a significant loss in activity, an important point for future drug development.

To conclude, KTGS is a very successful technique, which facilitates cheaper, faster and more efficient hit identification. The primary goal of the thesis is the expansion of the toolbox of protein-templated reactions in terms of reaction and target scope. In this thesis, we introduced three novel reactions, namely *situ* Ugi-4CR, reductive amination and esterification. In addition, we presented an efficient combination of techniques, fragment-linking and PTCC, which benefits from the advantages of both drug-discovery techniques and finally the design, synthesis and biochemical evaluation of stable bioisosteres of acylhydrazones as inhibitors of endothiapepsin. Moreover, we used Mdm2 for the first time as a target in KTGS, and demonstrated the benefit for exploring flexible binding pockets. We strongly believe that the methods that we developed could find many applications in hit-identification as well as in later stages, namely hit/lead optimization until the pre-clinical stage.

Samenvatting en perspectief

Ondanks recente ontwikkelingen in technologie en technieken in het ontwikkelen van nieuwe medicijnen, is het ontdekken van nieuwe biologisch actieve verbindingen zeer uitdagend. Het is van het hoogste belang om dit lange traject te verkorten voor medicinale chemici als wel voor farmaceutische bedrijven. In dit proefschrift hebben we ons geconcentreerd op het gebruik van nieuwe strategieën om sneller en efficiënter hits te identificeren, in een vroeg stadium van het proces van medicijnontwikkeling. Om dit doel te bereiken, hebben we gebruik gemaakt van doelwit-gerichte synthese (DGS) waarin het biologische doelwit, in ons geval een eiwit, de vorming van zijn eigen remmers faciliteert vanuit een aantal complementaire bouwstenen. Deze efficiënte techniek versnelt het medicijn-ontdekkingsproces, namelijk hit-identificatie en -optimalisatie, wat de selectie van beste binders mogelijk maakt zonder voorafgaande synthese, zuivering en biochemische evaluatie van elke individuele verbinding.

In Hoofdstuk 1, bespreken we de moeilijkheden in de huidige medicijn-ontwikkeling en hoe dit traject verbeterd en versneld kan worden door gebruik te maken van de DGS benadering. Ten eerste, geven we een algemene introductie over het medicijn-ontdekkingsproces en fragment-gebaseerd medicijn ontwerp (FGMO). Ten tweede, concentreren we op DGS in een breed perspectief en vervolgens op kinetisch DGS (KDGS) in detail. KDGS is een type van DGS benadering waarin de vorming en samenstelling van de remmers plaatsvindt op een irreversibele manier. Deze elegante benadering is al meerdere keren toegepast in medicijn-ontdekking, met name bij hit-identificatie. Desalniettemin staat het nog in de kinderschoenen voor wat betreft het spectrum aan doelwitten en het aantal compatibele reacties. In dit hoofdstuk, presenteren we een gedetailleerd literatuuroverzicht, het spectrum aan reacties en doelwitten als wel een aantal praktische aspecten in het uitvoeren van KDGS-experimenten.

In Hoofdstuk 2, demonstreren we een strategische combinatie van twee bestaande technieken, eiwit-gerichte klikchemie (EGKC) en FGMO, om hit-identificatie te faciliteren. FGMO begint met de identificatie van de fragmenten welke op twee manieren geoptimaliseerd worden: fragment groeien en fragment linken. Fragment-groeien is in de literatuur in meer detail bestudeerd, terwijl fragment-linken zelden gerapporteerd is vanwege de moeilijkheid in het vinden van een verbinding die optimaal past. Daarom streven we er naar om deze uitdaging aan te pakken, aangezien het erg aantrekkelijk is vanwege de potentie voor super-additiviteit. Om deze horde te overwinnen, hebben we gebruik gemaakt van KDGS, namelijk EGKC, om twee fragmenten te linken/optimaliseren tot een remmer van endothiaepsine. We hebben een bibliotheek gegenereerd van aziden en alkynen, afgeleid van fragmenten geïdentificeerd als remmers van endothiaepsine, en hebben de beste combinatie geselecteerd door gebruik te maken van het doelwit zelf in katalytische hoeveelheid. We hebben gebruik gemaakt van de

UPLC-TOF-SIM methode om vier triazolen vanuit 36 mogelijke triazool-combinaties als de beste binders te identificeren. De geïdentificeerde hits hebben IC_{50} waarden van 43–121 μM . Deze methode vertegenwoordigt het eerste voorbeeld van fragment-linken gefaciliteerd door EGKC, wat de eerste triazool-remmer van endothiasepsin oplevert. Aangezien dit werk een principieel bewijs en het doelwit een model-enzym is, hebben we geen verdere rondes van optimalisatie uitgevoerd. De volgende stap van deze methode zou het gebruik echte doelwitten zijn met meerdere rondes van optimalisatie.

In Hoofdstuk 3, presenteren we een nieuwe eiwit-gerichte reactie, de in situ Ugi-vier component reactie (Ugi-4CR), om hit-identificatie te faciliteren door het toegankelijk maken van een groter onderdeel van de chemische diversiteit. Door voor de eerste keer gebruik te maken van een eiwit-gerichte multicomponent reactie, hebben we een bibliotheek van complexe moleculen weten te screenen, afgeleid van twee multicomponent reacties (Ugi-4CR en Passerini) door gebruik te maken van een katalytische hoeveelheid aan eiwit in een zeer korte tijd. Deze nieuwe eiwit-gerichte reactie gaat verder dan EGKC aangezien het simultaan screenen van vier subpockets mogelijk maakt en de geïdentificeerde hits in één synthetische stap bereikbaar zijn. Om dit concept te bewijzen, hebben we een bibliotheek gegenereerd van bouwstenen bestaande uit aldehyden, amines, carbonzuren en isocyaniden en hebben we endothiasepsine in situ gebruikt, wat de selectie mogelijk maakt van twee hits met een laag-micromolaire activiteit (IC_{50} = 1.3 en 3.5 μM) in 18 uur. Dankzij het gebruik van de Ugi-4CR kunnen de remmers op een eenvoudige manier geoptimaliseerd worden. Aangezien dit het eerste voorbeeld in de literatuur is, zou het mechanisme in detail bestudeerd moeten worden en het spectrum van doelwitten en het aantal van multicomponente reacties in de KDGS gereedschapskist uitgebreid moeten worden.

In Hoofdstuk 4, demonstreren we de toepasbaarheid van KDGS in eiwit-eiwit interacties (EEIs) als wel in flexibele bindingsgebieden door gebruik te maken van een nieuwe reactie in KDGS. Door het creëren van een eiwit-gerichte reductieve aminatie konden we een bibliotheek aan potentiële remmers van de p53-Mdm2 EEI screenen en hebben we een hit met een K_i waarde van 0.76 μM geïdentificeerd. Mdm2 heeft een flexibel bindingsgebied, het Leu26 gebied, dat vergroot wordt door het binden van een ligand. Dit maakt de identificatie van remmers met behulp van structuur-gebaseerd medicijn ontwerp (SGMO) of andere computationele technieken zoals virtuele screening moeilijk. Imine chemie heeft meerdere toepassingen gevonden in de dynamische combinatorische chemie (DCC) context en het bevriezen van het evenwicht wordt meestal gedaan door in situ reductie met het risico op verlies aan affiniteit. In ons werk hebben we gebruik gemaakt van een reductieve aminatie reactie in een KDGS setting, door vanaf het begin gebruik te maken van een reductieve reagens, welke de reactie irreversibel maakt. Het

nadeel van het toepassen van deze techniek is de moeilijkheid om de affiniteiten te vergelijken na het reduceren, aangezien de selectie plaatsvindt op het imine stadium. Daarom zou zich toekomstig onderzoek naar deze reactie moeten focussen op het synthetiseren van een nabootser van de imine groep die gevormd wordt in de eiwit-gerichte reactie.

In Hoofdstuk 5, beschrijven we onze pogingen om de gereedschapskist van eiwit-gerichte reacties nog verder uit te breiden door voor de eerste keer in KDGS gebruik te maken van een verestering. In dit werk, hebben we het model enzym endothiapsine gebruikt en een bibliotheek gemaakt van bouwstenen met een alcohol en zuur-/ ester functionele groep. Onze eerste pogingen resulteerden in de gerichte vorming van een ester-remmer vanuit een carbonzuur en een alcohol bouwsteen. Uitgaande van de geoptimaliseerde condities voor de bibliotheekreactie resulteerde niet in de formatie van enig product onder verschillende condities. Er zijn veel nadelen in het gebruik van veresteringsreacties in deze context. Ten eerste is de reversibiliteit van de reactie, welke de vorming van producten in aantoonbare hoeveelheden voorkomt. Ten tweede zou hydrolyse van de esters in de reactiebuffer plaats kunnen vinden, welke nauwlettend in de gaten gehouden zou moeten worden voor de tijdsduur die voor de gerichte-reactie nodig is. Een ander belangrijke punt is het gebruik van een protease als modelenzym. Aangezien deze familie van enzymen verantwoordelijk is voor hydrolyse, zou het gebruik van een andere doelwitfamilie een betere keuze zijn voor toekomstige optimalisatiestudies.

In Hoofdstuk 6, beschrijven we het ontwerp, de synthese en de biochemische evaluatie van bioisosteren van de acylhydrazon-remmer ($IC_{50} = 12 \pm 0.4 \mu M$), welke reeds ontwikkeld was in onze groep en gebruikt werd als leidraadremmer voor hoofdstukken 3 en 5. We hebben drie nieuwe verbindingen gesynthetiseerd en geëvalueerd op hun remmende activiteit van endothiapsine, waaruit de beste bioisosteer endothiapsine remt met een IC_{50} waarde van $27 \pm 8 \mu M$. We hebben de kristalstructuur van één van de bioisosteren in complex met endothiapsine opgehelderd. Daarom hebben we gedemonstreerd dat het mogelijk is om de gevoelige en hydrolyseerbare acylhydrazon-linker zonder significant verlies in activiteit te vervangen, een belangrijk aspect voor toekomstige medicijn-ontwikkeling.

Samenvattend, is KDGS een zeer succesvolle techniek, die een goedkopere, snellere en efficiëntere hit-identificatie mogelijk maakt. Het hoofddoel van dit proefschrift is de uitbreiding van de gereedschapskist van eiwit-gerichte reacties in termen van reactie en spectrum aan doelwitten. In dit proefschrift, introduceren we drie nieuwe reacties, namelijk in situ Ugi-4CR, reductieve aminering en verestering. Als toevoeging presenteren we een efficiënte combinatie van de technieken fragment-linken en EGKC, die baat heeft bij de voordelen van beide medicijn-

ontdekkingstechnieken en uiteindelijk het ontwerp, de synthese en de biochemische evaluatie van stabiele bioisosteren van acylhydrazonen als remmers van endothiasepsine. Bovendien hebben we Mdmd2 voor de eerste keer als doelwit gebruikt in KDGS, en daarmee de toegevoegde waarde van de techniek aangetoond om flexibele bindingsgebieden te onderzoeken. We zijn er van overtuigd dat de methoden die we ontwikkeld hebben meerdere toepassingen kunnen vinden in hit-identificatie als wel in latere stadia, namelijk hit/ lead optimalisatie tot aan de preklinische fase.

Acknowledgements

I had a beautiful four years in this beautiful city and university. Now it is my turn to say goodbye and express my sincere gratitude to the people who made these four years of adventure unforgettable.

Anna, thank you very much for giving me the chance to work in your group, supporting and encouraging me in each phase of the PhD as well as for being there for all kinds of problems that I faced. You were not only an excellent supervisor but also a close friend to all of us. I am thankful for everything. Ben, thank you very much for being my co-supervisor and giving feedback on my thesis despite your busy schedule.

Adri, thank you very much for the help, discussions and your friendly approach to all kinds of problems I had during my whole PhD. Also, thank you for being in my reading committee and for your feedback.

Martin, I enjoyed our discussions about the projects a lot. They were very helpful and motivated me to read more and think. Thanks a lot!

I would like to thank the members of the assessment committee, Prof. Jörg Rademann, Prof. Adriaan Minnaard and Prof. Gerard Roelfes for accepting to be part of the committee and for the suggestions/comments.

Hilda and Tineke, thank you very much for the support especially in the last period of my PhD. Tineke, without your support I would not have been able to make this graduation on time. I would also like to thank Theodara for the help with the UPLC measurements, Marzia for the friendship and help in the lab especially for the diastereomeric separation nightmare.

I would like to thank my paranymphs Jonas and Ramon. Jonas, my dear friend, thanks for the sincere friendship in these four years. We had so many nice experiences and I will never forget them. I am wishing you the best of success in your upcoming defense. Ramon, thanks a lot for the nice friendship and for your help during the graduation period as well as the PPIs project. P.S: You are driving the car so slowly, man 😊

I would like to thank my students Fatima and Cagdas for being great and hardworking students. Also, Tommaso, I am so happy we met and thanks a lot for your great contribution to the PPIs project in a very short time period.

Thanks a lot to the people that I have collaborated with: Prof. Alex Dömling (for the discussions for my MCR and PPIs project), Dinos (for being available all the time for my questions related with MCRs), Prof. Tad Holak and Aleksandra (for the help in the FP assays and questions about MDM2). Prof. Gerard Klebe, and Leti & Varsha, thanks a lot for the support with the experiments (MCR and ester project) and nice friendships.

Turhan, The Basgan, thanks a lot for the great and sincere friendship from the first day I arrived in Groningen. You were always there for me even in the hardest times that I had here and helped me a lot in the writing period as well as in other things. Samimi dostluğun için çok teşekkür ederim baskan.

Alwin, we joined the group at the same time and later became colleagues. You have always been a very good friend of mine, supportive and listening. Thanks for everything and helping me with the Dutch summary as well. Nabil, kardesh, thanks a lot for the warm friendship and nice times together in Groningen. Tiziana and Milon, thanks for the nice welcome and friendship when I joined the group and helping me to settle down in this city. Kaja, thanks a lot for your nice friendship and the perfect cover design. You are so talented!

Yun, you are a great friend, I really miss our chats and discussions. You have such a big heart and I am very grateful that I met you. I wish you all the best in your future career. Jeffrey, my lab mate and inspiring chemist, it was very nice to work with you in the same lab. I miss our music discussions about Tarkan ☺

I would like to thank past and present members of the Hirsch group, Elenora, Judy, Robin, Spyros, Sebastian, Barbara, Michele, Vlad, Roos, Rita, Matthijs, Manuel and Renata, also the new people in Saarbrücken, Walid and Federica for the nice group environment.

Thanks to the whole chemical biology group, Marthe, Manhattan, Miriam, Vijay, Niek, Paul, Ashmir, Andreas, Ruben, Chris, Nittert, Bas, Liubov, Simona, Mira, Guillaume, Steven, Ji, Vivek and Gongbao for the nice atmosphere in the lab.

I would like to thank the whole Stratingh institute members, and in particular Ruben (my teaching buddy), Manuela, Dowine, Giulia, Michael, Mickel, Anna, Mark, Dusan, Liliana, Frederike and Wiktor.

I would like to thank the Turkish community and Groningen friends that I met here, Kanat Hocam (bu süreçteki desteğiniz ve abiliğiniz için çok teşekkürler hocam), Berfu, Volkan, Ercan, Bookie, Sanne (samimi dostluğunuz için teşekkürler başkanlar), Marco, Seyda, Ismail, Gokhan, Funda, Agah (kardeş gibi olduk artık), Massimo (you and Anna were always very warm and nice friends for me) and the people that I worked with in GUTSA board and finally Groningen Indoor Soccer team.

Nour, hayete ente, I do not know how I would have managed this period without you. Your support and presence in every phase of this thesis made this possible and I know that you will be even happier than me after I get my diploma. Thank you for millions of things I could not count and express here...

Finally, I am grateful to my dearest family and dedicating this thesis to them.

Canım ailem, sizlerin sayesinde bu diplomayı alıyorum. Birgun bile arkamdan desteęinizi esirgemediniz ve hep yanımda oldunuz. Sizin gibi bir ailem olduğu için çok şanslıyım...

Thanks to everyone that I could not mention here, my teachers and supervisors who played an important role in my education/career.

With best wishes,

Yagiz