Synthesis, Identification, Kinetic, and Structural Characterization of Inhibitors of the Aspartic Proteases HTLV-1 Protease and Endothiapepsin

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie der

Philipps-Universität Marburg

vorgelegt von

Maren Sophia Kuhnert

aus München

Marburg/Lahn 2015

Erstgutachter: Prof. Dr. Wibke E. Diederich

Zweitgutachter: Prof. Dr. Klaus Reuter

Eingereicht am 15.05.2015

Tag der mündlichen Prüfung: 26.06.2015

Hochschulkennziffer: 1180

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregungen von Frau Prof. Dr. Wibke E. Diederich am Institut für Pharmazeutische Chemie des Fachbereiches Pharmazie der Philipps-Universität Marburg in der Zeit von Februar 2011 bis Mai 2015 durchgeführt.

Publications

Journal Articles

- **Kuhnert, M.**, Blum, A., Steuber, H., Diederich, W. E., Privileged Structures Meet Human T-Cell Leukemia Virus-1 (HTLV-1): C_2-Symmetric 3,4-Disubstituted Pyrrolidines as Non-Peptidic HTLV-1 Protease Inhibitors. *J. Med. Chem.* 2015, in press.

- Schiebel, J., Radeva, N., Köster, H., Metz, A., Krotzky, T., **Kuhnert, M.**, Diederich, W. E., Heine, A., Neumann, L., Atmanene, C., Renaud, J.-P., Meinecke, R., Schlinck, N., Popp, F., Zeeb, M., Klebe, G., One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. To be submitted.

- **Kuhnert, M.**, Köster, H., Bartholomäus, R., Park, A.Y., Shahim, A., Heine, A., Steuber, H., Klebe, G., Diederich, W. E., Tracing Binding Modes in Hit-to-Lead Optimization: Chameleon-Like Poses of Aspartic Protease Inhibitors. *Angew. Chem., Int. Ed.* 2015, 54(9), 2849-2853; *Angew. Chem.* 2015, 127(9), 2891-2896.

- **Kuhnert, M.**, Steuber, H., Diederich, W. E., Structural Basis for HTLV-1 Protease Inhibition by the HIV-1 Protease Inhibitor Indinavir. *J. Med. Chem.* 2014, 57(14), 6266-72.

- **Kuhnert, M.**, Diederich, W. E., HIV Protease Inhibitors. *PHARMAKON* 2014, 2(4), 262-269.

Conference Contributions

- **Kuhnert, M.**, Köster, H., Bartholomäus, R., Park, A.Y., Shahim, A., Heine, A., Steuber, H., Klebe, G., Diederich, W. E., Tracing Binding Modes in Hit-to-Lead Optimization: Chameleon-Like Poses of *Gewald* Reaction-Based Aspartic Protease Inhibitors. *Frontiers in Medicinal Chemistry*, Marburg, Germany, 2015, Poster presentation.

- **Kuhnert, M.**, Köster H., Bartholomäus, R., Park, A.Y., Shahim, A., Heine, A., Steuber, H., Klebe,G., Diederich, W. E., *Gewald* Reaction-Derived Aspartic Protease Inhibitors: Challenging Current Paradigms in Medicinal Chemistry. *Frontiers in Medicinal Chemistry*, Tübingen, Germany, 2014, Poster presentation.

- **Kuhnert, M.**, Steuber, H., Diederich, W. E., Attacking Novel Targets with Known Drugs: HTLV-1 Protease in Complex with Indinavir. *Frontiers in Medicinal Chemistry*, München, Germany, 2013, Poster presentation.

Abbreviations

%-inhib.	percent inhibition value
°	degree
1-HOBT	1-hydroxybenzotriazole
Abz	2-aminobenzoyl
ACE	angiotensin-converting enzyme
AIDS	acquired immune deficiency syndrome
ATLL	adult T-cell leukemia
BICINE	*N*,*N*-bis(2-hydroxyethyl)glycine
Boc	*tert*-butyloxycarbonyl
Boc_2O	di-*tert*-butyl dicarbonate
C	celsius
CA/NC	capsid/nucleocapsid cleavage site
cal	calorie
CHES	2-(cyclohexylamino)ethanesulfonic acid
ClogP	calculated logarithm of the octanol-water partition coefficient
CNS	central nervous system
conc.	concentration
Da	dalton
DABCYL	4-(4-dimethylaminophenylazo)benzoyl)
DCM	dichloromethane
DIPEA	ethyldiisopropylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DpmINV	*des*-3-pyridylmethyl-indinavir
DTT	1,4-dithiothreitol
E. coli	*Escherichia coli*
EDANS	5-[(2-aminoethyl)amino]naphthalene-1-sulfonic acid
EDCI	*N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide
EDTA	ethylenediaminetetraacetate
EI	electron ionization
EP	endothiapepsin
ESI	electrospray ionization

EtOAc	ethyl acetate
EtOH	ethanol
FBDD	fragment-based drug discovery
FDA	food and drug administration
FPLC	fast protein liquid chromatography
FRET	fluorescence resonance energy transfer; also named “Förster resonance energy transfer”
h	hour
HA	number of non-hydrogen atoms, (i.e. heavy atoms)
HAID	HTLV-1-associated infective dermatitis
HAM	HTLV-1-associated myelopathy
H-bond	hydrogen bond
HCV	hepatitis C virus
HEPES	2-[4-(2-hydroxyethyl)-piperazin-1-yl]ethanesulfonic acid
HIV-1	human immunodeficiency virus-1
HPLC	high performance liquid chromatography
HTLV-1	human T-cell leukemia virus-1
HTS	high-throughput screening
Hz	hertz
INV	indinavir
IPTG	isopropyl-β-D-thiogalactopyranoside
l	liter
LB medium	lysogeny broth medium
LE	ligand efficiency
M	mol/l
m/z	mass-to-charge ratio
MeOH	methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
min	minute
mp	melting point
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PEI	percentage efficiency index

pH	negative decimal logarithm of the hydrogen ion activity
PIPES	1,4-piperazinediethanesulfonic acid
PIs	protease inhibitors
PR	protease
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
qHNMR	quantitative ^{1}H NMR
QSAR	quantitative structure-activity relationship
R	ideal gas constant
RFU	relative fluorescence unit
rmsd	root-mean-square deviation
rpm	rounds per minute
RT	room temperature
SAR	structure-activity relationship
SBDD	structure-based drug design
SDS	sodium dodecyl sulfate
SDS page	sodium dodecyl sulfate polyacrylamide gel electrophoresis
T	temperature [K]
TBAI	tetrabutylammonium iodide
TEA	triethylamine
TFA	trifluoroacetic acid
TFE	trifluoroethanol
THF	tetrahydrofuran
TLC	thin layer chromatography
T_m	melting temperature
TMS	tetramethylsilane
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
TSA	thermal shift assay
ΔG	Gibbs free energy
σ	standard deviation

Table of Contents

1. Introduction

1.1 Drug Design and Structure-Based Drug Design

In the interdisciplinary field of drug design that has continuously emerged over the last three decades, the rational development of ligands for a biological target with the overall long-term aim to develop a novel drug for the respective target is pursued. Two main concepts are followed in this respect: the ligand-based and the structure-based drug design.

According to the statement of the Nobel Prize winner James Black that "The most fruitful basis for the discovery of a new drug is to start with an old drug",[1,2] the ligand-based drug design utilizes previously identified molecules that bind to the target, and, subsequently, quantitative structure-activity relationships (QSAR) and pharmacophore models are derived and used for further lead optimization.[3]

Structure-based drug design (SBDD) relies on the knowledge of the three-dimensional structure of the biological target, which provides the basis for optimization of a lead structure. The first examples which reflect the central paradigm of SBDD were described in the 1970s and since then SBDD has become an important concept in medicinal chemistry.[4] Knowledge about the structure of the target is mainly obtained by X-ray crystallography or NMR spectroscopy, in addition, homology models might also be applied to derive initial hypotheses if the experimental approach to determine the target structure does not succeed within the anticipated time frame. In addition to the information about the structure of the protein, the identification of ligands as starting point for the SBDD process is essential. Ligands can be identified by computer-aided as well as experimental methods:[5] computer-aided methods include the virtual screening, where large databases of compounds are screened *in silico* against the structural model of the target and various *de novo* design methods. Among the experimental methods, which comprise lead finding strategies such as high-throughput screening (HTS), combinatorial chemistry approaches and fragment-based discovery of ligands, the concept of privileged structures provides a further alternative option. This concept was first described in 1988 by Evans *et al.*[6] Privileged structures in a broader definition are ligands or scaffolds that show a biological effect on more than one target protein, often on related proteins of the same gene family.[7] Once such a privileged scaffold has been identified, the selectivity profile towards the desired target needs to be optimized e.g. by exploiting sterical and/or electrostatic differences in the binding pockets of the related enzymes. In an ideal case, close structural analogues of the identified scaffold are already available allowing not only the deduction of a

preliminary SAR but also of a preliminary selectivity profile. In most cases, the knowledge of previously investigated compound characteristics such as data regarding the bioavailability and other pharmacokinetic properties are beneficial for the further development. Examples of privileged scaffolds in approved drugs comprise for instance the indole moiety (e.g. sumatriptane, indometacin, etodolac, ondansetron), the benzimidazole ring system (e.g. thiabendazole, rabeprazole, clemizole), or the benzothiophene scaffold (e.g. raloxifene, zileuton).[8,9]

To embark on an SDBB approach, after identification of a first ligand hit, the subsequent structure determination of the target protein in complex with the detected hit or lead molecule provides valuable information about key interactions of the small molecule to the protein. Based on this information, the further hit-to-lead- and lead optimization processes, which can be described as an iterative process of design, synthesis, affinity-, and structure determination of promising hits (Figure 1.1), is pursued. In the best case, after several cycles of optimization, a set of lead compounds with sufficient potency and various other suitable properties required for drug-likeliness, such as e.g. a low molecular weight, good solubility, cell permeability, metabolic stability, and compliance to the Lipinski rules (oral bioavailability) are obtained.

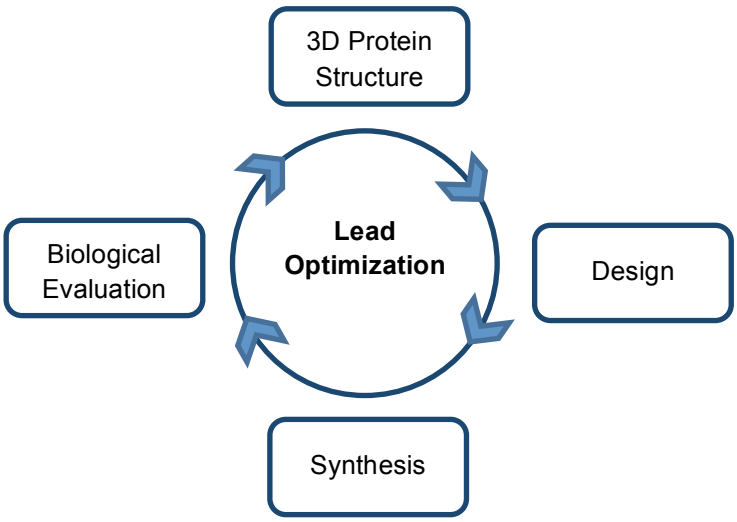


Figure 1.1. Drug design cycle in structure-based drug design.

The identification of these lead compounds for the respective target is the first step on the long way towards an approved drug. Further years of research and development are generally necessary to meet the requirements of toxicology and the ADME properties (Absorption, Distribution, Metabolism, Excretion), to start and pass the pre-clinical testings as well as clinical trials which finally might end up in an approved drug. The time span from the drug development to the approved drug nowadays takes about

15 years.[10] Between 2005 and 2013, on average 25 new molecular entities and therapeutic biologics were approved per year.[11]

As straightforward as the described lead-optimization process seems to be in theory, it is far-off from being an established standard procedure as it bears a lot of challenges: the first step of the directed optimization of a lead structure is the design of novel inhibitors, which is mostly performed by *in silico* methods while making use of the determined 3D structure of the target to identify the most promising candidates for prioritizing medicinal chemistry activities. In docking studies ligands are fitted into the active site of the target protein and the predicted binding pose is scored and ranked based on the interactions of the ligand to the potential key interaction partners on the protein side.

There are various limitations associated with this procedure: at the beginning, the correct binding pose of the inhibitor needs to be predicted. This might fail and also binding mode switches within the same inhibitor series might occur. A striking example for such a rather unexpected behavior is given in chapter 6 of this thesis. The correct prediction of a ligand`s binding pose can be aggravated as well by the presence or absence of the ubiquitously available water within and in the neighborhood of the binding pocket, as by any induced-fit adaptations occurring on the protein side upon ligand binding. Once a suitable placement of ligands within the expected binding site has been achieved, the subsequent affinity estimation (“scoring”) is a further challenge. In general, most scoring functions attempt to predict the binding affinity by evaluation of assumed polar and hydrophobic protein-ligand contacts. However, the strength of those interactions depends on various aspects. Knowledge on the protonation states of the ligand’s and the protein`s functional groups is an essential prerequisite to assess the H-bond network, but, however, relies in most cases on assumptions, as the experimental validation is rather resource-intensive. In addition, thermodynamic features such as enthalpy- or entropy-driven binding are mostly not considered by scoring functions. Desolvation of ligand and binding site are ultimate prerequisites to the binding process, however, difficult to simulate for a meaningful affinity ranking. Moreover, features such as buriedness of the interaction site, which influences the strength of H-bonds, or compensating effects such as enthalpy-entropy compensation are challenging to assess in absence of experimental methods.[12–14]

Hence, the “holy grail of structure-based drug design” to identify potent and specific ligands only by means of computational methods and structural information of the protein target of interest by reasonably predicting binding affinities has not yet been achieved.[15]

As a further prerequisite, also the synthesis of the *in silico*-identified ligands should be feasible, while still possessing sufficient options for structural variation.
The biological evaluation and especially the impact for the clinical success is uncertain due to the complexity of a whole number of various interdependent parameters. As recently collected evidence demonstrates, the determination of the affinity of a ligand to the target does not represent the solely relevant parameter describing the interaction between target and ligand. The transferability towards the biological system is often insecure, as aspects like the residence time of the ligand to the target protein seem to play an important role in the clinical performance.[16] Hence, the prediction of *in vivo* effects of a drug candidate is quite difficult.[13] Also the target selectivity is poorly predictable, but represents an important aspect for toxicology studies and side effects.
These aspects explain why drug discovery, despite a continuously increasing knowledge base, is a highly challenging as well as cost- and time-demanding endeavor.

1.2 Aspartic Proteases

1.2.1 Aspartic Proteases – an Overview

In general, proteases are enzymes, which cleave proteins by hydrolysis of peptide bonds. Depending on the mechanism of catalysis, they can be divided into seven major classes:[17] the aspartic-, cysteine, glutamic-, metallo-, asparagine-, serine-, and threonine-proteases. In addition, some peptidases of unknown catalytic type and one class of mixed peptidases are notated in the Merops database.[17]
The focus of this work lays on aspartic proteases, the history of which is nicely reflected in a publication from Szesci *et al*.:[18]

> *"The aspartic proteases (EC 3.4.24) were the first type of enzymes known to mankind, the first protease type described, the second protein to be crystallized and subjected to numerous early investigations by the pioneers of modern enzymology and protein chemistry."*[18]

Aspartic proteases possess two catalytic aspartates in the active site and usually require an acidic pH-value for efficient catalysis. They can be divided into endogenous and exogenous proteinases: endogenous aspartic proteases occur in the human organism, like the gastric proteinases pepsin and gastricsin, renin, which is involved in the regulation of blood pressure, or cathepsin D. Exogenous proteinases are found in bacteria, fungi, yeasts and viruses.[19] Therefore, aspartic PR are involved in physiological as well as pathophysiological processes.

Eukaryotic aspartic proteases are produced as catalytically inactive precursor proteins, so-called zymogenes, and are converted to their proteolytically active counterparts depending on their physiological role. This mechanism allows to regulate the location of action of the proteases and therefore *inter alia* protects the producing cells and the surrounding environment against damages. To auto-inhibit their catalytic activity, zymogenes contain an N-terminal extension (prosegment), while the mode of inhibition differs among the aspartic protease zymogenes: in gastric zymogenes, the substrate binding pocket is blocked by the prosegment, which is therefore inaccessible to substrates. In the case of proplasmepsin II, the zymogene of the malaria target plasmepsin II, the prosegment severely distorts the active site and therefore prevents the formation of a functional active site. The activation of zymogenes might occur by auto-activation at acidic pH, like for example for the gastric zymogene pepsinogene, or through assisted catalysis of other proteases, like for example in the case of prorenin.[20] According to the Merops database, proteases are further subdivided into clans and families. A clan contains proteases which show the same evolutionary origin. Proteins of one clan are divided into different protein families; proteins of the same family typically exhibit similarity in their amino acid sequence and their tertiary protein structure. Currently, the aspartic family is classified into six different clans, with clan AA being the biggest one, which itself contains eight different protein families. Among them the aspartic protease families A1 (pepsin family) and A2 (retropepsin family) bearing 183 identifiers (9381 sequences) and 51 identifiers (1150 sequences), respectively, are of utmost importance.[17] [status: 23.04.2015]

1.2.2 Catalytic Mechanism of Peptide Cleavage

The proposed mechanism of the cleavage of the peptide bond in the active site is shown in Figure 1.2.[21] Herein, the lytic water molecule, which is located close to the two catalytic aspartates, plays a key role. Before the onset of the cleavage reaction, one of the two aspartates is deprotonated and acts as H-bond acceptor to the water molecule, while the other aspartate remains protonated, thus destabilizing the C=O bond of the peptide bond by forming an H-bond to the carbonyl-oxygen. The now polarized nucleophilic water molecule then attacks the electrophilic carbonyl function of the peptide bond resulting in a geminal diol as tetrahedral intermediate, which in the following decomposes under reconstruction of the more stable carbon-oxygen-double bond rendering the respective carboxylic acid and amine.

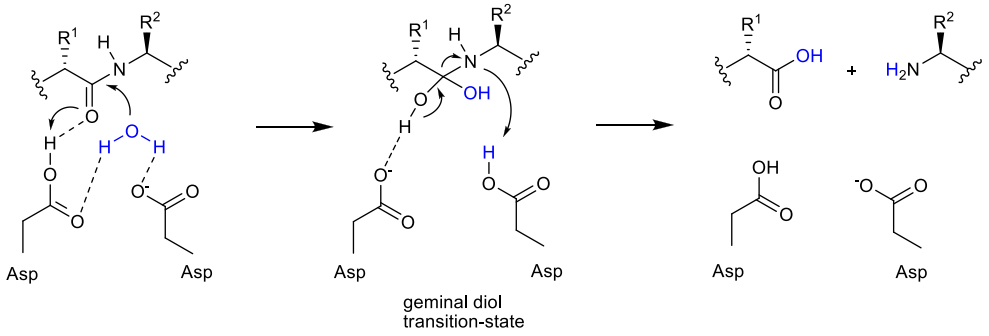


Figure 1.2. Schematic representation of the catalytic mechanism of aspartic proteases. The lytic water molecule is shown in blue.

1.2.3 Aspartic Proteases in Drug Design

Although in comparison to other peptidase families the group of aspartic proteases is a small group of enzymes, it is involved in several physiological and pathophysiological processes. Hence, its members represent attractive targets for inhibitor development. Famous examples of aspartic proteases show their significance in human diseases: they are *inter alia* involved in the regulation of the blood pressure (renin), in cancer-related processes (cathepsin D), in neurological diseases like Alzheimer's (β-secretase), in mycosis processes (SAP proteins), in malaria (plasmepsins) or in infectious diseases like HIV infections (AIDS).[22]

One common strategy for protease inhibition is to mimic the structure of the natural substrate and to prohibit its cleavage by blocking the active site of the enzyme with so-called transition-state analogs. In 1949, Pauling already stated that the interaction of an enzyme with its substrate is the strongest at the transition state.[23,24] In the mid-1970s this concept was applied to the aspartic proteases by the analysis of the activity of pepstatin.[25] It turned out that pepstatin is a potent inhibitor for a number of enzymes of the aspartic protease family. Pepstatin (Figure 1.3) represents a hexa-peptide and was originally isolated from actinomyces. The statine moiety as non-proteinogenic amino acid with the leucine side chain in P_1 position and the hydroxyethylene moiety in direction to the C-terminus mimics the scissile peptide bond and enables pepstatin to function as transition-state analogs. The two catalytic aspartates are addressed by its central hydroxyl group.

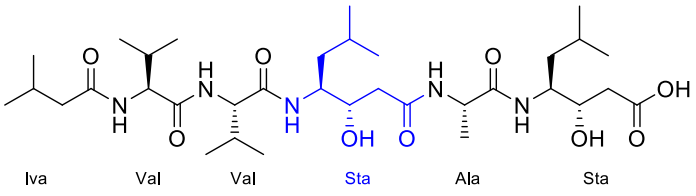


Figure 1.3. Chemical structure of pepstatin.

Besides the statin motif, other isosters of the transition-state, which are shown in Figure 1.4, have successfully been exploited.[26,27] For the first attempt to develop inhibitors for a respective target protein, the modification of natural substrates often leads to potent inhibitors. Nevertheless, as straightforward as it may sound, highly potent inhibitors of a target protein do not necessarily fulfil the requirements of an approved drug, but nevertheless, provide the indispensable basis for further studies.

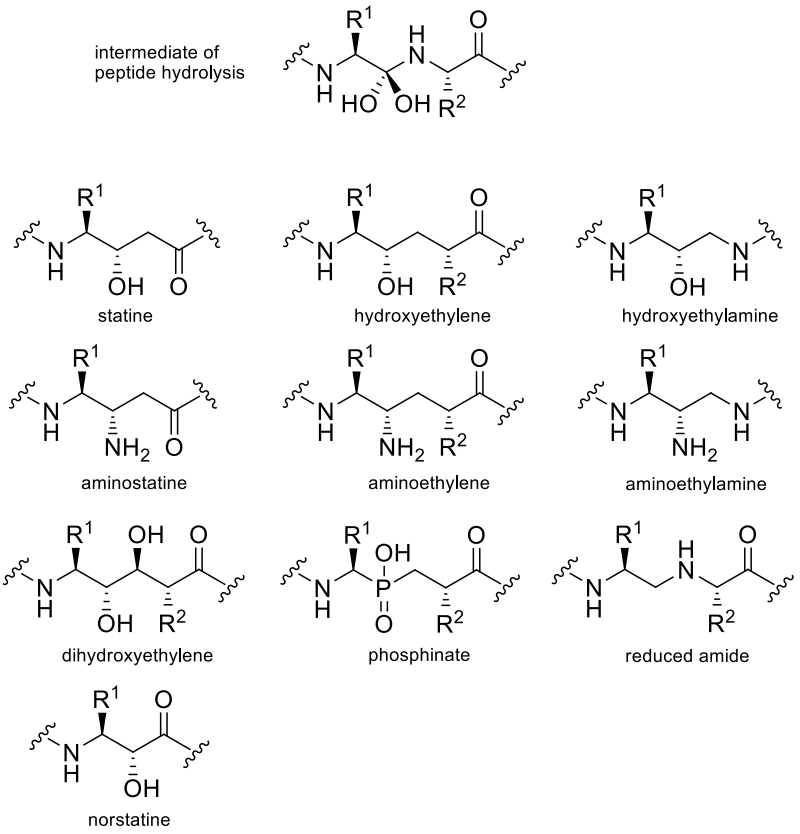


Figure 1.4. Examples of transition-state isosters of aspartic protease inhibitors.

The main challenge is therefore the development of so-called drug-like molecules. To estimate if the developed ligand might be orally bioavailable, the Lipinski “Rule of five”[28] might give a first indication. In many cases peptides show a high potency to the target of interest, but their most striking disadvantages are their poor bioavailability, metabolic instability, and poor drug-likeness. Therefore, research activities mostly aspire the development of inhibitors with a low peptidic character, namely peptidomimetic or non-peptidic molecules.

Currently approved aspartic protease inhibitors comprise renin- as well as HIV-1 protease inhibitors. The first approved HIV-1 PR inhibitor saquinavir (Figure 1.5) represents a transition-state isoster: the central scaffold is the hydroxyethylamine moiety which replaces the substrate cleaving site Phe-Pro of the protease/reverse

transcriptase autocleavage site.[29] In addition, the peptidominetic inhibitors indinavir, nelfinavir, (fos)amprenavir, and darunavir are also based on the hydroxyethylamine motif. Aliskiren, the first renin inhibitor, was approved in 2006 and is a non-peptidic inhibitor. Its development was inspired by the peptide analog of angiotensinogen, the natural substrate of renin, and eventually led via peptitomimetic compounds to the non-peptide-like compounds, with aliskiren as the approved representative.[30]

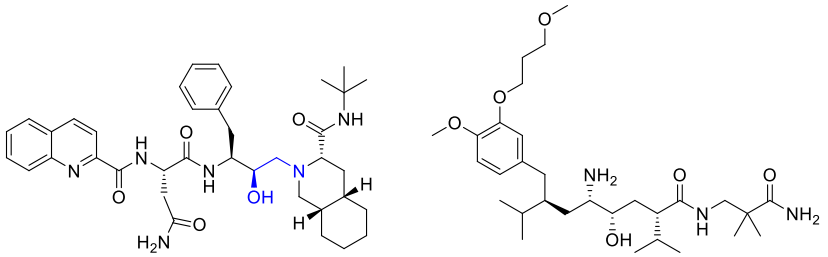


Figure 1.5. Chemical structures of saquinavir (HIV-1 PR inhibitor) and aliskiren (renin inhibitor). The hydroxyethylamine motif of saquinavir is colored blue.

Besides the above mentioned significance in drug design, nearly everybody is somewhat linked to the history of an aspartic protease in one's everyday life: the trademark "Pepsi Cola®" originally derived from the aspartic protease pepsin. Caleb Bradham, a pharmacist, renamed his drink, which was supposed to be more than a refreshment but should also support the digestion (like pepsin), in 1898 to the nowadays well-known "Pepsi-Cola".[31]

1.2.4 Structural Features of Aspartic Proteases

In 1972, the first amino acid sequence of an aspartic protease was determined by Tang *et al*. from porcine pepsin.[32] The first three-dimensional structures of aspartic proteases with a nowadays moderate crystallographic resolution around 3 Å were reported in 1977 and comprise the three fungal enzymes rhizopuspepsin, endothiapepsin and penicillopepsin, which all belong to the pepsin family A1.[23,33,34]

The structure of the pepsin-like proteases consists of two lobes, the N- and C-terminal one, with the substrate binding pocket in between the two lobes, where the two catalytic aspartates are located in a centric manner (Figure 1.6a). At the top of the active site a hairpin-like beta structure, the so called "*flap*", is present being a characteristic of most aspartic proteases. The first retroviral protease 3D structures (family A2) were determined from the rous sarcoma virus (RSV)[35,36] and from the HIV PR[37] in 1989.

The retroviral proteases exhibit similar characteristics (Figure 1.6b), while they consist of two identical monomers which form a homodimer as functional proteolytic unit. The

active site is located between both monomers and above the dimer interface and covered by the *flap*. Each monomer contributes one of the catalytic aspartates and one *flap*.

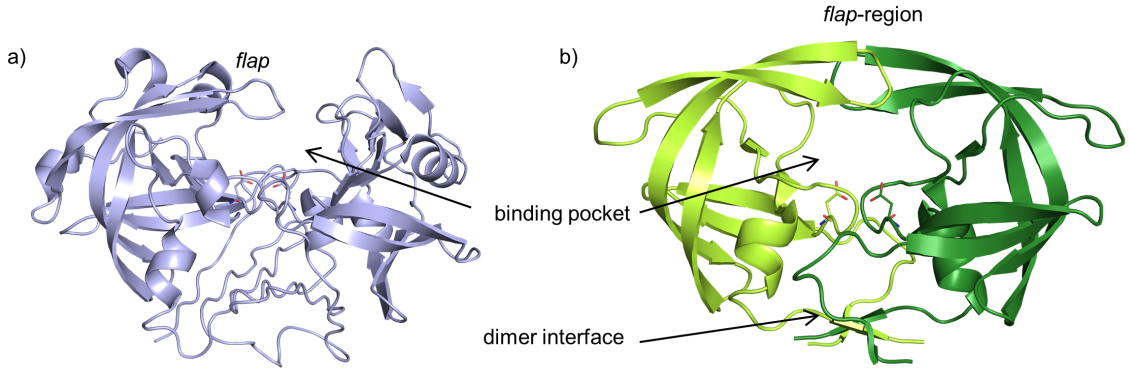


Figure 1.6. a) X-ray structure of endothiapepsin as representative of the pepsin family (A1), PDB entry: 1OEW. The protein consists of one amino acid chain, shown in blue. b) X-ray structure of HIV-1 PR, PDB entry: 1SDT, as member of the retroviral family (A2). The HIV-1 PR functions as a homodimer; each monomer is colored differently (light and dark green). Proteins are shown as cartoon, the catalytic aspartates in stick representation.

The similarity between pepsin-like (A1) and retroviral (A2) aspartic proteases cannot be dismissed: the functional proteolytic units are similar in size and in the arrangement of the active site, while the amino acid chain of proteases of the pepsin-like family are in general larger than those of the retroviral proteases.[38]

The main difference between the structures of the pepsin-like (A1) and retroviral (A2) aspartic proteases is reflected in the symmetrical arrangement of the amino acid chain or rather of the two monomers in case of the A2 family, which, however, does not cast doubt on their overall similarity. The pepsin-like proteases consist of one amino acid chain, but their C- and the N-terminal lobes show high homology in their secondary and tertiary structure.[23,39] On the contrary, the retroviral proteases bear not only two similar lobes, they instead feature two identical monomers which function as a homodimer.

The catalytic aspartates in the active site are in all aspartic proteases embedded in a conserved Asp-Thr/Ser-Gly sequence. This sequence is located within a loop and stabilized by a rigid hydrogen network, often called "*fireman's grip*", stabilizing the active site (Figure 1.7). Therefore the *fireman's grip* is important for the conformation of the active site and in case of the retroviral proteases additionally for the dimerization and dimer stability.[36,40] Between the two catalytic aspartates, a water molecule can be located, forming H-bonds to the carboxylate groups of the aspartates, acting as the lytic water molecule which is involved in the cleavage of the peptide bond, however this water molecule can be displaced by binding of certain inhibitors.

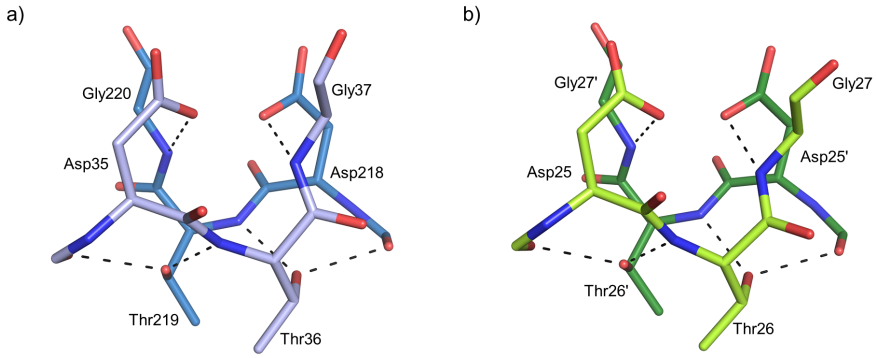


Figure 1.7. Stabilizing, rigid H-bond network (*fireman's grip*) of the conserved Asp-Thr-Gly sequence of aspartic proteases. a) Endothiapepsin (PDB entry: 1OEW). The amino acids of the N- and C-terminal lobe are colored light blue and blue, respectively. b) HIV-1 PR (PDB entry: 1SDT). The amino acids of each monomer are colored differently (light green and green).

Due to the proximity to the active site and the associated flexibility the *flap*, another characteristic feature of aspartic proteases,[23] plays an important role in substrate recognition and ligand specificity. During binding of substrates or inhibitors, the *flap* mostly changes the orientation and forms interactions to the ligand, and is therefore crucial in binding events to aspartic proteases. Most inhibitors address the closed *flap* conformation. Interestingly, in case of the HIV-1 PR it was shown that novel inhibitors can also target the protein in the open *flap* conformation,[41] which represents a different binding mode in comparison to all approved HIV-1 PR inhibitors which might thus be beneficial for the avoidance of drug resistance. Figure 1.8 shows the open and closed *flap*-conformation of the HIV-1 protease.

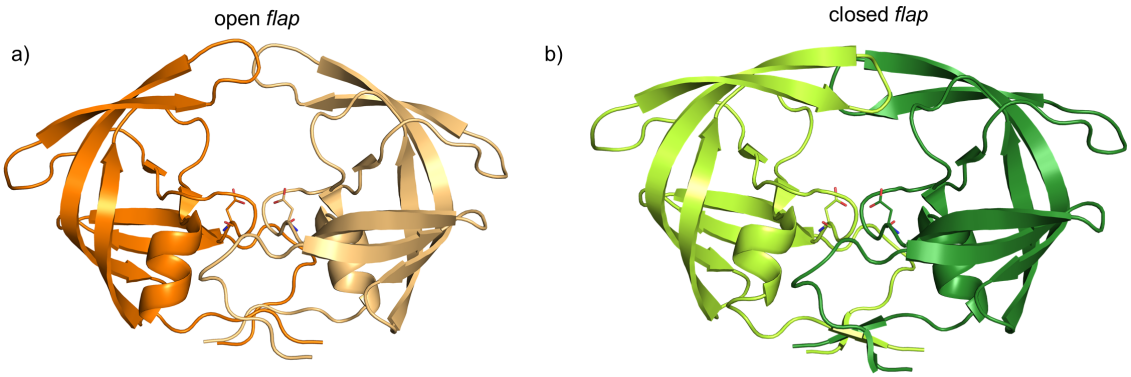


Figure 1.8. HIV-1 PR with open (a) and closed (b) *flap* conformation. PDB entries: (a) 2PC0; (b) 1SDT.

1.3 HTLV-1 and its Virus-Encoded Protease

1.3.1 Virus, Infections, Epidemiology and Related Diseases

The Human T-cell lymphotropic virus type 1 (HTLV-1) was the first human retrovirus isolated in the early 1980s.[42] The discovery of HTLV-1 by Robert Gallo and coworkers confirmed the existence of human retroviruses, which had been discussed controversially until then. The technological developments leading to the discovery of HTLV turned out to be a milestone for the discovery of further retrovirus like the human immunodeficiency virus (HIV) and its identification as causative agent of AIDS.[42] Worldwide approximately 10-20 million people are currently infected with HTLV-1.[43,44] Areas with high prevalence include Japan, Africa, the Caribbean islands, South America, and Iran.[45] In Europe and North America, the majority of HTLV-1 infected humans are immigrants or drug-addicted persons mainly dependent on intravenously applied drugs.[46]

The virus predominantly infects CD4+ T-cells. The transmission of HTLV-1 predominantly occurs via cell-cell contacts. The virus is mainly transmitted via breast-feeding, through contaminated blood products as well as by sexual contacts. As typical for a retrovirus, HTLV-1 persists in the host cell in form of a provirus: the viral DNA is inserted in the human DNA of the host cell and thus causing a lifelong infection. The diseases which are associated with HTLV-1 infections are quite divers: they comprise cancer-developing processes like the adult T-cell leukemia (ATLL) as well as neurological (HTLV-1-associated myelopathy (HAM)) and dermatological (HTLV-1-associated infective dermatitis (HAID)) diseases.[44,47] Due to the low incidence of these diseases within the whole population, these three types of disorder are classified as orphan diseases (< 5 out of 10,000 people).[48] Currently, no causative treatment options to cure these HTLV-1 associated diseases are available.

Even though most people remain free of symptoms after an infection with HTLV-1, still about 10 % of the infected people develop one of the typical fatal diseases among which 1-5 % acquire the adult T-cell leukemia, which represents an aggressive form of leukemia with a very poor life-time prognosis. About 0.3-4 % of all infected people suffer from the chronical inflammatory and neurodegenerative disorder of the medulla (HAM). Other HTLV-1-associated manifestations of the infection include HAID, uveitis, polymyositis, and arthropathy.[44] Until now, it has not been understood, why only a minor part of the infected people develop one of the related diseases, while others have an unaffected life.

The carcinogenic potential of HTLV-1 may origin from the oncogenic protein named tax. In addition to the virus-encoded structural and other non-structural proteins, the

genome of HTLV-1 encodes the tax gene product. The nuclear protein tax promotes the transcription of its own proviral genome, but also that of human genes, like for example anti-apoptotic genes and cytokines. It influences several signaling pathways like the NF-κB signal pathway and *inter alia* represses the transcription of genes that play an important role in the activation of apoptosis as well as in DNA repair. Tax also directly interacts with proteins that are involved in tumor suppression and DNA repair. As a consequence, genetic damages are not repaired, apoptosis of damaged cells does not occur and mutations may be accumulated in infected T-cells. In total, these negative effects of the tax protein promote the development of ATLL.[44]
Patients with ATLL have a low life expectation, the median survival in dependence of the subtype is shown in Table 1.1.

Table 1.1. Subtypes of ATLL, prevalence, and survival time.[46]

subtype	ratio	median survival
acute	55 %	6 months
lymphomatous	20 %	10 months
chronic	20 %	2 years
smouldering	5 %	-

The HTLV-1-associated myelopathy (HAM) leads to neurological dysfunctions in infected people, resulting from an inflammation of the spinal cord. Infected T-cells infiltrate the CNS. Most likely, the immunological response against HTLV-1 antigens leading to an enhanced production of inflammatory cytokines results in lesions in the spinal cord, in this context axons as well as myelin sheaths were affected. However, the complete mechanism how HTLV-1 causes HAM is not yet fully understood today.[49]

1.3.2 HTLV-1 Protease – an Overview

Like other retrovirus, HTLV-1 encodes a protease which is essential for the replication of the virus.[50] The HTLV-1 protease was identified and isolated in the late 1980s.[50,51]
Structurally, the HTLV-1 PR comprises the typical fold as observed for other proteases of the retroviral family (A2), and functions as a homodimer with each monomer consisting of 125 amino acids.
As successful drug discovery approaches have shown in the past, attacking the retroviral protease, which is essential for the replication of the virus, provides a

promising concept to combat viral infections. As a prominent example of this strategy, hitherto ten HIV-1 protease inhibitors have been approved and successfully used in the treatment of the acquired immune deficiency syndrome (AIDS). Therefore, and in addition to its essential role in the maturation of the virus, the HTLV-1 PR represents an attractive target for the development of HTLV-1 PR inhibitors that could overcome the still existing lack of a curative treatment against HTLV-1 infections.[50]

For directed development of novel HTLV-1 PR inhibitors by structure-based drug design, knowledge about the structure of the protease is an indispensable requirement. In view of this fact that the virus and the protease have been known for a long time, the first X-ray structure of the HTLV-1 PR was determined rather late. Its elucidation was reported in 2005,[52] five years later additional six HTLV-1 crystal structures were published.[53] All these seven structures were determined in complex with peptidic inhibitors, which means, up to now, to the best of our knowledge, no structure in absence of a bound ligand has been determined. Furthermore, all these structures were not determined with the full-length protein, but rather with a construct C-terminally truncated by about nine amino acids, consisting of 116 amino acids per monomer. The role of the C-terminal residues is still being discussed controversially in the literature and described in more detail in chapter 2 of this thesis.

The HTLV-1 protease shows a high sequence similarity to other protease structures of the retroviral family. As therapeutic target of the retroviral aspartic PR family A2, particularly the HIV-1 PR is subject of successful and present research approaches. The overall protein fold of the HIV-1 and HTLV-1 PR is very similar (Figure 1.9), their overall sequence identity is about 28 %, their ligand binding regions are even more conserved (45 %), but it is well known, that they differ remarkably from each other in their substrate specificity as well as in their inhibition profile.[54] The approved HIV-1 PR inhibitors show strongly reduced affinity against the HTLV-1 PR,[54,55] and to the best of our knowledge, no highly potent (≤ sub-μM) non-peptidic HTLV-1 PR inhibitors have been described so far. Considering the current state of the research regarding the discovery of HTLV-1 PR inhibitors, the development of novel protease inhibitors is strongly necessary in order to enhance the chance of identifying treatment options for HTLV-1 infections in the future.

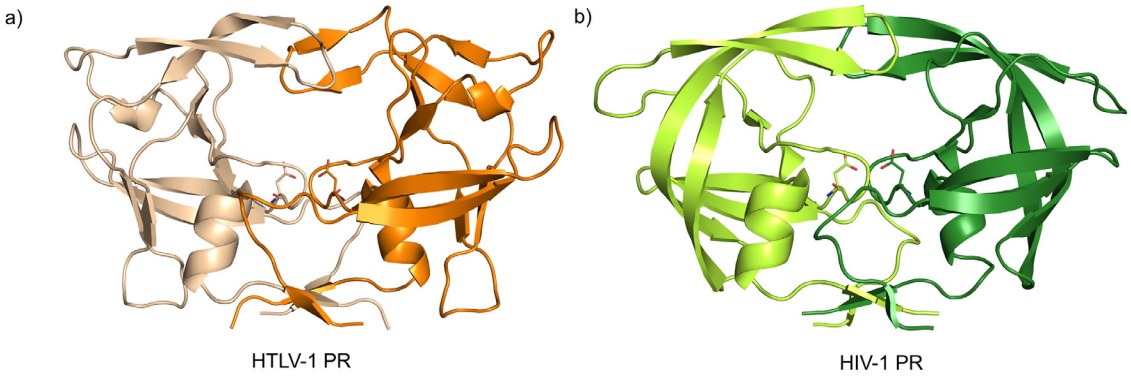


Figure 1.9. The overall structure of the HTLV-1 PR (PDB entry: 3LIX) and the HIV-1 PR (PDB entry: 1SDT) is very similar.

1.4 Endothiapepsin as Model System for Members of the A1 Family

Endothiapepsin (EP), the aspartic protease of *Endothia parasitica* belongs to the aspartic protease family A1. In the daily industrial business, EP is typically used as fungal rennet in cheese production.[56] In drug discovery, EP is commonly used as model enzyme and has successfully been exploited as surrogate in the structure-based drug design of renin as well as β-secretase inhibitors.[57,58] Major advantages of using endothiapepsin as model enzyme include its commercial availability, its biochemical stability, and the well-established crystallization conditions to name just a few. Therefore, EP is also well suited for the establishment of other biochemical and biophysical methods to characterize small molecules further regarding their inhibitory activity.

1.5 Aims of the Thesis

The strategy of protease inhibition and identification of suitable lead structures has been successfully established for quite some time.[59] Examples comprise among others the angiotensin-converting enzyme (ACE), the Hepatitis C virus (HCV) NS3/4A protease, and the HIV-1 protease. Ten HIV-1 protease inhibitors are still being used or have been used as approved drugs to date.[59–61]

Within this thesis, the two aspartic proteases human T-cell leukemia virus type 1 protease (HTLV-1 PR) and endothiapepsin (EP) are investigated in detail, particularly with focus on the identification and synthesis of small-molecule-type, nonpeptidic protease inhibitors as well as the kinetic and structural characterization of these in complex with the respective protease.

1.5.1 HTLV-1 Protease

The initial aims of this project focused on the generation of the target-related laboratory platform, particularly the establishment of suitable and robust protocols for protein expression, purification and refolding in the in-house laboratory in order to provide continuous access to the required target proteins. In addition, a reliable assay protocol, which enables the detection of the protease activity and its inhibition by various test compounds had to be established (**chapter 2**).

The overall purpose of the first project of this thesis was the identification of novel, non-peptidic small molecules for HTLV-1 protease inhibition.

For the validation of the kinetic assay as well as for the establishment of a suitable crystallization protocol, the commercially available HIV-1 protease inhibitor indinavir was initially selected. According to the literature, indinavir possesses, in contrast to several other approved HIV-1 PR inhibitors that were shown to be inactive against HTLV-1 PR, a K_i-value against the latter of 3.5 µM,[55] and therefore represents an easily accessible inhibitor with moderate affinity against the HTLV-1 protease. Hence, it turned out to be a well-suited inhibitor for the validation of the in-house assay establishment.

Interestingly, despite more than 10 years of structural research on HTLV-1 PR, prior to this thesis, only crystal structure complexes with peptide-type inhibitors had been described thus neglecting non-peptide-type ligands despite their increased relevance for structure-guided lead discovery in the recent past. Hence, the first central aim of this thesis as detailed in **chapter 3** was the generation of a crystal structure complex of HTLV-1 PR with bound indinavir in order to get insights into its binding mode, to investigate the structural basis of the different binding affinities with respect to HIV-1 and HTLV-1 PR, and finally to suggest appropriate structure-derived modifications for the next chemical candidates with improved properties.

Chapter 4 and 5 of this thesis focuses on the generation of novel chemical starting points for HTLV-1 PR lead discovery activities by means of the privileged scaffold approach, i.e. to exploit knowledge on inhibitor scaffolds already available for other representatives of the same enzyme family, and to generate novel chemical matter by characterizing their binding properties towards the particular target of interest. For this purpose, the aspartic protease inhibitor library synthesized in our research group over the last years, should be evaluated to identify suitable scaffolds with HTLV-1 PR inhibitory activity. Identified hits were ought to be investigated regarding their inhibitory properties and in structural terms with respect to their binding modes to the HTLV-1 PR by X-ray crystallography. The results of these studies were supposed to be interpreted

with respect to the deduction of an initial SAR from the library subset with appropriate inhibitory activity, and in terms of deviating binding features between their original target and HTLV-1 PR.

1.5.2 Endothiapepsin

The general aim of the second project **(chapter 6)** was the detailed investigation of the binding modes of 2-amino-thiophene ligands in the active site of endothiapepsin.

In the preliminary work of Dr. Ruben Bartholomäus, Amir Shahim und Dr. Helene Köster, *Gewald* reaction-based 2-amino-thiophene inhibitors of endothiapepsin were synthesized and characterized in terms of affinity measurements and X-ray structure determination.[62,63]

The crystal structures of three inhibitors in complex with EP were successfully determined by Dr. Helene Köster.[62] Most interestingly, each of these inhibitors exhibits a different binding mode (Figure 1.10): the orientation of the inhibitors in the active site as well as the mode of addressing the catalytic aspartates significantly differs among the three inhibitors. For the further investigation and interpretation of these unexpected findings, the synthesis of a novel series of 2-amino-thiophene ligands should be performed, as well as their kinetic characterization and crystallization.

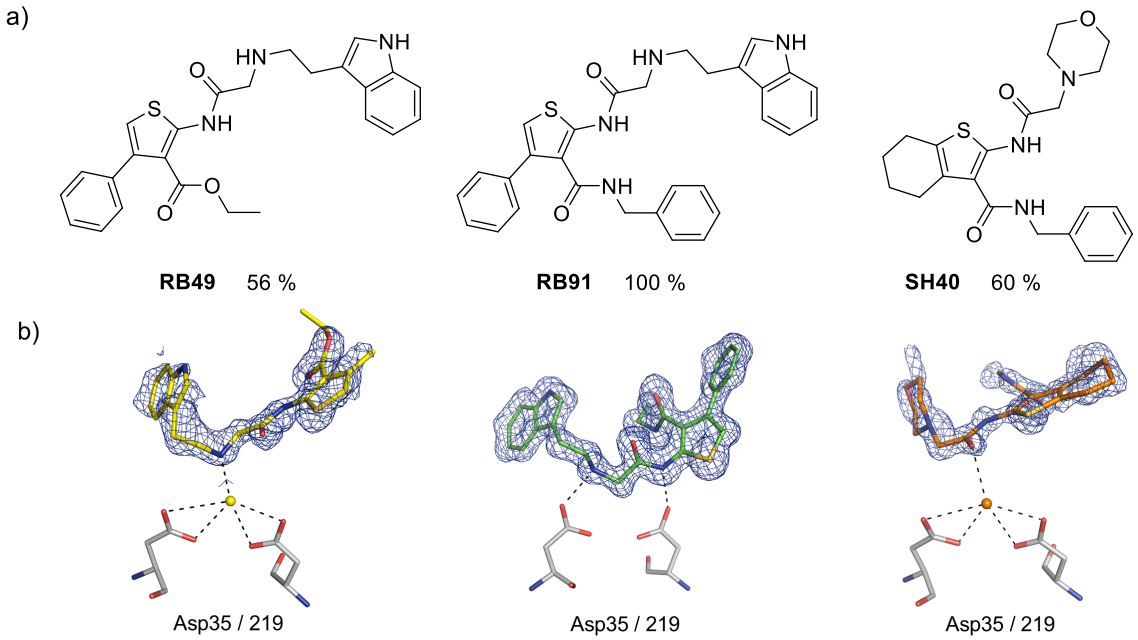


Figure 1.10. a.) Chemical structures of the *Gewald* reaction-based inhibitors RB49, RB91 and SH40, as well as %-inhibition data at 100 µM inhibitor concentration. b.) The X-ray structures of the inhibitors in complex with endothiapepsin exhibit different binding modes. Interactions of the inhibitors to the catalytic aspartates are shown, as well as the $2F_o$-F_c electron densities for the ligands (blue mesh, 1σ).

In addition, within a cooperation project with the research group of Prof. Klebe (Philipps-Universität Marburg) regarding different methods for fragment screening, the in-house fragment library should be screened against endothiapepsin utilizing the thermal shift assay (TSA) **(chapter 7)**. This TSA was supposed to contribute to the overall aim of comparing and analyzing the hit-overlap with hit lists of other fragment screening methods, such as a fluorescence-based assay, fragment screening by X-ray crystallography, or STD-NMR experiments. For this purpose, an appropriate robust experimental TSA setup had to be established for this protein, followed by the screening of the fragment library[64] of 364 existing fragments to identify stabilizing fragment-type binders.

1.6 References

(1) Zhao, H.; Guo, Z. Medicinal chemistry strategies in follow-on drug discovery. *Drug discovery today* **2009,** *14*, 516–522.

(2) Raju, T.N.K. (2000) The nobel chronicles. Lancet 355, 1022.

(3) Acharya, C.; Coop, A.; Polli, J. E.; Mackerell, A. D. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr. Comput.-Aided Drug Des.* **2011,** *7*, 10–22.

(4) Hubbard, R. E. 3D Structure and the Drug Discovery Process. In *Structure-Based Drug Discovery.* Hubbard, R. E., Ed.; Royal Society of Chemistry: Cambridge; pp. 1–31.

(5) Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **2003,** *10*, 787-797.

(6) Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **1988,** *31*, 2235-2246.

(7) Boehm, M. Virtual screening of chemical space: from generic compound collections to tailored screening libraries. In *Virtual Screening.* Sotriffer, C., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany; pp. 1–33.

(8) Welsch, M. E.; Snyder, S. A.; Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* **2010,** *14*, 347–361.

(9) DeSimone, R. W.; Currie, K. S.; Mitchell, S. A.; Darrow, J. W.; Pippin, D. A. Privileged structures: applications in drug discovery. *Comb. Chem. High Throughput Screening* **2004,** *7*, 473–494.

(10) Robuck, P. R.; Wurzelmann, J. I. Understanding the drug development process. *Inflammatory Bowel Dis.* **2005,** *11 Suppl 1*, S13-6.

(11) FDA/CDER. Novel New Drugs 2014 Summary. Available from: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm20025676.htm. [access: 26.02.2015].

(12) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004,** *3*, 935–949.

(13) Bajorath, J.; Jiang, H.; Shoichet, B. K.; Walters, W. P. Computational methods for medicinal chemistry. *J. Med. Chem.* **2015,** *58*, 1019.

(14) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006,** *65*, 15–26.

(15) Rami Reddy, M.; Erion, M. D. *Free energy calculations in rational drug design*; Kluwer Academic/Plenum Publishers: New York, 2001.

(16) Vauquelin, G. Determination of drug–receptor residence times by radioligand binding and functional assays: experimental strategies and physiological relevance. *Med. Chem. Commun.* **2012,** *3*, 645.

(17) Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014,** *42*, D503-9.

(18) Szecsi, P. B. The aspartic proteases. *Scand. J. Clin. Lab. Invest., Suppl.* **1992,** *210*, 5–22.

(19) Hibbetts, K.; Hines, B.; Williams, D. An overview of proteinase inhibitors. *J. Vet. Intern. Med.* **1999,** *13*, 302–308.

(20) Bernstein, N. K.; Cherney, M. M.; Loetscher, H.; Ridley, R. G.; James, M. N. Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from plasmodium falciparum. *Nat. Struct. Biol.* **1999,** *6*, 32–37.

(21) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci.* **1992,** *1*, 322–328.

(22) Dash, C.; Kulkarni, A.; Dunn, B.; Rao, M. Aspartic peptidase inhibitors: implications in drug development. *Crit. Rev. Biochem. Mol. Biol.* **2003,** *38*, 89–119.

(23) Tang, J. Aspartic proteases: structure, function, and inhibition. In *Aspartic Acid Proteases as Therapeutic Targets.* Ghosh, A. K., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany; pp. 23–41.

(24) Pauling, L. Nature of forces between large molecules of biological interest. *Nature* **1948,** *161*, 707–709.

(25) Marciniszyn, J.; Hartsuck, J. A.; Tang, J. Mode of inhibition of acid proteases by pepstatin. *J. Biol. Chem.* **1976,** *251*, 7088–7094.

(26) Torrado, A. Stereoselective synthesis of aminoethylamine aspartyl protease transition state isosteres. *Tetrahedron Lett.* **2006,** *47*, 7097–7100.

(27) Wlodawer, A.; Erickson, J. W. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **1993,** *62*, 543–585.

(28) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001,** *46*, 3–26.

(29) Virgil, S. C. First-Generation HIV-1 protease inhibitors for the treatment of HIV/AIDS. In *Aspartic Acid Proteases as Therapeutic Targets.* Ghosh, A. K., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany; pp. 139–168.

(30) Jensen, C.; Herold, P.; Brunner, H. R. Aliskiren: the first renin inhibitor for clinical treatment. *Nat. Rev. Drug Discovery* **2008,** *7*, 399–410.

(31) The birthplace of Pepsi Cola (Pepsi Store). Available from: http://www.pepsistore.com/history.asp. [access: 16.01.2015].

(32) Tang, J.; Sepulveda, P.; Marciniszyn, J.; Chen, K. C.; Huang, W. Y.; Tao, N.; Liu, D.; Lanier, J. P. Amino-acid sequence of porcine pepsin. *Proc. Natl. Acad. Sci. U. S. A.* **1973,** *70*, 3437–3439.

(33) Subramanian, E.; Swan, I.D.; Liu, M.; Davies, D. R.; Jenkins, J. A.; Tickle, I. J.; Blundell, T. L. Homology among acid proteases: comparison of crystal structures at 3Å resolution of acid proteases from Rhizopus chinensis and Endothia parasitica. *Proc. Natl. Acad. Sci. U. S. A.* **1977,** *74*, 556–557.

(34) Hsu, I.-N.; Delbaere, Louis T. J.; James, Michael N. G.; Hofmann, T. Penicillopepsin from Penicillium janthinellum crystal structure at 2.8 Å and sequence homology with porcine pepsin. *Nature* **1977,** *266*, 140–145.

(35) Miller, M.; Jaskólski, M.; Rao, J. K.; Leis, J.; Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **1989,** *337*, 576–579.

(36) Dunn, B. M.; Goodenow, M. M.; Gustchina, A.; Wlodawer, A. Retroviral proteases. *Genome Biol.* **2002,** *3*, reviews3006-reviews3006.7.

(37) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B.; Baldwin, E.; Weber, I.; Selk, L.; Clawson, L.; Schneider, J.; Kent, S. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989,** *245*, 616–621.

(38) Wlodawer, A.; Gustchina, A. Structural and biochemical studies of retroviral proteases. *Biochim. Biophys. Acta* **2000,** *1477*, 16–34.

(39) Tang, J.; James, M. N. G.; Hsu, I. N.; Jenkins, J. A.; Blundell, T. L. Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **1978,** *271*, 618–621.

(40) Ingr, M.; Uhlíková, T.; Strísovský, K.; Majerová, E.; Konvalinka, J. Kinetics of the dimerization of retroviral proteases: the "fireman's grip" and dimerization. *Protein Sci.* **2003,** *12*, 2173–2182.

(41) Böttcher, J.; Blum, A.; Dörr, S.; Heine, A.; Diederich, W. E.; Klebe, G. Targeting the open-*flap* conformation of HIV-1 protease with pyrrolidine-based inhibitors. *ChemMedChem* **2008,** *3*, 1337–1344.

(42) Gallo, R. C. History of the discoveries of the first human retroviruses: HTLV-1 and HTLV-2. *Oncogene* **2005,** *24*, 5926–5930.

(43) Edlich, R. F.; Arnette, J. A.; Williams, F. M. Global epidemic of human T-cell lymphotropic virus type-I (HTLV-I). *J. Emerg. Med.* **2000,** *18*, 109–119.

(44) Verdonck, K.; González, E.; van Dooren, S.; Vandamme, A.-M.; Vanham, G.; Gotuzzo, E. Human T-lymphotropic virus 1: recent knowledge about an ancient infection. *Lancet Infect. Dis.* **2007,** *7*, 266–281.

(45) Proietti, F. A.; Carneiro-Proietti, Anna Bárbara F; Catalan-Soares, B. C.; Murphy, E. L. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* **2005,** *24*, 6058–6068.

(46) Bangham, C. R M. HTLV-1 infections. *J. Clin. Pathol.* **2000,** *53*, 581–586.

(47) McGill, N.-K.; Vyas, J.; Shimauchi, T.; Tokura, Y.; Piguet, V. HTLV-1-associated infective dermatitis: updates on the pathogenesis. *Exp. Dermatol.* **2012,** *21*, 815–821.

(48) Orphanet, a reference portal for information on rare diseases and orphan drugs. Available from: http://www.orpha.net/consor/cgi-bin/index.php. [access: 26.02.2015].

(49) Fuzii, H. T.; da Silva Dias, G. A.; de Barros, R. J. S.; Falcão, L. F. M.; Quaresma, J. A. S. Immunopathogenesis of HTLV-1-assoaciated myelopathy/tropical spastic paraparesis (HAM/TSP). *Life sciences* **2014,** *104*, 9–14.

(50) Shuker, S. B.; Mariani, V. L.; Herger, B. E.; Dennison, K. J. Understanding HTLV-I protease. *Chem. Biol.* **2003,** *10*, 373–380.

(51) Hatanaka, M.; Nam, S. H. Identification of HTLV-I gag protease and its sequential processing of the gag gene product. *J. Cell. Biochem.* **1989,** *40*, 15–30.

(52) Li, M.; Laco, G. S.; Jaskolski, M.; Rozycki, J.; Alexandratos, J.; Wlodawer, A.; Gustchina, A. Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 18332–18337.

(53) Satoh, T.; Li, M.; Nguyen, J.-T.; Kiso, Y.; Gustchina, A.; Wlodawer, A. Crystal structures of inhibitor complexes of human T-cell leukemia virus (HTLV-1) protease. *J. Mol. Biol.* **2010,** *401*, 626–641.

(54) Kádas, J.; Weber, I. T.; Bagossi, P.; Miklóssy, G.; Boross, P.; Oroszlan, S.; Tözsér, J. Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell Leukemia virus type 1 protease. *J. Biol. Chem.* **2004,** *279*, 27148–27157.

(55) Bagossi, P.; Kádas, J.; Miklóssy, G.; Boross, P.; Weber, I. T.; Tözsér, J. Development of a microtiter plate fluorescent assay for inhibition studies on the HTLV-1 and HIV-1 proteinases. *J. Virol. Methods* **2004,** *119*, 87–93.

(56) Rawlings, N. D.; Salvesen, G. *Handbook of Proteolytic Enzymes*; Elsevier Science: San Diego, 2012, p. 147.

(57) Blundell, T. L.; Cooper, J.; Foundling, S. I.; Jones, D. M.; Atrash, B.; Szelke, M. On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition-state analogs. *Biochemistry* **1987,** *26*, 5585–5590.

(58) Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; Beer, T. de; Folmer, Rutger H A. Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J. Med. Chem.* **2007,** *50*, 5903–5911.

(59) Drag, M.; Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discovery* **2010,** *9*, 690–701.

(60) Mehellou, Y.; Clercq, E. de. Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go? *J. Med. Chem.* **2010,** *53*, 521–538.

(61) Venkatraman, S. Discovery of boceprevir, a direct-acting NS3/4A protease inhibitor for treatment of chronic hepatitis C infections. *Trends Pharmacol. Sci.* **2012,** *33*, 289–294.

(62) Köster, H. Endothiapepsin und Proteinkinase A: Komplexstrukturen mit neuartigen Inhibitoren, Durchmustern einer Fragmentbibliothek sowie Inhibitordesign ausgehend von einer Sonde. Dissertation, Philipps-Universität Marburg, 2012.

(63) Bartholomäus, R. Synthese und Struktur-Wirkungsbeziehungen substituierter Thiophene als neuartige Leitstrukturen zur Inhibition von Plasmepsin II und IV. Diplomarbeit, Philipps-Universität Marburg, 2007.

(64) Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011,** *54*, 7784–7796.

2. In-House Establishment of a HTLV-1 Protease Technology Platform: Protein Production, TSA, Fluorescence-Based Assay, and Crystallographic Setup

2.1 HTLV-1 Protease Constructs

As described in more detail in chapter 1.3 of this thesis, the HTLV-1 PR is an aspartic protease which functions as a homodimer, each monomer consisting of 125 residues. A comprehensive literature search revealed that, mainly two different protease constructs were used to characterize the HTLV-1 PR as well as to investigate the substrate and inhibition profile: for X-ray structure determinations of the HTLV-1 PR, a construct truncated by nine amino acids (HTLV-1 PR 1-116) was used, while for biochemical assays the full-length PR (HTLV-1 PR 1-125) as well as the HTLV-1 PR 1-116 were applied.[1–4] In order to identify small molecules which show inhibitory activity against the HTLV-1 PR as well as to get a deeper insight into their mode of action by determining the X-ray structures in complex with the protease, both constructs are of considerable interest.

Noteworthy, for all seven X-ray structures of the HTLV-1 PR available in the PDB,[5,6] the shorter HTLV-1 PR 1-116 construct was used.[1,2] However, there is an ongoing debate regarding the activity of the shortened constructs (truncated by 9 or 10 amino acids): the results vary from exhibiting similar activity in comparison to the full-length protein to reports describing the truncated ones to be inactive.[1,7–9] Overall, the importance of the C-terminal residues of the HTLV-1 PR is still not fully understood. Besides the HTLV-1 protease also the protease of the bovine leukemia virus (BLV) belonging likewise to the class of the deltaretroviruses possesses such a C-terminal extension, which is not present in most of the other retroviral proteases.[7]

Kádas *et al*.[7] performed purposeful experiments to prove the importance of the C-terminus of the HTLV-1 PR especially with respect to the stability of the enzyme dimer. Besides kinetic parameters they performed *inter alia* gel filtration experiments and determined K_d-values to investigate the dimer stability and rationalized their results by building a molecular model of the full-length protease based on the X-ray structure of the shortened construct followed by manual addition of the missing C-terminal residue and appropriate *in silico* energy minimization. According to Kádas *et al*.,[7] the C-terminal residues seem to play a critical role in dimerization as the shorter protease constructs have a significantly lower dimer stability. Also the aggregation tendency of the protease seems to be dependent on the C-terminal residues: the full-length protein shows a

higher aggregation tendency than the shortened constructs. The X-ray structure of the HTLV-1 PR 1-116 reveals that a hydrophobic surface is present close to the terminal β-sheets. The proposed molecular model of the full-length protein suggests that the C-terminal residues might form new strands for the terminal β-sheet and interacts with this hydrophobic surface thus rationalizing the observation of the contribution of the C-terminus towards the dimer stability and the aggregation tendency. In addition, it is reported that the C-terminus is also critical for the infectivity of the HTLV-1 virions, as the by about ten amino acids truncated C-terminal PR abolished the infectivity. Hence, the dimerization interface may also provide an additional target site of the HTLV-1 protease.[7]

Considering the state of research and the aims of this thesis, the protein expression, purification, and refolding for the two most commonly used HTLV-1 PR constructs should be established in the in-house laboratory: the shortened HTLV-1 PR 1-116 especially for crystallization experiments and the full-length protein mainly for kinetic measurements.

Besides the variability of the length of the HTLV-1 PR constructs used in various experiments, the applied ones also vary to some degree in their amino acid sequence. These mutations have mainly been introduced to prevent autoproteolysis. Mostly, as also in our case the single mutant L40I was used. Louis *et a*l.[10] mapped the site of autoproteolytic cleavage and described the stabilization against autoproteolysis caused by the L40I mutant. This mutant was shown to exhibit nearly identical kinetic parameters in comparison to the wild type protease.

2.2 Expression, Purification and Refolding of the HTLV-1 Proteases

Dr. Alexander Wlodawer, National Cancer Institute, Frederick, Maryland, USA, kindly provided the plasmid encoding the HTLV-1 PR 1-116 harboring the mutations L40I and Y114N for our studies. First, the asparagine was re-mutated to tyrosine by site-directed mutagenesis to obtain the favored plasmid only containing the L40I mutation.

The plasmid encoding the full-length protein was ordered from Life Technologies/GeneArt (Germany).

2.2.1 Site-Directed Mutagenesis of the HTLV-1 PR 1-116 Plasmid by PCR

The polymerase chain reaction (PCR) is one of the most important methods in molecular biology. The PCR technique was invented and developed by Kary Mullis in the 1980s and awarded with the Nobel Prize in Chemistry in 1993.[11] The PCR is e.g. used in the diagnostics of diseases, in forensic analysis, and as an important technique

in molecular biology research, for example for DNA cloning.[12]

The PCR is a three-step reaction: first the double-stranded DNA is denatured resulting in two single strands (95 - 98°C). In the second step, the temperature is decreased (≈50 - 60°C) allowing the primer to anneal to the DNA. In the third step, typically performed at 72°C, the applied thermo-stable polymerase finally creates the new DNA strand complementary to the DNA template strand. These three steps were repeated (about 25-35 cycles) resulting in the exponential amplification of the DNA.[12]

Within this thesis, the PCR technique was used for the site-directed mutagenesis of the HTLV-1 PR 1-116 (L40I, Y114N) construct: for the identification of the following optimized PCR conditions, different ratios of the components of the PCR solution as well as different hybridization temperatures were tested.

Finally, the following composition of the PCR solution and parameters of the PCR program were used:

10 µl 5x HiFi-buffer (BioCat)
1 µl PRECISOR high-fidelity DNA Polymerase, 250 U (BioCat)
5 µl plasmid (encoding HTLV-1 PR 1-116, L40I, Y114N)
1 µl primer 1 (20 nM)
1 µl primer 2 (20 nM)
1 µl dNTP-Mix (each 25 mM)
31 µl water (sterile)

The PCR was performed in a Thermocycler PCR Mini Cycler (MJ Research). Steps 2-4 were repeated 30 times (30 PCR cycles).

step	time	temperature	
1	60s	98°C	
2	30s	98°C	denaturation
3	30s	63°C	hybridisation
4	7min	72°C	extension
5	∞	4°C	

Before transforming the new plasmid into *E. coli* cells, first a DpnI digestion must be performed to remove the starting plasmid and the PCR product has to be purified. The DpnI digestion was performed by incubating 30 µL of the PCR product solution and 3 µL DpnI (10 U/µL, Fermentas) for 2 h at 37°C. For purification of the PCR product a purification kit (peQLab) was used.

After the plasmid minipreparation, the new plasmid (L40I) was sequenced (Eurofins, Germany) to confirm the mutation.

2.2.2 Expression System

The full-length protease (HTLV-1 PR 1-125) as well as the truncated protease (HTLV-1 PR 1-116) were expressed in the *E. coli* strain Rosetta 2 (DE3).

First, the respective plasmids, both bearing the L40I mutation, were transformed via heat shock into the Rosetta 2 (DE3) competent cells according to a standard protocol and a glycerol stock was prepared and stored at -80°C.

2.2.3 Inclusion Bodies

According to the literature, the HTLV-1 protease is expressed as inclusion bodies in *E. coli*. Inclusion bodies often result from overexpression of recombinant proteins through aggregation of the expressed and un- or partially folded protein. For the recovery of active protein from inclusion bodies, they firstly need to be extracted from the *E. coli* cells, followed by the solubilization of the aggregated protein, typically using strong denaturants, such as for example highly concentrated urea or guanidine hydrochloride, thiocyanate salts or detergents causing denaturation. Finally, the protein usually has to be refolded, for example by dilution or by chromatography. In this step the denaturing agent is removed to enable the refolding of the protein. Finding suitable refolding conditions is often quite challenging, due to aggregation and misfolding of the protein. Therefore it seems in general to be easier to work with soluble proteins instead of insoluble inclusion bodies. Process parameters that might influence the production of soluble protein are e.g. the media composition, the expression temperature, the production rate, and the availability of chaperones. To address the problem of protein insolubility, also special *E. coli* cell lines are available, providing a further approach to increase the yield of soluble protein, like e.g. the ArcticExpress cells.[13] However, working with inclusion bodies might also have some advantages: soluble impurities can easily be separated by washing steps and centrifugation. Usually the expression level is higher than in soluble systems, the protein product is protected from proteolytic degradation, and in case of target proteins with toxic properties towards the used expression organism, the cells are usually protected against their toxicity.[14,15]

2.2.4 Protein Expression

The protein expression protocol is based on the literature procedure from Li *et al*.[1] The pre-culture was prepared in 10 ml LB medium containing the antibiotics ampicillin (100 mg/l) and chloramphenicol (20 mg/l) at 37°C and 220 rpm shaking overnight. For the main culture two pre-cultures (20 ml) were transferred into 1.6 l LB medium containing the same antibiotics as in the pre-culture and shaken at 37°C until an OD_{600}

of about 0.7 was observed (≈ 3.5 h). The protein expression was induced by addition of isopropyl-β-D-thiogalactopyranosid (IPTG) (1 mM). The temperature was decreased from 37°C to 14°C to enable a slower protein production, after ca. 15 hours the cells were harvested by centrifugation (5500 rpm, 15 min, 4°C). The cells were resuspended in 70 ml buffer A (Table 2.1) and for cell disruption this solution was first incubated with lysozyme (1 h) followed by sonication on ice (10 times for 90 seconds). For removal of nucleic acids, the solution was incubated for 1 h at 4°C with Benzonase® Nuclease (1.5 µl). After that a washing step of the inclusion bodies was performed: first, urea was added to the inclusion body solution to finally obtain an urea concentration of 0.5 M, while the solution was again stirred for 1 h at 4°C. Second, the solution was centrifuged (20000 rpm, 30 min, 4°C) and the obtained inclusion bodies were washed three times with buffer B. For solubilization of the inclusion bodies, the pellet was resuspended in buffer C, containing 8 M urea as strong denatured, and stirred at 4°C for 15 h.

2.2.5 Protein Purification

The purification was performed according to the procedure from Li *et al.*[1] and comprises two purification steps by fast protein liquid chromatography (FPLC). First, the inclusion body solution (supernatant) was passed through a Q Sepharose (anion exchange column) equilibrated with buffer D. Before loading the eluate onto an SP Sepharose (cation exchange columns) equilibrated with buffer E, the pH of the eluate was adjusted to 3.0. The bound protein was eluted with buffer E containing 0.3 M NaCl.

2.2.6 Refolding

The refolding of both protease constructs was performed via dialysis; however, a different refolding protocol was used for each protein.

The refolding of the HTLV-1 PR 1-116 was performed according to the publication from Li et al.[1] using the sodium acetate buffer F. The protein was stored at -80°C until use.

The refolding of the full-length protease (HTLV-1 PR 1-125) was performed as published[9] and comprised three different refolding steps: first, the protein was dialyzed against buffer G for 15-24 h, followed by buffer H for 12-24 h, and, finally, against buffer I for 24 h. The HTLV-1 PR 1-125 was stored at -20°C until use.

Table 2.1. Used buffers for HTLV-1 PR expression, purification and refolding.

buffer	buffer composition	buffer	buffer composition
A	10 mM Tris, pH 7.5 5 mM EDTA	**F**	15 mM sodium acetate, pH 3.0 5 % PEG 300 5 mM DTT
B	0.5 M urea 10 mM Tris, pH 7.5 5 mM EDTA	**G**	25 mM formic acid, pH 2.8
C	8 M urea 10 mM Tris, pH 7.5 5 mM EDTA 10 mM DTT	**H**	50 mM sodium acetate, pH 5.0 1 mM EDTA 1 mM DTT
D	6 M urea 10 mM Tris, pH 7.5 5 mM EDTA 5 mM DTT	**I**	20 mM PIPES, pH 7.0 150 mM sodium chloride 10 % glycerol 1 mM EDTA 2 mM DTT 0.5 % Nonidet P-40
E	6 M urea 20 mM sodium acetate, pH 3.0 5 mM EDTA 5 mM DTT		

2.2.7 Conclusion and Outlook

Within this thesis a robust protocol for the expression, purification, and refolding of HTLV-1 PR was successfully established in our laboratory which provides continuous access to both HTLV-1 protease constructs. As the SDS gels (Figure 2.1 and 2.2) indicate the expression of both proteins worked well. Via the inclusion body washing steps most of the protein impurities were removed; only two weak protein bands between 35 and 40 kDa were present before the purification via Q and SP Sepharose. After the purification by the Q and SP Sepharose hardly any impurities were detectable (Figure 2.1 and 2.2).

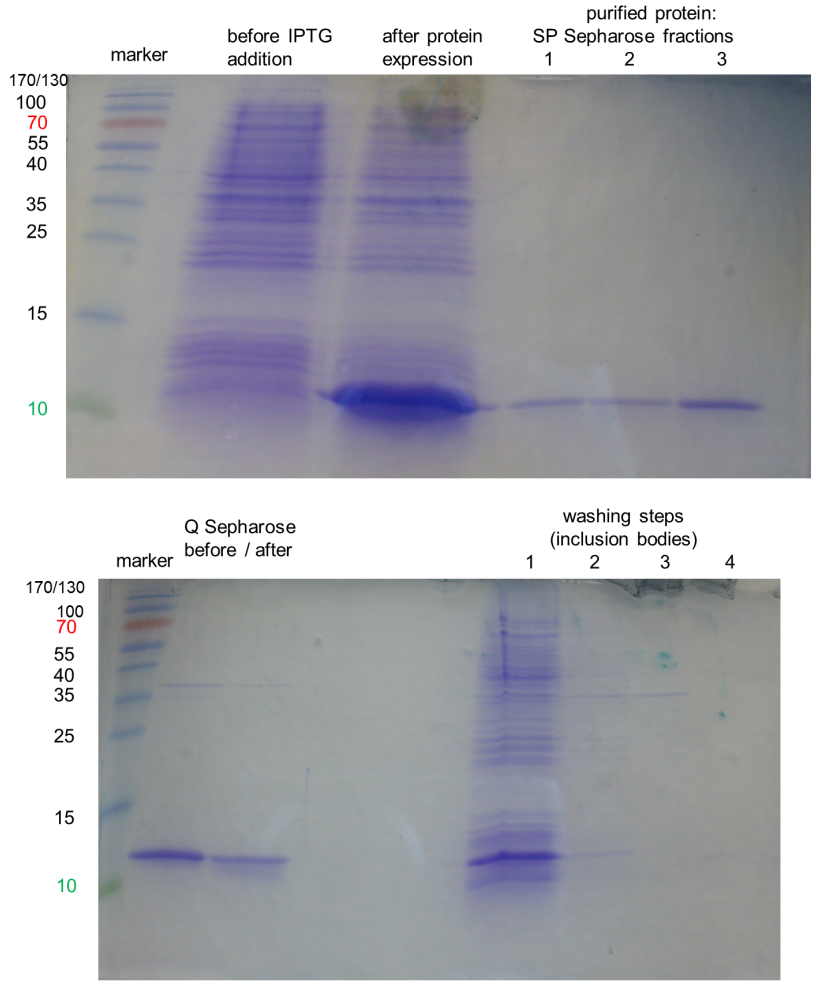


Figure 2.1. SDS gels of the HTLV-1 PR 1-116.

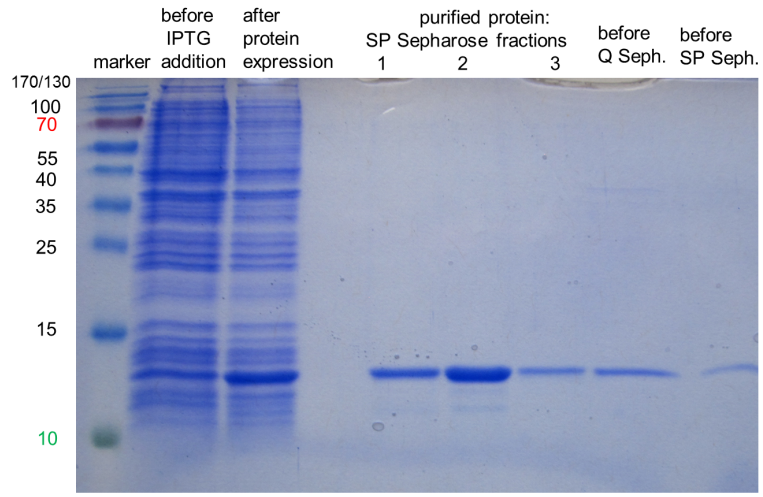


Figure 2.2. SDS gel of the HTLV-1 PR 1-125.

2.3 Thermal Shift Assay

Thermal shift assay (TSA) measurements are in general easy to perform and suitable as screening method to identify novel inhibitor scaffolds as well as stabilizing buffer conditions (e.g. required to proceed into crystallization trials), while neither substrate nor knowledge about the protein function or even enzymatic activity is needed. Therefore, this assay was chosen for the identification of novel inhibitor scaffolds which could serve as starting points for the discovery of HTLV-1 PR inhibitors. Crystallization of these should, in the following step, provide valuable insights into the binding mode of these novel inhibitor scaffolds in complex with the protease and thus facilitate the further structure-based optimization of the initial screening hits. The principle and general application of the TSA is described in detail in chapter 7.1.

Within this thesis, the development of a suitable TSA assay protocol was envisaged for the truncated as well as for the full-length protein constructs (HTLV-1 PR 1-116 and HTLV-1 PR 1-125, respectively).

2.3.1 HTLV-1 PR 1-116

The first measurements were performed in an assay buffer published for the truncated construct (buffer 1, Table 2.2) using different protein concentrations and protein batches which had been stored at different conditions as well as pepstatin as tool inhibitor. In addition, the buffer was used without DTT (buffer 2) and the heating temperature was expanded from the classical range of 20 - 80°C to 5 - 100°C. Unfortunately, no evaluable melting curves were obtained under these various conditions.

To investigate the feasibility of the TSA setup in general, the assay was performed with endothiapepsin as well as with the related HIV-1 protease as model systems.

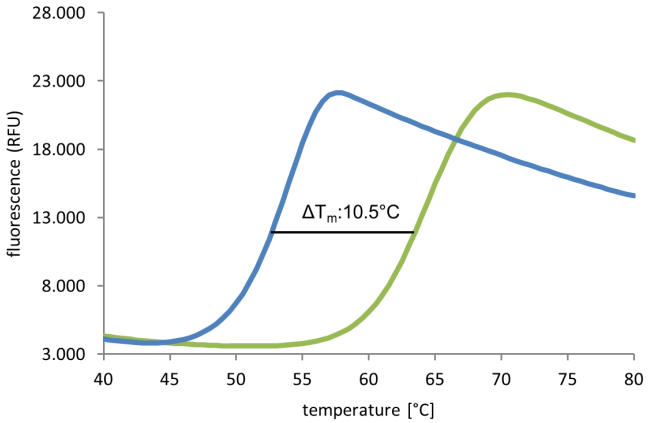


Figure 2.3. TSA curves of the HIV-1 PR without inhibitor (blue) and with pepstatin (63 µM) (green). The curves possess the shape of typical TSA curves and a clear shift of the melting temperature due to protein stabilization through pepstatin as inhibitor is observed.

In both cases the typical TSA curves were obtained (Figure 2.3) indicating that the assay setup itself in principle was functional.

Attempts to optimize the assay setup for HTLV-1 PR included measurements in different buffers (Table 2.2) having been used in the literature for HTLV-1 PR in different assay systems as well as in water. Besides the refolded protein, also protein before the refolding step as well as protein which had been denatured through heating after the refolding was used for TSA experiments. In Figure 2.4, TSA curves from the HTLV-1 PR as well as from the related HIV-1 PR are shown, for comparison each one with denatured protein as well. In case of the HTLV-1 PR no difference between the denatured and refolded protein as well as the protein before the refolding step can be seen, while in case of the HIV-1 PR a typical curve was obtained for the active protease whereas a curve similar to that of the HTLV-1 PR was observed for the denatured protein.

Noteworthy, the solvent exchange of buffer for water yielded different melting curves (Figure 2.5). However, this curve progression was only observed with protein before refolding, and could not be interpreted in a meaningful way while the curves recorded from protein batches after the refolding step resembled the curves obtained for the denatured protein (Figure 2.4). Also a pH-screen (pH-values from 4.0 to 9.5 (pH-screen I) did not result in any improvement of the curves.

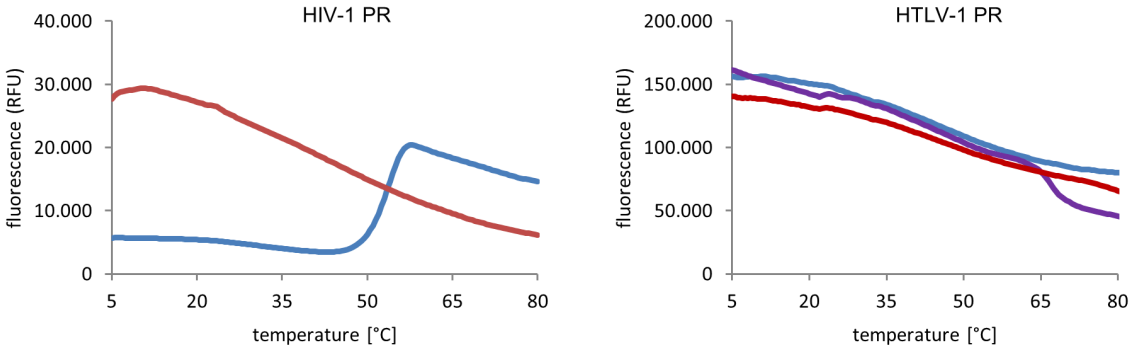


Figure 2.4. TSA curves of the HIV-1 PR (left) and HTLV-1 PR (right) of refolded protein (blue) and denatured protein (red), in the case of the HTLV-1 PR also protein before the refolding step is shown (purple).

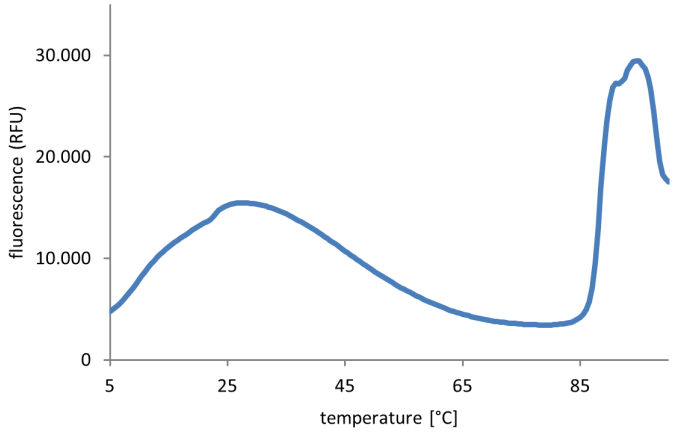


Figure 2.5. HTLV-1 PR before the refolding measured in water as “assay buffer”.

After the first successful crystallization trials of the refolded protein using the truncated construct, once again some additional tests were performed to find suitable buffer conditions for the TSA measurements utilizing the following screening conditions:

- A pH-screen (pH-screen II) probing the pH-range from 4.5 - 9.5 with in total 23 different buffers, combined with different sodium chloride concentrations (without NaCl, 100 mM NaCl and 500 mM NaCl), was performed on the basis of the publication from Ericsson *et al*.[16]
- A PACT Screen (Quiagen), originally composed to test the effect of anions, cations, and pH-values on crystallization, was performed based on a published protocol.[17]

However, also these attempts did not yield typical TSA curves. Thus, the establishment of a TSA protocol for the HTLV-1 1-116 protease construct was discontinued as this construct is obviously not suited for this biophysical assay.

Table 2.2. Different buffers used in the TSA of the HTLV-1 PR 1-116.

buffer (annotation)	buffer composition	buffer (annotation)	buffer composition
buffer 1 (HTLV-1 1-116 assay)[1]	0.5 M NaCl 50 mM Na-acetate, pH 5.5 5 mM DTT	**buffer 6** (crystallization)[1]	0.1 M Na-acetate, pH 5.2 17 % PEG 8000 10 % PEG 300 10 mM DTT
buffer 2 (buffer 1 without DTT)	0.5 M NaCl 50 mM Na-acetate, pH 5.5	**buffer 7** (HTLV-1 PR assay)[3]	0.4 M citric acid, pH 5.3 2.1 mM NaCl 10.4 mM EDTA 12.5 % gylcerol 2.1 mM DTT
buffer 3 (HTLV-1 PR refolding)[3]	20 mM PIPES, pH 7.0 1 mM EDTA 150 mM NaCl 10 % glycerol 2 mM DTT	**buffer 8** (HIV-1 PR assay)	100 mM MES, pH 5.5 300 mM KCl 5 mM EDTA
buffer 4 (HTLV-1 PR 1-116 refolding)[2]	15 mM Na-acetate, pH 3.0 5 % PEG 300 5 mM DTT	**buffer 9** (endothiapepsin assay)	0.1 M Na-acetate, pH 4.6 0.01 % Tween20
buffer 5 (HTLV-1 PR 1-116 SP Sepharose)[1]	20 mM Na-acetate, pH 3.0 6 M urea 5 mM EDTA 5 mM DTT 0.3 M NaCl	water	

2.3.2 HTLV-1 PR 1-125

TSA measurements for the full-length protein were performed in numerous different buffers (Table 2.3), and in presence of a variety of inhibitors as tool compounds (pepstatin, NK102, NK192, NK232, (see appendix)) using different protein concentrations. Also attempts were made using other protein buffers (buffer exchange after refolding of the full-length protein immediately before the TSA measurements into water or sodium acetate buffer (50 mM, pH 5.0, containing 1 mM EDTA)).

Unfortunately, none of these tested conditions resulted in evaluable TSA curves, which led to the termination of the establishment efforts towards a TSA setup for the full-length protease as well. Similarly to other target proteins, not all biophysical methods are necessarily suitable to study particular proteins of interest, hence, the investigation of ligand binding towards HTLV-1 PR was pursued using other methods.

Table 2.3. Different buffers used in the TSA of the HTLV-1 PR 1-125.

buffer (annotation)	buffer composition
buffer 8 (HIV-1 PR assay)	see Table 2.2.
buffer 9 (endothiapepsin assay)	see Table 2.2.
buffer 10 (HTLV-1 PR 1-125 refolding buffer 2)[7]	50 mM Na-acetate, pH 5.0 100 mM NaCl
buffer 11 (HTLV-1 PR 1-125 refolding buffer 3)[9]	20 mM PIPES, pH 7.0 150 mM NaCl 10 % glycerol 1 mM EDTA 2 mM DTT 0.5 % Nonidet P-40
buffer 12 (HTLV-1 PR 1-125 fluorescence assay)[18]	250 mM potassium hydrogen phosphate, pH 5.6 5 % glycerol 1 mM EDTA 5 mM DTT 500 mM NaCl

2.3.3 Conclusion and Outlook

Unfortunately, all attempts to establish the thermal shift assay for the HTLV-1 protease did not succeed after testing many various conditions for both protease constructs (full-length protein and HTLV-1 PR 1-116). As subsequently the full-length protein showed activity in the fluorescence-based assay and the shortened construct could successfully be used for crystallization of the protease, hence confirming reasonably folded protein preparations for both constructs, these negative TSA results are, retrospectively, rather surprising. One possible explanation could be that the refolding rate is too low and that the non-properly folded amounts of protein thus disturbs the detection of the unfolding process. Additionally, the used fluorescence dye (SPYRO Orange) might not be compatible with the protein. The usage of other fluorescence dyes, like for example 1.8-ANS (1-anilino-8-naphthalene sulfonate), bis-ANS, nile red or dapoxyl sulfonic acid (see chapter 7.2) could perhaps enable TSA measurements. However, it is also described in literature that the terminal β-sheets of the HTLV-1 PR are comparatively hydrophobic and exposed to the solvent, while in the HIV-1 PR for example theses residues are hydrophilic and might interact with the solvent.[7] Thus, maybe the resulting and also experimentally proven high aggregation tendency of the HTLV-1 PR as well as the surface characteristics might also impede TSA measurements in the folded form.

2.3.4 Experimental Section

General TSA Protocol

TSA measurements were performed using the iCycler IQ5 Real Time Detection System (Bio-Rad). For the assay 96-well plates (Multiplate™ Low-Profile 96-well unskirted PCR Plates, BioRad) were used. Unless specified otherwise, the total volume of the well solution was 40 µL and the plate was heated from 20°C to 80°C with a heating rate of 0.5°C/min. SYPRO Orange (protein gel stain 5000x concentration in DMSO, Invitrogen) was used as fluorescence dye.

Detailed TSA Procedures

pH-Screen I

12 different buffer solutions (see below) covering pH-values from 4.0 - 9.5 were used. 2 µl of each buffer solution (1 M) were added to 38 µl of the water-protein-SYPRO Orange-mixture (800 µl water, 1.7 µl SYPRO Orange, 22 µl HTLV-1 PR 1-116 before refolding (10 mg/ml). For comparison, the screen was also performed with the protein being denatured by heating.

1	1 M sodium acetate, pH 4.0	**7**	1 M HEPES, pH 7.0
2	1 M sodium acetate, pH 4.5	**8**	1 M HEPES, pH 7.5
3	1 M sodium acetate, pH 5.0	**9**	1 M Tris, pH 8.0
4	1 M MES, pH 5.5	**10**	1 M Tris, pH 8.5
5	1 M MES, pH 6.0	**11**	1 M CHES, pH 9.0
6	1 M MES, pH 6.5	**12**	1 M CHES, pH 9.5

pH-Screen II (Ericsson)

The pH-screen was similarly performed as described in the publication by Ericsson *et al*.[16] In total, 23 different buffer conditions were tested (see below). The well solution consisted of 19 µl water-SYPRO Orange solution (water with 0.2 % SYPRO Orange), 19 µl twofold buffer solution (200 mM), and 1.5 µl protein solution (HTLV-1 PR 1-116, 10 mg/ml).

1	Sodium acetate, pH 4.5	**13**	Potassium phosphate, pH 7.0
2	Sodium citrate, pH 4.7	**14**	HEPES, pH 7.0
3	Sodium acetate, pH 5.0	**15**	Ammonium acetate, pH 7.3
4	Potassium phosphate, pH 5.0	**16**	Sodium phosphate, pH 7.5
5	Sodium phosphate, pH 5.5	**17**	Tris, pH 7.5
6	Sodium citrate, pH 5.5	**18**	Imidazole, pH 8.0
7	MES, pH 5.8	**19**	HEPES, pH 8.0
8	Potassium phosphate, pH 6.0	**20**	Tris, pH 8.0
9	MES, pH 6.2	**21**	BICINE, pH 8.0
10	Sodium phosphate, pH 6.5	**22**	Tris, pH 8.0
11	Sodium cacodylate, pH 6.5	**23**	BICINE, pH 9.0
12	MES, pH 6.5		

PACT buffer screen

The PACT buffer TSA screen was performed according to a published protocol:[17] to a TBS (Tris-buffered saline) buffer pH 7.5, 35 µl HTLV-1 PR 1-116 (10 mg/ml) as well as 2 µl SYPRO Orange were added to a final volume of 2 ml. 19 µl of this mixture were pipetted in each well of the 96-well plate, followed by the addition of 19 µl of the respective PACT screen solution (Quiagen). This procedure was also repeated using the sodium acetate buffer pH 5.5 (final concentration: 50 mM) as protein solution, based on buffer 1, instead of the TBS buffer.

2.4 Fluorescence-Based Assay

2.4.1 Assay Establishment and Validation

For the establishment of a fluorescence-based assay aiming to further characterize inhibitors affinity-wise, different substrates were tested, all based on the fluorescence resonance energy transfer (FRET), which is also referred to as Förster resonance energy transfer.

FRET describes the emission-light-free energy transfer of a fluorophore (donor) to a quencher molecule (acceptor). Therefore, the emission spectrum of the donor must overlap with the absorption spectrum of the acceptor. For this purpose so-called FRET-substrates have a harmonized fluorophore- and quencher molecule in close proximity (each at one site of the scissile bond). The excitation of the fluorophore does not result in a significant fluorescence signal if the fluorophore and quencher molecule are in close proximity. Caused by the catalytic activity of the enzyme, the cleavage of the substrate induces a separation of the fluorophore from the quencher molecule and results thereby in an increase in fluorescence. Under some conditions the fluorescence

can also be quenched by the quencher group of neighboring cleavage products or substrate molecules, which results in a lower fluorescence signal not reflecting the whole amount of the cleaved substrate (inner filter effect). The magnitude of the inner filter effect is dependent on the substrate concentration, the wavelength range, and the path length, and therefore might vary in dependence of the instrumental configuration.[19]

Within this thesis, in total three different substrates were tested using the full-length protein: the HIV-1 PR substrate (substrate A, purchased from Bachem), and two substrates derived from the natural HTLV-1 CA/NC cleavage site, possessing the Abz/*p*-nitro-Phe (substrate B) or the EDANS/DABCYL (substrate C)[18] as fluorophore/quencher pair. Substrate B and C were kindly synthesized by Dr. Kornelia Hardes (AK Steinmetzer, Philipps-Universität Marburg), substrate C was also reordered from Bachem (Germany). ↓ denotes the cleavage site.

Substrate A	Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH_2
HTLV-1 CA/NC cleavage site	Lys-Thr-Lys-Val-Leu↓Val-Val-Gln-Pro-Lys
Substrate B	Abz-Lys-Thr-Lys-Val-Leu↓Val-Val-Gln-Pro-Lys-*p*-nitro-Phe-NH_2
Substrate C[18]	Arg-Glu-(EDANS)-Thr-Lys-Val-Leu↓Val-Val-Gnl-Pro-Lys(DABCYL)-Arg-NH_2

As expected, the full-length HTLV-1 PR did not cleave the HIV-1 PR substrate (substrate A), which underlines the known differences in the substrate specificity of both related proteases. Substrate B, an anthraniloyl-substrate derived from the natural HTLV-1 cleavage site, is cleaved by the protease, but the slopes of the curves are relatively low. However, with the EDANS/DABCYL HTLV-1 substrate (substrate C) the obtained results enabled the determination of K_m-values and kinetic inhibitor characterization.

In addition to the full-length protein, also measurements with the shortened construct (HTLV-1 PR 1-116) were performed using substrate C. These measurements revealed that the shortened protease construct possesses to some extent catalytic activity, however, higher protein concentrations are required (ca. factor 20) to obtain comparable slopes of the recorded curve in comparison to the full-length protein. Therefore, the full-length protein was subsequently used for kinetic measurements and affinity determinations via the fluorescence-based assay.

For validation of the K_i-value determination via the fluorescence-based assay, the HIV-1 PR inhibitor indinavir was used as reference ligand. A K_i-value against HTLV-1 PR of 3.48 µM was reported in literature,[18] which is in good agreement with the K_i-

value determined in our laboratory (2.5 ± 0.7 µM, average of in total 7 measurements (chapter 4 and 5)). The following assay setup was used for affinity determination of HTLV-1 PR inhibitors which is based on the protocol from Bagossi *et al.*:[18]

The fluorescence-based assay was performed in 96-well plates using a Tecan Safire II plate reader. Each well contained a reaction mixture consisting of 186 µL assay buffer, 2 µL substrate C (500 µM/300 µM (concentration is specified in the corresponding chapter of this thesis)), 2 µL DMSO/inhibitor. Immediately before the measurement, 10 µL HTLV-1 PR 1-125 (≈ 24 µg/ml, diluted with the protein buffer (buffer I, Table 2.1)) were added to start the substrate cleavage. The assay buffer consisted of 250 mM potassium hydrogen phosphate, pH 5.6, 5 % glycerol, 1 mM EDTA, 5 mM DTT and 500 mM sodium chloride. The increase of fluorescence was detected at 490 nm after excitation at 340 nm. IC_{50}-values as well as percent inhibition values were determined at a substrate concentration of 5 µM, for K_m-determination 12 different substrate concentrations were used (50 µM - 24 nM). From the increase in fluorescence, the IC_{50}-values were calculated with the program GraFit 4.09 (Erithacus Software). The corresponding K_i-values were calculated by means of the Cheng-Prusoff-equation[20] using the determined K_m-value (Lineweaver-Burk) and the substrate concentration.

2.4.2 Conclusion

Within this thesis, a fluorescence-based assay procedure could be successfully established enabling the affinity estimation by means of either percent inhibition values or K_i-value determination of small molecule inhibitors to the HTLV-1 PR. Testing different substrates, the EDANS/DABCYL substrate (substrate C) was found to be the most suitable one. The assay was validated for the full-length protein using indinavir as reference. Also the shortened protease construct HTLV-1 PR 1-116 exhibited catalytic activity in our hands, however, due to the enhanced activity of the full-length construct, the latter was chosen for the enzymatic assay and inhibitor characterization.

2.5 Crystallization

Within this thesis, elucidation of the binding modes of small molecule inhibitors to the HTLV-1 PR was of great interest. To achieve this, extensive efforts were undertaken focusing on the crystallization of the shortened construct HTLV-1 PR 1-116 that in the past had already been successfully used for crystallization; however, also attempts with the catalytically more active HTLV-1 PR 1-125 were performed.

2.5.1 HTLV-1 PR 1-116

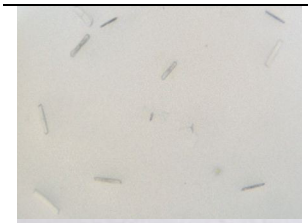
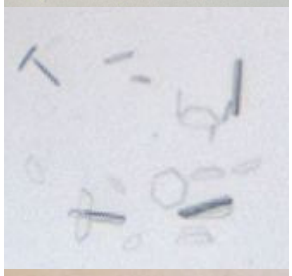
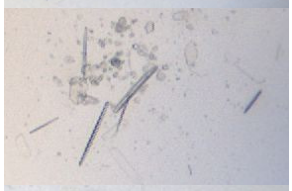
As it was expected that the C-terminal residues, which are truncated in this protease construct, are essential for the dimerization, the focus was laid on cocrystallization experiments in order to induce a stabilizing effect by the ligand towards the dimerization hence compensating for the loss of the C-terminal residues relevant for dimer formation.

Initial Crystallization Screen

As the identification of successful crystallization conditions for a protein does not follow any predictable rules to identify promising starting conditions and succeeds often by serendipity, we started the crystallization project with a robot screening at the Core Facility Structural Biology, Philipps-Universität Marburg: crystallization attempts were performed with two inhibitors, namely with pepstatin, a well-known and unselective aspartic protease inhibitor, and with the in-house developed HIV-1 PR inhibitor NK101 (Figure 2.7, appendix). In total, 1248 conditions were screened in 96-well plates by the sitting-drop method.

Fortunately, this initial screen yielded promising crystals in presence of NK101 as inhibitor in the protein solution at five different conditions, for pepstatin two conditions afforded crystal-like rods (Table 2.4). The crunch test, a first easily accessible indicator for protein crystals, was used to qualify the obtained crystals before the initiation of initial diffractions experiments. Measurements of the obtained crystals at the synchrotron (Bessy II, Berlin, Germany) confirmed that the crystals were indeed protein crystals, and for some HTLV-1 PR-NK101 crystals initial datasets with resolutions of 3.5 - 4 Å could be collected. Therefore, the identified conditions provided the basis for further optimization.

Table 2.4. Crystallization hits from the initial screen (HTLV-1 PR 1-116).

crystal photo	inhibitor	screen (Quiagen)	condition
	NK101	$AmSO_4$ Suite	0.2 M potassium sodium tartrate 2.2 M ammonium sulfate
	NK101	$AmSO_4$ Suite	0.1 M tri-sodium citrate, pH 4.0 1.6 M ammonium sulfate
	NK101	Cryos Suite	0.16 M magnesium acetate 0.08 M sodium cacodylate, pH 6.5 20 % glycerol 16 % polyethylene glycol 8,000
	NK101	JCSG Core Suite III	0.1 M MES, pH 6.0 2.4 M ammonium sulfate
	NK101	JCSG Core Suite III	0.1 M tri-sodium citrate, pH 5.0 1 M lithium chloride
	pepstatin	JCSG Core Suite II	0.1 M tri-sodium citrate, pH 5.0 2.4 M ammonium sulfate
	pepstatin	JCSG Core Suite II	0.07 M Na-acetate, pH 4.6 0.14 M calcium chloride 30 % glycerol 14 % 2-propanol

Reproduction and Optimization

Attempts to reproduce these hits manually did not succeed which prompted us to reproduce these initial conditions at the same screening facility again. Indeed, crystal growth for drops containing NK101 was again observed, however, at different conditions (Table 2.5)

Table 2.5. Crystallization screens and successful conditions used for the second robot screen for the HTLV-1 PR 1-116.

screen (Quiagen)	condition
$AmSO_4$ Suite	0.2 M ammonium nitrate 2.2 M ammonium sulfate
$AmSO_4$ Suite	0.2 M potassium iodide 2.2 M ammonium sulfate
$AmSO_4$ Suite	0.1 M Tris, pH 8.5 1.5 M ammonium sulfate 15 % (v/v) glycerol
JCSG Core Suite II	0.14 M calcium chloride 0.07 M sodium acetate, pH 4.6 30 % (v/v) glycerol 14 % (v/v) 2-propanol

The comparison of conditions which yielded crystals of the HTLV-1 PR showed that most conditions contain ammonium sulfate (1.6 - 2.4 mM) as precipitant.

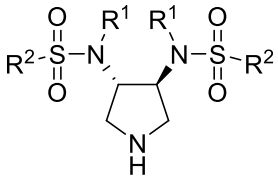
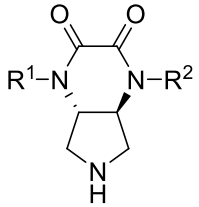
Further attempts to reproduce the screening hits by varying the pH-value, PEG, glycerol, and salt concentration of the hit conditions, using sitting as well as hanging drops, with and without seeding, or trying to crystallize HTLV-1 PR in presence of other inhibitors did not succeed. All these attempts were performed with a drop volume of 2 or 4 µl, containing reservoir/protein-inhibitor solution at a ratio of 1:1.

Further crystallization attempts comprised the variation of the ratio of reservoir solution to protein-inhibitor solution, again using grids around the found conditions varying the pH, the precipitant concentrations as well as the crystallization temperature (291K as well as 277K). Surprisingly, crystals appeared in some drops only containing the protein-inhibitor solution in absence of the reservoir solution. The thereby obtained crystals of HTLV-1 PR in complex with NK101 finally resulted in an X-ray structure with a resolution of 3.15 Å (chapter 5). The following crystallization conditions were successful for HTLV-1 PR-NK101 crystals, which grew in hanging drops at 291K, using 1 ml reservoir solution (0.1 M Tris, pH 7.5, 1.5 M ammonium sulfate, 20 % glycerol)

and a drop volume of 1 µl, containing 5 mM inhibitor.

Using a grid screen around the successfully determined ammonium sulfate conditions varying the ammonium sulfate as well as the glycerol concentration, and pH-values, a set of 21 additional HTLV-1 PR inhibitors were tried to crystallize by cocrystallization (Table 2.6).

Table 2.6. Protease inhibitors used for crystallization trials of the HTLV-1 PR 1-116 using ammonium sulfate crystallization conditions.

inhibitor class	**inhibitors**
aspartic protease inhibitors	pepstatin, saquinavir, ritonavir, *des*-3-pyridylmethyl-indinavir
C_2-symmetric 3,4-disubstituted pyrrolidines*: 	AB69, AB83, AB84, AB91, AB101, AB111
pyrrolidine-based bicyclic inhibitors:** 	NK102, NK192, NK232, NK237, NK251, NK252, NK253, NK343 (one carbonyl group)
other 3,4 disubstituted pyrrolinde-based HIV-1 PR inhibitors:**	NK88, NK200

*) inhibitors were developed and synthesized as HIV-1 PR inhibitors by Dr. Andreas Blum.

**) inhibitors were developed and synthesized as HIV-1 PR inhibitors by Dr. Nina Klee.

Via these trials, X-ray structures of the pyrrolidine-based C_2-symmetric inhibitors AB83 and AB84 in complex with the HTLV-1 PR were obtained providing important insights into the SAR of this inhibitor series, which are discussed in detail in chapter 4.

Additionally, a screen in 96-well plate format was performed using $AmSO_4$ Suite, Cryos Suite, PACT Suite, and JCSG Core Suite III purchased from Quiagen in presence of the inhibitors indinavir and NK253. Crystals were obtained for both inhibitors: the protein crystals obtained in the drops containing NK253, however, did not result in an evaluable diffraction pattern at the synchrotron (Bessy II, Berlin, Germany), while for the HTLV-1 PR-indinavir crystals an X-ray structure with a resolution of 2.4 Å was obtained. The results are discussed in detail in chapter 3.

2.5.2 HTLV-1 PR 1-125

The effort of crystallizing the full-length protein was started using conditions which yielded crystals for the truncated protease construct including the successful ammonium sulfate conditions (0.1 M Tris, pH 7.5, 1.5 M ammonium sulfate, 20 % glycerol) containing only protein-inhibitor solution in the crystallization drop with indinavir, NK101, NK253 as inhibitors, and in absence of any inhibitor. As these attempts did not yield crystals for the full-length construct, again a screen at the Core Facility Structural Biology, was performed; for this screen a total of 1248 conditions (Table 2.7) with indinavir and NK253 were tested. Crystals which grew in this screening were evaluated by the crunch test, staining trials and at the synchrotron (Bessy II, Berlin, Germany), but did turn out not to consist of protein.

Table 2.7. Applied crystallization screens for the HTLV-1 PR 1-125.

screen (Quiagen)	crystallization temperature
$AmSO_4$ Suite	277K and 291K
Anions Suite	291K
Classics Suite	291K
Cryos Suite	291K
JCSG Core Suite I	277K and 291K
JCSG Core Suite II	277K and 291K
JCSG Core Suite III	277K and 291K
JCSG Core Suite IV	277K and 291K

2.5.3 Conclusion and Outlook

In conclusion, four novel X-ray structures of the HTLV-1 protease 1-116, being the first HTLV-1 PR structures with non-peptidic inhibitors (chapters 3, 4, and 5), could successfully be obtained. The HTLV-1 PR-indinavir complex was obtained directly from one screening hit, while the optimization of crystallization conditions found in the initial screening of the HTLV-1 PR resulted in X-ray structures in complex with NK101, AB83, and AB84. To the best of our knowledge, these results provided the first X-ray structure-supported SAR interpretation of small molecule HTLV-1 PR inhibitors described so far.

Unfortunately, despite various crystallization efforts for the full-length protein, no protein crystals were obtained within the screens performed during this thesis. In further

crystallization trials, variation of the crystallization parameters such as usage of different protein buffers or different ratio of protein- and reservoir solution in the crystallization drop could be beneficial. Additionally, the usage of another protein construct, e.g. the expression of the homodimer as a single-chain construct, while connecting both monomers by a linker, or, furthermore, the usage of inhibitors with other scaffolds or higher affinity could eventually be successful. As described at the beginning of this chapter, the full-length protein has a high aggregation tendency, which might explain the difficulties observed during the crystallization process.

2.6 Appendix

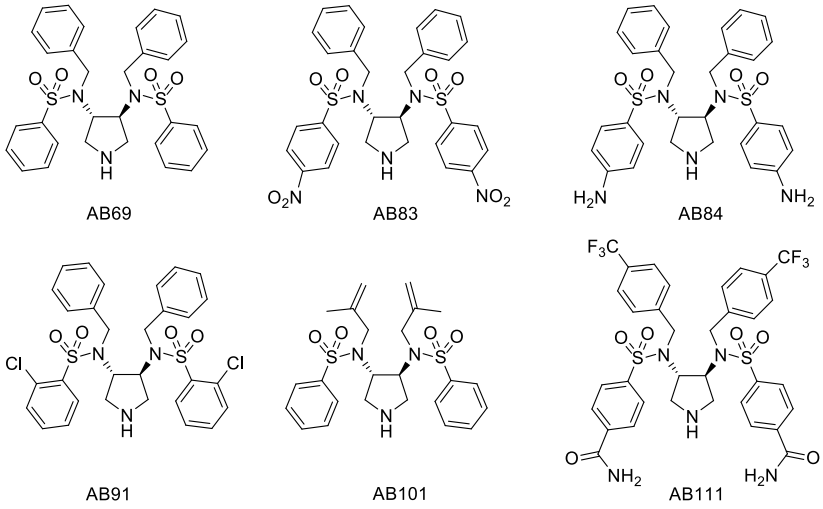


Figure 2.6. HIV-1 PR inhibitors used for HTLV-1 PR crystallization trials synthesized by Dr. Andreas Blum, AK Diederich, Philipps-Universität Marburg.

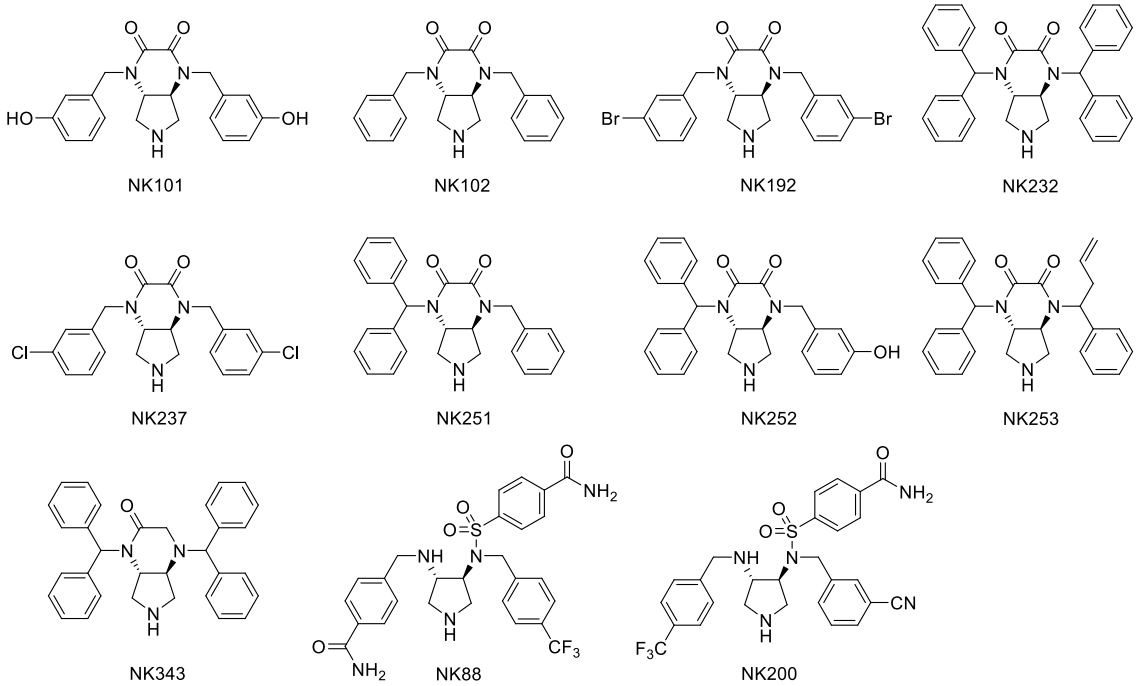


Figure 2.7. HIV-1 PR inhibitors used for HTLV-1 PR crystallization trials synthesized by Dr. Nina Klee, AK Diederich, Philipps-Universität Marburg.

2.7 References

(1) Li, M.; Laco, G. S.; Jaskolski, M.; Rozycki, J.; Alexandratos, J.; Wlodawer, A.; Gustchina, A. Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 18332–18337.

(2) Satoh, T.; Li, M.; Nguyen, J.-T.; Kiso, Y.; Gustchina, A.; Wlodawer, A. Crystal structures of inhibitor complexes of human T-cell leukemia virus (HTLV-1) protease. *J. Mol. Biol.* **2010,** *401*, 626–641.

(3) Kumada, H.-O.; Nguyen, J.-T.; Kakizawa, T.; Hidaka, K.; Kimura, T.; Hayashi, Y.; Kiso, Y. Development of [Ile^{40}]HTLV-I protease inhibition assay using novel fluorogenic and chromogenic substrate. *J. Pept. Sci.* **2011,** *17*, 569–575.

(4) Demir, A.; Oguariri, R. M.; Magis, A.; Ostrov, D. A.; Imamichi, T.; Dunn, B. M. Kinetic characterization of newly discovered inhibitors of various constructs of human T-cell leukemia virus-1 (HTLV-1) protease and their effect on HTLV-1-infected cells. *Antiviral Ther.* **2012,** *17*, 883–892.

(5) RCSB PDB available from: www.rcsb.org.

(6) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000,** *28*, 235–242.

(7) Kádas, J.; Boross, P.; Weber, I. T.; Bagossi, P.; Matúz, K.; Tözsér, J. C-terminal residues of mature human T-lymphotropic virus type 1 protease are critical for dimerization and catalytic activity. *Biochem. J.* **2008,** *416*, 357–364.

(8) Herger, B. E.; Mariani, V. L.; Dennison, K. J.; Shuker, S. B. The 10 C-terminal residues of HTLV-I protease are not necessary for enzymatic activity. *Biochem. Biophys. Res. Commun.* **2004,** *320*, 1306–1308.

(9) Li, C.; Li, X.; Lu, W. Total chemical synthesis of human T-cell leukemia virus type 1 protease via native chemical ligation. *Biopolymers* **2010,** *94*, 487–494.

(10) Louis, J. M.; Oroszlan, S.; Tözsér, J. Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. *J. Biol. Chem.* **1999,** *274*, 6660–6666.

(11) "The Nobel prize in chemistry 1993". Nobelprize.org. Nobel Media AB 2014. Available from: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1993/. [access:04.03.2015]

(12) Dingermann, T.; Winckler, T.; Zündorf, I. *Gentechnik, Biotechnik*; Wiss. Verl.-Ges.: Stuttgart, 2011.

(13) Instruction Manual: ArcticExpress competent cells and ArcticExpress (DE3) competent cells. Available from: http://www.genomics.agilent.com/literature.jsp?crumbAction=push&tabId=AG-PR-1088&contentType=User+Manual. [access: 03.03.2015].

(14) Clark. Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* **1998,** *9*, 157–163.

(15) Su, Z.; Lu, D.; Liu, Z. Refolding of inclusion body proteins from E. coli. In *Protein Purification:* Principles, High Resolution Methods, and Applications. Third Edition. Janson, J.-C., Ed.; John Wiley & Sons, Inc: Hoboken, NJ, USA; pp. 319–338.

(16) Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **2006,** *357*, 289–298.

(17) Phillips, K.; de la Peña, A. H. The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Curr. Protoc. Mol. Biol.* **2011,** *94*, 10.28.1-10.28.15.

(18) Bagossi, P.; Kádas, J.; Miklóssy, G.; Boross, P.; Weber, I. T.; Tözsér, J. Development of a microtiter plate fluorescent assay for inhibition studies on the HTLV-1 and HIV-1 proteinases. *J. Virol. Methods* **2004,** *119*, 87–93.

(19) Liu, Y.; Kati, W.; Chen, C. M.; Tripathi, R.; Molla, A.; Kohlbrenner, W. Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **1999,** *267*, 331–335.

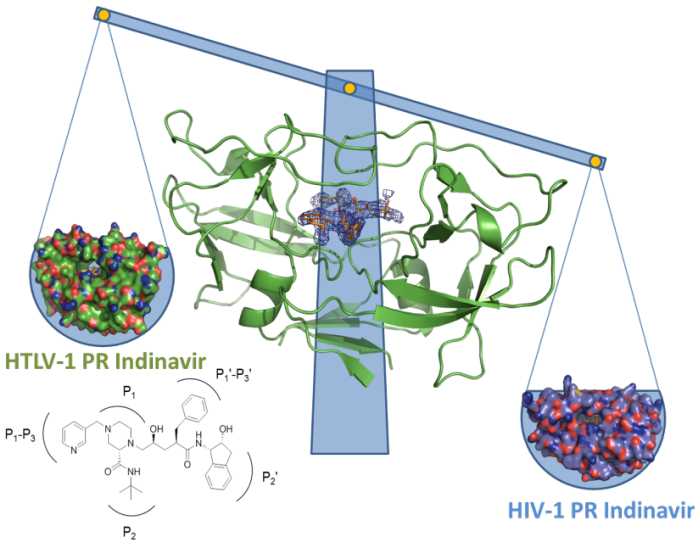
(20) Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973,** *22*, 3099–3108.

3. Structural Basis for HTLV-1 Protease Inhibition by the HIV-1 Protease Inhibitor Indinavir*

3.1 Introductory Remarks

The following chapter has been published in the *Journal of Medicinal Chemistry*: Kuhnert, M., Steuber, H., Diederich, W. E., *J. Med. Chem.* **2014**, 57(14), 6266-72.

The respective text is taken from the original publication, figures of the supporting information are implemented in the main text. The author of this thesis performed all experiments including the protein expression and purification, the kinetic assay, crystallization, synchrotron data collection and docking studies, the interpretation of the structure- and affinity data as well as drafting of the manuscript. This project was carried out in cooperation with Dr. Holger Steuber (LOEWE-Zentrum für Synthetische Mikrobiologie SYNMIKRO, Philipps-Universität Marburg), who refined the X-ray structure.



* Reproduced with permission from *J. Med. Chem.* **2014**, *57*, 6266–6272. Copyright 2014 American Chemical Society.

3.2 Abstract

HTLV-1 protease (HTLV-1 PR) is an aspartic protease which represents a promising drug target for the discovery of novel anti-HTLV-1 drugs. The X-ray structure of HTLV-1 PR in complex with the well-known and approved HIV-1 PR inhibitor Indinavir was determined at 2.40 Å resolution. In this contribution, we describe the first crystal structure in complex with a nonpeptidic inhibitor that accounts for rationalizing the rather moderate affinity of Indinavir against HTLV-1 PR and provides the basis for further structure-guided optimization strategies.

3.3 Introduction

Human T-cell leukemia virus type 1 (HTLV-1) is a retrovirus involved in several cancer-developing processes, neurological and dermatological diseases such as the adult T-cell leukemia (ATL), tropical spastic myelopathy (HAM), and HTLV-1-associated infective dermatitis (HAID).[1,2] Approximately 10−20 million people of the world's population are infected with the virus.[3] Because all currently available treatment options (cytotoxic chemotherapy, interferons, zidovudine) show only limited therapeutic efficiency,[1,4] the development of novel treatment alternatives is still of utmost importance. As the successful drug discovery programs against the human immunodeficiency virus-1 (HIV-1) protease as well as the hepatitis C virus NS3-4A protease have impressively shown, attacking the viral protease being essential for the maturation of the virus provides a promising therapeutic concept.[5,6] Thus, the development of new HTLV-1 PR inhibitors could overcome the still existing lack of a curative treatment against HTLV-1 infections.

The HTLV-1 PR is an aspartic protease and functions as a homodimer, each chain consisting of 125 residues.[1,7,8] The protease is closely related to the well-known HIV-1 PR, the latter representing a highly validated and exploited drug target. The strategy of developing protease inhibitors to attack the HIvirus has now been successfully pursued for almost 30 years, leading to hitherto 10 HIV protease inhibitors that are in clinical use.[5,9] While the overall sequence identity between both proteases is about 28 %, their binding regions are even more conserved (45 %).[10] Even though both proteases are very similar in their overall protein fold, they, however, substantially differ in their substrate specificity as well as inhibition profile.[10] Considerable differences were also observed with respect to the effect of known protease inhibitors: some of the highly potent HIV protease inhibitors, namely Saquinavir, Ritonavir, Nelfinavir, and Amprenavir, did not show any relevant affinity against the HTLV-1 protease, thus

disqualifying them for further lead discovery campaigns (K_i > 20 µM).[10,11] Interestingly, only the likewise approved HIV protease inhibitor Indinavir was able to inhibit the protease at low micromolar concentration (K_i = 3.5 µM).[11] Albeit its affinity against HTLV-1 PR is still strongly reduced in comparison to HIV-1 PR (K_i = 540 pM),[12] this moderate activity suggests Indinavir as putative starting point for the development of design hypotheses toward the first nonpeptidic HTLV-1 PR lead candidate. To pursue this approach on a rational basis, knowledge of the interaction geometry is indispensable, especially as the strong affinity deviation challenges a conserved binding mode of Indinavir toward both proteases. To answer this question and thereby provide the basis for lead design hypotheses, we determined the crystal structure of the clinically used, peptidomimetic HIV-1 PR inhibitor Indinavir in complex with HTLV-1 PR. In addition, founding on the results of this study, the influence of the pyridyl moiety of the inhibitor was investigated in more detail.

3.4 Results and Discussion

3.4.1 Structure Determination and Binding Mode of Indinavir to HTLV-1 PR

The X-ray structure of HTLV-1 PR in complex with the HIV-PR inhibitor Indinavir was determined at 2.40 Å resolution in the hexagonal space group $P6_322$ (Table 3.3, appendix), presenting a well-defined density for the inhibitor structure (Figure 3.1, left). The obtained crystal structure unambiguously confirms Indinavir with its clearly defined stereochemistry bound in the substrate binding pocket of HTLV-1 PR. As expected, the two catalytic aspartates Asp32 and Asp32' are directly addressed by the ligand's central hydroxyl group. As the core of the inhibitor resides in the S_1/S_1' pockets, the four side chain substituents of Indinavir are located in the S_2/S_2' and S_1-S_3/S_1-S_3' subpockets, respectively (Figure 3.1, right).

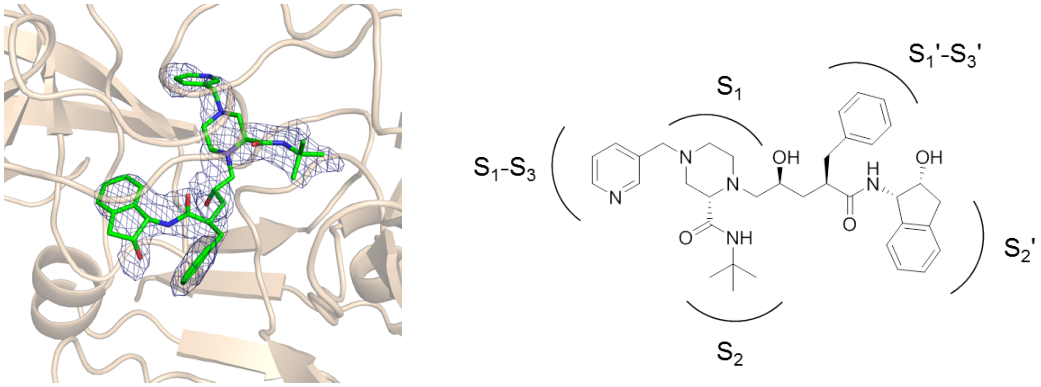


Figure 3.1. Left: the omit F_o-F_c electron density map contoured at 2.5σ (blue) for Indinavir (green) is well defined. HTLV-1 protease is shown as cartoon in wheat. Right: chemical structure of Indinavir and schematic representation of the binding mode.

While the polar piperazinyl and pyridyl moiety point into the hydrophilic S_1 and S_1-S_3 pockets, respectively, the rather hydrophobic S_2 and S_2' pockets are addressed by the *tert*-butyl and indanol substituents of Indinavir. Its benzyl group occupies the likewise hydrophobic surrounding of the S_1'-S_3' subpocket. The H-bond inventory established by Indinavir is shown in Figure 3.2.

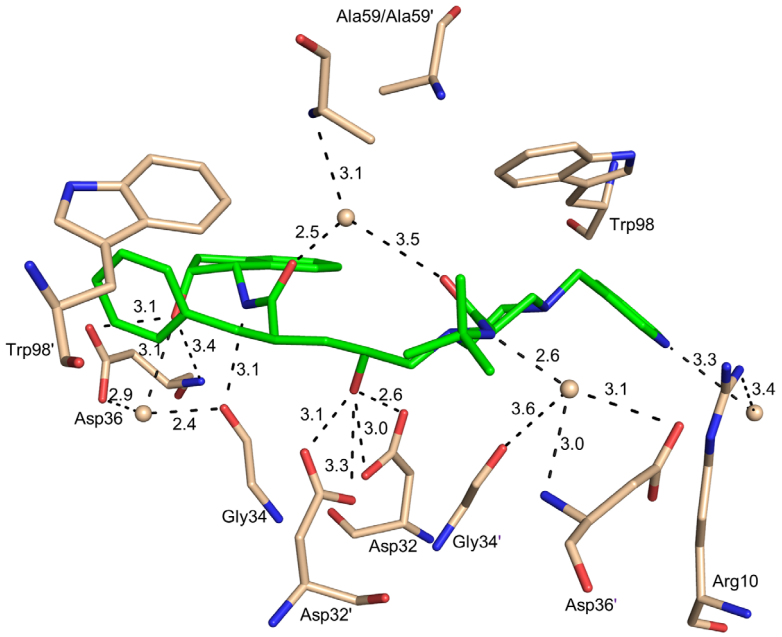
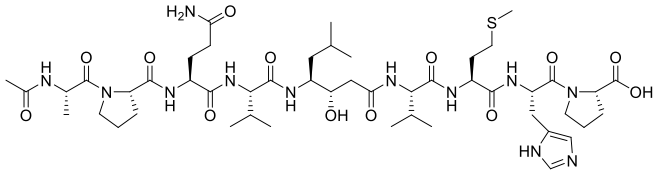


Figure 3.2. Observed hydrogen bonds between Indinavir (green) and active-site HTLV-1 PR residues (wheat). Distances are in Å, ligand and amino acids are displayed in stick representation, color-coded by atom types.

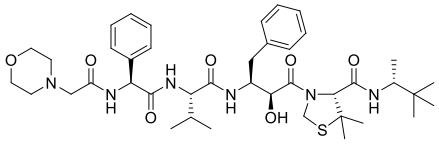
Except for the two nitrogen atoms of the piperazine moiety, all other ligand heteroatoms form H-bonds to the protease, either directly or water-mediated. For both protease monomers of the asymmetric unit, the *flaps* exhibit significant mobility, as indicated by elevated B-factors and a less well-defined electron density compared to the remaining protease. The interactions to the *flap* region are formed between two Indinavir carbonyl atoms and the Ala59 amide nitrogen via a water molecule, also known as "*flap*" water. The two tryptophanes 98 and 98' enable edge-to-face π−π interactions between the protein and the benzyl and pyridyl moiety of the inhibitor, which are oriented at a distance of 3.5 Å to Trp98' and 3.4 Å to Trp98, respectively. Even though the pyridyl-nitrogen of Indinavir is in distance to form a weak direct polar contact to Arg10 (3.4 Å), the geometry for such an interaction is rather unfavorable, while the presence of an interstitial water molecule mediates a polar contact between the ligand nitrogen and Arg10.

3.4.2 Comparison with Currently Known HTLV-1 PR X-Ray Structures

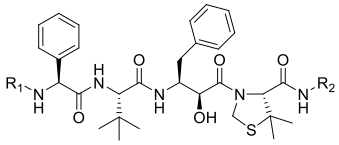
To date, seven HTLV-1 PR crystal structures with six different substrate analogues or peptide-like inhibitors have been published.[1,7] The first one represents a cocrystallized statine-based inhibitor, whereas all other ligands are so-called KNI-inhibitors, which share a hydroxyl-phenyl-butane-amide moiety at a similar position (Figure 3.3).



statine inhibitor (pdb: 2B7F, 3LIY)



KNI-10729 (pdb: 3LIX)



Inhibitor	R1	R2
KNI-10562 (pdb: 3LIN)		
KNI-10673 (pdb: 3LIQ)	H	
KNI-10681 (pdb: 3LIT)	H	
KNI-10683 (pdb:3LIV)	H	

Figure 3.3. Chemical structures of the statine- and KNI-inhibitors, of which crystal structures with the HTLV-1 PR exist.[1,7]

Comparable to Indinavir, all these inhibitors address the two catalytic aspartates 32 and 32' by their central hydroxyl functionality.

Because the statine-based inhibitor (Ac-Ala-Pro-Gln-Val-Sta-Val-Met-His-Pro) is considerably larger than all published KNI-inhibitors, it occupies significantly more space in the binding pocket, even reaching into the S_5/S_5' pocket. On the contrary, the smaller KNI-inhibitors only occupy the S_3-S_2' pocket (Figure 3.4).

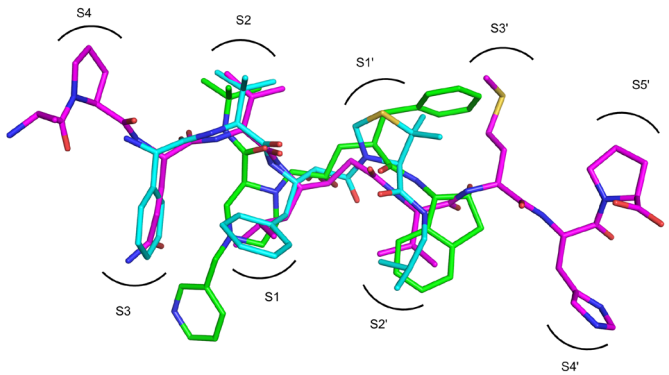


Figure 3.4. Occupancy of the HTLV-1 PR binding pocket. Superposition of the statine-based inhibitor (magenta), which occupies the S_5-S_5' pocket, with KNI-10673 (cyan) and Indinavir (green), which are smaller and occupy the S_3-S_2' and S_1-S_3-S_1'-S_3' pocket, respectively.

In comparison, the occupancy of the subpockets by the KNI-inhibitors as well as Indinavir is quite similar, except for the important fact that Indinavir addresses the S_3/S_3' pockets only partially. Interestingly, all inhibitors, including Indinavir, have a *tert*-butyl or *iso*-propyl moiety in common, either of which reside in the S_2 pocket. Indinavir's larger 3-pyridylmethyl-piperazinyl moiety reaches into the S_1-S_3 pocket, whereas for all other KNI-inhibitors, only the S_1 pocket is occupied by the smaller benzyl group. The S_1' pocket is addressed by the benzyl substituent of Indinavir instead of the KNI's dimethylthiazolidine residue, whereas the S_2' pocket hosts an indanol ring in place of the hydrophobic alkyl substituent of the KNI-inhibitors.

The *flap* region of the protease, namely Ala59 and Ala59', is typically addressed via a water molecule in all structures, which seems to be conserved at this position. The B-factors of the *flap* water molecules in the published X-ray structures range from 14 to 37 Å^2,[2,13] a similar value of 42 Å^2 is observed in our Indinavir complex. Considerable differences between these structures regarding the protein mainly concern the tryptophanes 98 and 98' that both exhibit pronounced mobility with respect to the orientation of the aromatic ring (Figure. 3.5a). In comparison to the larger statine-based inhibitor, the Trp98 and 98' in the HTLV-1-PR−Indinavir complex are flipped away around their χ2 angle by about 100° and 160°, respectively, for the two monomers.

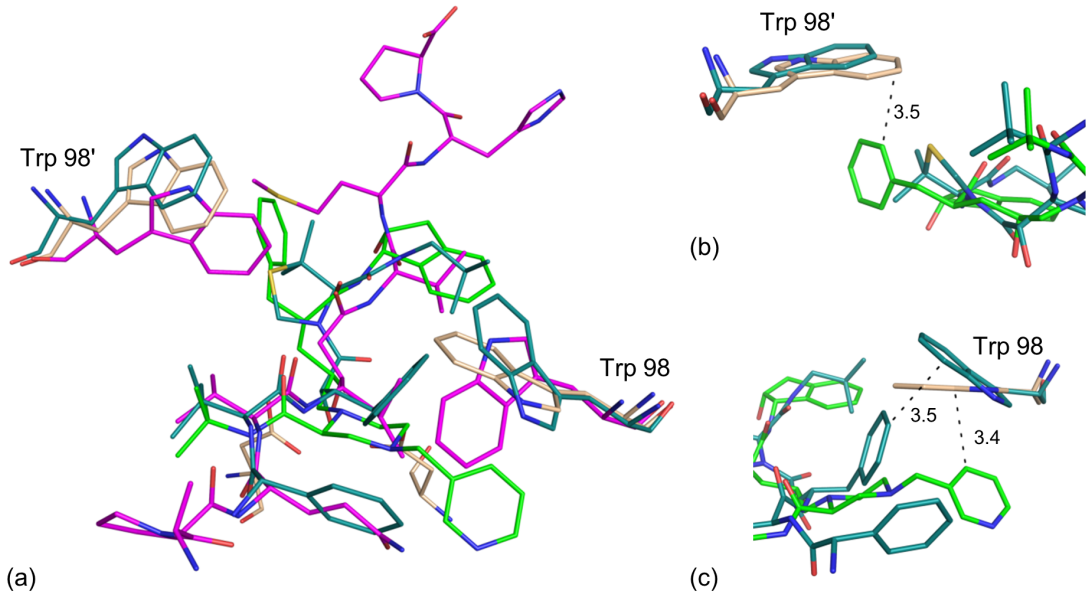


Figure 3.5. (a) Different conformations of Trp98/98' observed in the HTLV-1 PR (wheat: Indinavir structure, cyan: KNI-10673 (PDB: 3LIQ), magenta: statine inhibitor (PDB: 3LIY)) with the correspondent PR inhibitors (green: Indinavir). The tryptophanes show an inhibitor-dependent flexibility to avoid clashes with the inhibitors thus enabling π-π-interactions. (b) and (c) π-π-interactions of KNI-10673 and Indinavir to Trp98'/98, respectively. Distances are in Å. Ligands and amino acids are displayed in stick representation, color-coded by atom types.

Even though this tryptophan conformation observed in our complex resembles those found in most of the structures of the KNI-inhibitors (KNI-10673, KNI-10683, KNI-10562, KNI-10681), a slight adaptation by ~40° is still observed for the indole moiety of Trp98. Several reasons might rationalize this conformational change, two of which are discussed in the following. The first one concerns the avoidance of clashes with the protein. Indinavir's pyridyl and benzyl moiety, located in the S_1-S_3/S_1'-S_3' pockets, as well as the thiazolidine and benzyl residue of the KNI-inhibitors would result in clashes with the Trp98/Trp98', if these amino acids were to reside at the same position as observed in the statine-inhibitor complex structure. Thus, the observed adaptations are strongly required for inhibitor binding of KNI-type inhibitors as well as for Indinavir. Another important fact speaks for a changed conformation: the tryptophanes are turned in a way that π-π interactions are established. Edge-to-face stacking is observed between Indinavir's benzyl and pyridyl group and Trp98' and Trp98, respectively, and likewise between the benzyl moiety of the KNI-inhibitors and Trp98 (Figure 3.5b,c).

3.4.3 Comparison of Indinavir Binding to HTLV-1 PR and HIV-1 PR

After superposition of both retroviral proteases bound to Indinavir, the overall orientation of the inhibitor structure in the binding pocket is almost identical in both proteases. Regarding the inhibitor, the only pronounced difference is the relocation of the pyridyl ring (Figure 3.6).

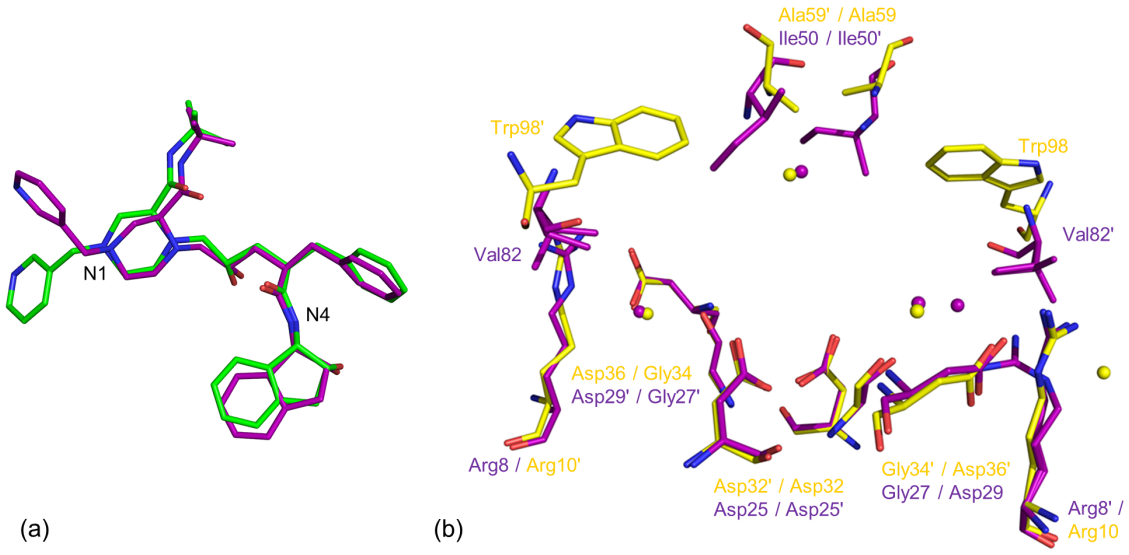


Figure 3.6. (a) Different conformations of Indinavir`s pyridyl ring in HTLV-1 PR, PDB: 3WSJ (green) and HIV-1 PR, PDB: 1SDT (purple). (b) Amino acids in the binding pocket which interact with Indinavir in HTLV-1 PR (yellow) und HIV-1 PR (purple), color-coded by atom types. Water molecules which are involved in interactions are shown as spheres.

Proteinwise, the overall location as well as orientation of most amino acids in the binding pocket of HIV-1 PR interacting with the inhibitor resembles very much that of HTLV-1 PR. The first slight structural alteration can be observed in the *flap* region: while in the case of HIV-1 PR, Ile50/50' are found there, in the HTLV-1 PR, the less space-demanding amino acids Ala59/59' are located at nearly the same position. A major difference can be determined in the so-called 80s loop, where in the HTLV-1 PR Trp98/98' are present, whereas in the HIV-1 PR, the smaller amino acids Val82/82' appear at this location. Nearly identical positions are observed for Asp32 and Asp25, Gly34 and Gly27, Asp36 and Asp29, as well as Arg10 and Arg8 in both the HTLV-1 PR and HIV-1 PR, respectively. In addition, slight differences are also observed concerning the water molecule pattern that directly mediates interactions between the protein and the inhibitor. While in the HIV-1 PR, four water molecules (including the *flap* water) are involved in H-bond interactions, in the HTLV-1 PR three of them mediate similar contacts. Interestingly, the arrangement of H-bond interactions between Indinavir and the protein are also very similar for both proteins (Figure 3.2 and Figure 3.7).

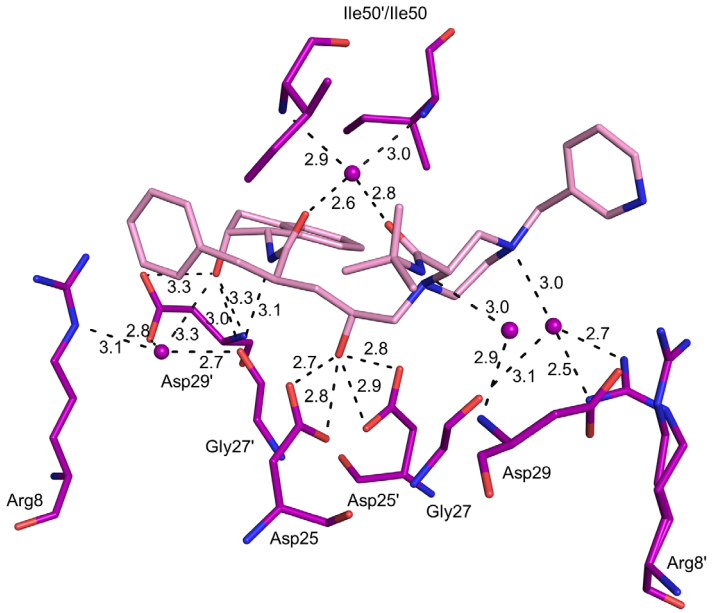


Figure 3.7. Hydrogen bond network of Indinavir bound to HIV-1 PR (PDB: 1SDT), color-coded by atom types. Involved water molecules are shown as spheres, distances are in Å.

The H-bonds between the N4 atom of Indinavir and Gly34/27' are nearly identical in both proteases, as are those to the catalytic aspartates, where only little differences regarding the geometry of the H-bonds are observed. In both proteases, the hydroxyl group of the indanol moiety forms an H-bond to Asp36/29'. Significant differences can be observed in the *flap* region concerning the interactions between the protein and the

pyridyl as well as the piperazine moiety of Indinavir: in the HIV-1 PR, both carbonyl atoms of Indinavir form H-bonds to the *flap* water, which at the same time addresses both Ile50/50' amide nitrogens of the *flap*. In contrast, the *flap* water of the HTLV-1 PR forms just one H-bond to one of the alanines, namely Ala59. In the HIV-1 PR, one of the nitrogens of the piperazine ring (N1) forms, mediated via a water molecule, an H-bond to the carbonyl oxygen of Gly27 and a charge-reinforced H-bond to the Arg8' side chain nitrogen. On the contrary, in the HTLV-1 PR complex, both piperazine nitrogens do not contribute to polar interactions, and the shift of this scaffold by about 1 Å toward the bottom compared to the HIV-1 PR complex prevents the insertion of a similar water molecule. The orientation of the pyridyl group of Indinavir is the most pronounced difference between the binding modes toward the respective targets. In the HTLV-1 PR, the pyridyl moiety forms a water-mediated H-bond of moderate strength to Arg10. Furthermore, a π−π interaction to Trp98 is observed that cannot be established in HIV-1 PR due to the absence of the Trp moiety. At a first glance, this pyridyl moiety might be detrimental to binding affinity, as in the HTLV-1 PR-bound conformation, it requires a rearrangement of Trp98 in HTLV-1 PR, thereby possibly reducing the affinity gain of the additionally established π−π interaction.

To investigate the above-mentioned influence of Indinavir's pyridyl substituent in more detail, we determined percentage inhibition values for Indinavir as well as for *des*-3-pyridylmethyl-Indinavir (DpmINV) (Figure 3.8) at 80, 40, and 8 µM in a fluorescence-based assay (Table 3.1). The percentage efficiency index (PEI) was calculated from the percentage inhibition values and the molecular weight.[14] In fact, the affinity of the *des*-pyridylmethyl-derivative is significantly reduced compared to Indinavir possessing the pyridyl moiety, which is also reflected in the higher PEI values for the latter one. Even though the pyridyl ring does not establish pronounced polar interactions, its presence in Indinavir obviously significantly contributes to the ligand affinity, although slight protein adaptations are required to enable Indinavir binding.

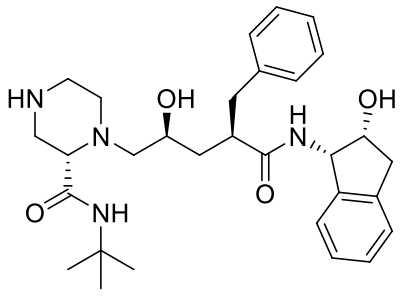


Figure 3.8. Chemical structure of *des*-3-pyridylmethyl-Indinavir (DpmINV).

Table 3.1. Percent inhibition values (%-inhib) for Indinavir (INV) and DpmINV at 80 µM, 40 µM, and 8 µM and the correspondent PEI values.

[inhibitor]	80 µM	40 µM	8 µM
INV			
%-inhib PEI	95.4 ± 0.8 1.55	91.2 ± 1.1 1.49	71.8 ± 3.7 1.17
DpmINV			
%-inhib PEI	60.8 ± 3.9 1.16	45.1 ± 5.2 0.86	16.8 ± 3.0 0.32

3.5 Implications for Further Lead Design

3.5.1 Structural Origin for Affinity Deviation

As the affinity of Indinavir against the proteases HIV-1 PR and HTLV-1 PR deviates strongly by about 4 orders of magnitude (K_i: 3.5 µM for HTLV-1 PR vs 540 pM for HIV-1 PR),[11,12] one naturally would assume significant differences in the interaction mode toward the two targets. Thus, the detection of those and the determination of the actual binding mode of Indinavir is a prerequisite for the development of novel design hypotheses toward improved leads. Surprisingly, as our contribution on the crystal structure determination of the HTLV-1 PR-Indinavir complex elucidates, the interaction patterns in both proteases at a first glance seem rather conserved. The most pronounced difference regarding the inhibitor concerns the orientation of its pyridyl moiety, thus enabling additional π-π stacking interactions to Trp98 in case of the HTLV-1 PR. The main difference proteinwise are the amino acids being involved in van der Waals interactions to the inhibitor in the S_1−S_3 pocket, in which Trp98/98' in the HTLV-1 PR is exchanged by the smaller Val82/82' in the HIV-1 PR. The required side chain relocation for Indinavir binding of the Trp moiety indeed leads to additional interactions, albeit this still appears to be entropically unfavorable. As the comparison between the crystal structures in complex with the statine-type ligand, the KNI-inhibitors, and Indinavir clearly demonstrates, Trp98/98' is able to adopt multiple conformations. Assuming that these persist as well in a ligand-free state, inhibitor binding will be associated with an entropic penalty. On the contrary, binding to HIV-1 PR should not be affected by this issue because of the less mobile Val residue and hence lower entropic costs.

As described above, our crystal structure provides insights into the changed inventory of affinity-relevant polar contacts (interaction to *flap* water in HTLV-1 PR only via one instead of two H-bonds in HIV-1 PR; loss of one H-bond of the piperazine moiety). In addition, the HTLV-1 PR-Indinavir complex enables for the first time a structure-guided interpretation of previously determined mutagenesis data, particularly as it provides experimental evidence of the exact positioning of the inhibitor in the HTLV-1 PR binding site. Interestingly, in HIV-1 PR, the mutation V82F, among others, was found to contribute to drug resistance during the treatment with Indinavir.[12,15,16] This mutation leads to a reduced susceptibility toward Indinavir and other PIs. Particularly, for Indinavir, the HIV-1 PR V82F mutant shows a reduced inhibitory activity by 2 orders of magnitude.[17] These results suggest that the incorporation of even larger aromatic moieties at this position such as Trp98 in HTLV-1 PR most likely would act even more detrimental on the binding affinity and thus helps to rationalize the loss in inhibitory potency by about 2 orders of magnitude. Unfortunately, the experimental validation of this mutational effect suffers from a strongly reduced catalytic activity of the HTLV-1 PR W98V mutant.[10]

Former studies of substrate and inhibitor specificity of HTLV-1 and HIV-1 PR suggest that the *flap* region, which differs between both proteases, also plays an important role in substrate specificity as well as for inhibitor susceptibility. This has been unambiguously demonstrated by HTLV-1 PR mutants (e.g., for the HTLV-1 PR triple mutant V56I, L57G, A59I), which particularly increase the affinity against HIV-PR inhibitors by about 1 order of magnitude.[10] This is presumably caused by the establishment of additional van der Waals contacts to the inhibitor facilitated by the exchange of the smaller Ala/Val side chain against the bulkier Ile, which consequently leads to more shape-defined subpockets. Even though it has to be stated that the above-mentioned mutational results do not necessarily cooperate in an additive manner,[18] deviations in the 80s loop (Trp98) and in the *flap* region (Val56, Leu57, Ala59) account for 2 and 1 orders of magnitude of the affinity deviation, respectively. The influence of these residues contributing essential van der Waals contacts to the inhibitor on the binding affinity seems to be more pronounced than the effect of the slightly changed network of polar interactions. This view is also supported by scoring the two Indinavir structures in complex with HIV-1 and HTLV-1 PRs with the scoring tool DSX.[19] Here, nonpolar atoms of Indinavir show a significantly more favorable contribution to binding in HIV-1 PR compared to HTLV-1 PR (Figure 3.9) and overall a significantly better scoring value (−245 for HIV-1 PR−Indinavir versus −172 units for HTLV-1 PR-Indinavir). In particular, the *tert*-butyl and the benzyl moieties of Indinavir allow a more efficient van der Waals

contact inventory to HIV-1 PR compared to HTLV-1 PR. The structural basis for this deviation consists in a larger S_2 pocket harboring the *tert*-butyl substituent (in HTLV-1 PR extended in particular by the Val56/Ala59 deviation to HIV-1PR Ile47/Ile50) and a larger S_3' pocket harboring the benzyl moiety (caused by a reorganized conformation of the 80s loop in which the space occupied by Val82 in HIV-1 PR is unoccupied in HTLV-1 PR due to the flipped Trp98 conformer). Caused by the Val56/Ile47 and Met37/Asp30 deviation in HTLV-1 PR vs HIV-1 PR, respectively, also the inhibitor's indanol moiety more favorably contributes to the overall binding affinity, as these replacements result in a narrower and more polar S_2' pocket in the HIV-1 PR (Figure 3.9).

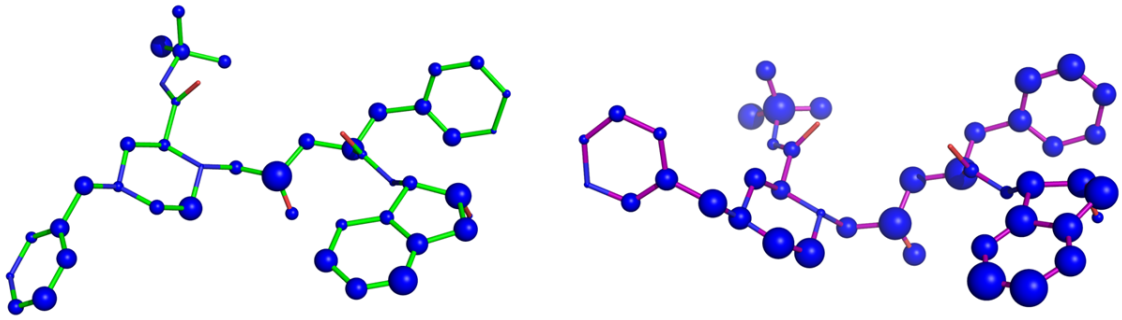


Figure 3.9. Visualization of favorable (blue spheres) interactions with DSX[19] for Indinavir in complex with HTLV-1 PR (PDB: 3WSJ, left) in green and HIV-PR (PDB: 1SDT, right) in purple.

3.5.2 Binding Properties of Further HIV-1 PR Inhibitors

We have selected Indinavir for our study as this HIV-1 PR inhibitor exhibited, among a set of experimentally tested marketed drugs, solely sufficient potency.[11] The cocrystal structure of HTLV-1 PR-Indinavir determined in this study raises the question whether this novel protein conformer helps to rationalize the lower potencies of other HIV-1 PR inhibitors. We have chosen Saquinavir, Ritonavir, Nelfinavir, and Amprenavir as representative examples to compare their binding geometries to HIV-1 PR with putative binding modes to HTLV-1 PR. For this purpose, two approaches were followed: first, we superimposed the corresponding HIV-1 PR-ligand complexes[20] with the novel HTLV-1 PR conformer and searched for suitable or detrimental features toward binding affinity. This analysis reveals for all four ligands significantly unfavorable interactions and close contacts (Figure 3.10), thus suggesting poor affinity for these inhibitors.

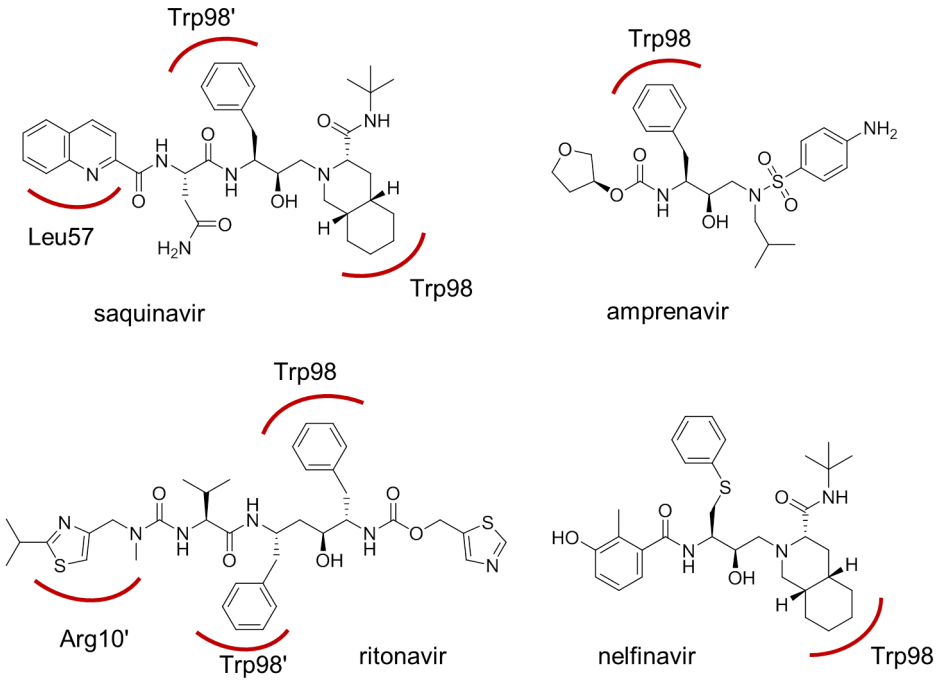


Figure 3.10. Clashes of the HIV-1 PR inhibitors Saquinavir, Amprenavir, Ritonavir, and Nelfinavir after superposition with the HTLV-1 PR (PDB: 3WSJ) are schematically shown with red curves.[20]

Second, as this approach would disqualify also Indinavir as inhibitor to HTLV-1 PR,[1] we next accounted for compensating adaptations in ligand geometry by docking the four ligands into the novel HTLV-1 PR conformer and compared the resulting DSX scores with those of the corresponding HIV-1 PR complexes (Table 3.2).

Table 3.2. Calculated dsx_csd scores[19] for the HIV-1 PR inhibitors Saquinavir, Nelfinavir, Ritonavir, and Amprenavir for the corresponding HIV-1 PR X-ray structures[20] and the best scored docking solution in HTLV-1 PR (PDB: 3WSJ).

	dsx_csd scores	
Inhibitor	**HIV-1 PR** X-ray	**HTLV-1 PR** docking in 3WSJ
Saquinavir	-215.362	-96.119
Nelfinavir	-207.444	-120.946
Ritonavir	-205.813	-135.306
Amprenavir	-236.789	-92.382

Here, all four ligands showed a significantly reduced (mostly about halved) score in HTLV-1 PR compared to HIV-1 PR, while the redocked Indinavir (rmsd 1.0 Å) virtually retains its original scoring value. In accordance with previous experimental data, these results rationalize that Indinavir is able to maintain moderate inhibitory potency to HTLV-1 PR after inducing some conformational changes while other HIV-1 PR drugs are not able to compensate for detrimental binding features in HTLV-1 PR.

3.5.3 Design Hypotheses for Next-Generation Indinavir Derivatives

In absence of any knowledge of Indinavir's binding geometry, rational design of potentially improved derivatives was challenging, and the strong affinity deviation aggravated those efforts even more, as the reduced affinity might easily have originated from a changed binding mode. The structural insights into the first nonpeptidic HTLV-1 PR inhibitor complex now enable the development of design hypotheses for further affinity improvement. The observed strong deviation in affinity of Indinavir toward the two analyzed retroviral proteases possessing an almost conserved polar contact inventory suggests that introduction of more suitable van der Waals contacts shall provide a key element for affinity improvement. The comparison of the visualized atom-by-atom scores reveals the indanol, the *tert*-butyl, and the benzyl moiety as promising sites for chemical variation toward larger substituents, as these moieties contribute less favorably to the van der Waals contact inventory in HTLV-1 PR compared to HIV-1 PR. First approaches for the design of next-generation Indinavir derivatives derived from our HTLV-1 structure are shown in Figure 3.11. The benzyl moiety for example could be replaced by larger aromatic residues (e.g., naphthyl derivatives) because of the bigger S_3' pocket of the HTLV-1 PR compared to the HIV-1 PR due to the absence of Val82 and different conformations of the 80s loop. Likewise the S_2 pocket is larger than in the HIV-1 PR, so that the exchange of the *tert*-butyl residue for other branched aliphatic side chain moieties might increase van der Waals contacts. We have already shown that the elimination of the pyridyl ring results in a significant decrease in affinity. On the basis of unoccupied space in the S_3 pocket, a more space-demanding residue thus should increase the affinity. In addition, the larger S_2' pocket suggests to explore the influence of halogen or methyl substitution at the indane moiety.

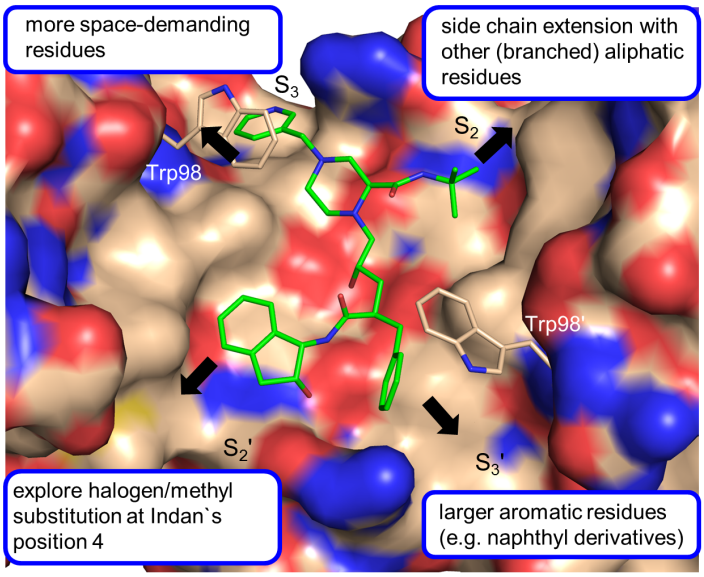


Figure 3.11. Design hypotheses for next-generation Indinavir derivatives. Wheat: HTLV-1 PR (PDB: 3WSJ) shown as surface [excluding the amino acids of the *flap*: 55-67/55'-67', 95-96, 95'-96']; green: Indinavir bound to the HTLV-1 PR shown in stick representation.

3.6 Conclusions

In this contribution, we present the first crystal structure of HTLV-1 PR in complex with Indinavir, an approved, nonpeptidic HIV-protease inhibitor in clinical use. Even though the binding modes of Indinavir in HTLV-1 and HIV-1 protease seem to be quite similar at a first glance, the affinity of this inhibitor against the HIV-1 PR is about 4 orders of magnitude higher than that toward HTLV-1 PR, which suggests more significant deviations with respect to the interaction network. While some conformational adaptations are required in HTLV-1 PR to enable Indinavir binding, which, however, might be penalized by unfavorable entropic costs, it has to be challenged whether these differences alone might sufficiently explain the affinity difference between the contended proteases. Nevertheless, the comparison of our crystal structure with the HIV-1 PR−Indinavir complex supports the interpretation of previous mutagenesis data accounting for contributions of a 2 and a 1 order of magnitude drop in affinity when considered in isolated mutant enzymes. In addition, we could show by visualization of scoring contributions that the *tert*-butyl, the indanol, and the benzyl moieties contribute less favorable contacts to HTLV-1 PR compared to HIV-1 PR, suggesting these moieties as candidates for variation. The strong deviation in affinity combined with an almost conserved inventory of polar interactions confirms a well-known paradigm that the establishment of suitable van der Waals interactions is of major importance, and their (e.g., mutational) modification shall significantly affect ligand affinity.[18]

In summary, among the broad structural variety of HIV-1 PR inhibitors that have been launched to the market, only Indinavir exhibits moderate affinity against HTLV-1 PR at low micromolar concentrations in biochemical assays. Albeit this residual affinity was sufficient for our successful crystallization studies, it, notwithstanding, has to be taken into account that different enzymes of different viruses are being compared here. Although the proteins show a similar fold, their sequence identity of 28 % overall and 46 % for the binding pocket also accounts for differences concerning protein dynamics as well as entropic and desolvation properties. In consequence, single amino acid variations not directly involved in inhibitor binding might additionally be important for inhibitor susceptibility. Nevertheless, the above presented novel HTLV-PR-Indinavir complex represents the first nonpeptidic inhibitor bound to this virus-obligatory enzyme and therefore now enables development of design hypotheses for a prospective rational drug design process toward the first nonpeptidic HTLV-1 PR lead.

3.7 Experimental Section

3.7.1 Protein Expression and Purification

For crystallization, a nine amino acids-truncated construct of the HTLV-1 PR (HTLV 1-116), which has also been utilized for all seven published HTLV-1 PR X-ray structures,[1,7] was used. The protease assay was performed with the full-length protein (HTLV 1-125). Both protein constructs harbor the mutation L40I to prevent autoproteolysis.[21]

The plasmid encoding the truncated protease as L40I and Y114N mutant was kindly provided by Dr. Alexander Wlodawer (NCI, Frederick/USA). To receive a plasmid only containing the L40I protease, Y114 was remutated to N by site-directed mutagenesis. The PCR product was sequenced (Eurofins, Germany) to confirm the mutation. The plasmid encoding the full-length protein was ordered from Life Technologies/GeneArt (Germany). The received plasmids were transformed into Rosetta 2 (DE3) cells. Cultures in LB-Media were induced with 1 mM IPTG (OD ≈ 0.6) for about 14 h (14 °C), cells were pelleted and resuspended in 10 mM Tris and 5 mM EDTA solution at pH 7.5. Cell disruption was performed with lysozyme and by sonification, the obtained inclusion bodies were washed with solutions of 0.5 M urea, 10 mM Tris and 5 mM EDTA.[1] The protein purification was performed as previously described,[7] however, using DTT instead of 2-mercaptoethanol.

The refolding conditions diverge for both protein constructs. The HTLV 1-116 protein was refolded via dialysis in one refolding step against a sodium acetate puffer as published[1,7] and stored at -80°C. The full-length protein (HTLV 1-125) was dialyzed in

three different buffers consecutively as described before:[22] 25 mM formic acid at pH 2.8 for 15 h, 50 mM sodium acetate at pH 5.0, 1 mM DTT and 1 mM EDTA for 12 h, and finally against a PIPES buffer at pH 7.0 consisting of 20 mM PIPES, 150 mM NaCl, 10 % Glycerol, 1 mM EDTA, 2 mM DTT, and 0.5 % Nonidet P-40 for 24 h. The protein was stored at -20°C.

3.7.2 Indinavir and *Des*-3-pyridylmethyl-Indinavir

The HIV-1 protease inhibitor Indinavir was purchased from Sigma Aldrich (Germany) as its sulfate salt with a purity of 98 % determined by HPLC. *Des*-3-pyridylmethyl-Indinavir was ordered from Toronto Research Chemicals Inc (TRC). The purity was determined by combustion analysis and quantitative ^{1}H NMR (qHNMR)[23] and accounts for 80 %. Elemental combustion analyses were performed on a vario MICRO cube (Elementar Analysensysteme GmbH, Hanau, Germany). The qHNMR was recorded on a Jeol ECX-400 spectrometer.

3.7.3 Protease Assay

Percentage inhibition values for Indinavir and *des*-3-pyridylmethyl-Indinavir were determined via a fluorescence assay. The assay buffer (250 mM potassium hydrogen phosphate, 5 % glycerol, 1 mM EDTA, 5 mM DTT, 500 mM NaCl, pH 5.6) and fluorescence substrate (RE(Edans)TKVLVVQPK(Dabsyl)R) were used as published,[11] however, the protocol was modified: each well contained a reaction mixture consisting of 186 µL assay buffer, 2 µL substrate (500 µM), 2 µL DMSO/inhibitor (8 mM/ 4 mM/ 0.8 mM). For *des*-3-pyridylmethyl-Indinavir 10 mM / 5 mM/ 1 mM stock solutions were used which corresponds to 8 mM/ 4 mM/ 0.8 mM inhibitor concentration taking the inhibitor's purity of about 80 % into account. The reaction was started with 10 µL protease HTLV 1-125 (≈ 24 µg/ml) per well. The increase of fluorescence was detected at 490 nm after excitation at 340 nm using a Tecan Safire II plate reader. Percent inhibition values and blanks were measured at least in triplicate at room temperature.

3.7.4 Crystallization

Crystals of the HTLV-1-PR-Indinavir complex were obtained from a HTLV-1 PR (1-116) solution at a concentration of 10 mg/ml in presence of 5 mM Indinavir sulfate using 10 % PEG 3000, 0.2 M lithium sulfate, and 0.1 M imidazole, pH 8.0 as reservoir solution. Crystals appeared after several days of incubation at 291 K.

3.7.5 Data Collection, Structure Determination and Refinement

Crystals of the HTLV-1-PR-Indinavir complex were harvested from mother liquor and frozen without cryoprotectant in liquid nitrogen. Data were collected at 100K at the Bessy II (Helmholtz-Zentrum Berlin, Germany), Beamline 14.1.[24] Data processing and scaling was performed using the program XDS[25] as implemented into the graphical user interface XDSAPP.[26] The coordinates of HTLV-1 PR as deposited under PDB access code 3LIX[7] were used as search model after removal of ligand and water molecules. Molecular replacement was carried out via Phaser[27] as implemented in Phenix.[28] Refinement was performed under repeated cycles of manual model building using Coot[29] and crystallographic refinement with the program phenix.refine. The final model was validated using PROCHECK.[30] Data collection and refinement statistics are shown in Table 3.3 (appendix).

3.7.6 Docking of HIV-1 PR Inhibitors

The docking of Amprenavir, Saquinavir, Ritonavir, and Nelfinavir in the HTLV-1 PR-Indinavir complex (PDB: 3WSJ) was performed with FlexX[31] (version 2.0.2.), after minimization of the ligand with MOE[32] (version 2012.10). In each case, the top-ranked solution was scored with DSX[19] using csd potentials and visualized with the DSX-visualization tool.

3.7.7 Figure Preparation

Figure 3.1, 3.2, 3.4, 3.5, 3.6, 3.7, 3.9, and 3.11 were prepared with PyMol.[33]

3.8 Appendix

Table 3.3. Data collection and refinement statistics.

Data Collection and Refinement Statistics (PDB entry 3WSJ)	
Data collection and Processing	
No. of crystals used	1
Wavelength [Å]	0.91841
Space group	P 6_3 2 2
Unit cell parameters	
a, b, c [Å]	77.3; 77.3; 159.3
α, β, γ [°]	90 ; 90 ; 120
Matthews coefficient [Å³/Da]	2.66
Solvent content [%]	53.8
Diffraction data	
Resolution range [Å]	42.0 – 2.40 (2.55-2.40)*
Unique reflections	11 561 (1 813)
$R(I)_{sym}$ [%]	11.7 (64.8)
Completeness [%]	99.3 (100.0)
Redundancy	12.5 (10.6)
I/σ(I)	20.4 (3.8)
Refinement	
Resolution range [Å]	41.6 – 2.40
Reflections used in refinement (work/free)	10 740 / 807
Final R values for all reflections (work/free) [%]	20.5 / 25.6
Protein residues	232
Inhibitor atoms	45
Water molecules	68
RMSDs	
Bonds [Å]	0.003
Angles [°]	0.704
Ramachandran plot	
Residues in most favoured regions [%]	89.9
Residues in additional allowed regions [%]	10.1
Residues in generously allowed regions [%]	-
Mean B factors [$Å^2$]	
Protein	47.9
Inhibitor	46.4
Water	43.3

* Numbers in parentheses characterize the highest resolution shell.

3.9 References

(1) Li, M.; Laco, G. S.; Jaskolski, M.; Rozycki, J.; Alexandratos, J.; Wlodawer, A.; Gustchina, A. Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 18332–18337.

(2) McGill, N.-K.; Vyas, J.; Shimauchi, T.; Tokura, Y.; Piguet, V. HTLV-1-associated infective dermatitis: updates on the pathogenesis. *Exp. Dermatol.* **2012,** *21*, 815–821.

(3) Edlich, R. F.; Arnette, J. A.; Williams, F. M. Global epidemic of human T-cell lymphotropic virus type-I (HTLV-I). *J. Emerg. Med.* **2000,** *18*, 109–119.

(4) Ishikawa, T. Current status of therapeutic approaches to adult T-cell leukemia. *Int. J. Hematol.* **2003,** *78*, 304–311.

(5) Fernández-Montero, J. V.; Barreiro, P.; Soriano, V. HIV protease inhibitors: recent clinical trials and recommendations on use. *Expert Opin. Pharmacother.* **2009,** *10*, 1615–1629.

(6) Schaefer, E. A K; Chung, R. T. Anti-hepatitis C virus drugs in development. *Gastroenterology* **2012,** *142*, 1340-1350.

(7) Satoh, T.; Li, M.; Nguyen, J.-T.; Kiso, Y.; Gustchina, A.; Wlodawer, A. Crystal structures of inhibitor complexes of human T-cell leukemia virus (HTLV-1) protease. *J. Mol. Biol.* **2010,** *401*, 626–641.

(8) Shuker, S. B.; Mariani, V. L.; Herger, B. E.; Dennison, K. J. Understanding HTLV-I protease. *Chem. Biol.* **2003,** *10*, 373–380.

(9) Mehellou, Y.; De Clercq, E. Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go? *J. Med. Chem.* **2010,** *53*, 521–538.

(10) Kádas, J.; Weber, I. T.; Bagossi, P.; Miklóssy, G.; Boross, P.; Oroszlan, S.; Tözsér, J. Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell leukemia virus type 1 protease. *J. Biol. Chem.* **2004,** *279*, 27148–27157.

(11) Bagossi, P.; Kádas, J.; Miklóssy, G.; Boross, P.; Weber, I. T.; Tözsér, J. Development of a microtiter plate fluorescent assay for inhibition studies on the HTLV-1 and HIV-1 proteinases. *J. Virol. Methods* **2004,** *119*, 87–93.

(12) Mahalingam, B.; Wang, Y.-F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004,** *271*, 1516–1524.

(13) Coordinates from the following PDB codes were used for the comparison of B-factors: 2B7F, 3LIN, 3LIV, 3LIX, 3LIT, 3LIQ, 3LIY.

(14) Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005,** *10*, 464–469.

(15) Sa-Filho, D. J.; Costa, L. J.; de Oliveira, C. F.; Guimarães, A. P. C.; Accetturi, C. A.; Tanuri, A.; Diaz, R. S. Analysis of the protease sequences of HIV-1 infected individuals after indinavir monotherapy. *J. Clin. Virol.* **2003,** *28*, 186–202.

(16) Rhee, S.-Y.; Taylor, J.; Fessel, W. J.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shirvani, V.; Zolopa, A.; Shafer, R. W. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob. Agents. Chemother.* **2010,** *54*, 4253–4261.

(17) Gulnik, S. V.; Suvorov, L. I.; Liu, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **1995,** *34*, 9282–9287.

(18) Ala, P. J.; Huston, E. E.; Klabe, R. M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y.; Hodge, C. N.; Chang, C. H. Molecular basis of HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with cyclic urea inhibitors. *Biochemistry* **1997,** *36*, 1573–1580.

(19) Neudert, G.; Klebe, G. DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J. Chem. Inf. Model.* **2011,** *51*, 2731–2745.

(20) Coordinates from the following PDB codes were used for the comparision: 3OXC, 3EKX, 3NU3, 1HXW.

(21) Louis, J. M.; Oroszlan, S.; Tözsér, J. Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. *J. Biol. Chem.* **1999,** *274*, 6660–6666.

(22) Li, C.; Li, X.; Lu, W. Total chemical synthesis of human T-cell leukemia virus type 1 protease via native chemical ligation. *Biopolymers* **2010,** *94*, 487–494.

(23) Pauli, G. F.; Gödecke, T.; Jaki, B. U.; Lankin, D. C. Quantitative ^{1}H NMR. Development and potential of an analytical method: an update. *J. Nat. Prod.* **2012,** *75*, 834–851.

(24) Mueller, U.; Darowski, N.; Fuchs, M. R.; Förster, R.; Hellmig, M.; Paithankar, K. S.; Pühringer, S.; Steffien, M.; Zocher, G.; Weiss, M. S. Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J. Synchrotron Radiat*. **2012,** *19*, 442–449.

(25) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 125–132.

(26) Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U. XDSAPP. A graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.* **2012,** *45*, 568-572.

(27) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40*, 658–674.

(28) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 213–221.

(29) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 486–501.

(30) Laskowski, R. A. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, 26, 283-291.

(31) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. **1996,** *261*, 470–489.

(32) Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.

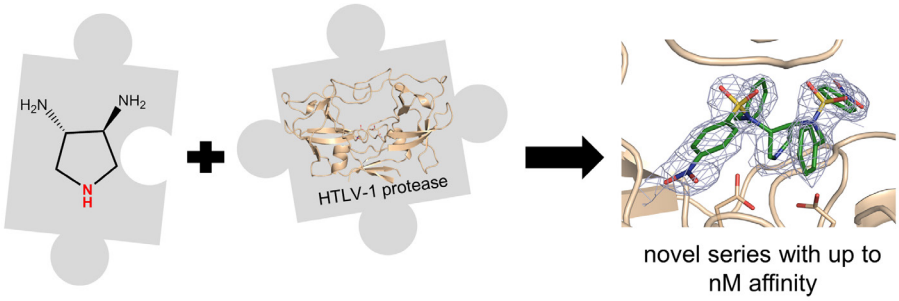
(33) DeLano, W. L.; The PyMol Molecular Graphics System, 0.99; DeLano Scientific: San Carlos, CA, **2002**.

4. Privileged Structures Meet Human T-Cell Leukemia Virus-1 (HTLV-1): C_2-Symmetric 3,4-Disubstituted Pyrrolidines as Non-Peptidic HTLV-1 Protease Inhibitors*

4.1 Introductory Remarks

The following text is accepted for publication in the *Journal of Medicinal Chemistry*: Kuhnert, M., Blum, A., Steuber, H., Diederich, W. E., *J. Med. Chem.* **2015**, in press.

The respective text is taken from the original manuscript, figures of the supporting information are implemented in the main text. The author of this thesis performed the expression and purification of the HTLV-1 protease, the kinetic assay, crystallization, synchrotron data collection, the interpretation of the structure- and affinity data as well as drafting of the manuscript. The set of evaluated inhibitors was synthesized by Dr. Andreas Blum within a structure-guided design project of HIV-1 protease inhibitors. The project of this chapter was carried out in cooperation with Dr. Holger Steuber (LOEWE-Zentrum für Synthetische Mikrobiologie SYNMIKRO, Philipps-Universität Marburg), who refined the HTLV-1 PR X-ray structures.



* Reproduced with permission from *J. Med. Chem.*, in press. Unpublished work copyright 2015 American Chemical Society.

4.2 Abstract

A series of 3,4-disubstituted pyrrolidines originally designed to inhibit the closely related HIV-1 protease was evaluated as privileged structure against HTLV-1 protease (HTLV-1 PR). The most potent inhibitor of this series exhibits two-digit nanomolar affinity and represents, to the best of our knowledge, the most potent non-peptidic inhibitor of HTLV-1 PR described so far. The X-ray structures of two representatives bound to HTLV-1 PR were determined and the structural basis of their affinity is discussed.

4.3 Introduction

Human T-cell leukemia virus type 1 (HTLV-1) was the first human retrovirus being isolated in the early 1980s.[1,2] Infections with HTLV-1 are associated with neurological and dermatological diseases as well as cancer-related processes.[3,4] Worldwide, approximately 10-20 million people are currently infected with HTLV-1.[5] Although the virus has been known for quite a long time, so far no causative means to attack the virus are known. All currently available treatment options to cure such an infection or rather the related diseases show only limited therapeutic efficacy.[3,6] One promising target to combat this therapeutic lack is the virus-encoded protease, an approach which in case of the related human immunodeficiency virus type 1 protease (HIV-1 PR) has led to a number of approved protease inhibitors within the last decades. The HTLV-1 protease (HTLV-1 PR), an aspartic protease which is essential for the maturation of the virus, represents a homodimer with each subunit consisting of 125 residues.[2] The protease is closely related to the thoroughly studied HIV-1 PR, which belongs to the A2 family of aspartic proteases.[7] Both proteins show a very similar fold and share a sequence identity within the binding region of about 45% (overall 28%).[8] However, it is well known that their substrate specificity as well as their inhibition profile differs remarkably from each other.[8] As a consequence, approved HIV-1 PR inhibitors show only a strongly reduced affinity against the HTLV-1 PR, thus hampering their clinical use. Hence, rational design as well as synthesis strategies directed towards the discovery of novel HTLV-1 PR inhibitors are strongly required to overcome the lack of treatment options in the future. We have recently described the complex of the HIV-1 PR inhibitor indinavir as the first peptidomimetic inhibitor bound to HTLV-1 PR,[9] which should facilitate the further structure-based optimization of this and related peptidomimetic inhibitors. By *in-silico* screening of about 140,000 compounds from the National Cancer Institute Developmental Therapeutics Program, thirteen structurally

diverse non-peptidic inhibitors of HTLV-1 PR with affinities ranging from two-digit to 0.8 µM were identified.[10] However, to the best of our knowledge, no structural data for any of these screening hits are available so far, which complicates the further optimization of these inhibitors.

In this contribution, we therefore embarked on a privileged structure approach[11,12] to identify novel HTLV-1 PR inhibitors and screened an in-house library of non-peptidic aspartic protease inhibitors that had originally been designed to target either members of the pepsin-like (A1) or the viral aspartic protease family (A2).[7,13–17] As common feature in all these inhibitors the two aspartates of the catalytic dyad are addressed via a most likely protonated amino functionality. Through decoration of this key element with appropriate side chains, selectivity towards particular family members could be achieved in previous studies.[13,14] The screening finally led to the identification of a novel series of HTLV-1 PR inhibitors based on a 3,4-disubstituted pyrrolidine core motif. X-ray structures for two representatives bound to the target protein were determined to elucidate their binding mode and to provide the structural basis for meaningful SAR interpretations.

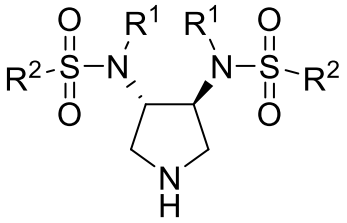
4.4 Results and Discussion

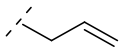
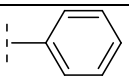
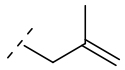
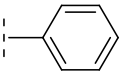
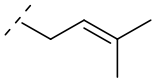
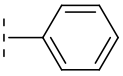
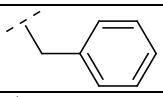
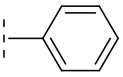
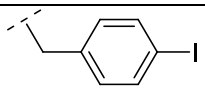
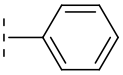
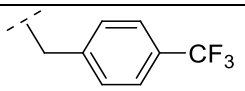
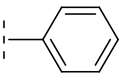
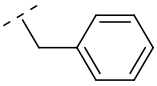
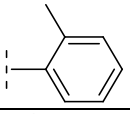
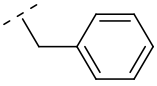
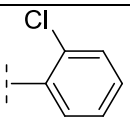
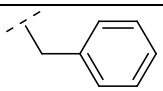
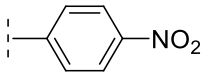
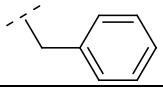
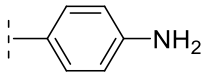
4.4.1 Biological Evaluation

The initial screening of a diverse subset of our in-house aspartic PR inhibitor library against HTLV-1 PR in a fluorescence-based assay at 100 µM inhibitor concentration revealed that, except for 3,4-substituted pyrrolidine derivatives, most of the compounds displayed only weak inhibitory activity (data not shown). Interestingly, some representatives of a series of C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines which had initially been developed as potent inhibitors against wildtype and Ile84Val mutant of HIV-1 PR[13,18] showed promising inhibitory activity against the HTLV-1 PR.

Based on these initial single-concentration screening results, K_i-values were determined for a congeneric series of ten compounds (Table 4.1).

Table 4.1. K_i-values of 1-10 towards HTLV-1 PR as well as HIV-1 PR.



	R^1	R^2	HTLV K_i [µM]	HIV K_i [µM]*
1			172.5 ± 16.3	12.3
2			17.8 ± 7.7	74.7
3			11.6 ± 1.9	1.57
4			2.0 ± 0.4	2.15
5			>200[a]	0.39
6			>250[a]	0.80
7			2.6 ± 0.1	0.67
8			2.4 ± 0.2	0.77
9			7.9 ± 1.7	1.72
10			0.015 ± 0.0004	0.27

* K_i-values taken from reference[13]

[a] K_i-value could not be determined accurately, due to inhibitor precipitation at high concentrations.

As observed by crystal structure analysis of HIV-1 and HTLV-1 PR-inhibitor complexes (see below), the R^1 substituents reside within the S_1/S_1' pockets, while the R^2 moieties populate the S_2/S_2' pockets. Six compounds of the selected library are altered in their P_1/P_1' substituent (R^1) while sharing a phenyl ring in P_2/P_2' position (R^2), whereas the remaining four compounds differ in the decoration of the P_2/P_2' position and are equipped with a benzyl moiety as invariant P_1/P_1' substituents. As expected, despite

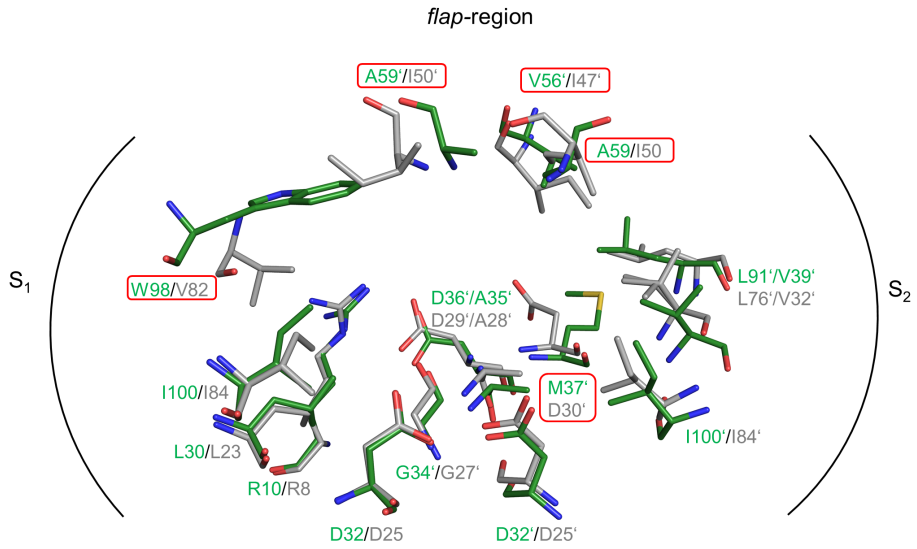


Figure 4.1. Superposition of the S_1 and S_2 pocket, as well as Arg10/Arg8, Asp36/29 and the catalytic dyad of HTLV-1 PR (green, PDB code: 4YDF) and HIV-1 PR (grey, PDB code: 2PWC) shows the similarity in the active site. Differences are indicated with red frames, amino acids are shown in stick representation, color coded by atom type.

some similarity in between the active sites of the HIV-1 and HTLV-1 PR (Figure 4.1), the inhibition profile against both proteases differs for the presented inhibitor series.

Within the first subset of inhibitors (**1**-**6**) being varied only in P_1/P_1' position, a steady increase in affinity from the allyl-substituted derivative (**1**) via sterically more demanding allylic substituents (**2** and **3**) to the benzyl-substituted derivative (**4**) can be observed, the latter representing the most potent derivative of this subset with an affinity of 2.0 µM. However, the introduction of hydrophobic residues in *para*-position of the benzyl moiety such as an iodo (**5**) or a CF_3 substituent (**6**) clearly reduces the affinity.

In the second subset of inhibitors, the P_2/P_2' position is altered: compounds **7** and **8** are decorated with a phenyl ring bearing hydrophobic *ortho* substituents; in **9** and **10**, the phenyl ring is substituted in *para*-position with a nitro or amino group, respectively. Neither the *ortho*-methyl (**7**) nor the *ortho*-chloro (**8**) substituent in P_2/P_2' position has any significant influence on the affinity against the HTLV-1 PR in comparison to the unsubstituted derivative **4** (2.6 and 2.4 µM, respectively, versus 2.0 µM). While the introduction of a *para*-nitro residue in P_2/P_2' position (**9**) decreases the affinity by a factor of 4, the presence of a *para*-amino moiety (**10**) results in a remarkable increase in affinity by about two orders of magnitude (2 µM versus 15 nM). Consequently, **10** not only represents the most potent compound within this inhibitor series, but, moreover, to the best of our knowledge, the most potent non-peptidic inhibitor of HTLV-1 PR described to date.

4.4.2 Binding Mode Analysis

Since knowledge of the binding mode of an inhibitor in the active site of a target protein is on the one hand a prerequisite for structure-based drug design and on the other hand facilitates the deduction of a meaningful SAR, we embarked on cocrystallization experiments and successfully determined the X-ray structures of **9** and **10** in complex with HTLV-1 protease at a resolution of 2.8 Å and 3.25 Å, respectively (Table 4.2, appendix). In both complex structures, the observed difference density allows an unambiguous placement of the ligand molecules (Figure 4.2a). However, the better resolved HTLV-1 PR-**9** complex offers a more comprehensive picture, since in the HTLV-1 PR-**10** complex the hairpin-like loops of the *flap* region are disordered.

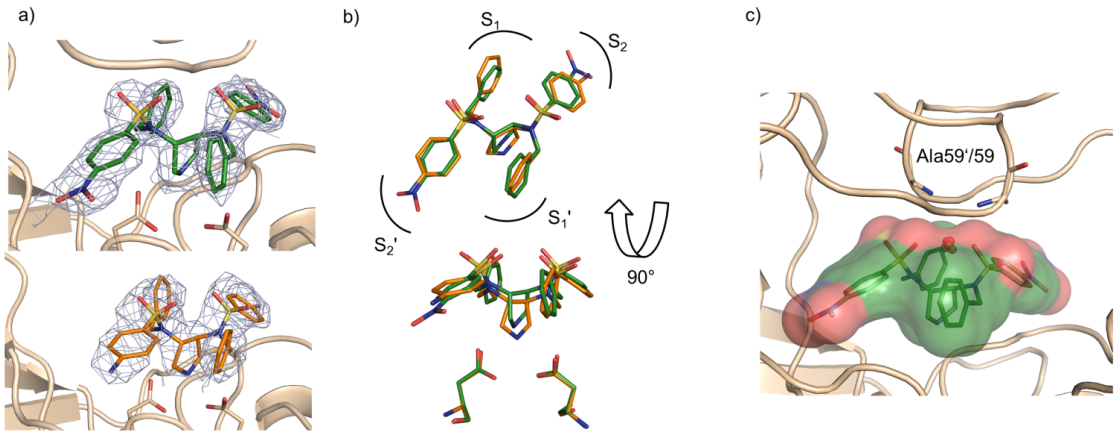


Figure 4.2. a) The $2F_o-F_c$ electron density maps for **9** (green) and **10** (orange) bound to the HTLV-1 PR contoured at 1.3σ (blue mesh) are well-defined. b) Superposition of **9** (green) and **10** (orange) bound to the HTLV-1 PR. c) The flap water[19] (red spheres) is replaced by the ligand`s sulfonyl-oxygens, here exemplarily shown for **9** (green, surface representation). Ligands are shown in stick representation, color-coded by atom types.

At first glance, the crystal structure analysis shows a quite similar binding mode for both inhibitors (Figure 4.2b): the two aspartates of the catalytic dyad are addressed by the endocyclic amine, the R^1 substituents reside in the S_1/S_1', while the R^2 moieties populate the S_2/S_2' pockets. Interestingly, **9** and **10** both adopt a slightly unsymmetrical binding mode within the active site.

However, a more detailed look into the structures yet reveals slight differences in the binding poses: although, in both structures the presumably protonated pyrrolidine nitrogen addresses the two catalytic aspartates 32/32', the position of the central pyrrolidine ring of **10** seems to be slightly shifted towards the catalytic aspartates (Figures 4.2b, 4.3, 4.4). However, these slight deviations might to some degree be attributed to the coordinate uncertainties at the given resolutions.

The interactions of **9** to the HTLV-1 PR are shown in Figure 4.3: either sulfonyl group of **9** forms direct hydrogen bonds to the backbone NH of Ala59 and Ala59' (3.1 and 3.4 Å, respectively), which are located in the *flap* region, while the *flap* water, which mediates the interaction to the *flap* region throughout all previously published HTLV-1 PR X-ray structures, is in this case released upon inhibitor binding (Figure 4.3).

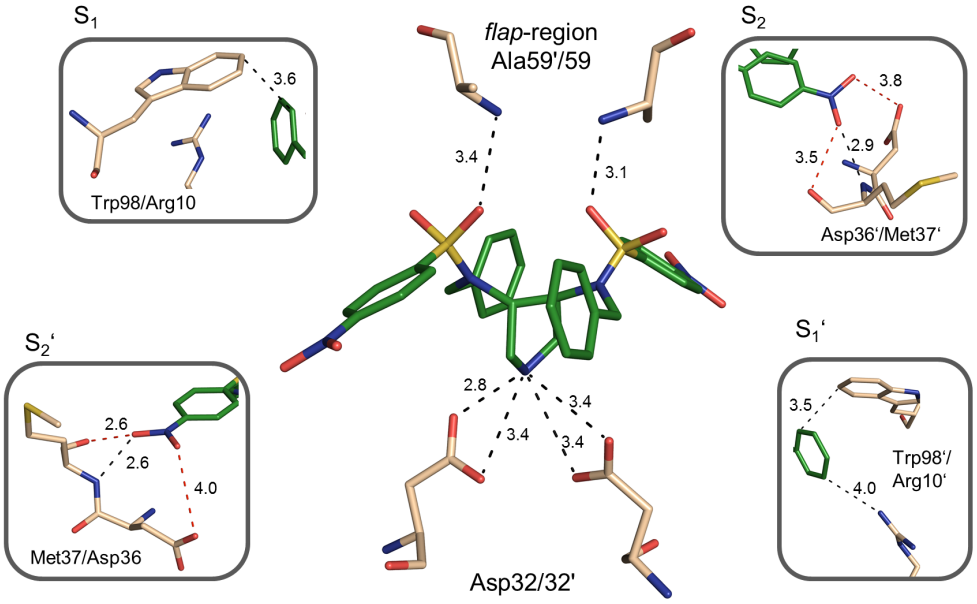


Figure 4.3. Binding mode of **9** to HTLV-1 PR. Amino acids and **9** are shown in stick representation, color-coded by atom type, distances are in Å. Hydrogen bonds are shown as black dashes. Unfavorable polar contacts in the S_2/S_2' pocket are shown as red dashes.

A superposition of our novel X-ray complexes with those previously determined clearly shows that the original position of the *flap* water is occupied by either of the bound inhibitors (Figure 4.2c), hence excluding the possible presumption that the *flap* water might not be observed in our complexes due to the deviating crystallographic resolutions of the compared crystal structures. The P_1/P_1' benzyl moieties occupy the S_1/S_1' pockets mainly formed by Leu30/30', Asp32/32', Gly34'/34, Ile100/100', Ala59'/59, and Trp98/98', and the P_2/P_2' substituents reach into the S_2/S_2' pockets, where Asp36/36', Met37'/37, Val39'/39, Leu91'/91, and Ile100'/100 are involved in key interactions. The P_1/P_1' benzyl moieties form π-π stacking interactions to Trp98/98' via distances of 3.6 and 3.5 Å, respectively and also intramolecular π-π edge-to-face stacking interactions to the aromatic residues in P_2/P_2' position. In the S_1' pocket the phenyl ring forms an additional cation-π interaction to Arg10'. In the S_2/S_2' pockets the nitro groups form an H-bond to the backbone amides of Met37'/Met37 (2.9 and 2.6 Å, respectively, Figure 4.3).

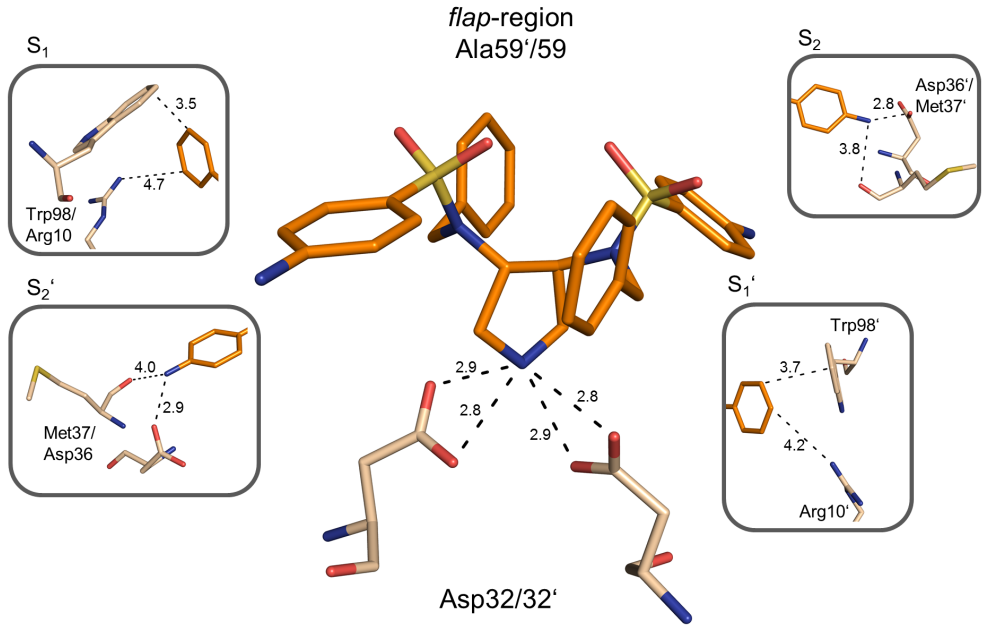


Figure 4.4. Binding mode of **10** bound to the HTLV-1 PR. Amino acids and **10** are shown in stick representation, color-coded by atom type, distances are in Å.

Besides the slightly shifted position of the pyrrolidine core structure of **10** in the active site in comparison to **9**, the interactions of **10** in the S_1/S_1' pocket are overall similar to that described for **9**: both phenyl rings establish π-π interactions to the tryptophanes 98/98' as well as cation-π interactions to Arg10/10' (Figure 4.4).

However, in the S_1 pocket this interaction is with a distance of 4.7 Å rather weak. A major difference between the two HTLV-1 structures can be observed on the protein side: the tryptophanes 98/98' in the S_1/S_1' pocket show, especially in the S_1' pocket, distinct flexibility in their conformation (Figure 4.5). A pronounced flexibility of Trp98/98' and seemingly adaptation through inhibitor binding was also observed in the past.[9] In the S_2/S_2' pocket, each amino group of **10** forms one charge-assisted H-bond to Asp36'/36 (2.8 and 2.9 Å, respectively) and weak polar contacts to the backbone amides of Met37'/Met37 (3.8 and 4.0 Å, respectively, Figure 4.4).

The *flap* region in the HTLV-1 PR complex of **10** is, probably due to a reduced internal order of the crystal accompanied by only moderate resolution, not defined and was therefore omitted from the structural model. Hence, we could not observe the interaction of **10** to the *flap* at the molecular level, however, the high structural similarity suggests similar interactions as observed in the HTLV-1 PR complex with **9**.

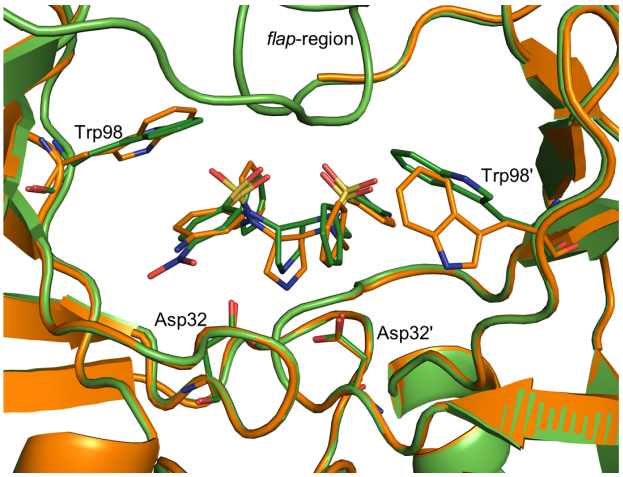


Figure 4.5. Superposition of **9** (green) and **10** (orange) bound to the HTLV-1 PR, their pyrrolidine-cores are slightly shifted. Tryptophanes 98/98' show distinct flexibility in their conformation, especially in the S_1' pocket (Trp98').

4.4.3 Comparison of Binding Modes and SAR Between HTLV-1 PR and HIV-1 PR

Overall, the described binding mode in the HTLV-1 PR is quite similar to that observed for this ligand series in the HIV-1 PR, except for the interaction to the *flap* region.[13] Due to the unsymmetrical binding mode observed in the HIV-1 PR complexes as well, in the latter only one of the sulfonamide moieties is involved in the interaction to the *flap* region (Ile50/50'). **9** and **10** also adopt a slightly unsymmetrical binding mode within the active site of the HTLV-1 PR, yet the deviation from C_2-symmetry is not as pronounced as already observed in the HIV-1 PR complexes.

Assuming a similar binding mode for the inhibitors of this lead series in HTLV-1 PR, the SAR follows a coherent picture: with increasing steric demand of the R^1 substituent, the available space in the S_1/S_1' pockets is filled more properly, thus an aromatic ring system is favored over allylic substituents. Moreover, the phenyl ring is able to form π-π interactions to Trp98/98' and cation-π interactions to Arg10/10' more efficiently compared to the allylic substituents. The comparison of the inhibition data determined for **1**, **2**, **3**, and **4** against HTLV-1 PR and HIV-1 PR suggests that the introduction of an aromatic moiety is more beneficial in case of HTLV-1 PR, as the establishment of a π-π stacking to Trp98/98` is due to the absence of an appropriate stacking partner in case of HIV-1 PR not relevant for the latter. **4**, which bears an unsubstituted benzyl moiety in P_1/P_1' position, possesses nearly the same affinity towards HTLV-1 PR (2.0 µM) and HIV-1 PR (2.15 µM).

The elongation of the P_1/P_1' substituents in *para*-position with hydrophobic moieties (**5**: iodo and **6**: CF_3) yet has an opposing effect: in the HTLV-1 PR this structural modification leads to a dramatic drop in affinity by about two orders of magnitude,

whereas the affinity against HIV-1 PR is increased about 3-fold. This deviating behavior is challenging to explain, however, in case of HTLV-1 PR it might be related to deviating dynamic properties of Arg10/10'. To accommodate these larger P_1/P_1' substituents, a flip of Arg10/10' would be required.

In the S_2/S_2' pockets the effect of the substitution pattern of the phenyl rings on the affinity deviates again between HTLV-1 PR and HIV-1 PR: in case of HIV-1 PR the introduction of hydrophobic residues in *ortho*-position (**7**: methyl, **8**: chloro) results in an increase of affinity by a factor of 3. In case of HTLV-1 PR, however, to accommodate these substituents, presumably a slight change of the rotational state of the Ile100/100' side chains is required, which might explain the observation that these substituents exert only a marginal influence on the potency.

Introduction of polar substituents in *para*-position of the benzene moiety (**9**: NO_2, **10**: NH_2) addressing the S_2/S_2' pockets has the strongest impact on the affinity amongst the derivatives tested in our set: in **9** the affinity towards HTLV-1 PR is decreased about 4-fold, whereas that towards HIV-1 PR is nearly unaffected. The introduction of a *para*-amino group (**10**) increases the affinity in both PR: in case of HIV-1 PR by about one order, and in case of HTLV-1 PR even more significantly by about two orders of magnitude.

In general, the nitro group is a weak H-Bond acceptor. For **9** an H-bond to the backbone NH of Met37 and Met37', respectively, is formed. However, especially in the S_2' pocket the interacting oxygen atom of the nitro group comes in close contact (2.6 Å) to the backbone carbonyl group of Met37 (Figure 4.3). In the S_2 pocket the distance of this unfavorable contact amounts to 3.5 Å. In addition, an adverse contact of the nitro group to the carbonyl oxygen of the side chain of Asp36'/36 is observed (3.8 Å and 4.0 Å, respectively). These unfavorable repulsive interactions rationalize well the loss in affinity compared to the unsubstituted derivative **4** (7.9 vs 2.0 µM).

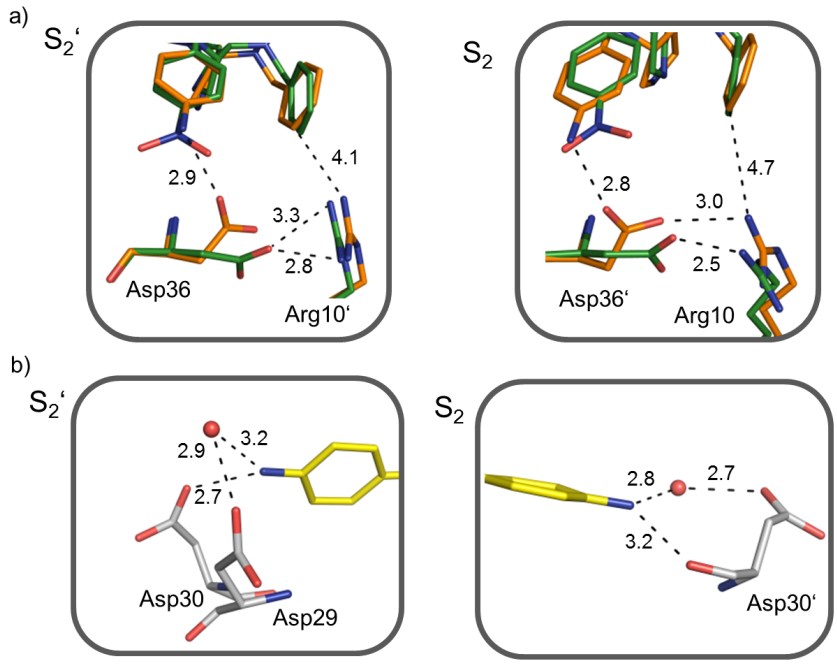


Figure 4.6. a) Asp36/36' turned towards **10** (HTLV-1 PR, orange), whereas in the structure of **9** (HTLV-1 PR, green) the aspartates form an internal salt bridge to Arg10/Arg10'. b.) H-Bond network of **10** (yellow) in the HIV-1 PR (grey). Amino acids and ligands are shown in stick representation, involved water molecules as spheres, distances in Å.

On the contrary, the *para*-amino group in **10** is capable of acting as H-bond donor, and two charge-assisted H-bonds to Asp36'/36 via distances of 2.8 and 2.9 Å, respectively, are established (Figure 4.4). The contacts of the amino to the carbonyl groups of Met37'/37 are, in contrast to the nitro derivative, in this case not detrimental towards binding. However, with distances of 3.8 and 4.0 Å, respectively, these contacts are not expected to have a major impact on the binding affinity. Noteworthy, to enable the charge-assisted interactions, Asp36'/36 turn towards the ligand's amino groups, thus disrupting the internal salt bridges towards Arg10/10' observed in the complex with **9** (Figure 4.6a). Most likely, the deviating interaction pattern of the amino and nitro groups of **9** and **10** account for their different inhibitory activity towards HTLV-1 PR. In comparison to **4**, **10** is able to establish favorable polar contacts in the S_2/S_2' pockets of the protein. As the formation of a charge-assisted H-bond is accompanied by a gain in binding enthalpy of 5-30 kJ/mol per H-bond,[20–22] the increase in affinity by about two orders of magnitude (2.0 µM vs 15 nM) can likely be attributed to the establishment of two charge-assisted H-bonds in the S_2/S_2' pockets.

Interestingly, the introduction of an amino function in *para*-position is beneficial for binding to both proteases, however, as described, this substituent effect is much more pronounced for HTLV-1 PR. Fortunately, for a direct comparison X-ray structures of **10**

in complex with both proteases are available. While in the HIV-1 PR **10** exhibits an unsymmetrical binding mode[13] which results in a different interaction pattern of the ligand's amino groups in the S_2 and S_2' pocket, in the HTLV-1 PR the interactions of the amines are similar (Figure 4.4 and 4.6b). The structural comparison provides an explanation for the observed distinct difference in affinity of **10** against the two proteases: whereas in HTLV-1 PR two charge-assisted H-bonds, one in the S_2 and one in the S_2' pocket, are established, in HIV-1 PR such a direct charge-assisted H-bond is only formed in the S_2' pocket.

As noted before, in our study we detect remarkable changes in affinity caused solely by the variation of the substituents interacting to the S_2/S_2' pockets. That minor structural alterations can result in distinctive shifts in affinity is also well-described in literature: as in our case, these are mostly related to an unfavorable switch regarding the electrostatic contact inventory such as the replacement of an H-bond acceptor by a donor functionality or *vice versa*. An example is represented by a series of phosphorus-containing thermolysin inhibitors:[23] the exchange of a phosphonamide to a phosphonate reduces the affinity by about three orders of magnitude. Besides the lack of one H-bond, in this example also two H-bond acceptors face each other resulting in an unfavorable repulsion of the ligand oxygen and a carbonyl oxygen on the protein side. Another example is described for a series of hydroxypyrroline-based thrombin inhibitors,[24] where the exchange of an amide ($N(R_2)$-$\underline{CO}$-CH_3) for an amine residue ($N(R_2)$-$\underline{CH_2}$-CH_3) has a dramatic effect on the affinity, likely due to the loss of an H-bond and the change of the electronic properties thus resulting in an affinity decrease by four orders of magnitude.[25]

4.5 Conclusion

Despite various definitions, the fundamental idea of the "privileged structure" concept suggests to identify inhibitor scaffolds for a target of interest by utilizing knowledge on attractive scaffolds that are clearly established within a particular target class instead of performing (ultra-)high-throughput screenings of large chemical libraries. In this contribution, we applied this concept to HTLV-1 PR and identified C_2-symmetric pyrrolidine-based 3,4-bis-*N*-alkylsulfonamides making this compound class a novel, non-peptidic scaffold for HTLV-1 protease inhibition. A series of ten inhibitors exerted affinities from the submillimolar down to two-digit nanomolar range against HTLV-1 protease, which, in combination with the successful crystal structure determination, lays the foundation for further structure-guided design of HTLV-1 protease inhibitors

based on the presented pyrrolidine scaffold. Undoubtedly, our proof-of-concept study only monitors the evolving SAR by means of a strictly limited set of ten compounds selected to explore the suitability of the pyrrolidine scaffold for HTLV-1 PR inhibition in general, accompanied by structure analysis. Hence, a broader collective set of compounds should certainly allow to study and interpret substituent effects in much more detail than reliably possible with our set. For this purpose, the two crystal structures determined in our study provide some guidance for further lead optimization: first, an appropriate substitution of the so far non-decorated P_1/P_1' benzyl moiety shall be attempted. Here, addressing the Gly34/34' carbonyl group and the side chain of Arg10/10' by appropriate functional groups such as e.g. *meta*-carboxamide or *para*-methoxy/carboxylate are options to be pursued. For P_2/P_2', Asp36/36' and Met37/37' may be addressed more efficiently by e.g. a *para*-carboxamide substituent. In addition, even combinatorial approaches have been considered for this purpose at this stage, once a suitable "privileged structure" core motif has been identified.[11] Nevertheless, among the tested compounds **10** bearing a *para*-amino-substituted phenyl moiety as P_2/P_2' substituent exhibits a K_i-value of 15 nM and thus represents not only the most potent compound of this inhibitor series, but, moreover, to the best of our knowledge, the most potent non-peptidic inhibitor of HTLV-1 PR described so far.

4.6 Experimental Section

4.6.1 Protein Expression and Purification

The protein expression, purification, and refolding was performed as previously described.[9,26] For the protease assay the full-length protein (HTLV 1-125) was used, while the crystallization was performed with a nine amino acids-truncated construct of the protease (HTLV 1-116), which has been used for all eight HTLV-1 PR X-ray structures published to date. To prevent autoproteolysis, both constructs harbor the stabilizing mutation L40I.

4.6.2 Kinetic Assay

The fluorescence-based assay was in general performed as previously published.[9]
IC_{50}-values were determined at a substrate concentration of 5 µM, for K_m-determination 12 different substrate concentrations were used (50 µM- 24 nM). From the increase in fluorescence the IC_{50}-values were calculated with the program GraFit 4.09 (Erithacus Software). The corresponding K_i-values were calculated by means of the Cheng-Prusoff-equation[27] under inclusion of the determined K_m-value and substrate

concentration. Each K_m- and K_i-values were determined at least in duplicate.
The assay was validated using indinavir as reference inhibitor: using our settings we achieve a K_i-value of 3.1 ± 0.6 µM, which is in acceptable agreement as found in the literature (3.48 µM).[26] In addition, we have confirmed exemplarily that our inhibitor series (**10**) does not act as fluorescence quencher in the relevant wavelength range.

4.6.3 Crystallization of HTLV-1 Protease Inhibitor Complexes

Crystals of the HTLV-1 protease inhibitor complexes were obtained by cocrystallization from a HTLV-1 PR (1-116) solution at a concentration of 10 mg/ml in presence of 5 mM inhibitor using 1.0-1.3 M ammonium sulfate, 20 % glycerol (v/v), 0.1 M Tris, pH 7.0 as reservoir solution. Crystals appeared after several days of incubation at 291K.

4.6.4 Data Collection and Refinement

Crystals of the HTLV-1-PR-Inhibitor complexes were harvested from mother liquor and frozen without cryoprotectant in liquid nitrogen. Data were collected at 100K at the Bessy II (Helmholtz-Zentrum Berlin, Germany), Beamline 14.1.[28] Data processing and scaling was performed using the program XDS[29] as implemented into the graphical user interface XDSAPP.[30] The coordinates of HTLV-1 PR as deposited under PDB access code 3LIX[31] were used as search model after removal of ligand and water molecules. Molecular replacement was carried out via Phaser[32] as implemented in Phenix.[33] Refinement was performed under repeated cycles of manual model building using Coot[34] and crystallographic refinement with the program phenix.refine. The final model was validated using PROCHECK.[35] Data collection and refinement statistics are shown in Table 4.2 (appendix).

4.6.5 Synthesis

The synthesis and analytical data of compound **1**-**10** are described in Blum *et al.*[13] All test compounds (**1**-**10**) were at least 95 % pure, determined by combustion analysis.

4.7 Appendix

Table 4.2. Data collection and refinement statistics.

Data Collection and Refinement Statistics		
Complex/PDB entry	**9/4YDF**	**10/4YDG**
Data collection and Processing		
No. of crystals used	1	1
Wavelength [Å]	0.91841	0.91841
Space group	P 6_3 2 2	P 6_3 2 2
Unit cell parameters		
a, b, c [Å]	77.4; 77.4; 160.1	77.6; 77.6; 160.2
Matthews coefficient [Å³/Da]	2.55	2.56
Solvent content [%]	51.7	52.0
Diffraction data		
Resolution range [Å]	50.0 – 2.80 (3.05-2.80)*	50.0 – 3.25 (3.5-3.25)
Unique reflections	7 097 (1 573)	4 746 (907)
CC(1/2)	99.8 (91.5)	98.8 (83.1)
Completeness [%]	94.6 (96.1)	96.3 (97.4)
Redundancy	9.9 (10.3)	14.3 (14.9)
I/σ(I)	15.5 (3.7)	8.6 (3.2)
Refinement		
Resolution range [Å]	38.7 – 2.80	41.9 – 3.25
Reflections used in refinement (work/free)	6 735 / 355	4 499 / 238
Final R values for all reflections (work/free) [%]	24.2 / 29.7	27.9 / 31.7
Protein residues	232	219
Inhibitor atoms	45	41
Water molecules	4	0
RMSDs		
Bonds [Å]	0.006	0.003
Angles [°]	1.3	0.7
Ramachandran plot		
Residues in most favoured regions [%]	89.4	88.8
Residues in additional allowed regions [%]	10.1	10.7
Residues in generously allowed regions [%]	0.5	-
Residues in disallowed regions [%]	-	0.5
Mean B factors [$Å^2$]		
Protein	25.9	47.4
Inhibitor	35.4	42.4
Water	9.9	-

* Numbers in parentheses characterize the highest resolution shell

4.8 References

(1) Poiesz, B. J.; Ruscetti, F. W.; Gazdar, A. F.; Bunn, P. A.; Minna, J. D.; Gallo, R. C. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **1980,** *77*, 7415–7419.

(2) Shuker, S. B.; Mariani, V. L.; Herger, B. E.; Dennison, K. J. Understanding HTLV-I protease. *Chem. Biol.* **2003,** *10*, 373–380.

(3) Li, M.; Laco, G. S.; Jaskolski, M.; Rozycki, J.; Alexandratos, J.; Wlodawer, A.; Gustchina, A. Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 18332–18337.

(4) McGill, N.-K.; Vyas, J.; Shimauchi, T.; Tokura, Y.; Piguet, V. HTLV-1-associated infective dermatitis: updates on the pathogenesis. *Exp. Dermatol.* **2012,** *21*, 815–821.

(5) Edlich, R. F.; Arnette, J. A.; Williams, F. M. Global epidemic of human T-cell lymphotropic virus type-I (HTLV-I). *J. Emerg. Med.* **2000,** *18*, 109–119.

(6) Gonçalves, D. U.; Proietti, F. A.; Ramos Ribas, J. G.; Araújo, M. G.; Pinheiro, S. R.; Guedes, A. C.; Carneiro-Proietti, A. B. F. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. *Clin. Microbiol. Rev.* **2010,** *23*, 577-589.

(7) Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014,** *42*, D503-D509.

(8) Kádas, J.; Weber, I. T.; Bagossi, P.; Miklóssy, G.; Boross, P.; Oroszlan, S.; Tözsér, J. Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell leukemia virus type 1 protease. *J. Biol. Chem.* **2004,** *279*, 27148–27157.

(9) Kuhnert, M.; Steuber, H.; Diederich, W. E. Structural basis for HTLV-1 protease inhibition by the HIV-1 protease inhibitor indinavir. *J. Med. Chem.* **2014,** *57*, 6266–6272.

(10) Demir, A.; Oguariri, R. M.; Magis, A.; Ostrov, D. A.; Imamichi, T.; Dunn, B. M. Kinetic characterization of newly discovered inhibitors of various constructs of human T-cell leukemia virus-1 (HTLV-1) protease and their effect on HTLV-1-infected cells. *Antiviral Ther.* **2012,** *17*, 883–892.

(11) Müller, G. Medicinal chemistry of target family-directed masterkeys. *Drug Discovery Today* **2003,** *8*, 681–691.

(12) Welsch, M. E.; Snyder, S. A.; Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* **2010,** *14*, 347–361.

(13) Blum, A.; Böttcher, J.; Heine, A.; Klebe, G.; Diederich, W. E. Structure-guided design of C_2-symmetric HIV-1 protease inhibitors based on a pyrrolidine scaffold. *J. Med. Chem.* **2008,** *51*, 2078–2087.

(14) Luksch, T.; Blum, A.; Klee, N.; Diederich, W. E.; Sotriffer, C. A.; Klebe, G. Pyrrolidine derivatives as plasmepsin inhibitors: binding mode analysis assisted by molecular dynamics simulations of a highly flexible protein. *ChemMedChem* **2010,** *5*, 443–454.

(15) Luksch, T.; Chan, N.-S.; Brass, S.; Sotriffer, C. A.; Klebe, G.; Diederich, W. E. Computer-aided design and synthesis of nonpeptidic plasmepsin II and IV inhibitors. *ChemMedChem* **2008,** *3*, 1323–1336.

(16) Blum, A.; Böttcher, J.; Sammet, B.; Luksch, T.; Heine, A.; Klebe, G.; Diederich, W. E. Achiral oligoamines as versatile tool for the development of aspartic protease inhibitors. *Bioorg. Med. Chem.* **2008,** *16*, 8574–8586.

(17) Böttcher, J.; Blum, A.; Dörr, S.; Heine, A.; Diederich, W. E.; Klebe, G. Targeting the open-*flap* conformation of HIV-1 protease with pyrrolidine-based inhibitors. *ChemMedChem* **2008,** *3*, 1337–1344.

(18) Böttcher, J.; Blum, A.; Heine, A.; Diederich, W. E.; Klebe, G. Structural and kinetic analysis of pyrrolidine-based inhibitors of the drug-resistant Ile84Val mutant of HIV-1 protease. *J. Mol. Biol.* **2008,** *383*, 347–357.

(19) Superposition of the *flap* water of all published HTLV-1 PR X-ray structures, PDB-codes: 2B7F, 3LIT, 3LIN, 3LIX, 3LIY, 3LIQ, 3LIV, 3WSJ.

(20) Holdgate, G. A. Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *BioTechniques* **2001,** *31*, 164–184.

(21) Martin, T. W.; Derewenda, Z. S. The name is bond-H bond. *Nat. Struct. Biol.* **1999,** *6*, 403-406.

(22) Steiner, T. The hydrogen bond in the solid state. *Angew. Chem., Int. Ed.* **2002,** *41*, 48–76.

(23) Morgan, B.; Scholtz, J. M.; Ballinger, M. D.; Zipkin, I. D.; Bartlett, P. A. Differential binding energy: a detailed evaluation of the influence of hydrogen-bonding and hydrophobic groups on the inhibition of thermolysin by phosphorus-containing inhibitors. *J. Am. Chem. Soc.* **1991,** *113*, 297–307.

(24) Klein, S. I.; Dener, J. M.; Molino, B. F.; Gardner, C. J.; D'Alisa, R.; Dunwiddie, C. T.; Kasiewski, C.; Leadley, R. J. O-Benzyl hydroxyproline as a bioisostere for Phe-Pro: novel dipeptide thrombin inhibitors. *Bioorg. Med. Chem. Lett.* **1996,** *6*, 2225–2230.

(25) Kubinyi, H. Hydrogen bonding: the last mystery in drug design? In *Pharmacokinetic Optimization in Drug Research.* Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R., Eds.; Verlag Helvetica Chimica Acta: Zürich, 2007; p. 513–524.

(26) Bagossi, P.; Kádas, J.; Miklóssy, G.; Boross, P.; Weber, I. T.; Tözsér, J. Development of a microtiter plate fluorescent assay for inhibition studies on the HTLV-1 and HIV-1 proteinases. *J. Virol. Methods* **2004,** *119*, 87–93.

(27) Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973,** *22*, 3099–3108.

(28) Mueller, U.; Darowski, N.; Fuchs, M. R.; Förster, R.; Hellmig, M.; Paithankar, K. S.; Pühringer, S.; Steffien, M.; Zocher, G.; Weiss, M. S. Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J. Synchrotron Radiat.* **2012,** *19*, 442–449.

(29) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 125–132.

(30) Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U. XDSAPP. A graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.* **2012,** *45*, 568-572.

(31) Satoh, T.; Li, M.; Nguyen, J.-T.; Kiso, Y.; Gustchina, A.; Wlodawer, A. Crystal structures of inhibitor complexes of human T-cell leukemia virus (HTLV-1) protease. *J. Mol. Biol.* **2010,** *401*, 626–641.

(32) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40*, 658–674.

(33) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 213–221.

(34) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 486–501.

(35) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993,** *26*, 283-291.

5. Pyrrolidine-Based Bicyclic Compounds as HTLV-1 Protease Inhibitors

5.1 Introduction

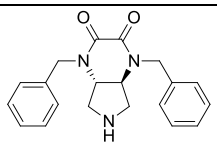
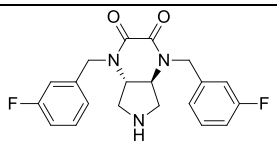
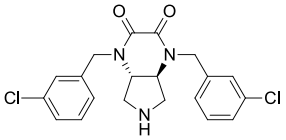
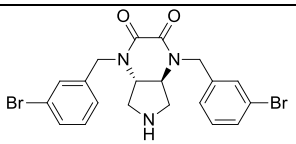
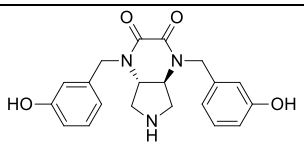
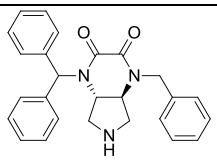
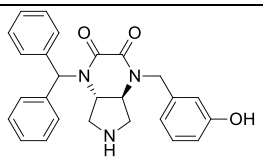
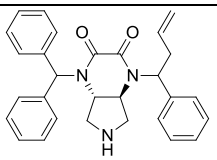
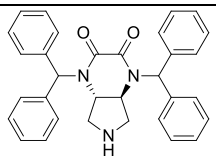
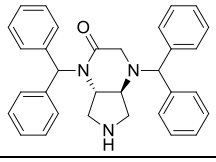
The initial screening of a diverse subset of our in-house aspartic protease inhibitor library yielded besides the C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines (chapter 4) also pyrrolidine-based bicyclic inhibitors, which showed promising affinity against the HTLV-1 protease at 100 µM inhibitor concentration. The series of this compound class was originally designed and synthesized as HIV-1 PR inhibitors by Dr. Nina Klee.[1] In total, ten bicyclic inhibitors (Table 5.1) were investigated in more detail regarding their inhibitory profile and binding mode to the HTLV-1 protease.

5.2 Affinity Data

The K_i-values of the bicyclic inhibitor series determined via a FRET assay are given in Table 5.1. Five of these ten inhibitors, in the following named two-armed inhibitors, are C_2-symmetric being equipped with differently substituted benzyl moieties (NK101, NK102, NK191, NK192, NK237). Three other compounds (NK251, NK252, and NK253) are further substituted at the benzylic position in order to address additional specificity pockets of the enzyme and are in the following referred to as three-armed and four-armed inhibitors, respectively. Common to all of the latter ones is a benzyhydryl moiety. NK343 and NK232 are both decorated with two benzhydryl substituents, however, in case of NK343 one carbonyl group of the bridging ring system was replaced by a methylene moiety resulting in a glycine-like bridging.

The K_i-values for this inhibitor series range from three- up to one-digit micromolar, with NK232 representing the most potent inhibitor of this series with a K_i-value of 1.4 µM.

Table 5.1. Affinity data as well as ligand efficiency (LE) for a series of pyrrolidine-based bicyclic inhibitors against the HTLV-1 PR.

	chemical structure	%-inhibition values at 100 µM [%]	K_i-values [µM]	LE [kcal/mol][2,3]
NK102		25.9 ± 9.3	215.6 ± 83.5	0.20
NK191		50.8 ± 21	45.6 ± 0.9	0.22
NK237		83.0 ± 0.9	9.0 ± 0.5	0.26
NK192		83.4 ± 0.4	5.0 ± 1.0	0.27
NK101		66.7 ± 0.5	29.7 ± 4.4	0.23
NK251		71.1 ± 3.1	15.0 ± 1.8	0.21
NK252		81.2 ± 1.8	10.5 ± 3.3	0.21
NK253		91.7 ± 2.1	3.7 ± 0.7	0.22
NK232		89.2 ± 0.4	1.4 ± 0.4	0.22
NK343		64.8 ± 1.0	20.7 ± 1.8	0.18

5.3 SAR Interpretation and Binding Mode Analysis

Considering at first only the inhibitors bearing unsubstituted benzyl- or benzhydryl-moieties, it becomes evident that the affinity increases with the size of the molecule: the two-armed benzyl-substituted compound NK102 possesses the lowest affinity (K_i-value ≈ 215 µM), the analogous three-armed derivative NK251 shows an increase by one order of magnitude, and the four-armed compound NK232 bearing two benzhydryl-moieties exhibits improved affinity by two orders of magnitude in comparison to NK102 and represents with a K_i-value of 1.4 µM the most potent compound of this ligand series. However, the ligand efficiency, which describes the gained binding energy per heavy atom of a ligand,[3] and therefore indirectly considers the contribution of the size of the molecule towards the affinity, is similar for these three compounds (NK102, NK251 and NK232) (Table 5.1). NK343, deviating from NK232 only in the bridging of the 3,4-diamine-pyrrolidine scaffold (carbonyl versus methylene), exhibits in comparison to NK232 a significantly reduced affinity by one order of magnitude (20.7 vs 1.4 µM) and exhibits the worst ligand efficiency of this inhibitor series against the HTLV-1 PR (0.18 kcal/mol).

Within the symmetric two-armed inhibitors, an affinity difference by a factor of about 40 is observed in dependence on the substitution pattern of the benzyl moiety. The affinity increases from the unsubstituted benzyl moiety (NK102) via the *meta*-fluoro- (NK191) and *meta*-hydroxyl- (NK101) substituted compounds to the *meta*-chloro (NK237) and *meta*-bromo (NK192) derivatives, the latter two showing a K_i-value in the one-digit µM range (9.0 and 5.0 µM, respectively). Interestingly, despite the slightly lower affinity, but considering the reduced number of heavy atoms in comparison to NK232, NK192 and NK237 exhibit the best ligand efficiency within this inhibitor series (0.27 kcal/mol and 0.26 kcal/mol, respectively).

A direct comparison of NK252 and NK251 reveals that the introduction of a hydroxyl substituent in *meta*-position of the benzyl moiety does not significantly affect the affinity. The removal of one carbonyl group (NK343) impairs the affinity by one order of magnitude.

Overall, the SAR seems to be consistent, however, for a more comprehensive understanding further structural investigations of the binding mode have been applied to complement this SAR interpretation.

The X-ray structure of NK101 in complex with the HTLV-1 protease was successfully determined at a resolution of 3.15 Å. The observed difference density allows an unambiguous placement of NK101 in the active site (Figure 5.1a). Unfortunately, despite extensive efforts (chapter 2), no diffracting crystals of HTLV-1 PR in complex

with other bicyclic inhibitors could be obtained. The X-ray structure of NK101 in complex with HTLV-1 PR was refined by Dr. Holger Steuber, Philipps-Universität Marburg.

The binding mode of NK101 in the HTLV-1 PR is shown in Figure 5.1b. The pyrrolidine nitrogen addresses the two catalytic aspartates, while the *meta*-hydroxyl-substituted benzyl moieties reside in the S_2 and S_2' pockets. Two H-bonds are formed between the NH groups of Ala59 and Ala59' located in the *flap* region and the two carbonyl oxygens of the inhibitor. However, as the hairpin loops of the *flap* region are not well defined, the interactions need to be interpreted with care. As also observed for the C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines (chapter 4), the *flap* water is replaced by the inhibitor. The hydroxyl substituents of the benzyl moiety occupying the S_2/S_2' pockets form H-bonds to Asp36'/36 at a distance of 2.9 and 3.2 Å, respectively.

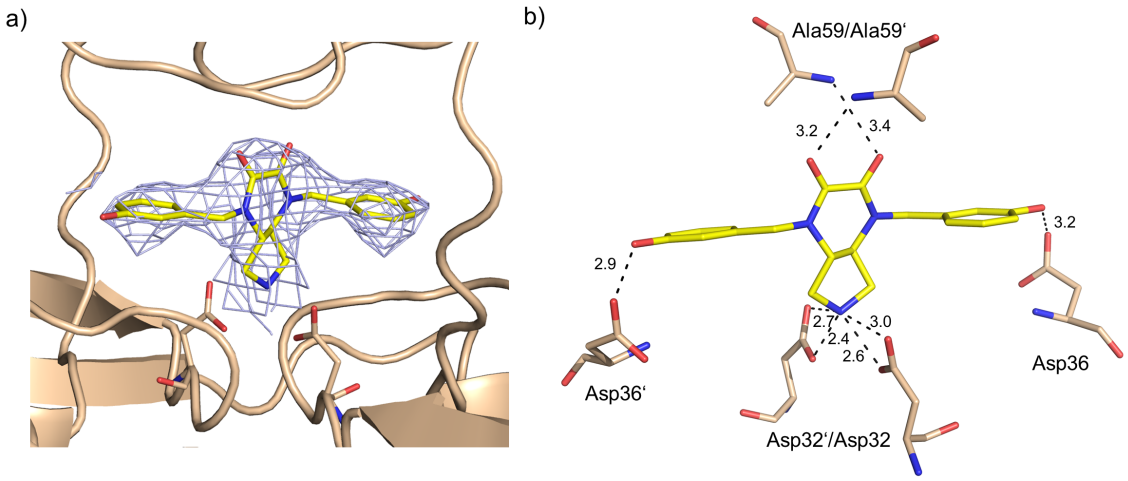


Figure 5.1. a) $2F_o-F_c$ electron density map for NK101 bound to the HTLV-1 PR contoured at 1.0σ (blue mesh). b) Binding mode of NK101 in the active site of the HTLV-1 PR.

Compared to the unsubstituted analogue (NK102), the introduction of halogen atoms in *meta*-position of the benzyl substituent yields equal- or more potent derivatives with increasing size of the halogen. As no stacking partner for the aromatic moieties is located in adequate proximity, this effect is probably not mediated by the electronic properties of the latter, but rather dependent on the halogen substituents themselves: with increasing size they might contribute more efficiently to van der Waals contacts within the S_2/S_2' pockets and their increasing polarizability (e.g. by Asp36/36') might create a reinforcing electrostatic contribution growing with larger atom radius.

In contrast to the C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines (chapter 4), the bicyclic inhibitors are significantly less potent. However, the most active compound of the bicyclic inhibitors being equipped with two benzhydryl substituents (NK232) was not

specifically optimized to form directed interactions to the protein. NK232 exhibits, similar to the benzyl/phenyl-substituted 4-armed 3,4-bis-*N*-alkylsulfonamido-pyrrolidine AB69 (compound **4**, chapter 4), a K_i-value in the one digit µM range (1.4 and 2.0 µM, respectively). As the molecular weight or, more precisely, the number of heavy atoms, is only slightly lower for NK232, the ligand efficiency of the two compounds is comparable (0.20 vs 0.22 kcal/mol). Thus, the pyrrolidine-based bicyclic inhibitors represent as well a promising scaffold for HTLV-1 PR inhibition and provide a suitable basis for lead optimization by introduction of tailor-made substituents.

The superposition of AB83 and AB84 (**9** and **10**, respectively, chapter 4) as representatives of the C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines and the bicyclic pyrrolidine-based inhibitor NK101, all bound to the HTLV-1 PR (Figure 5.2) reveals that all three inhibitors occupy the S_2/S_2' pocket with the substituted phenyl or benzyl moiety.

The affinity gain most likely originates from the introduction of one or two additional aromatic "arms", as these most probably address the so far unoccupied S_1/S_1' pockets hence giving rise to an affinity improvement.

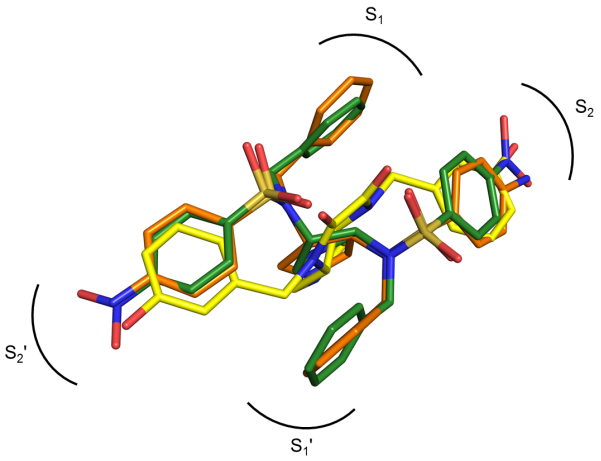


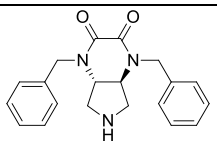
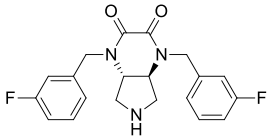
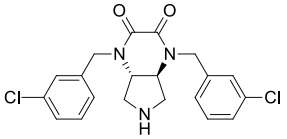
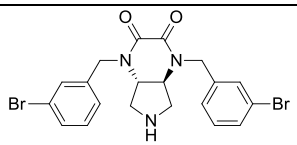
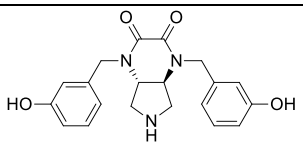
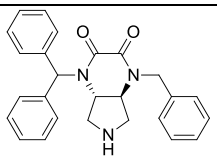
Figure 5.2. Superposition of AB83 (green), AB84 (orange), and NK101 (yellow) bound to the HTLV-1 PR.

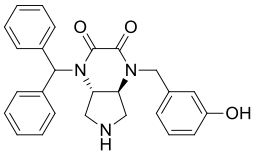
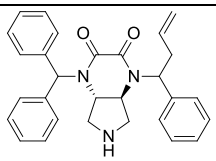
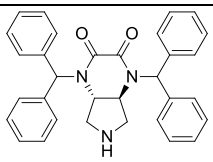
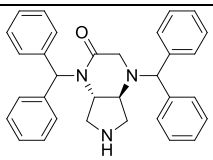
Even though the bicyclic scaffold is more rigid, which in general is associated with a reduced conformational entropy loss upon binding, the inhibition data obtained for the two scaffolds do not support a clear affinity improvement for the more constrained scaffold in our case.

5.4 Comparison to the HIV-1 Protease

The comparison of the inhibition profile of the pyrrolidine-based bicyclic inhibitors against the HTLV-1 PR and the HIV-1 PR (Table 5.2) reveals that these inhibitors are in general more potent against the HIV-1 PR than the HTLV-1 PR, which is not too surprising as these inhibitors were originally designed as HIV-1 PR inhibitors by structure-based drug design.[1] A closer look reveals that the two-armed inhibitors only show a slightly enhanced inhibitory effect up to a factor of 6 against the HIV-1 PR whereas within the three- and four-armed inhibitors the difference in affinity is more pronounced. While the general tendency of the inhibition profile within each particular target is similar, the inhibitors are by factors between 20 - 200 more potent against the HIV-1 PR than the HTLV-1 PR.

Table 5.2. K_i-values of a series of pyrrolidine-based bicyclic compounds against the HTLV-1 and HIV-1 protease.

	chemical structure	**K_i-values [µM] HTLV-1 PR**	**K_i-values [µM] HIV-1 PR***	**$K_{i(HTLV)}/K_{i(HIV)}$**
NK102		215.6 ± 83.5	35	6
NK191		45.6 ± 0.9	21	2
NK237		9.0 ± 0.5	23	0.4
NK192		5.0 ± 1.0	7.3	0.7
NK101		29.7 ± 4.4	6.4	5
NK251		15.0 ± 1.8	0.9	17

NK252		10.5 ± 3.3	0.21	50
NK253		3.7 ± 0.7	0.024	154
NK232		1.4 ± 0.4	0.007	200
NK343		20.7 ± 1.8	0.160	129

* K_i-values taken from reference.[1]

The superposition of the X-ray structures of NK101 in complex with the HIV-1 PR and HTLV-1 PR shows that this inhibitor occupies virtually the identical space within both proteases (Figure 5.3). Besides various water-mediated interactions in the HIV-1 PR-NK101 complex, the ligand's two hydroxyl groups both form an H-bond to the NH backbone of Asp29'/29 and Asp30'/30 in the S_2/S_2' pockets, respectively.

Due to the moderate resolution of the HTLV-1 PR-NK101 complex, no information on interacting or mediating water molecules could be deduced. Hence, a more detailed comparison of the binding modes of NK101 bound to HTLV-1 PR and HIV-1 PR, respectively, appeared rather speculative and was therefore omitted in this respect.

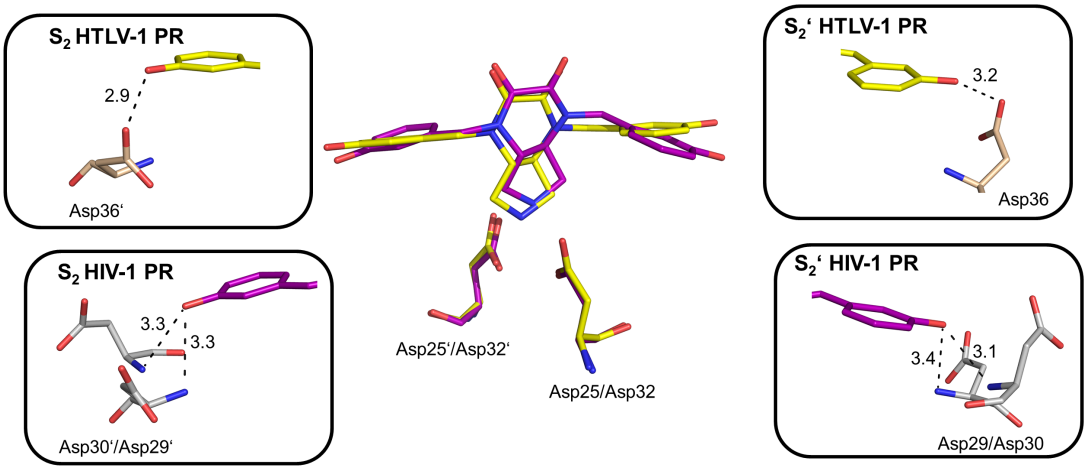


Figure 5.3. Superposition of NK101 bound to the HIV-1 PR (purple, PDB code: 3QN8) and HTLV-1 PR (yellow), as well as interactions of the *meta*-hydroxyl moiety in the S_2 and S_2' pocket.

5.5 Conclusion and Outlook

The pyrrolidine-based bicyclic core structure represents an additional promising small molecule scaffold for HTLV-1 protease inhibition. Herein, the unsubstituted benzhydryl-substituted inhibitor NK232 is the most potent compound of the investigated series with a K_i-value of 1.4 µM and therefore represents a well-suited starting point of further structure-based inhibitor design for this enzyme. The further decoration of the benzhydryl-moieties with appropriate substituents might enable the formation of additional directed interactions to the protein thus potentially enhancing the inhibitory effect of these newly designed ligands.

As the *para*-phenyl substituents of AB83 and AB84 (**9** and **10**, chapter 4) and the *meta*-benzyl substituent of NK101 all occupy nearly the same position in the active site (Figure 5.2), especially the introduction of *meta*-amino substituents at the benzyl or rather benzhydryl moieties of the pyrrolidine-based bicyclic compounds could be beneficial, since the introduction of an amino functionality resulted in a significant increase in affinity by two orders of magnitude in case of the C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines (chapter 4).

5.6 Experimental Section

The experimental procedure for the protein expression and purification, FRET assay as well as crystallization was in general performed as described in chapter 4. Little modifications regarding the substrate concentration in the fluorescence-based assay as well as the crystallization conditions were included:

For IC_{50}-determination of the pyrrolidine-based bicyclic inhibitors, the FRET substrate RE(Edans)TKVLVVQPK(Dabcyl)R (Bachem, Germany) was used in a concentration of 3 µM. Under these conditions the K_i-values of indinavir (as reference) was determined to 2.0 ± 0.3 µM, which is in acceptable agreement to the literature value (3.48 µM).[4]

Crystals of the HTLV-1 protease-NK101 complex were obtained by cocrystallization from a HTLV-1 PR (1-116) solution at a concentration of 10 mg/ml in presence of 5 mM inhibitor (NK101) using 1.5 M ammonium sulfate, 20 % glycerol (v/v), 0.1 M Tris, pH 7.5 as reservoir solution.

5.7 References

(1) Klee, N. Strukturbasiertes Design, Synthese und Affinitätsbestimmung neuartiger HIV-1-Protease-Inhibitoren. Dissertation, Philipps-Universität Marburg, 2013.

(2) The ligand efficiency was calculated by following equation: LE=-R*T*ln(K_i)/HA. R: ideal gas constant, T: temperature [K], K_i: K_i-value, HA: number of non-hydrogen atoms.

(3) Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discovery* **2014,** *13*, 105–121.

(4) Bagossi, P.; Kádas, J.; Miklóssy, G.; Boross, P.; Weber, I. T.; Tözsér, J. Development of a microtiter plate fluorescent assay for inhibition studies on the HTLV-1 and HIV-1 proteinases. *J. Virol. Methods* **2004,** *119*, 87–93.

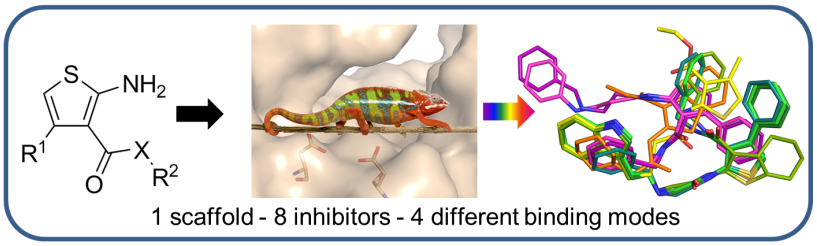
6. Tracing Binding Modes in Hit-to-Lead Optimization: Chameleon-Like Poses of Aspartic Protease Inhibitors

6.1 Introductory Remarks

The following chapter has been published in *Angewandte Chemie Int. Ed*. (M. Kuhnert *et al.*, **2015**, 54, 2849-2853) and *Angewandte Chemie* (M. Kuhnert *et al.*, **2015**, 127, 2891-2896).

The respective text is taken from the original publication, figures of the supporting information are implemented in the main part of the manuscript. The experimental procedures of compounds which were not synthesized within this thesis are not included in the experimental section but can be found in the original publication. In addition to the published manuscript, a detailed discussion of the synthesis is included (chapter 6.4).

The preliminary work of Dr. Ruben Bartholomäus, Dr. Helene Köster, and Amir Shahim related to this project is described in chapter 1.5.2 and its results are incorporated in this chapter. The synthesis of the majority of the studied compounds, i.e. **3**, **4**, **5**, **6**, **8**, **9**, and **10**, the determination of all biophysical and affinity data presented in this study (TSA and fluorescence based assay), and the crystallization of compounds **5**, **6** and **8** have been performed by the author of this thesis, along with the interpretation of the data, as well as drafting of the manuscript. The X-ray structures of **5**, **6**, and **8** were refined within a cooperation project by Dr. Holger Steuber (LOEWE-Zentrum für Synthetische Mikrobiologie SYNMIKRO, Philipps-Universität Marburg). The X-ray structures of **3** and **9** were determined in cooperation with the group of Prof. Dr. G. Klebe (Philipps-Universität Marburg) by Dr. Ah Young Park and Prof. Dr. Andreas Heine.



6.2 Abstract

Successful lead optimization in structure-based drug discovery depends on the correct deduction and interpretation of the underlying structure-activity relationships (SAR) to facilitate efficient decision-making on the next candidates to be synthesized. Consequently, the question arises, how frequently a binding mode (re)-validation is required, to ensure not to be misled by invalid assumptions on the binding geometry. We present an example in which minor chemical modifications within one inhibitor series lead to surprisingly different binding modes. X-ray structure determination of eight inhibitors derived from one core scaffold resulted in four different binding modes in the aspartic protease endothiapepsin, a well-established surrogate for e.g. renin and β-secretase. In addition, we suggest an empirical metrics that might serve as an indicator during lead optimization to qualify compounds as candidates for structural revalidation.

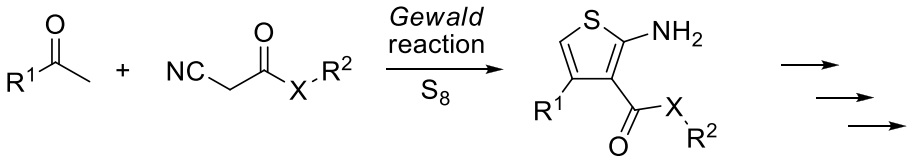
6.3 Introduction, Results and Discussion, Conclusion

In the field of drug discovery and medicinal chemistry, the concept of target-structure-based design considerations in the optimization process from hit-to-lead and (pre-) clinical candidate to an eventually approved drug has been well established over the last three decades.[1,2] Success stories of this structure-guided design approach range from the discovery of various HIV-1 protease inhibitors via the design of the neuraminidase inhibitor oseltamivir, the BCR-Abl kinase inhibitor imatinib, and various HCV protease inhibitors to a virtually unlimited number of current design campaigns that make use of structure-educated decision-making in prioritizing medicinal chemistry activities.[3–6]

This optimization typically starts with one or more identified hits sharing a common basic scaffold. The binding mode of one representative with this scaffold and the target protein is determined by crystallography, and subsequently a series of derivatives is synthesized and their activity is monitored by appropriate assays. Typically, structural modifications are realized assuming that an identified binding mode of a given parent scaffold is maintained during structure-activity relationship (SAR) exploration and optimization. However, despite the meanwhile long history of structure-guided lead discovery there is an ongoing debate on how frequently structural support is required to establish a reliable SAR by monitoring putative changes of ligand binding geometry. Moreover, what criteria can be applied to indicate the requirement to assess putative conservation or changes in the binding mode throughout the optimization campaign?

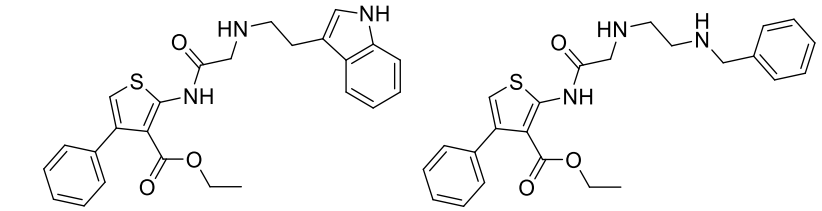
Herein, we follow these questions by utilizing a well-suited model system and suggest advice that may support the decision as to whether the (re)-validation of a binding mode is required.

In lead discovery, the aspartic protease endothiapepsin (EP) commonly serves as a model enzyme and has successfully been exploited as a surrogate for renin and β-secretase in structure-based inhibitor design.[7,8] Based on the *Gewald* reaction (Scheme 6.1),[9] we have synthesized a series of substituted 2-aminothiophene-type EP inhibitors that cover a broad range of affinities from two-digit µm up to submicromolar potency.



Scheme 6.1. General synthetic route to the 2-aminothiophene core through the *Gewald* reaction.

Starting from the two-digit micromolar inhibitors **1** and **2** (Figure 6.1) we embarked on a structure-guided SAR exploration of this ligand series.



1 K_i: 18.2 ± 7.4 µM **2** K_i: 21.2 ± 9.7 µM

Figure 6.1. Chemical structures of **1** and **2** and their affinity data against EP.

The crystal structure of EP in complex with **1**, determined at 1.30 Å resolution, unambiguously shows the ligand's tryptamine moiety bound into the S_1 pocket, while the thiophene core resides between the S_1' and S_2 pockets (Figure 6.2a). Key hydrogen bonds are formed between the ligand's sp^3-hybridized nitrogen and the two aspartates of the catalytic dyad (Asp35 and Asp219), mediated by the lytic water molecule. Likely, this amine is protonated and serves as hydrogen-bond donor to the carbonyl group of Gly221. The indole NH contributes a charge-assisted H-bond to Asp81 located at the tip of the hairpin-type *flap* region. Noteworthy, for the 4-phenyl substituent of the ligand, no properly defined difference electron density could be observed, suggesting a distribution over multiple conformational states in the binding site, and therefore it was omitted from the structural model. Interestingly, the amide NH

and the ester carbonyl oxygen of the ligand establish an intramolecular H-bond, keeping the ester and the thiophene plane approximately coplanar (dihedral angle ca. 18°) and the spatially demanding substituents (ethyl ester and phenyl ring) in an eclipsed-type conformation.

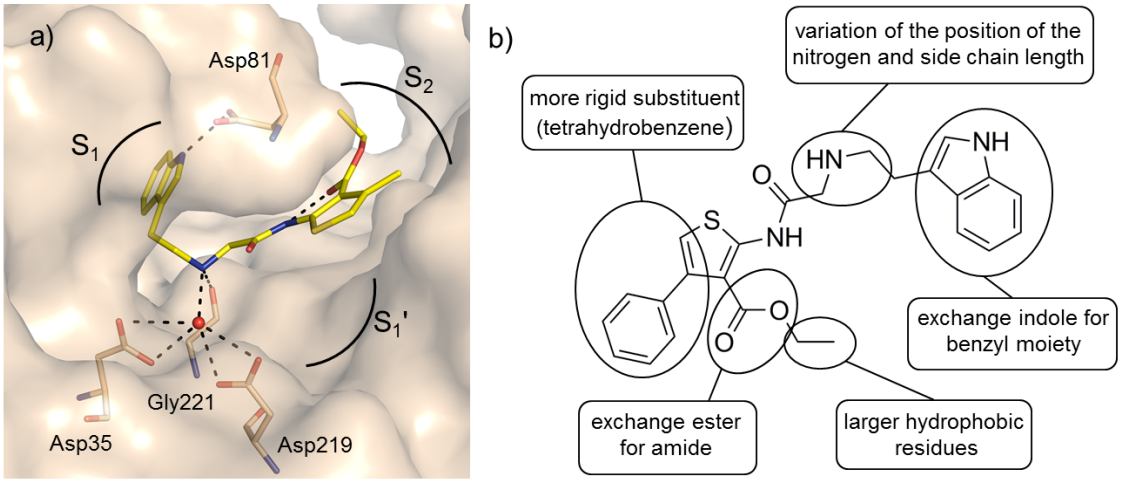


Figure 6.2. a) Observed binding mode of **1** in complex with EP. Amino acids involved in key hydrogen bonds (dashed lines) are shown as rods. For the detailed interaction pattern, see Figure 6.12 (appendix). b) Sites for structural variations derived from compound **1**.

Based on the observed binding mode we designed and synthesized a series of compounds with different structural variations to explore the underlying SAR of this inhibitor scaffold (Figure 6.2b). Via extension of the ethyl ester with larger hydrophobic substituents aimed to address the hydrophobic S_2/S_4 pocket, a gain in affinity was anticipated. In addition, to modulate the electronic properties of the intramolecular H-bond, the influence of an exchange of the ester for an amide moiety was investigated. A thereby induced change in the rotameric state of the carbonyl group might additionally enable a polar contact to Tyr226 (Figure 6.3). Further sites for variation consisted in changing the length of the tryptamine side chain as well as the positioning of the basic nitrogen. The replacement of the indole by morpholine was considered to enhance solubility. Finally, the observation that the 4-phenyl substituent is disordered prompted us to evaluate a more rigid scaffold such as a tetrahydrobenzothiophene.

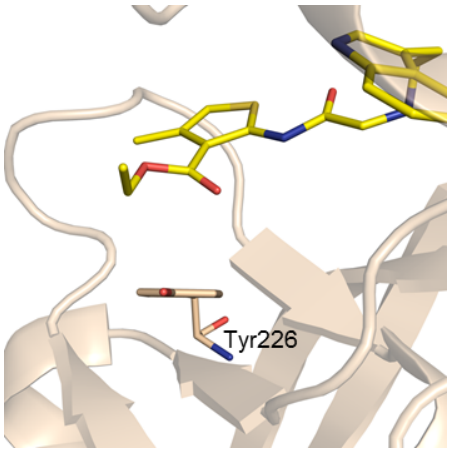


Figure 6.3. The exchange of the ester moiety to an amide in **1** (yellow, stick representation) might enable a polar contact to Tyr226.

The synthesized compounds and their inhibitory activities are shown in Figure 6.4, with **4** and **9** being the most potent representatives (460 and 545 nM, respectively). However, the interpretation of the underlying SAR turned out to be challenging based on the initially determined binding geometry of **1**. Surprisingly, the simple ester-to-amide exchange resulted in an affinity improvement by a factor of five (**3**, 4.0 µM). Is this gain solely attributed to the changed interaction of the introduced amide function, and how does this relate to the intramolecular H-bond observed in **1**? The replacement of the ethyl moiety in **3** by a benzyl substituent leads to the most potent compound of our series, **4**, with a K_i value of 460 nM and thus an improvement in the inhibitory activity of one order of magnitude. Removal of the 4-phenyl substituent and annulation

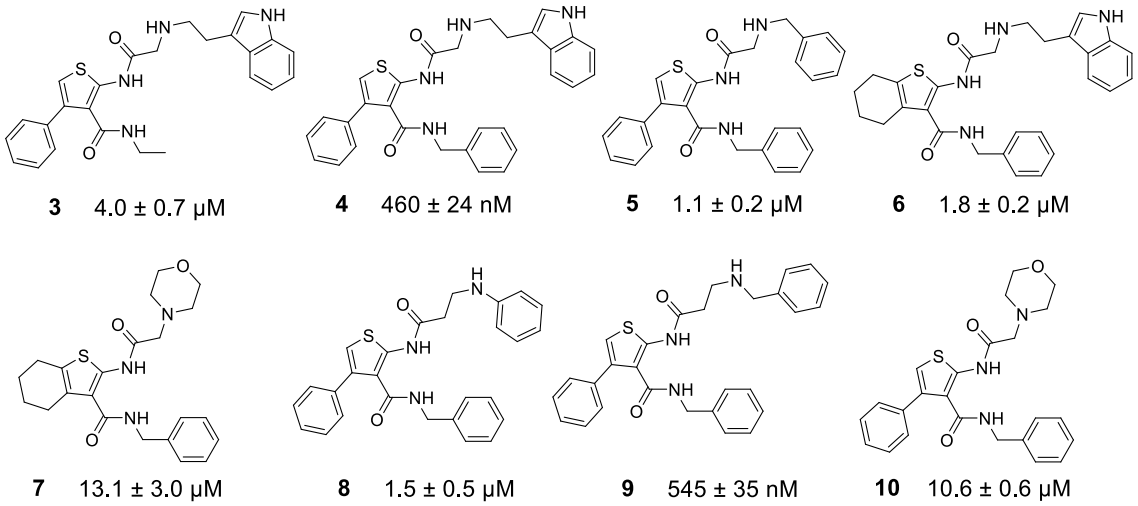


Figure 6.4. Chemical structures of the synthesized compounds **3**-**10** and their corresponding K_i values against EP.

of a tetrahydrobenzene moiety at the [b] site of the thiophene ring (**6**) reduces affinity by about fourfold, although in **1** this substituent appeared disordered and was not expected to contribute significantly to the establishment of a directional interaction to EP. The replacement of the indole moiety (**4**) by a benzyl group (**5**), presumably accompanied by the loss of the charge-assisted H-bond to Asp81, reduces the affinity by approximately twofold (1.1 μM). A positional shift of the aliphatic nitrogen that was involved as a donor in an H-bond to the lytic water resulted in **8** (1.5 μM) and additionally a one-carbon insertion rendered **9**, which, with an activity of 545 nM, was unexpectedly found to be nearly as potent as **4**.

Overall, this rather incongruent SAR prompted us to perform a comprehensive structural analysis of the underlying binding modes by crystal structure analysis. To our surprise, the binding modes of the synthesized ligands successfully determined for **1** and **3**-**9**, split into four clusters of binding poses, and our starting compound **1** turned out to represent a structural singleton, in other words, the sole representative of this binding mode. Figure 6.5a shows the crystal structures determined in this study and the affiliation for each inhibitor to the corresponding cluster, as well as the superposition of one cluster representative with the binding pose of our initial lead **1**. Among the eight crystal structures, all adopting the same crystal packing, the protein itself manifests nearly identical conformation: rmsd values for the all-atom fitting are below 0.5 Å, backbone fitting yielded rmsd values below 0.3 Å.[10,11] Hence, the deviating binding modes are neither caused by significant rearrangements of the protein environment, nor are they induced by different crystal forms that might create a different packing environment. Surprisingly, the two most potent compounds belong to different binding mode clusters. The largest cluster (green) comprises four derivatives (**3**, **4**, **5**, and **6**), followed by a two-membered cluster (**8** and **9**) and two structural singletons (**1** and **7**).

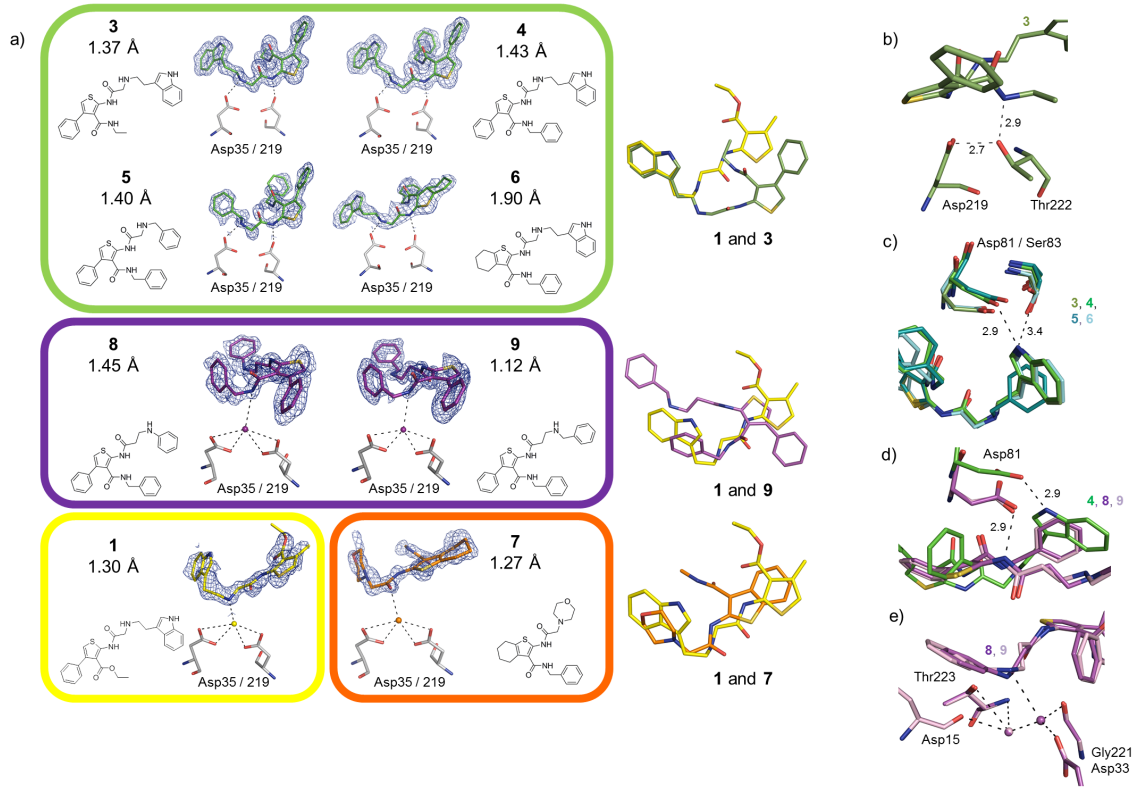


Figure 6.5. a) Left: Binding modes and assignment to binding mode clusters. Green, purple, yellow, and orange frames indicate the four different binding modes. Resolutions of the X-ray structures are shown in Å. The $2F_o$-F_c electron densities are depicted for the ligands as a blue mesh at 1σ. Right: Superposition of **1** (yellow) with one representative of each binding mode cluster in the active site. b) H-bond of the amide NH of **3** to Thr222 Oγ. Likely Thr222 forms an H-bond as an H-bond donor to Asp219 and therefore can only act as an H-bond acceptor to the ligand. Distances in Å. c) H-bond of the indole NH of **3**, **4**, and **6** to Asp81 and Ser83. This interaction is not possible for **5** due to the replacement of the indole by benzyl. d) The charge-assisted H-bond interaction to Asp81, which in the green cluster (here shown for **4**) is formed with the indole NH, is established in **8** and **9** by means of the 2-amino function of the thiophene. e) H-bond network of **8** and **9** formed with the ligands' aniline NH or benzyl NH, respectively. Involved water molecules are shown in spheres.

Noteworthy, in between the four clusters, the binding modes of the inhibitors with respect to the two catalytic aspartates and the occupancy of the remaining binding pockets differ substantially. The lytic water molecule adopts a versatile role: it acts either as an H-bond donor or an H-bond acceptor or is released from the complex. Inhibitor **1** addresses the Asp35/219 with its sp^3-hybridized nitrogen via the lytic water molecule, the latter acting as an H-bond acceptor (Figure 6.5a). In the structures of **8** and **9** (purple cluster) the benzyl-substituted amide nitrogen forms a polar contact to the lytic water molecule, even though the interaction geometry appears rather unfavorable for a typical H-bond. Inhibitor **7** interacts through its morpholinomethylene-substituted carbonyl oxygen mediated by the lytic water molecule with the catalytic

aspartates; in this case, however, the water molecule acts as an H-bond donor. All representatives of the green cluster directly address the catalytic dyad, while the lytic water molecule is released upon binding of the inhibitors. The significantly different binding modes provoke substantially deviating H-bond networks between the inhibitors and the protein, which are shown in Figures 6.10-6.13 (appendix).

The binding mode of **1** apparently relies on the presence of the ester moiety, whereas the amide derivative (**3**) adopts an altered binding mode. Interestingly, the intramolecular H-bond observed in **1** (Figure 6.2a) breaks up in favor of the establishment of an H-bond to Thr222 Oγ through the ethyl-substituted amide nitrogen. How does the mere replacement of the ester by an amide translate into such a predominant change in binding geometry? In all the EP crystal structures determined in our study, the interaction pattern between Thr222 Oγ and the Asp219 carboxylate (2.7 Å) is conserved. Likely, this Thr side chain donates an H-bond to the carboxylate of Asp219 keeping the proton of Thr222 Oγ oriented towards the aspartate (Figure 6.5b). This observation supports the conclusion that Thr222 Oγ itself can only be involved in favorable polar interactions if addressed by appropriate H-bond donors on the ligand site. This is impossible for the ester **1** as its oxygen can only serve as an H-bond acceptor; hence the ester prefers to saturate its polar contact inventory by an intramolecular H-bond. However, when the ester is changed to an amide (**3**), the introduced NH function is able to address Thr222 Oγ as a donor rather than maintaining the intramolecular H-bond, while the concomitantly nonsaturated glycine amide NH rotates by ≈100° to establish an H-bond to the Asp219 carboxylate (Figure 6.5a).

Although the overall SAR interpretation is inconclusive in the absence of structural data, the SAR within the clusters reveals a consistent picture, which will be discussed only exemplarily. In the green cluster, **3** bearing an ethyl group at P_2 has the lowest affinity (4.0 µM); this most likely results from the reduced hydrophobic interactions, compared to those of the inhibitors equipped with a benzyl group at this position. Furthermore, **3** and **6** lack the intramolecular π–π interaction between the two phenyl rings observed in **4** and **5**, which leads to a loss of preorganization corresponding to the bound conformation. In **5**, due to the exchange of the indole moiety for the benzyl group, the H-bond to Asp81 and the weak polar contact to Ser83 as observed in **3**, **4**, and **6** cannot be established (Figure 6.5c). In **4**, the representative with the highest affinity (460 nM) of this cluster, the four varied substituents seem to be optimally chosen among all investigated variations, as the indole moiety is able to form an H-bond to Asp81 whereas the benzyl moiety interacts via π–π stacking to Tyr226.

Inhibitor **8** (purple cluster) differs from **5** (green cluster) only by the position of the secondary amino nitrogen in the P_1 substituent of the inhibitor. Based on the X-ray structure of **5**, a lower affinity caused by a less efficient interaction of the relocated nitrogen to the catalytic aspartates would have been expected for **8**. Interestingly though, its affinity is nearly unchanged. The X-ray structure of **8** provides the explanation for this initially surprising observation: the switch in binding geometry between **5** and **8** seems to be mainly triggered by the loss of the H-bond to Asp35 that would be expected for an unchanged geometry of **5**. The thiophene moiety formerly located in the S_1' pocket (green cluster) flips by about 180° in order to populate now the S_2 pocket. Likewise, inhibitor **9**, in which the P_1 substituent is extended by one carbon atom, shows this altered binding geometry, which enables **8** and **9** to maintain the charge-assisted H-bond interaction to Asp81, although the indole moiety that previously accomplished this task is replaced by the phenyl/benzyl group with the amide NH in thiophene 2-position (Figure 6.5d). Even though the aryl- and benzyl-substituted nitrogens are located at virtually identical positions, the increase in affinity by about threefold from **8** to **9** might be explained by the different pK_a values of the aniline NH compared to the benzyl NH. Thus, the latter, which is being most likely protonated to a larger extent, is capable of forming a more efficient charge-assisted H-bond to the water molecule that mediates the H-bond to Asp33 and Gly221 (Figure 6.5e).

In our inhibitor series also less or inactive compounds were synthesized which are essential for SAR interpretations as well. However, in the present case the interpretation of these data is almost impossible, as the assignment to one of the binding mode clusters remains speculative.

The current example highlights the complexity of binding events and their strong dependence on seemingly minor effects of scaffold decoration and modifications. That **1**, **7**, **8**, and **9** do not adopt the same binding geometry as observed in the green cluster can be rationalized retrospectively, but without the crystal structure determinations the adopted binding modes would have been difficult to predict, and, if so, would have hardly been believed or attracted sustained attention without our experimental evidence.

Such an unexpected flipping of binding modes after modification of the scaffold substitutions has been detected sporadically, however, mostly by serendipity.[12–15] As the commonly accepted hypothesis that similar ligands bind in a similar fashion[16] apparently does not hold true on a comprehensive basis, the question arises, how frequently a (re-)validation of an assumed binding mode is required. Are there any

easily available indicative hints to estimate whether a crystallographic (re-)validation of the assumed binding mode used as working hypothesis is required throughout the lead exploration campaign?

In addition to the affinity determination via a fluorescence-based assay (K_i values), we performed a thermal shift assay (TSA) to investigate the extent to which the studied compounds stabilize the protein, especially with respect to the different binding modes. TSAs are commonly used for hit identification and affinity ranking of inhibitors.[17–21]

The results for our inhibitor series (Table 6.1) clearly suggest that the observed shifts in the melting temperature do not generally correlate with the K_i values.

Table 6.1. TSA data ordered by decreasing ΔT_m values, as well as assignment to the binding mode clusters (color coded) and K_i values.

inhibitor [100 µM]	T_m [°C]	ΔT_m [°C]	K_i
DMSO	58.7 ± 0.2		
4	63.3 ± 0.2	4.6	460 ± 24 nM
9	62.5 ± 0.0	3.8	545 ± 35 nM
6	61.9 ± 0.1	3.2	1.8 ± 0.2 µM
3	61.5 ± 0.2	2.8	4.0 ± 0.7 µM
5	60.9 ± 0.1	2.2	1.1 ± 0.2 µM
1	59.4 ± 0.2	0.7	18.2 ± 7.4 µM
8	59.1 ± 0.1	0.4	1.5 ± 0.5 µM
7	58.8 ± 0.1	0.1	13.1 ± 3.0 µM

However, within one binding mode cluster, the ΔT_m values are ranked correctly, with **5** being an exception: this might be attributed to the loss of the charge-assisted H-bond to Asp81 due to the absence of the indole moiety, while the other members of this cluster establish this interaction. This observation is in agreement with the concept that TSA measurements require similar binding enthalpies for correct affinity ranking.[21] However, how do TSA data perform for ligands of identical chemotype that fall in the same affinity range, but exhibit different binding geometries?

Only two of our four binding mode clusters qualify for such a comparison, as the singletons **1** and **7** are slightly off-range with respect to their affinity. However, the affinities of the green and the purple clusters cover a similar affinity range thus permitting such a comparison. Interestingly, for these two clusters, ligands of rather identical affinity deviate with respect to their associated TSA profile: the two most

potent ligands **4** and **9** possess affinities of 460 nM and 545 nM, respectively, but differ by 0.8°C in their TSA profiles (4.6 versus 3.8°C) which corresponds to a deviation of more than threefold of the standard deviation of the measurement. Analogously, **5** and **6** inhibit EP with a virtually identical affinity of 1.1 and 1.8 µM, respectively, and **8** is equally potent (1.5 µM). However, the latter exhibits a TSA shift of only 0.4°C, while the first two ligands significantly enhance the stability of the protein as indicated by a shift of ΔT_m of 2.2 and 3.2°C, respectively (with **5** being the abovementioned internal outlier within the green cluster). These observations suggest that protein stabilization by a particular ligand depends not only on affinity, but also on its adopted interaction pattern. This is also true for the comparison of **1** and **8**, both of which exhibit a similar TSA shift (0.7 vs. 0.4°C, respectively) but possess a substantial difference in affinity (18 vs. 1.5 µM), thus suggesting that they stabilize the target protein by deviating interaction geometries.

Hence, discrepancies between affinity data based on enzyme kinetics and TSA results within one inhibitor series may be exploited as a hint for putative changes in the adopted binding geometry, which may qualify the seeked “outlier”-type ligands as promising candidates for revalidation of the assumed binding mode.

The presented example impressively points out that minor structural changes within one lead series, such as the replacement of an ester by an amide or the introduction of an additional carbon atom, may result in entirely different binding modes. An inconsistent affinity-TSA correlation may serve as an easily available indicator to qualify candidates for structural revalidation, even if the interaction mode is believed to be known. In the absence of such structural data, a reasonable interpretation of substituent effects or assumed bioisosteric replacements at a given chemotype will be rather meaningless and will likely drive scaffold optimization in the wrong direction. We expect that our case study and the application of the suggested validation protocol will encourage drug designers to further pursue and study chemically attractive scaffolds albeit an at first glance seemingly inconclusive SAR. This is likely to be rewarded with a more comprehensive understanding and an increasing number of starting options for further lead optimization.

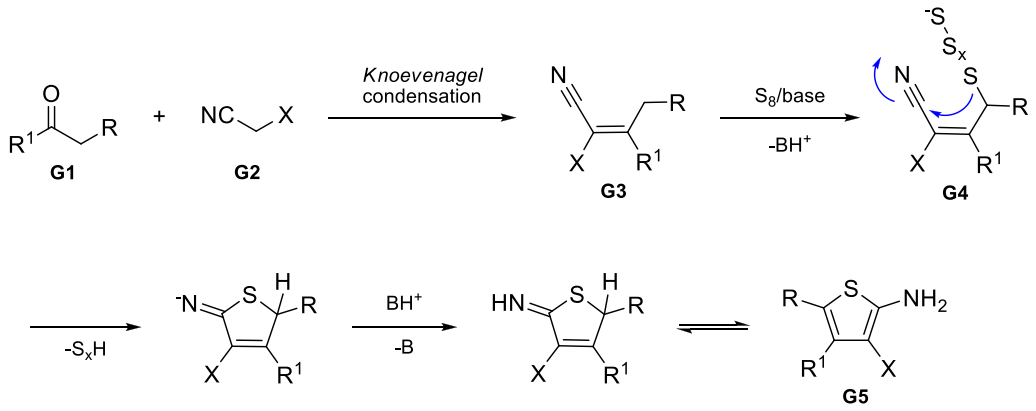
Protein Databank codes for this manuscript:
1 (3T7Q), **3** (4L6B), **4** (3PSY), **5** (3WZ6), **6** (3WZ7), **7** (3T7X), **8** (3WZ8), **9** (4LAP).

6.4 Detailed Discussion of the Synthesis of *Gewald* Reaction-Based Inhibitors

As the manuscript as published in *Angewandte Chemie* is mainly related to the lessons learned for drug discovery and lead optimization, the synthesis of the compounds is described in more detail in this subchapter.

6.4.1 *Gewald* Reaction

The *Gewald* reaction, named after the German chemist Karl Gewald, is an organic name reaction in which the treatment of an α-methylen-activated nitrile with an α-methylene carbonyl compound and elemental sulfur gives rise to substituted 2-aminothiophenes.[9] In the first step, *Knoevenagel* condensation of the CH-acidic α-carbon of nitrile **G2** with the carbonyl function of **G1** furnishes the acrylonitrile intermediate **G3**. The subsequent mechanism finally leading to the 2-aminothiophene scaffold **G5** is yet not fully understood, but it most likely proceeds via the thiolated intermediate **G4**, followed by its ring closure and prototropic rearrangement to **G5** in the last step (Scheme 6.2).[22,23]



Scheme 6.2. Synthesis of the 2-aminothiophene scaffold by the *Gewald* reaction.

The 2-aminothiophene scaffold has also attracted some interest in the field of drug discovery: the atypical antipsychotic drug olanzapine (Eli Lilly) represents a well-known successful example of a *Gewald* reaction-derived small molecule being approved to the market. In addition, the 2-aminothiophene scaffold can also be found in the non-steroidal anti-inflammatory drug tinoridine (nonflamin®). Other drug candidates like T-62, a selective allosteric enhancer of the adenosine A1 receptor (neuropathic pain), or AX20017, an inhibitor of the protein kinase G (tuberculosis), also bear this structure motif. (Figure 6.6).[23,24]

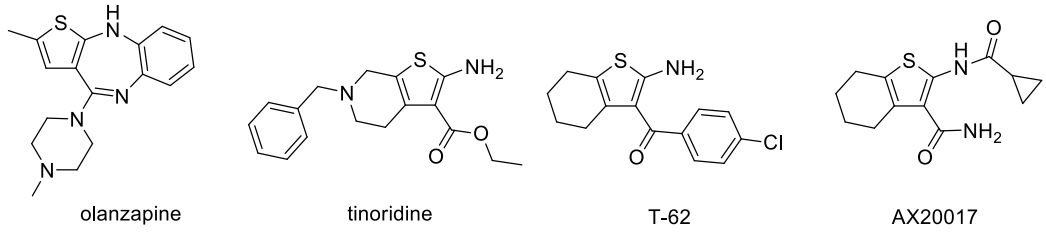


Figure 6.6. Chemical structures of 2-aminothiophene compounds in drug research.

6.4.2 Previous Work

The *Gewald* reaction-based compounds synthesized previously in our research group (Dr. Ruben Bartholomäus and Amir Shahim) all show the common basic scaffold shown below (Figure 6.7).

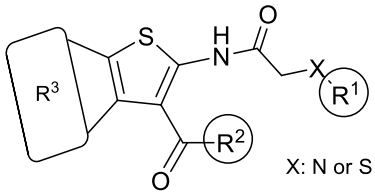


Figure 6.7. Basic scaffold of the 2-aminothiophene compounds synthesized as aspartic protease inhibitors in our group.

Within the scope of this thesis as well as with respect to the motivation described in chapter 1.5 and 6.3, seven *Gewald* reaction-based inhibitors were synthesized (Figure 6.8).

Except for **3**, all inhibitors bear a benzylamine moiety in R^2-position and are varied in their R^1- and/or R^3-substituents. While the R^3-substituent is introduced via the ketone component at the beginning of the synthesis, a synthetic route allowing the introduction of the variable R^1-substituent in a later step was envisioned to be advantageous as it would reduce the number of intermediates to be characterized.

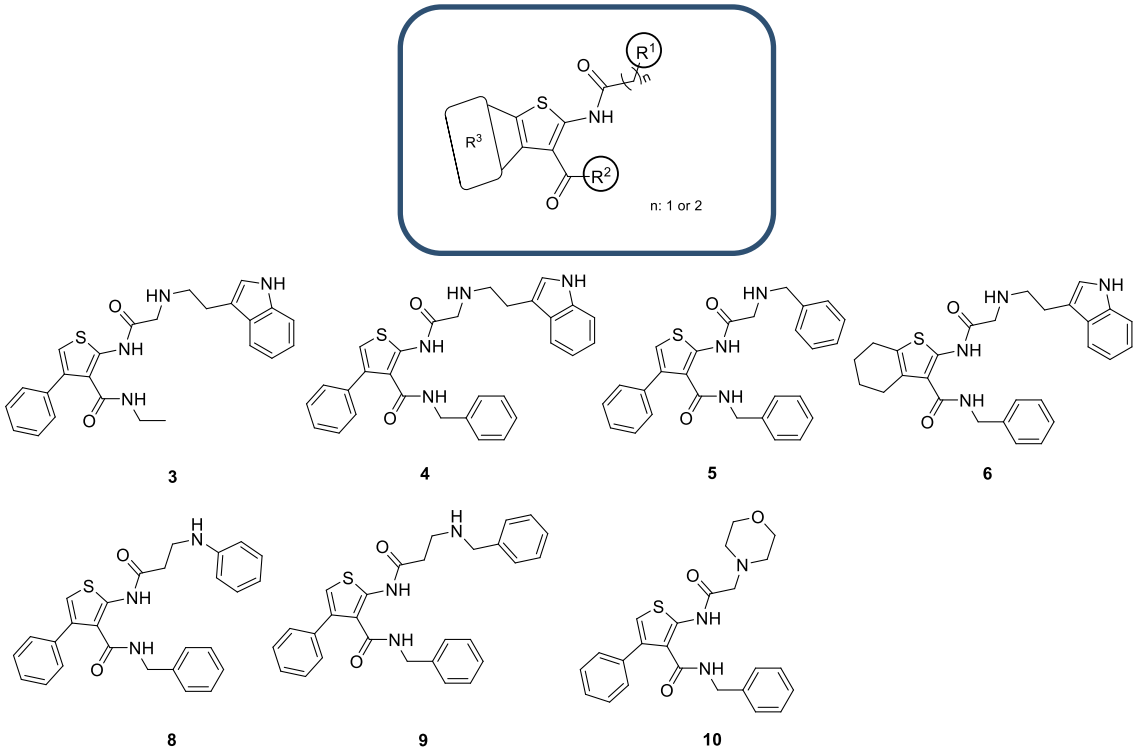
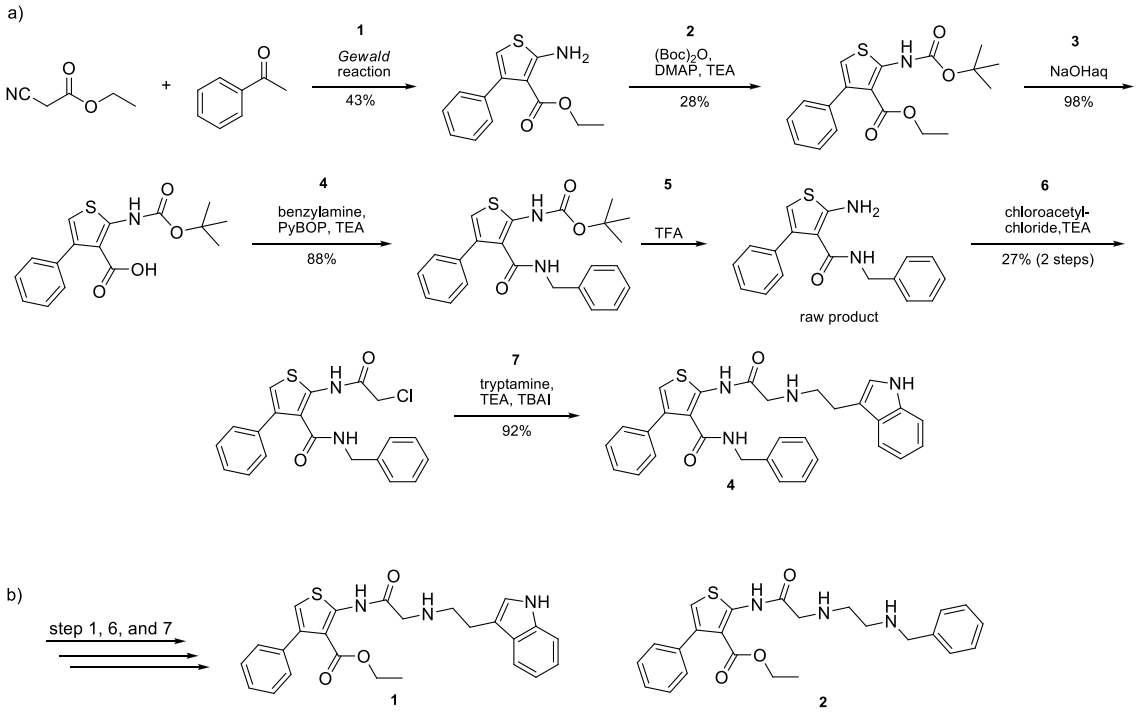


Figure 6.8. Chemical structures of compounds **3**-**6** and **8**-**10**, which were synthesized within the scope of this thesis.

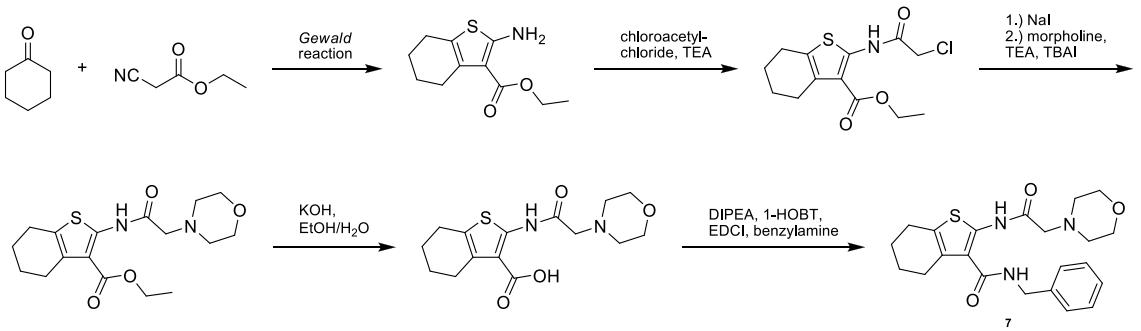
During his Diploma thesis,[25] Dr. Ruben Bartholomäus developed a synthetic strategy for the synthesis of *inter alia* compounds **1**, **2** and **4** (Scheme 6.3). This route, leading in three consecutive steps to compounds **1** and **2** and in seven steps to compound **4**, is quite straightforward for synthesizing inhibitors which bear an ethylester function in R^2-position and vary in their R^1-substituent, such as in compounds **1** and **2**. The synthesis of compound **4** being equipped with a benzylamine moiety in R^2-position, however, was intricate as it included a low-yielding Boc-protection-step of the aromatic amine at the beginning of the synthesis (step 2), and hence an additional deprotection step at a later stage of the protocol.

In conclusion, this route is quite feasible for the synthesis of inhibitors featuring an ethylester functionality as R^2-residue, but for the introduction of other R^2-residues, like e.g. a benzylamine group or the variation of the R^1-substituent, it is less suitable.



Scheme 6.3. Synthetic route for compound **4** (a) as well as **1** and **2** (b) developed by Dr. R. Bartholomäus.[25] Compounds **1** and **2** were synthesized analogously, however, the reaction steps 2, 3, 4, and 5 for introducing the R^2-substituent were skipped and in step 7 the respective amines were used instead of tryptamine.

As shown for compound **7** (Scheme 6.4), Amir Shahim synthesized derivatives with a morpholino residue in R^1- and varying substituents in R^2-position. Therefore, the synthesis of these compounds was designed in such a way that enabled the introduction of the R^1-substituent at first and the variation of the R^2-moiety in the last step, which also disqualified this route for the efficient synthesis of the inhibitor series with modified R^1-moieties planned to be synthesized at the onset of this thesis (Figure 6.8).

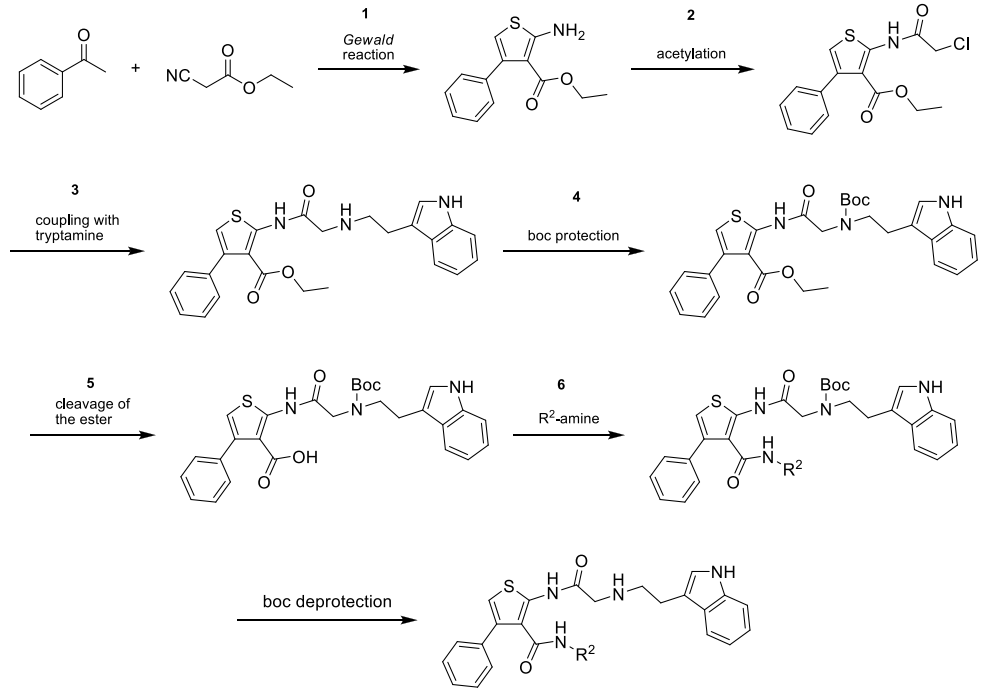


Scheme 6.4. Synthesis route developed by Amir Shahim for synthesizing compound **7**.

A previous synthesis project carried out during my practical year prior to the initiation of this thesis also comprised *Gewald* reaction-based inhibitors: a series of inhibitors with variation of the R^2-moiety was planned to be synthesized as inhibitors for aspartic proteases, while the tryptamine moiety as R^1-substituent should be present in all inhibitors. For synthetic reasons, in this case it was considered to be favorable to introduce the common tryptamine R^1-substituent at first and the different variable R^2-residues at a later stage. The proposed route of the synthesis is shown in Scheme 6.5.

Interestingly, in practice this synthetic pathway, however, beard unexpected challenges: the cleavage of the ester (step 5) for instance resulted in an additional cleavage of the amide function (main product) and therefore in the loss of the R^1-substituent. Also different reaction conditions (varying bases and solvents) for the ester cleavage did not give rise to the desired product. Pursuit of other protection strategies instead of the Boc-protection like usage of the benzyl-, triphenylmethyl- or *p*-nitrobenzyl group did not lead to the final compound due to difficulties in the amine protection (*p*-nitrobenzyl protection group), cleavage of the ester after protection (triphenylmethyl group) or the final deprotection of the amine function (benzyl protection group).

From a synthetic point of view, the described challenges were rather unexpected, especially when keeping in mind that a similar route (Scheme 6.4) worked well for its closely related scaffold.



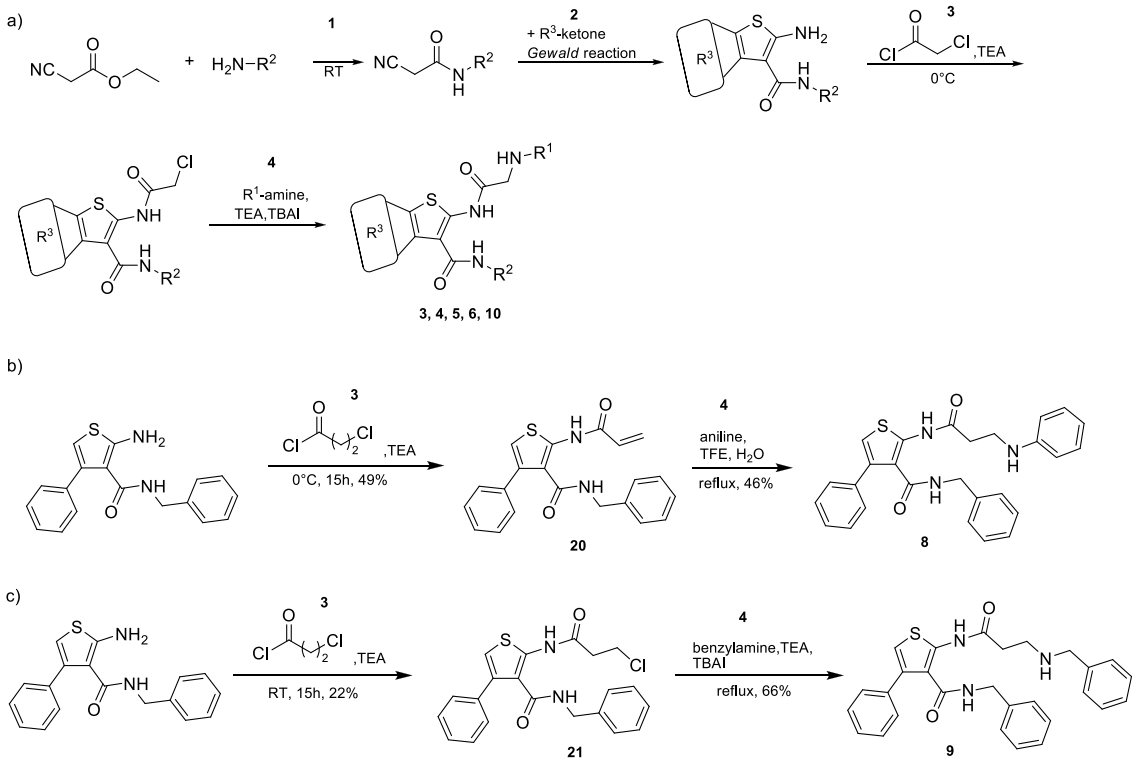
Scheme 6.5. Proposed synthesis route for the inhibitor series with the tryptamine moiety as R^1-substituent and variable R^2-residues.

6.4.3 Synthesis of Compounds 3-6 and 8-10

Based on the previous results, an appropriate route for synthesizing the above mentioned compounds **3**-**6** and **8**-**10** had first to be developed in order to efficiently enable the incorporation of a broader variety of R^1-substituents such as e.g. tryptamine, benzylamine and morpholine moieties.

First of all, the optimization of the Boc-protection (step 2, Scheme 6.3) of the aromatic amino functionality of the thiophene was pursued in order to follow up on the synthesis via the established route shown in Scheme 6.3. Different bases and catalysts like triethylamine (TEA) and 4-dimethylaminopyridine (DMAP) as well as different solvents were tested for optimizing the formation of the Boc-protected 2-aminothiophene, but, unfortunately, all these attempts did not enhance the yield of this step. Efforts like the iodine-catalyzed solvent-free Boc-protection[26] or the usage of guanidine-hydrochloride as an alternative catalyst[27] also did not result in a satisfactory conversion of the 2-aminothiophene starting material into the desired product.

Fortunately, the failure of optimizing the initial sequence lead to an elegant and efficient route (Scheme 6.6a) designed in the context of this work, which was successfully employed for most of the seven compounds to be synthesized.



Scheme 6.6. a) Overview of the synthesis of compounds **3**-**6** and **10**. b) The synthesis of **8** is similar, but proceeds *via* the acrylamide derivative **20** as intermediate product of step 3 (dehydrohalogenation). c) Synthesis of **9** *via* the 3-chloro-propanamide derivative **21** as intermediate product.

The *Gewald* reaction (step 2) was directly performed with the final R^2-moiety, being synthesized from ethylcyanoacetate and the respective amine prior to the *Gewald* reaction (step 1). Thus, the cleavage of the ester moiety and protection/deprotection of the reactive amino group was no longer required. The R^3-residue was also directly introduced via the *Gewald* reaction. Subsequent reaction of the 2-aminothiophene derivative with chloroacetyl chloride/ chloropropionyl chloride (step 3), followed by the introduction of various amines by a nucleophilic substitution of the terminal chloro-substituent (step 4) finally yielded the fully decorated inhibitors. In comparison to the initial route, the newly developed sequence significantly reduced the required steps of the synthesis of **4** from seven to four and concomitantly resulted in a by a factor of five increased overall yield (3 % over 7 steps versus 14 % over 4 steps). Compounds **3**, **5**, and **6** were synthesized analogously to **4**, while for compound **6**, cyclohexanone instead of acetophenone was used as ketone-component in the *Gewald* reaction.

The synthesis of compounds **8** and **9** (Scheme 6.6b, c) slightly deviates from the previous one because the chloropropionyl chloride turned out to be susceptible to terminal dehydrohalogenation giving rise to the respective acrylamide derivative **20** (step 3). In analogy to the synthesis of **3**, **4**, **5**, and **6**, **9** could be obtained *via* the 3-chloro-propanamide derivative, while for the generation of **8** the acrylamide intermediate **20** was subjected to a *Michael reaction* with aniline (step 4) rendering the final compound. The 3-chloro-propanamide- and the acrylamide-intermediates, **21** and **20** respectively, were isolated from the reaction mixture obtained under the denoted reaction conditions (room temperature and 0°C, respectively), however both reactions were not optimized for obtaining the respective intermediate product.

6.4.4 Conclusion

In summary, an efficient and successful synthetic route has been developed for the synthesis of compounds **3**-**6** and **8**-**10**. In this context it is noteworthy to mention that the chemical scaffold of 2-amino-3-carbonyl-thiophenes has recently been assigned to belong to the group of pan-assay interference compounds (PAINs)[28] and has therefore even been retracted from certain chemical libraries. However, as convincingly emphasized in our study, this structural motif nevertheless is able to establish specific interactions and inhibition mechanisms to enzyme targets thus qualifying a removal of such a particular scaffold to be seemingly premature. The utilization of the *Gewald* reaction as pursued in this thesis has successfully been exploited in order to make novel derivatives of this chemical compound class available.

6.5 Experimental Section

6.5.1 Synthesis - General Experimental Details

Unless stated otherwise, reagents and solvents that were purchased were used without further purification. Cyclohexane was purchased in p.a. quality (Grüssing, Germany) and distilled prior to use. Thin layer chromatography (TLC) was performed on precoated TLC plates purchased from Merck (silica gel 60 F_{254}). Purification by flash chromatography was either performed manually with glass columns (silica gel 60, 0.04-0.063mm, Macherey-Nagel) or with a Büchi separation system using prepacked flash chromatography columns purchased from Interchim (PuriFlash colums). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a JEOL ECX-400 or on a JEOL ECA-500 spectrometer. Chemical shifts (δ) are given in ppm with the residual solvent signal (unless otherwise noted) or TMS (Tetramethylsilane) signal used as reference [^{1}H: $CDCl_3$: s, δ=7.26ppm, DMSO-*d6*: quint, δ=2.50ppm; ^{13}C: $CDCl_3$: t, δ=77.16ppm, DMSO-*d6*: quint, δ=39.52ppm); MeOH-*d4*: sept, δ=49.00ppm; TMS: s, δ=0.0ppm].[29] Spectra with $CDCl_3$ as solvent were recorded at room temperature (RT) while spectra with DMSO-*d6* as solvent were recorded at 30.0°C. Coupling constants (*J*) are reported in hertz (Hz). Peak patterns were abbreviated as follows: s (singlet), d (doublet), dd (double doublet), ddd (doublet of doublet of doublet), t (triplet), m (multiplet), sm (symmetric multiplet), br (broad), ps (pseudo). Elemental combustion analyses were performed on a Vario MICRO cube (Elementar Analysensysteme GmbH, Hanau, Germany). Mass spectra were recorded on a double-focusing sector field spectrometer type 70-70H (VacuumGenerators) or on a double-focusing sector field spectrometer type AutoSpec (Micromass) using electron ionization (EI) or electrospray ionization (ESI). Melting points were determined using the melting point meter MPM-H2 (Schorpp Gerätetechnik, Germany) or FP62 (Mettler Toledo) and are uncorrected.

6.5.2 Synthesis - General Procedures

Procedure A: *Gewald* Reaction - Synthesis of the 2-Amino-Thiophene Scaffold

The *Gewald* reaction was performed following a slightly modified literature procedure:[30] To an ice-cooled solution of the respective ketone and nitrile dissolved in dry tetrahydrofuran (THF) or dichloromethane (DCM) titanium tetrachloride ($TiCl_4$) (2eq) was added dropwise. Pyridine (3.4eq) was added stepwise after 10min (1/3) and after 30min (2/3) and the reaction mixture was stirred at RT until full conversion of the ketone was observed, (≈15h), indicated by TLC. The solvent was removed in vacuo, the remaining residue was dissolved in ethyl acetate and washed with 2M HCl, water, and brine, dried over $MgSO_4$, filtered, concentrated in vacuo and finally redissolved in

dry THF. Sulfur (1.17eq) and diethylamine (3.5eq) were added and the reaction mixture was stirred at RT until TLC indicated the conversion of the intermediate (≈15h). THF was removed in vacuo, the residue was dissolved in DCM and washed with water and brine, dried over $MgSO_4$, filtered, and concentrated in vacuo. The remaining oily residue was purified by flash chromatography.

Procedure B: Acetylation of the 2-Amino-Thiophene Scaffold

The acetylation was performed in accordance to a publication from Kobayashi-Matsunaga *et al*.:[31]

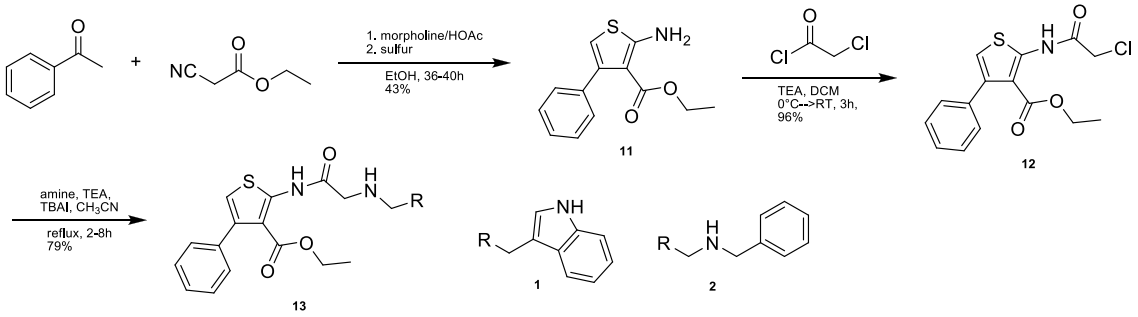
To an ice-cooled solution of the respective thiophene (1eq) in dry DCM, triethylamine (1.5eq) and chloroacetyl chloride or chloropropionyl chloride (1.5eq) were added dropwise. The reaction mixture was stirred at room temperature until TLC indicated full conversion of the starting material (4-14h). The reaction mixture was diluted with DCM, washed with an aqueous ammonium chloride solution (sat.) and brine, dried over $MgSO_4$, filtered, concentrated in vacuo and the remaining residue was purified by flash chromatography.

Procedure C: Nucleophilic Substitution with Primary Amines or Morpholine

According to a procedure from Jung *et al*.,[32] triethylamine (2eq), the respective amine (2-5eq) and tetrabutylammonium iodide (TBAI) (0.1eq) were added to a solution of the respective *N*-chloroacetyl [or *N*-chloropropionyl]-2-aminothiophene (1eq) in acetonitrile. The reaction mixture was stirred for 4-15h at 100°C (reflux) and concentrated in vacuo. The residue was dissolved in DCM, washed with water and brine, dried over $MgSO_4$, filtered, and concentrated in vacuo. The product was obtained after purification by flash chromatography.

6.5.3 Synthesis – Reaction Schemes and Characterization of the Synthesized Compounds

Synthesis of compounds 1 and 2*

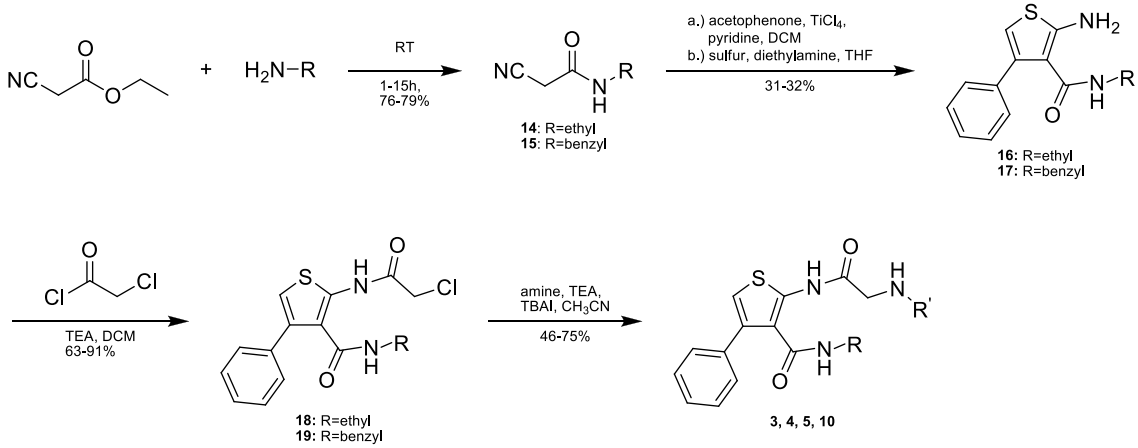


Scheme 6.7. Synthesis of compound **1** and **2**.

*) Compounds **1** and **2** were synthesized by Dr. Ruben Bartholomäus.

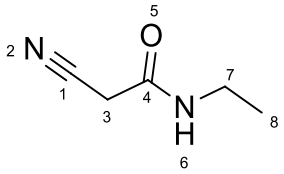
The characterization of **11**, **12**, **1**, and **2** as well as the experimental procedures can be found in the diploma thesis of Dr. Ruben Bartholomäus, Philipps-Universität Marburg and in the original publication.[25,33]

Synthesis of Compounds 3, 4, 5, 10



Scheme 6.8. Synthesis of **3**, **4**, **5**, **10**.

2-Cyano-*N*-ethyl-acetamide (14)



The compound was synthesized according to a literature procedure[34] utilizing ethylamine (68% solution in H_2O) (5mL, 61.1mmol) and ethyl cyanoacetate (3.25mL, 30.5mmol). The product was purified by flash chromatography (DCM, 2% MeOH) and **14** was obtained as colorless solid (2.7g, 24.1mmol, 79%).

^{1}H-NMR (399.78MHz, DMSO-*d6*):
δ=8.18 (brs, 1H, H-N(6)), 3.57 (s, 2H, H-C(3)), 3.09 (dq, ^3J=7.3Hz, ^3J=5.5Hz, 2H, H-C(7)), 1.03 (t, ^3J=7.2Hz, 3H, H-C(8))

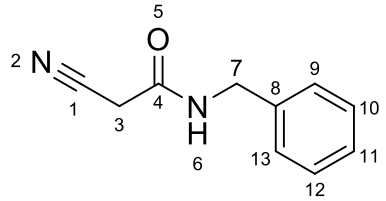
^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):
δ=161.7 (C(4)),116.2 (C(1)), 34.0 (C(7)), 25.2 (C(3)), 14.3 (C(8))

MS (ES+):
m/z (%): 113 (100) $[M+H]^+$, 130 (34) $[M+NH_4]^+$

CHN:

Anal. calcd for $C_5H_8N_2O$ (%):	C: 53.56	H: 7.19	N: 24.98
Found:	C: 54.01	H: 7.26	N: 24.77

***N*-Benzyl-2-cyano-acetamide (15)**



The compound was synthesized according to a literature procedure[35] employing benzylamine (10mL, 91.6mmol) and ethyl cyanoacetate (9.8mL, 91.6mmol). The product was recrystallized from ethanol. The obtained crystals were washed with cooled ethanol, the crystallization procedure was repeated three times from mother liquor. **15** was obtained as colorless crystals (12.05g, 69mmol, 76%).

^{1}H-NMR (399.79MHz, $CDCl_3$):

δ=7.40-7.27 (m, 5H, 5 phenyl-H), 6.49 (brs, 1H, H-N(6)), 4.46 (d, ^{3}J=5.5Hz, 2H, H-C(7)), 3.38 (s, 2H, H-C(3))

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):

δ=162.1 (C(4)), 138.5 (C(8)), 128.3 (2C, C(10), C(12)), 127.3 (2C, C(9), C(13)), 127.0 (C(11)), 116.1 (C(1)), 42.7 (C(7)), 25.3 (C(3))

MS (ES+):

m/z (%): 192 (100) $[M+NH_4]^+$, 366 (89) $[2M+NH_4]^+$, 371 (49) $[2M+Na]^+$

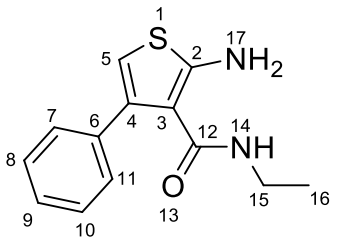
HRMS (EI):

Calcd for $C_{10}H_{10}N_2O$ $[M]^+$:	174.079313
Found:	174.078649

CHN:

Anal. calcd for $C_{10}H_{10}N_2O$ (%):	C: 68.95	H: 5.79	N: 16.08
Found:	C: 69.06	H: 5.71	N: 16.24

2-Amino-*N*-ethyl-4-phenyl-thiophene-3-carboxamide (16)



Compound **16** was synthesized according to procedure A, however, using dry DCM instead of THF in the first step. Utilization of 2-cyano-*N*-ethyl-acetamide (**15**) (1.6g, 14.3mmol), acetophenone (1.53mL, 13.0mmol), $TiCl_4$ (2.86mL, 26.0mmol), sulfur (0.49g, 15.2mmol), pyridine (3.6mL, 44.1mmol), and diethylamine (4.75mL, 45.5mmol) rendered after two purification steps by flash chromatography (1. cyclohexane:EtOAc 85:15, 2. DCM:MeOH 99:1) **16** as orange-brown solid (0.98g, 4.0mmol, 31%).

Mp: 135°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):

δ=7.42-7.30 (m, 5H, 5 phenyl-H), 6.74 (s, 2H, H-N(17)), 6.26 (s, 1H, H-C(5)), 5.98 (brt, ^{3}J=5.6Hz, 1H, H-N(14)), 3.02 (sm, 2H, H-C(15)), 0.80 (t, ^{3}J=7.2Hz, 3H, H-C(16))

^{13}C-NMR (100.53 Hz, DMSO-*d6*, APT):

δ=164.9, 159.7 (C(12), C(2)), 138.8, 136.7 (C(6), C(4)), 128.4 (2C, C(8), C(10)), 128.2 (2C, C(7), C(11)), 127.3 (C(9)), 108.6 (C(3)), 105.1 (C(5)), 33.2 (C(15)), 14.3 (C(16))

MS (ES+):

m/z (%): 247 (100) $[M+H]^+$), 202 (86) $[M-C_2H_6N]^+$

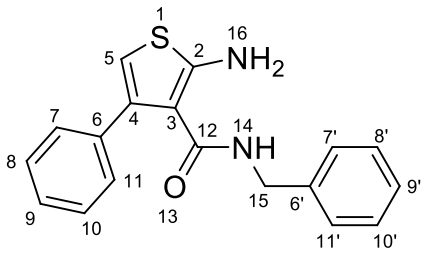
HRMS (EI):

Calcd for $C_{13}H_{14}N_2OS$ $[M]^+$:	246.082685
Found:	246.082233

CHN:

Anal. calcd for $C_{13}H_{14}N_2OS$ (%):	C: 63.39	H: 5.73	N: 11.37
Found:	C: 63.72	H: 5.88	N: 11.47

2-Amino-*N*-benzyl-4-phenyl-thiophene-3-carboxamide (17)



According to procedure A, utilization of acetophenone (5.3mL, 45.3mmol), *N*-benzyl-2-cyano-acetamide (**15**) (8.69mg, 49.9mmol), $TiCl_4$ (10mL, 90.7mmol), pyridine (12.5mL, 154.2mmol), and diethylamine (16.6mL, 159mmol) rendered, after two purification steps by flash chromatography (1. DCM:cyclohexane 8:2; 2. cyclohexane:EtOAc 8:2), compound **17**, as a light yellow solid (4.5g, 14.6mmol, 32%).

Mp: 121°C

^{1}H-NMR (500.16MHz, DMSO-*d6*):

δ=7.29 (pss, 5H, H-C(6-11)), 7.26-7.23 (m, 2H, H-C(8'), H-C(10')), 7.21-7.18 (m, 1H, H-C(9')), 7.04 (d, ^{3}J=7.2Hz, 2H, H-C(7'), H-C(11')), 6.82 (s, 2H, H-N(16)), 6.50 (t, ^{3}J=5.4 Hz, 1H, H-N(14)), 6.25 (s, 1H, H-C(5)), 4.22 (d, ^{3}J=5.7Hz, 2H, H-C(15))

^{13}C-NMR (125.77MHz, DMSO-*d6*):

δ=165.2, 160.1 (C(2), C(12)), 139.0, 138.8, 136.7 (C(6'), C(6), C(4)), 128.3 (2C, C(8), C(10)), 128.2 (2C, C(7), C(11)), 128.0 (2C, C(8'), C(10')), 127.23 (C(9)), 127.17 (2C, C(7'), C(11')), 126.6 (C(9')), 108.3 (C(3)), 105.3 (C(5)), 42.3 (C(15))

MS (ES+):

m/z (%): 309 (100) $[M+H]^+$, 331 (36) $[M+Na]^+$, 617 (38) $[2M+H]^+$

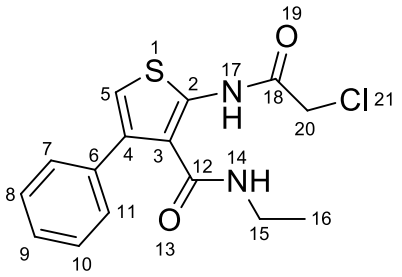
HRMS (EI):

Calcd for $C_{18}H_{16}N_2OS$ $[M]^+$:	308.098335
Found:	308.099874

CNH:

Anal. calcd for $C_{18}H_{16}N_2OS$ (%):	C: 70.10	H: 5.23	N: 9.08
Found:	C: 70.19	H: 5.37	N: 9.03

2-[(2-Chloroacetyl)amino]-*N*-ethyl-4-phenyl-thiophene-3-carboxamide (18)



According to procedure B, usage of 2-amino-*N*-ethyl-4-phenyl-thiophene-3-carboxamide (**16**) (0.8g, 3.3mmol), chloroacetyl chloride (0.4mL, 4.9mmol), and triethylamine (1.1mL, 8.1mmol) yielded after purification by flash chromatography (cyclohexane:EtOAc 9:1) **18** as light yellow solid (660mg, 2.0mmol, 63%).

Mp: 145°

^{1}H-NMR (399.79MHz, DMSO-*d6*):

δ=11.50 (s, 1H, H-N(17)), 7.43-7.33 (m, 5H, 5 phenyl-H), 7.23 (brt, ^{3}J=4.9Hz, 1H, H-N(14)), 7.01 (s, 1H, H-C(5)), 4.51 (s, 2H, H-C(20)), 3.14 (sm, 2H, H-C(15)), 0.89 (t, ^{3}J=7.3Hz, 3H, H-C(16))

^{13}C-NMR (100.53MHz, DMSO-*d6,* APT):

δ=164.03, 163.99 (C(12), C(2)), 141.0 (C(18)), 137.4, 135.5 (C(4), C(6)), 128.3 (4C, C(7), C(8), C(10), C(11)), 127.5 (C(9)), 119.4 (C(3)), 115.8 (C(5)), 42.4 (C(20)), 33.7 (C(15)), 13.9 (C(16))

MS (ES+):

m/z (%) = 323 (100) $[M^{35Cl}+H]^{+}$, 325 (43) $[M^{37Cl}+H]^{+}$

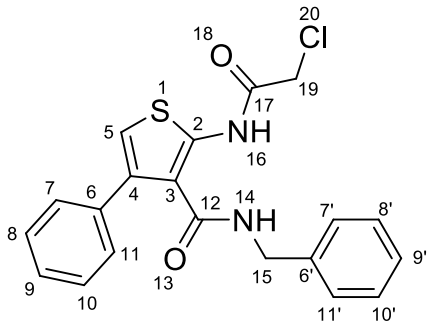
HRMS (ES+):

Calcd for $C_{15}H_{16}N_2O_2SCl^{35}$ $[M+H]^{+}$:	323.062102
Found:	323.060514

CHN:

Anal. calcd for $C_{15}H_{15}N_2O_2SCl$ (%):	C: 55.81	H: 4.68	N: 8.68
Found:	C: 56.22	H: 4.81	N: 8.66

***N*-Benzyl-2-[(2-chloroacetyl)amino]-4-phenyl-thiophene-3-carboxamide (19)**



According to procedure B, utilization of 2-amino-*N*-benzyl-4-phenyl-thiophene-3-carboxamide (**17**) (2.5g, 8.11mmol), chloroacetyl chloride (0.97mL, 12.16mmol), and triethylamine (1.69mL, 12.16mmol) rendered, after purification by flash chromatography (cyclohexane:DCM 7:3), **19** (2.8g, 7.37mmol, 91%) as light yellow solid.

Mp: 123°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):

δ=11.49 (s, 1H, H-N(16)), 7.82 (t, ^{3}J=5.8Hz, 1H, H-N(14)), 7.36-7-29 (m, 5H, H-C(7-11)), 7.28-7.20 (m, 3H, H-C(8‘), H-C(9‘), H-C(10‘)), 7.09 (m, 2H, H-C(7‘), H-C(11‘)), 7.07 (s, 1H, H-C(5)), 4.52 (s, 2H, H-C(19)), 4.33 (d, ^{3}J=6.0Hz, 2H, H-C(15))

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):

δ=164.4, 164.1 (C(2), C(12)), 141.2 (C(17)), 138.3, 137.4, 135.4 (C(4), C(6), C(6’)), 128.3 (4C, C(7), C(8), C(10), C(11)), 128.1 (2C, C(8’), C(10’)), 127.5 (3C, C(7’), C(11’), C(9)), 126.8 (C(9’)), 119.4 (C(3)), 115.9 (C(5)), 42.7, 42.4 (C(19), C(15))

MS (ES+):

m/z (%): 385 (100) $[M^{35Cl}+H]^{+}$, 387 (44) $[M^{37Cl}+H]^{+}$, 402 (43) $[M^{35Cl}+NH_4]^{+}$, 407 (52) $[M^{35Cl}+Na]^{+}$), 409 (31) $[M^{37Cl}+Na]^{+}$

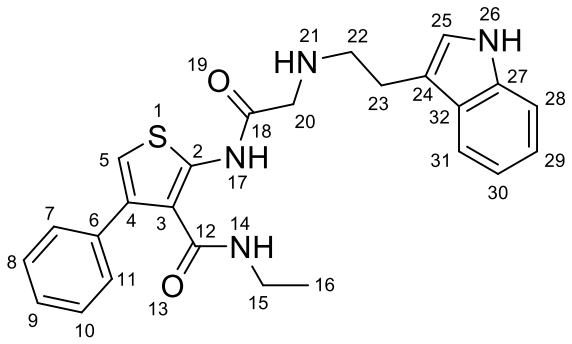
HRMS (EI):

Calcd for $C_{20}H_{17}N_2O_2SCl^{35}$ $[M]^{+}$:	384.069928
Found:	384.070887
Calcd for $C_{20}H_{17}N_2O_2SCl^{37}$ $[M]^{+}$:	386.066977
Found:	386.068372

CHN:

Anal. calcd for $C_{20}H_{17}N_2O_2SCl$ (%):	C: 62.41	H: 4.45	N: 7.28
Found:	C: 62.65	H: 4.53	N: 7.21

***N*-Ethyl-2-[[2-[2-(1*H*-indol-3-yl)ethylamino]acetyl]amino]-4-phenyl-thiophene-3-carboxamide (3)**



According to procedure C, employment of 2-[(2-chloroacetyl)amino]-*N*-ethyl-4-phenyl-thiophene-3-carboxamide (**18**) (0.30g, 0.93mmol), tryptamine (0.45g, 2.79mmol), triethylamine (0.26mL, 1.86mmol), and TBAI (34mg, 0.09mmol) furnished **3** after purification by flash chromatography (DCM:MeOH:$NH_{3(MeOH)}$ 99:1:0.1) as beige solid (310mg, 0.69mmol, 75%).

Mp: 149°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):
δ=10.77 (s, 1H, H-N(17)), 7.55 (d, ^{3}J=8.0Hz, 1H, H-C(31)), 7.44-7.35 (m, 6H, 5 phenyl-H, H-N(26)), 7.32 (d, ^{3}J=8.2Hz, 1H, H-C(28)), 7.16 (sm, 1H, H-C(25)), 7.07-7.03 (m, 2H, H-C(29), H-N(14)), 6.99 (s, 1H, H-C(5)), 6.95 (sm, 1H, H-C(30)), 3.41 (s, 2H, H-C(20)), 3.14 (sm, 2H, H-C(15)), 2.94-2.91 (m, 2H, H-C(22)), 2.87-2.82 (m, 2H, H-C(23)), 0.88 (t, ^{3}J=7.2Hz, 3H, H-C(16)).

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):
δ=169.9, 164.1 (C(2), C(12)), 142.4 (C(18)), 137.3 (C(4) or C(6)), 136.2 (C(27)), 135.6 (C(4) or C(6)), 128.4 (2C, C(8), C(10)), 128.3 (2C, C(7), C(11)), 127.4 (C(9)), 127.2 (C(32)), 122.5 (C(25)), 120.8 (C(30)), 118.2, 118.1 (C(29), C(31)), 117.6 (C(3)), 114.7 (C(5)), 112.1 (C(24)), 111.3 (C(28)), 51.7 (C(20)), 50.5 (C(22)), 33.6 (C(15)), 25.6 (C(23)), 14.1 (C(16))

MS (ES+):
m/z (%): 447 (100) $[M+H]^{+}$

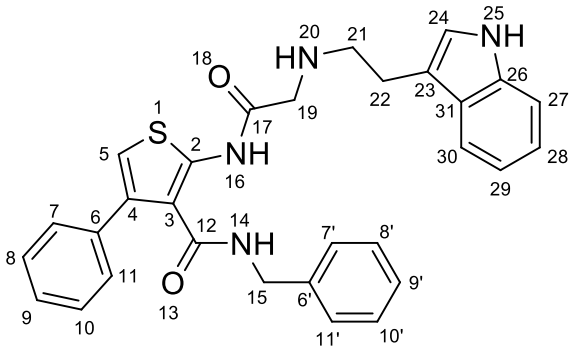
HRMS (ES+):

Calcd for $C_{25}H_{27}N_4O_2S$ $[M+H]^{+}$:	447.185473
Found:	447.182874

CHN:

Anal. calcd for $C_{25}H_{26}N_4O_2S$ (%):	C: 67.24	H: 5.87	N: 12.55
Found:	C: 66.98	H: 5.97	N: 12.50

_N_-Benzyl-2-[[2-[2-(1_H_-indol-3-yl)ethylamino]acetyl]amino]-4-phenyl-thiophene-3-carboxamide (4)



According to procedure C, employment of *N*-benzyl-2-[(2-chloroacetyl)amino]-4-phenyl-thiophene-3-carboxamide (**19**) (0.45g, 1.17mmol), tryptamine (0.56g, 3.5mmol), triethylamine (0.33mL, 2.34mmol), and TBAI (43mg, 0.12mmol) yielded, after purification by flash chromatography (gradient cyclohexane:EtOAc 8:2 – 6:4), **4** as orange-brown solid (385mg, 0.76mmol, 65%).

Mp: 71°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):
δ=10.78 (s, 1H, H-N(16)), 7.57-7.52 (m, 2H, H-C(30), H-N(14)), 7.35-7.29 (m, 7H, H-C(7-11), H-C(27), H-N(25)), 7.24-7.16 (m, 3H, H-C(8‘), H-C(9‘), H-C(10‘)), 7.16 (sm, 1H, H-C(24)), 7.07-7.03 (m, 3H, H-C(28), H-C(7‘), H-C(11‘)), 6.98-6.93 (m, 2H, H-C(5), H-C(29)), 4.32 (d, ^{3}J=6.0Hz, 2H, H-C(15)), 3.42 (s, 2H, H-C(19)), 2.95-2.91 (m, 2H, H-C(21)), 2.86-2.83 (m, 2H, H-C(22))

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):
δ=170.0, 164.5 (C(2), C(12)), 142.8 (C(17)), 138.4, 137.3 (2C, C(4) and/or C(6) and/or C(6‘)), 136.2 (C(26)), 135.7 (2C, C(4) or C(6) or C(6‘)), 128.5 (2C, C(8), C(10)), 128.3 (2C, C(7), C(11)), 128.1 (2C, C(8‘), C(10‘)), 127.4, 127.3 (3C, C(9), C(7‘), C(11‘)), 127.2 (C(31)), 126.7 (C(9‘)), 122.6 (C(24)), 120.8 (C(29)), 118.3, 118.1 (C(28), C(30)), 117.3 (C(3)), 114.9 (C(5)), 112.1 (C(23)), 111.3 (C(27)), 51.7 (C(19)), 50.4 (C(21)), 42.6 (C(15)), 25.5 (C(22))

MS (ES+):
m/z (%): 509 (100) $[M+H]^{+}$

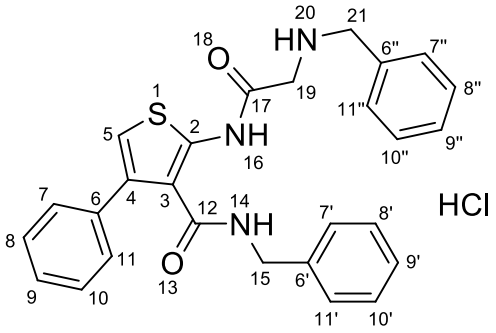
HRMS (EI):

Calcd for $C_{30}H_{28}N_4O_2S$ $[M]^+$:	508.193298
Found:	508.192909

CHN:

Anal. calcd for $C_{30}H_{28}N_4O_2S$ (%):	C: 70.84	H: 5.55	N: 11.01
Found:	C: 70.64	H: 5.7 3	N: 11.00

***N*-Benzyl-2-[[2-(benzylamino)acetyl]amino]-4-phenyl-thiophene-3-carboxamide hydrochloride (5)**



According to procedure C, usage of *N*-benzyl-2-[(2-chloroacetyl)amino]-4-phenyl-thiophene-3-carboxamide (**19**) (0.50g, 1.30mmol), benzylamine (0.43mL, 3.9mmol), triethylamine (0.36mL, 2.59mmol), and TBAI (0.05g, 0.13mmol) furnished, after purification by flash chromatography (cylcohexane:EtOAc 9:1-1:1) and precipitation as HCl salt with hydrogen chloride solution 2M in diethyl ether, **5** as colorless solid (340mg, 0.69mmol, 53%).

Mp: 188°C

^{1}H-NMR (500.16MHz, DMSO-*d6*):

δ=11.23 (bs, 1H, H-N(16)), 9.84 (bs, 2H, H-N(20)), 8.45 (t, ^3J=5.9Hz, 1H, H-N(14)), 7.62-7.57 (m, 2H, H-C(7''), H-C(11'')), 7.48-7.42 (m, 3H, H-C(8''), H-C(9''), H-C(10'')), 7.34-7.20 (m, 8H, H-C(7-11), H-C(8'), H-C(9'), H-C(10')), 7.13 (s, 1H, H-C(5)), 7.12-7.10 (m, 2H, H-C(7'), H-C(11')), 4.35 (d, ^3J=6.0Hz, 2H, H-C(15)), 4.24 (s, 2H, H-C(21)), 4.03 (s, 2H, H-C(19))

^{13}C-NMR (125.77MHz, DMSO-*d6*):

δ=164.0, 163.5 (C(2), C(12)), 138.6, 138.1, 137.6, 135.5, 131.6 (C(17), C(4), C(6), C(6'), C(6'')), 130.2 (2C, C(8''), C(10'')), 129.0 (C(9'')), 128.6, 128.2, 128.1, 127.8 (8C, C(7), C(8), C(10), C(11), C(8'), C(10'), C(7''), C(11'')), 127.5 (2C, C(7'), C(11')), 127.2 (C(9)), 126.7 (C(9')), 122.3 (C(3)), 115.9 (C(5)), 49.9 (C(21)), 47.0 (C(19)), 42.8 (C(15))

MS (ES+):

m/z (%): 456 (100) $[M+H]^+$

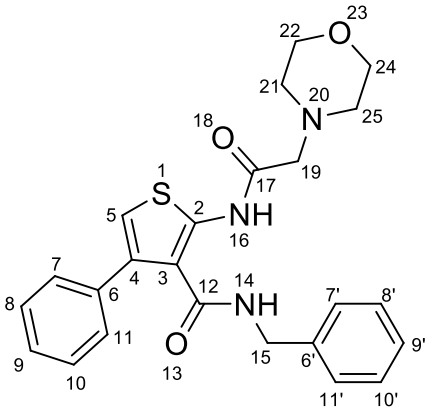
HRMS (EI):

Calcd for $C_{27}H_{25}N_3O_2S$ $[M]^+$:	455.166749
Found:	455.165004

CHN:

Anal. calcd for $C_{27}H_{25}N_3O_2S$*HCl (%):	C: 65.91	H: 5.33	N: 8.54
Found:	C: 65.95	H: 5.48	N: 8.47

***N*-Benzyl-2-[(2-morpholinoacetyl)amino]-4-phenyl-thiophene-3-carboxamide (10)**



According to procedure C, usage of *N*-benzyl-2-[(2-chloroacetyl)amino]-4-phenyl-thiophene-3-carboxamide (**19**) (0.76g, 1.97mmol), morpholine (0.52mL, 5.92mmol), triethylamine (0.55mL, 3.95mmol), and TBAI (0.07g, 0.20mmol) yielded, after purification by flash chromatography (cyclohexane:EtOAc 8:2) and recrystallization from methanol, **10** as colorless needles (403mg, 0.94mmol, 46%).

Mp: 123°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):

δ=11.74 (s, 1H, H-N(16)), 7.45 (t, ^{3}J=5.8Hz, 1H, H-N(14)), 7.37-7.31 (m, 5H, H-C(7-11)), 7.28-7.19 (m, 3H, H-C(8'), H-C(9'), H-C(10')), 7.10-7.08 (m, 2H, H-C(7'), H-C(11')), 6.99 (s, 1H, H-C(5)), 4.31 (d, ^{3}J=5.95Hz, 2H, H-C(15)), 3.65 (sm, 4H, H-C(22), H-C(24)), 3.24 (s, 2H, H-C(19)), 2.53* (sm, 4H, H-C(21), H-C(25)). *overlaid with the DMSO signal

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):

δ=167.7, 164.6 (C(2), C(12)), 142.9 (C(17)), 138.5, 137.3, 135.5 (C(4), C(6), C(6')), 128.5 (2C, C(8), C(10)), 128.3 (2C, C(7), C(11)), 128.1 (2C, C(8'), C(10')), 127.4 (C(9)), 127.3 (2C, C(7'), C(11')), 126.7 (C(9')), 117.2 (C(3)), 115.1 (C(5)), 66.1 (C(22), C(24)), 60.4 (C(19)), 53.2 (C(21), C(25)), 42.6 (C(15))

MS (ES+):

m/z (%): 436 (75) $[M+H]^+$, 458 (45) $[M+Na]^+$, 871 (100) $[2M+H]^+$, 893 (23) $[2M+Na]^+$

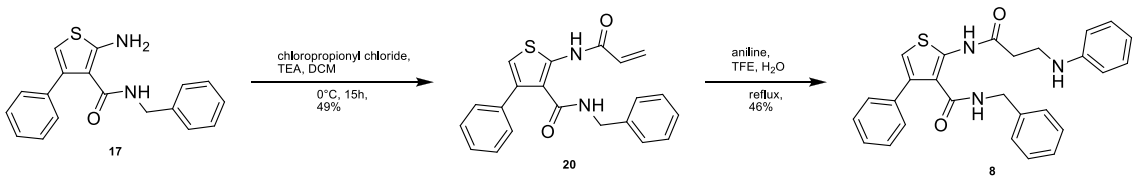
HRMS (EI+):

Calcd for $C_{24}H_{25}N_3O_3S$ $[M]^+$:	435.161664
Found:	435.160123

CHN:

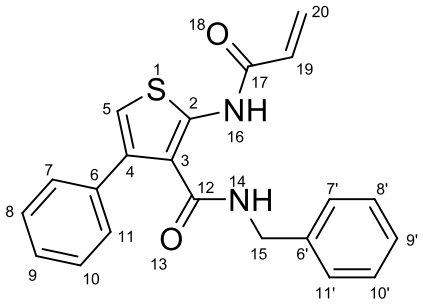
Anal. calcd for $C_{24}H_{25}N_3O_3S$ (%):	C: 66.18	H: 5.79	N: 9.65
Found:	C: 66.04	H: 5.87	N: 9.63

Synthesis of Compound 8



Scheme 6.9. Synthesis of **8**.

2-Acrylamido-*N*-benzyl-4-phenylthiophene-3-carboxamide (20)



Following procedure B, employment of 2-amino-*N*-benzyl-4-phenyl-thiophene-3-carboxamide (**17**) (1.5g, 4.86mmol), chloropropionyl chloride (0.7mL, 7.30mmol), and triethylamine (1.0mL, 7.30mmol) furnished, after two purification steps by flash chromatography (cyclohexane:*tert*-butyl methyl ether 8:2), **8** as light yellow solid (850mg, 2.3mmol, 49%).

Mp: 79°C

^{1}H-NMR (500.16MHz, DMSO-*d6*):
δ=11.04 (s, 1H, H-N(16)), 8.15 (t, ^{3}J=5.7Hz, 1H, H-N(14)), 7.38-7.32 (m, 2H, H-C(7), H-C(11)), 7.31-7.21 (m, 6H, H-C(8-10), H-C(8'), H-C(9'), H-C(10')), 7.13 (m, 2H, H-C(7'), H-C(11')), 7.06 (s, 1H, H-C(5)), 6.69 (dd, ^{3}J=10.2Hz, ^{3}J=17.0Hz, 1H, H-C(19)), 6.32 (dd, ^{2}J=1.7Hz, ^{3}J=16.9Hz, 1H, H_{trans}-C(20)), 5.86 (dd, ^{2}J=1.7Hz, ^{3}J=10.3Hz, 1H, H_{cis}-C(20)), 4.36 (d, ^{3}J= 6.0Hz, 2H, C(15))

^{13}C-NMR (125.77MHz, DMSO-*d6*):
δ=164.4, 162.3 (C(12), C(2)), 140.2, 138.6, 137.4, 135.6 (C(17), C(6), C(6'), C(4)), 130.2 (C(19)), 128.2, 128.1 (6C, C(7), C(8), C(10), C(11), C(8'), C(10')), 127.5, 127.2 (4C, C(7'), C(11'), C(9), C(20)), 126.7 (C(9')), 120.4 (C(3)), 115.9 (C(5)), 42.8 (C(15))

MS (ES+):
m/z (%): 108 (50) $[C_7H_9N+H]^+$, 256 (24) $[M-C_7H_8N]^+$, 363 (100) $[M+H]^+$

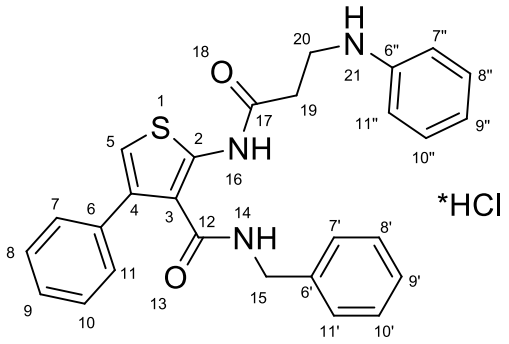
HRMS (EI):

Calcd for $C_{21}H_{18}N_2O_2S$ $[M]^+$:	362.108900
Found:	362.113617

CHN:

Anal. calcd for $C_{21}H_{18}N_2O_2S$*0.25 H_2O (%):	C: 68.74	H: 5.08	N: 7.63
Found:	C: 68.94	H: 5.14	N: 7.70

2-(3-Anilinopropanoylamino)-*N*-benzyl-4-phenyl-thiophene-3-carboxamide hydrochloride (8)



To a solution of 2-acrylamido-*N*-benzyl-4-phenylthiophene-3-carboxamide (**20**) (0.45mg, 1.24mmol) in trifluoroethanol (0.2mL, 2.48mmol) and water (1.5mL), aniline (1.1mL, 1.24mmol) was added. The reaction mixture was heated at 100°C until complete conversion of the starting material, detected by TLC. The reaction was allowed to reach RT and was diluted with ethyl acetate, washed with water and brine, dried over $MgSO_4$, filtered, and concentrated in vacuo. Purification by flash chromatography [cyclohexane:EtOAc 9:1 (10min), gradient to 8:2 (7min), 8:2 (5min)], followed by precipitation of the reaction product as its HCl salt with hydrogen chloride solution 2M in diethyl ether gave rise to **8** as beige solid (279mg, 0.57mmol, 46%).

Mp: 140°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):
δ=10.94 (s, 1H, H-N(16)), 8.04 (t, ^3J=5.8Hz, 1H, H-N(14)), 7.33-7.19 (m, 11H, H-C(7-11), H-C(8', 9',10'), H-C(8'',9'',10'')), 7.12-7.10 (m, 2H, H-C(7'), H-C(11')), 7.01 (s, 1H, H-C(5)), 6.91-6.77 (m, 2H, H-C(7''), H-C(11'')), 4.31 (d, ^3J=6.0Hz, 2H, H-C(15)), 4.12 (brs, 2H, H-N(21)), 3.43 (t, ^3J=6.9Hz, 2H, H-C(20)), 2.80 (t, ^3J=6.9Hz, 2H, H-C(19))

^{13}C-NMR (125.77MHz, MeOH-*d4*):
δ=168.9, 167.0 (C(2), C(12)), 145.0 (C(6'')), 139.2, 138.9, 137.5, 137.1 (C(17), C(6), C(6'), C(4)), 131.5, 130.1, 130.0, 129.8, 129.5, 129.1, 128.8, 128.3, 123.2 (15C, C(7-11), C(7'-11'), C(7''-11'')), 118.6 (C(3)), 117.0 (C(5)), 48.3 (C(15)), 44.4 (C(20)), 32.6 (C(19))

MS (ES+):

m/z (%): 456 (100) $[M+H]^+$, 478 (39) $[M+Na]^+$

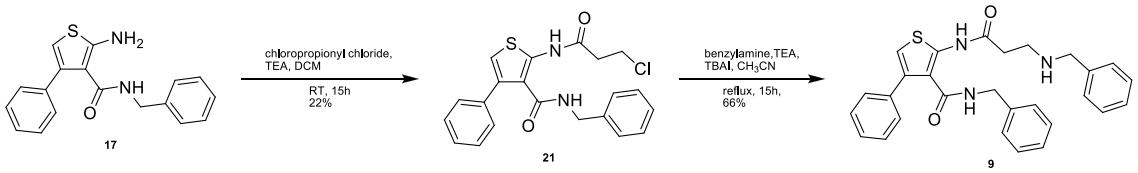
HRMS (EI):

Calcd for $C_{27}H_{25}N_3O_2S$ $[M]^+$:	455.166749
Found:	455.166775

CHN:

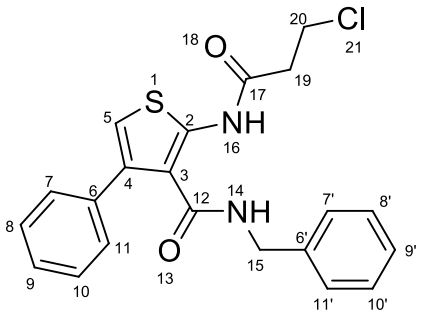
Anal. calcd for $C_{27}H_{25}N_3O_2S$*HCl (%):	C: 65.91	H: 5.33	N: 8.54
Found:	C: 65.92	H: 5.75	N: 8.12

Synthesis of Compound 9



Scheme 6.10. Synthesis of **9**.

***N*-Benzyl-2-(3-chloropropanamido)-4-phenylthiophene-3-carboxamide (21)**



To a solution of 2-amino-*N*-benzyl-4-phenyl-thiophene-3-carboxamide (**17**) (4.5g, 14.59mmol) in dry DCM (70mL), 3-chloropropionyl chloride (2.1mL, 21.89mmol) was added dropwise, followed by the addition of triethylamine (3.05mL, 21.89mmol). The reaction mixture was stirred at RT, until complete consumption of the 2-amino-thiophene (≈15h), detected by TLC. The reaction mixture was diluted with DCM, washed with an aqueous ammonium chloride solution (sat.), brine, dried over $MgSO_4$, filtered, and concentrated in vacuo. Purification by flash chromatography (twice, cyclohexane:*tert*-butyl methyl ether 8:2), yielded **21**, as colorless solid (1.3g, 3.3mmol, 22%).

Mp: 102°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):

δ=10.93 (s, 1H, H-N(16)), 8.14 (t, ^{3}J=5.6Hz, 1H, H-N(14)), 7.34-7.20 (m, 8H, H-C(7-11), H-C(8'), H-C(9'), H-C(10')), 7.13-7.11 (m, 2H, H-C(7'), H-C(11')), 7.03 (s, 1H, H-C(5)), 4.35 (d, ^{3}J=6.0Hz, 2H, H-C(15)), 3.89 (t, ^{3}J=6.3Hz, 2H, H-C(20)), 3.00 (t, ^{3}J=6.2Hz, 2H, H-C(19))

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):

δ=167.3, 164.4 (C(2), C(12)), 140.1, 138.6, 137.3, 135.7 (C(17), C(6), C(6'), C(4)), 128.2, 128.1, 128.0 (6C, C(8), C(10), C(8'), C(10'), C(7), C(11)), 127.5, 127.2 (3C, C(7'), C(11'), C(9)), 126.7 (C(9')), 120.1 (C(3)), 115.3 (C(5)), 42.8 (C(15)), 40.4 (C(20)), 38.2 (C(19))

MS (ES+):

m/z (%): 108 (60) $[C_7H_9N+H]^+$, 292 (18) $[M^{35Cl}-C_7H_8N]^+$, 399 (100) $[M^{35Cl}+H]^+$, 401 (39) $[M^{37Cl}+H]^+$

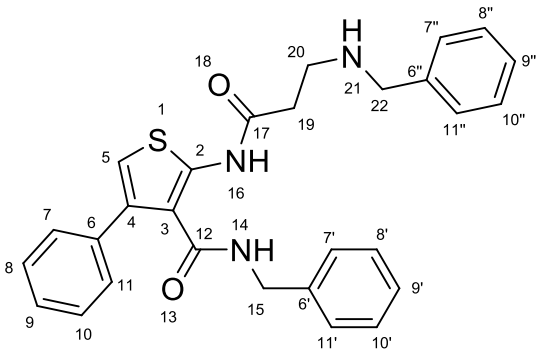
HRMS (EI):

Calcd for $C_{21}H_{19}N_2O_2SCl^{35}$ $[M]^+$:	398.085578
Found:	398.086965
Calcd for $C_{21}H_{19}N_2O_2SCl^{37}$ $[M]^+$:	400.082627
Found:	400.081696

CHN:

Anal. calcd for $C_{21}H_{19}N_2O_2SCl$ (%):	C: 63.23	H: 4.80	N: 7.02
Found:	C: 63.65	H: 5.02	N: 7.20

***N*-Benzyl-2-[3-(benzylamino)propanoylamino]-4-phenyl-thiophene-3-carboxamide (9)**



Following procedure C, utilization of *N*-benzyl-2-(3-chloropropanamido)-4-phenylthiophene-3-carboxamide (**21**) (0.40g, 1.00mmol), benzylamine (0.33mL, 3.02mmol), triethylamine (0.28mL, 2.01mmol), and TBAI (37mg, 0.10mmol), furnished, after two purification steps (1. gradient cyclohexane:*tert*-butyl methyl ether 7:3 - 0:100, 2. cyclohexane:*tert*-butyl methyl ether 4:6), **9** as colorless solid (310mg, 0.7mmol, 66%).

Mp: 120°C

^{1}H-NMR (399.79MHz, DMSO-*d6*):
δ=8.05 (bs, 1H, H-N(16)), 7.38-7.19 (m, 14H, H-C(7-11), H-C(8‘,9‘,10‘), H-(7‘‘-11‘‘), H-N(14)), 7.10-7.07 (m, 2H, H-C(7‘), H-C(11‘)), 6.97 (s, 1H, H-C(5)), 4.28 (d, ^{3}J=6.0Hz, 2H, H-C(15)), 3.73 (s, 2H, H-C(22)), 2.78 (t, ^{3}J=6.3Hz, 2H, H-C(20)), 2.57 (t, ^{3}J=6.2Hz, 2H, H-C(19))

^{13}C-NMR (100.53MHz, DMSO-*d6*, APT):
δ=169.8, 164.5 (C(2), C(12)), 141.0, 140.2, 138.7, 137.2, 135.8 (C(17), C(6‘’), C(6’), C(6), C(4)), 128.2 (2C, C(8), C(10)), 128.1 (5C, C(7), C(11), C(8’), C(10’), C(8‘’), C(10‘’)), 127.4 (2C, C(7’), C(11’)), 127.1 (C(9)), 126.7, 126.6 (4C, C(9’), C(7‘’), C(11‘’), C(9‘’)), 119.1 (C(3)), 114.9 (C(5)), 52.5 (C(22)), 44.1 (C(20)), 42.6 (C(15)), 35.2 (C(19))

MS (ES+):
m/z (%): 470 (100) $[M+H]^{+}$

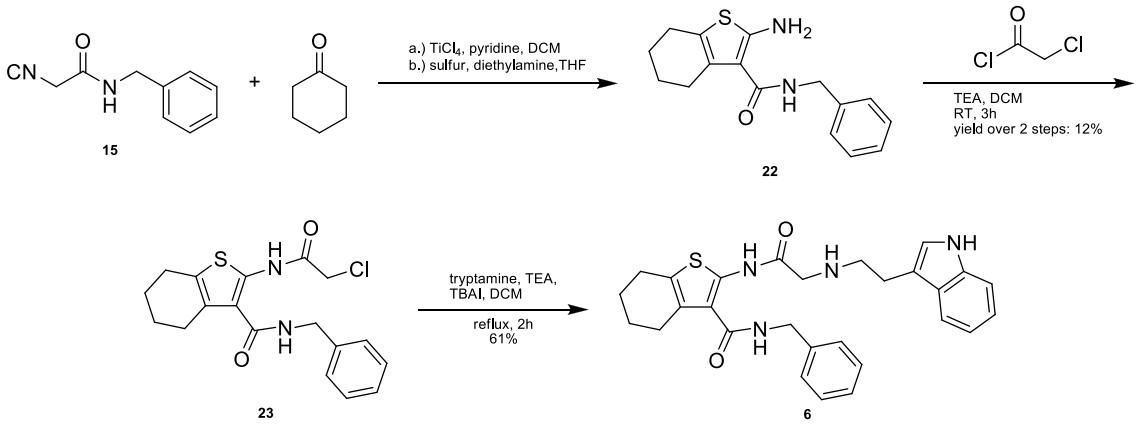
HRMS (EI):

Calcd for $C_{28}H_{27}N_3O_2S$ $[M]^{+}$:	469.182399
Found:	469.181301

CHN:

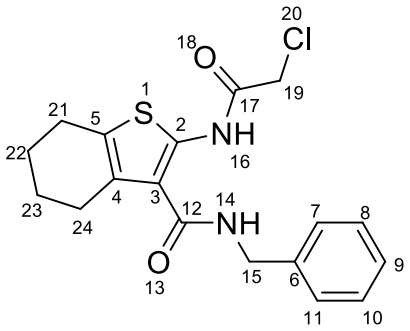
Anal. calcd for $C_{28}H_{27}N_3O_2S$ (%):	C: 71.61	H: 5.80	N: 8.95
Found:	C: 71.57	H: 5.97	N: 8.99

Synthesis of Compound 6



Scheme 6.11. Synthesis of **6**.

N-Benzyl-2-[(2-chloroacetyl)amino]-4,5,6,7-tetrahydrobenzothiophene-3-carboxamide (23)



According to procedure A, usage of **15** (1.5g, 8.6mmol), cyclohexanone (0.74mL, 7.2mmol), $TiCl_4$ (1.59mL, 14.4mmol), pyridine (1.98mL, 24.5mmol), sulfur (0.27g, 8.4mmol), and diethylamine (2.63mL, 25.2mmol) yielded after two purification steps by flash chromatography (cyclohexane:EtOAc 85:15) **22** as crude product which was used without further purification (710mg, not pure). According to procedure B, employment of crude **22** (130mg), chloroacetyl chloride (0.05mL, 0.63mmol), and triethylamine (0.09mL, 0.65mmol) rendered **23** as colorless solid (58mg, 0.16mmol, 12% over to steps) after purification by flash chromatography (cyclohexane:EtOAc 8:2).

Mass spectra data for **22**:

HRMS (EI):

Calcd for $C_{16}H_{18}N_2OS$ [M^+]:	286.113985
Found:	286.116558

Analytical data for **23**:

Mp: 188°C

^{1}H-NMR (500.16MHz, DMSO-*d6*):
δ=11.84 (s, 1H, H-N(16)), 8.03 (t, ^{3}J=5.7Hz, 1H, H-N(14)), 7.36-7.31 (m, 4H, 4 phenyl-H), 7.27-7.22 (m, 1H, 1 phenyl-H), 4.49 (d, ^{3}J=6.0Hz, 2H, H-C(15)), 4.48 (s, 2H, H-C(19)), 2.72 (sm, 2H, H-C(21)), 2.64 (sm, 2H, H-C(24)), 1.74 (sm, 4H, H-C(22), H-C(23))

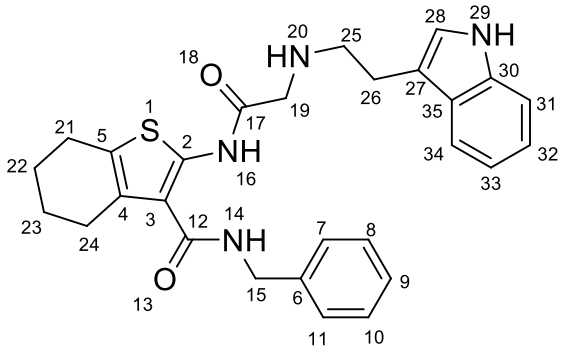
^{13}C-NMR (125.77MHz, DMSO-*d6*):
δ=164.9, 163.4 (C(2), C(12)), 140.0 (C(17)), 139.2 (C(6)), 128.9 (C(4)), 128.2 (2C, C(8), (10)), 127.1 (2C, C(7), C(11)), 127.0 (C(5)), 126.7 (C(9)), 117.9 (C(3)), 42.5, 42.4 (C(15), C(19)), 25.0 (C(21)), 23.8 (C(24)), 22.4, 22.3 (C(22), C(23))

MS (ES+):
m/z (%): 119 (100), 363 (77) $[M^{35Cl}+H]^{+}$, 365 (47) $[M^{37Cl}+H]^{+}$, 380 (45) $[M^{35Cl}+NH_4]^{+}$, 382 (17) $[M^{37Cl}+NH_4]^{+}$, 385 (63) $[M^{35Cl}+Na]^{+}$, 387 (24) $[M^{37Cl}+Na]^{+}$

HRMS (EI):

Calcd for $C_{18}H_{19}N_2O_2SCl^{35}$ $[M^{+}]$:	362.085578		
Found:	362.085863		
Calcd for $C_{18}H_{19}N_2O_2SCl^{37}$ $[M^{+}]$:	364.082627		
Found:	364.081173		
CHN:			
Anal. calcd for $C_{18}H_{19}N_2O_2SCl$ (%):	C: 59.58	H: 5.28	N: 7.72
Found:	C: 59.29	H: 5.37	N: 7.44

_N_-Benzyl-2-[[2-[2-(1_H_-indol-3-yl)ethylamino]acetyl]amino]-4,5,6,7-tetrahydro-benzothiophene-3-carboxamide (6)



According to procedure C, utilization of *N*-benzyl-2-[(2-chloroacetyl)amino]-4,5,6,7-tetrahydrobenzothiophene-3-carboxamide (**23**) (90mg, 0.25mmol), tryptamine (0.12g, 0.75mmol), triethylamine (69µL, 0.50mmol), and TBAI (9mg, 0.02mmol) furnished after purification by flash chromatography (DCM:MeOH 99:1), **6** as light orange solid (74mg, 0.15mmol, 61%).

Mp: 113°C

^{1}H-NMR (500.16MHz, DMSO-*d6*):
δ=10.76 (s, 1H, H-N(16)), 8.00 (t, ^{3}J=5.7Hz, 1H, H-N(14)), 7.54 (d, ^{3}J=7.7Hz, 1H, H-C(34)), 7.38-7.31 (m, 4H, phenyl-H), 7.29 (d, ^{3}J=7.5Hz, 1H, H-C(31)), 7.27 (s, 1H, H-N(29)), 7.23-7.17 (m, 1H, phenyl-H), 7.13 (sm, 1H, H-C(28)), 7.05 (dd, ^{3}J=7.7Hz, ^{3}J=7.2Hz, 1H, H-C(32)), 6.95 (dd, ^{3}J=7.7Hz, ^{3}J=7.2Hz, 1H, H-C(33)), 4.50 (d, ^{3}J=6.0Hz, 2H, H-C(15)), 3.38 (s, 2H, H-C(19)), 2.90 (t, ^{3}J=7.5Hz, 2H, H-C(25)), 2.81(t, ^{3}J=7.3Hz, 2H, H-C(26)), 2.73 (sm, 2H, H-C(21)), 2.63 (sm, 2H, H-C(24)), 1.74 (sm, 4H, H-C(22), H-C(23))

^{13}C-NMR (125.77MHz, DMSO-*d6*, APT):
δ=169.3, 164.9 (C(2), C(12)), 140.9 (C(17)), 139.5 (C(6)), 136.2 (C(30)), 128.6 (C(4)), 128.2 (2C, C(8), C(10)), 127.21 (C(5)), 127.17 (2C, C(7), C(11)), 126.6 (C(9)), 125.8 (C(35)), 122.5 (C(28)), 120.8 (C(33)), 118.2, 118.1 (C(32), C(34)), 116.8 (C(3)), 112.0 (C(27)), 111.2 (C(31)), 51.7 (C(19)), 50.3 (C(25)), 42.4 (C(15)), 25.3 (C(26)), 25.1 (C(21)), 23.8 (C(24)), 22.5, 22.4 (C(22), C(23))

MS (ES+):
m/z (%): 487 (100) $[M+H]^{+}$

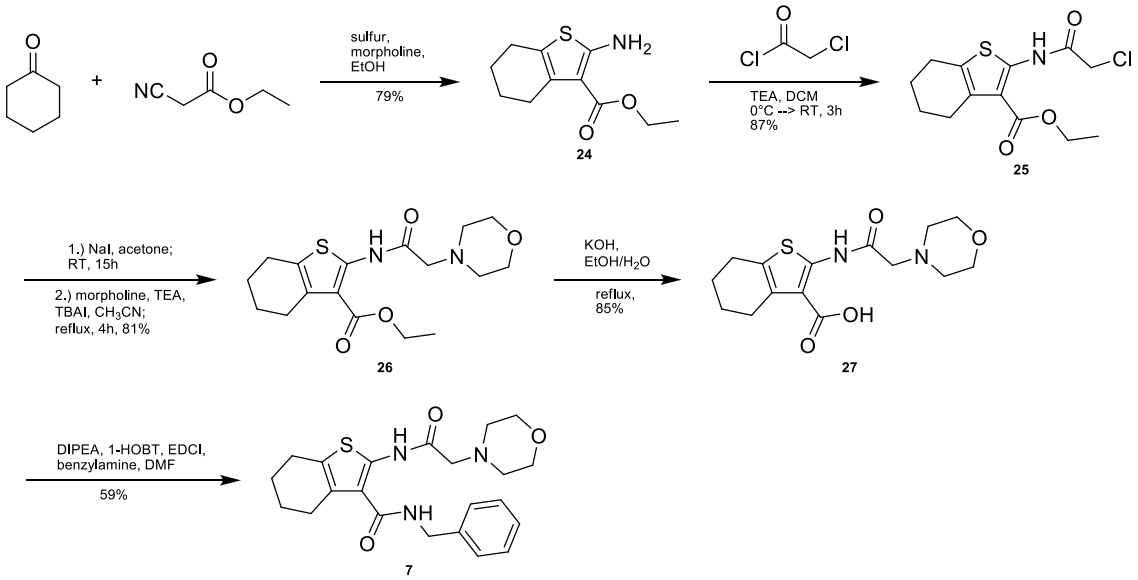
HRMS (EI):

Calcd for $C_{28}H_{30}N_4O_2S$ $[M]^+$:	486.208948
Found:	486.207907

CHN:

Anal. calcd for $C_{28}H_{30}N_4O_2S$ (%):	C: 69.11	H: 6.21	N: 11.51
Found:	C: 68.72	H: 6.17	N: 11.30

Synthesis of Compound 7



Scheme 6.12. Synthesis of **7**.*

*) Compound **7** was synthesized by Amir Shahim (AK Diederich, Philipps-Universität Marburg). The experimental procedures and characterization of **24**-**27** and **7** can be found in the original publication.[33]

6.5.4 Protein Purification

Endothiapepsin was purified as published previously.[36] The protein (5mg/mL) was stored at -20°C until use.

6.5.5 Fluorescence Assay

K_i-values were determined in a fluorescence based assay. To detect the activity of the enzyme, the FRET substrate Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH_2 (Bachem) was used. The assay was performed in the assay buffer containing 0.1M sodium acetate, pH 4.6 and 0.01% Tween20. Each well contained a reaction mixture of 178µL assay buffer or assay buffer-substrate-mix and 2µL DMSO or 2µL inhibitor (DMSO) or 2µL substrate (DMSO). Immediately before the measurements, 20µL endothiapepsin [≈ 2nM] were added to start the reaction. The increase of fluorescence was detected at 414nm after excitation at 337nm using a Tecan Safire II plate reader. For K_m-determination the assay was run with 12 different substrate concentrations (100µM-49nM). IC_{50} values were determined at a substrate concentration of 1.78µM, for each inhibitor 23 different inhibitor concentrations (1mM-238pM) were used. From the increase in fluorescence the IC_{50}-values were calculated with the program GraFit 4.09 (Erithacus Software). The corresponding K_i-values were calculated by means of the Cheng-Prusoff-equation[37] under inclusion of the determined K_m-value and substrate concentration. Each K_m- and K_i-values were determined at least in duplicate.

6.5.6 Thermal Shift Assay (TSA)

TSA measurements were performed in the iCycler IQ5 Real Time Detection System (Bio-Rad). For the assay, 96-well plates (Multiplate™ Low-Profile 96-well unskirted PCR Plates, BioRad) were used with a total volume of 40µL. The reaction mixture contains 36µL assay buffer-SyproOrange-mix, 3µL endothiapepsin (70µM) and 1µL DMSO/inhibitor. The assay buffer-SyproOrange-mix consists of the endothiapepsin-assay buffer (0.1M sodium acetate, pH 4.6, 0.01% Tween20) and 0.2% (V/V) SyproOrange (protein gel stain 5000x concentration in DMSO, Invitrogen). The plate was heated from 20°C to 80°C with a heating rate of 0.5°C/min. The final inhibitor concentration was 100µM. T_m values were manually determined from the first derivative of the resulting data (fluorescence intensity vs temperature). The unfolding temperature of endothiapepsin without a ligand ($T_{m(DMSO)}$) was subtracted from the unfolding temperature with ligand ($T_{m(lig)}$) to obtain the thermal shifts ΔT_m. Each T_m-value was at least measured in triplicate.

6.5.7 Crystallization

Crystals of endothiapepsin were obtained using the vapor diffusion method in drops containing 2µL protein solution (endothiapepsin 5mg/mL, 0.1M sodium acetate, pH 4.6) and 2µL reservoir solution (0.1M ammonium acetate, 24-30% PEG4000 (w/w) and 0.1M acetate buffer, pH 4.6). Crystals appeared after several days of incubation at 291K. The complex formation for compounds **5**, **6** and **8** was performed by soaking the obtained ligand-free endothiapepsin crystals for an appropriate time in reservoir solutions saturated with the respective ligand. For the other ligands the crystals were soaked for 1-2d in a mixture of 25% glycerol, 75% mother liquor and the ligand at a final concentration of 10mM.

6.5.8 Data Collection, Structure Determination and Refinement

Crystals were frozen in liquid nitrogen after prior cryo-protection treatment using 25% glycerol in reservoir solution. Data were collected at 100K at the synchrotron beamline as indicated in Table 6.2.

Data processing and scaling was performed using the programs HKL2000[38] (EP in complex with **1**, **3**, **4**, **7,** and **9**) or XDS (EP in complex with **5**, **6**, and **8**).[39] The coordinates of EP as deposited under PDB access code 1OEW[40] were used as search models after removal of ligands, ions and water molecules. Molecular replacement was carried out via Phaser[41] as implemented in Phenix.[42] Refinement was performed under repeated cycles of manual model building using Coot[43] and crystallographic refinement with the program phenix.refine or SHELXL[44] as indicated in Table 6.2. The final models were validated using PROCHECK.[45] Data collection and refinement statistics are shown in Table 6.2. The coordinates of the EP complexes determined in this study have been deposited in the Protein Data Bank under the PDB accession codes listed in Table 6.2.

Table 6.2. Data collection and refinement statistics.

Inhibitor	**1**	**3**	**4**	**5**
PDB Entry	3T7Q	4L6B	3PSY	3WZ6
Data collection and Processing				
Beamline Location	SLS (X06DA)	BESSY (BL14.2)	SLS (X06DA)	PETRA III (P14/MX 2)
Wavelength [Å]	1.000	0.91841	1.000	0.826606
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters				
a, b, c [Å]	45.4; 73.1; 52.8	45.4; 73.5; 53.0	45.5; 73.1; 52.8	45.4; 73.0; 53.2
β [°]	109.7	109.8	109.5	110.1
Matthews coefficient [$Å^3$/Da]	2.4	2.5	2.4	2.4
Solvent content [%]	49.6	50.0	49.6	48.9
Diffraction data				
Resolution range [Å]	40-1.30 (1.32-1.30)*	30-1.37 (1.39-1.37)*	40-1.43 (1.45-1.43)*	50-1.40 (1.60-1.40)*
Unique reflections	78609 (3339)	68886 (3442)	55570 (2578)	62145 (19124)
$R(I)_{sym}$ [%]	6.3 (41.6)	9.2 (48.6)	3.4 (28.5)	10.6 (64.6)
Completeness [%]	98.6 (83.6)	99.9 (100.0)	92.8 (87.3)	97.7 (93.2)
Redundancy	3.3 (2.4)	3.9 (3.9)	3.5 (3.3)	8.2 (7.6)
I/σ(I)	20.1 (2.2)	15.6 (2.7)	34.2 (4.2)	15.3 (3.0)
Refinement				
Program	PHENIX	PHENIX	SHELXL	PHENIX
Resolutions range [Å]	29.4-1.3	29.6-1.37	10.0-1.43	42.7-1.40
Reflections used in refinement (work/free)	71189/3845	65410/3452	51397/2741	60894/1242
Final R values for all reflections (work/free) [%]	16.2/17.5	16.6/19.7	12.8/16.7	14.3/16.3
Protein residues	330	330	330	330
Inhibitor atoms	32	32	37	33
Water molecules	420	324	235	614
RMSDs				
Bonds [Å]	0.005	0.009	0.010	0.005
Angles [°]	1.1	1.2	2.7	1.1
Ramachandran plot				
Residues in most favored regions [%]	94.2	94.2	93.9	93.9
Residues in additional allowed regions [%]	5.8	5.8	6.1	6.1
Residues in generously allowed regions [%]	-	-	-	-
Mean B factors [$Å^2$]				
Protein	11.1	10.9	14.3	9.5
Inhibitor	21.4	11.5	16.3	23.8
Water	24.5	22.8	26.9	26.0
Other ligands (DMSO,PEG, ...)	19.3	15.4	35.2	35.3
* Number in parentheses characterize the highest resolution shell				

Inhibitor	**6**	**7**	**8**	**9**
PDB Entry	3WZ7	3T7X	3WZ8	4LAP
Data collection and Processing				
Beamline Location	BESSY (BL14.1)	BESSY (BL14.2)	BESSY (BL14.1)	BESSY (BL14.2)
Wavelength [Å]	0.91841	0.91841	0.91841	0.91841
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters				
a, b, c [Å]	45.2; 73.6; 53.2	45.4; 73.2; 53.0	45.6; 73.9; 54.2	45.4; 73.1; 52.8
β [°]	110.2	109.8	110.7	109.7
Matthews coefficient [$Å^3$/Da]	2.4	2.5	2.5	2.4
Solvent content [%]	49.1	49.9	50.5	49.5
Diffraction data				
Resolution range [Å]	50-1.90 (2.05-1.90)*	25-1.27 (1.29-1.27)*	50-1.45 (1.60-1.45)*	30-1.12 (1.14-1.12)*
Unique reflections	23054 (3959)	84384 (3480)	59674 (15187)	119175 (5702)
$R(I)_{sym}$ [%]	7.3 (35.2)	4.0 (16.1)	8.1 (56.7)	4.5 (42.1)
Completeness [%]	89.1 (75.6)	98.0 (81.1)	99.2 (99.3)	95.9 (92.1)
Redundancy	3.2 (3.0)	3.1 (2.5)	3.7 (3.7)	3.9 (3.8)
I/σ(I)	17.2 (3.5)	23.8 (6.1)	12.9 (2.6)	30.2 (3.3)
Refinement				
Program	PHENIX	PHENIX	PHENIX	PHENIX
Resolutions range [Å]	42.4-1.90	21.6-1.27	40.5-1.45	17.6-1.12
Reflections used in refinement (work/free)	22030/920	78767/4206	57296/1770	113197/5978
Final R values for all reflections (work/free) [%]	24.0/28.0	15.4/16.9	18.9/22.4	12.9/14.4
Protein residues	330	330	330	330
Inhibitor atoms	35	29	33	34
Water molecules	196	409	566	321
RMSDs				
Bonds [Å]	0.007	0.005	0.004	0.012
Angles [°]	2.0	1.1	0.97	1.5
Ramachandran plot				
Residues in most favored regions [%]	90.6	93.5	93.9	93.9
Residues in additional allowed regions [%]	9.4	6.5	6.1	6.1
Residues in generously allowed regions [%]	-	-	-	-
Mean B factors [$Å^2$]				
Protein	15.7	9.2	8.4	11.3
Inhibitor	22.5	17.2	18.2	19.8
Water	17.6	21.9	21.0	25.0
Other ligands (DMSO,PEG, …)	18.5	21.9	26.6	19.2
* Numbers in parentheses characterize the highest resolution shell				

6.6 Appendix

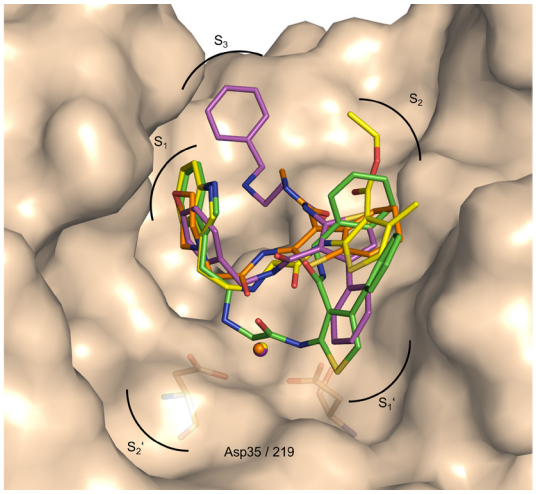


Figure 6.9. Schematic overview of the binding modes of **1** (yellow), **4** (green), **7** (orange) and **9** (purple) in the active site of endothiapepsin (wheat, surface representation, PDB code: 3T7Q). Ligands are shown in stick representation. The *flap* amino acids 77-83 are excluded.

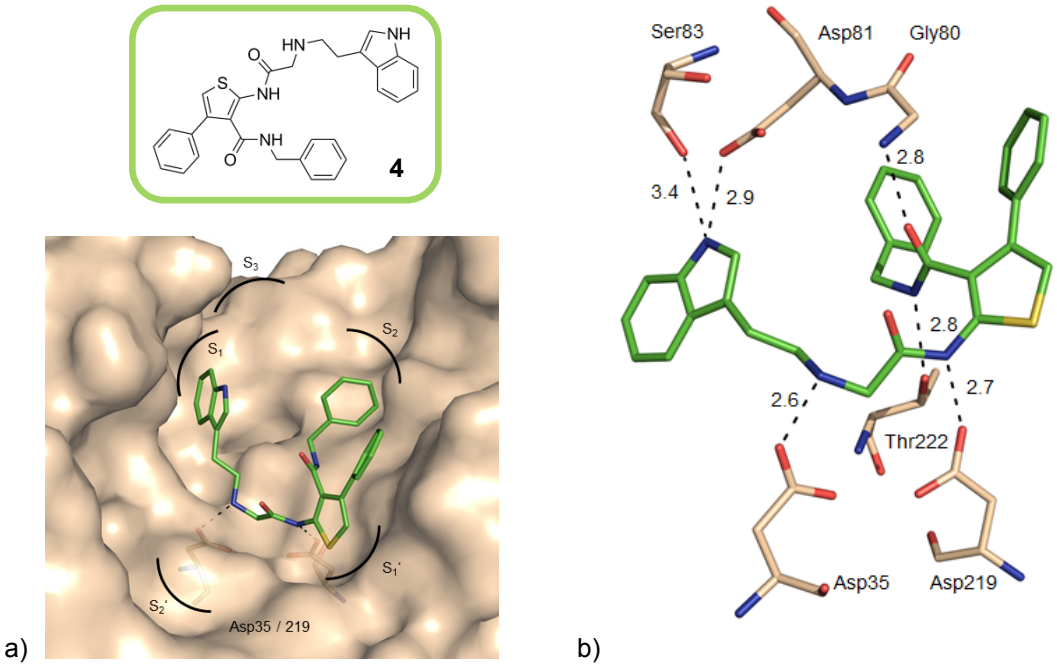


Figure 6.10. a) The schematic overview of the binding mode of the green cluster is exemplarily shown for **4** (stick representation). Endothiapepsin is shown as surface, the *flap* amino acids 77-83 are excluded. b) Detailed H-bond network of **4**. Ligands and amino acids are shown in stick representation, color coded by atom type, distances in Å.

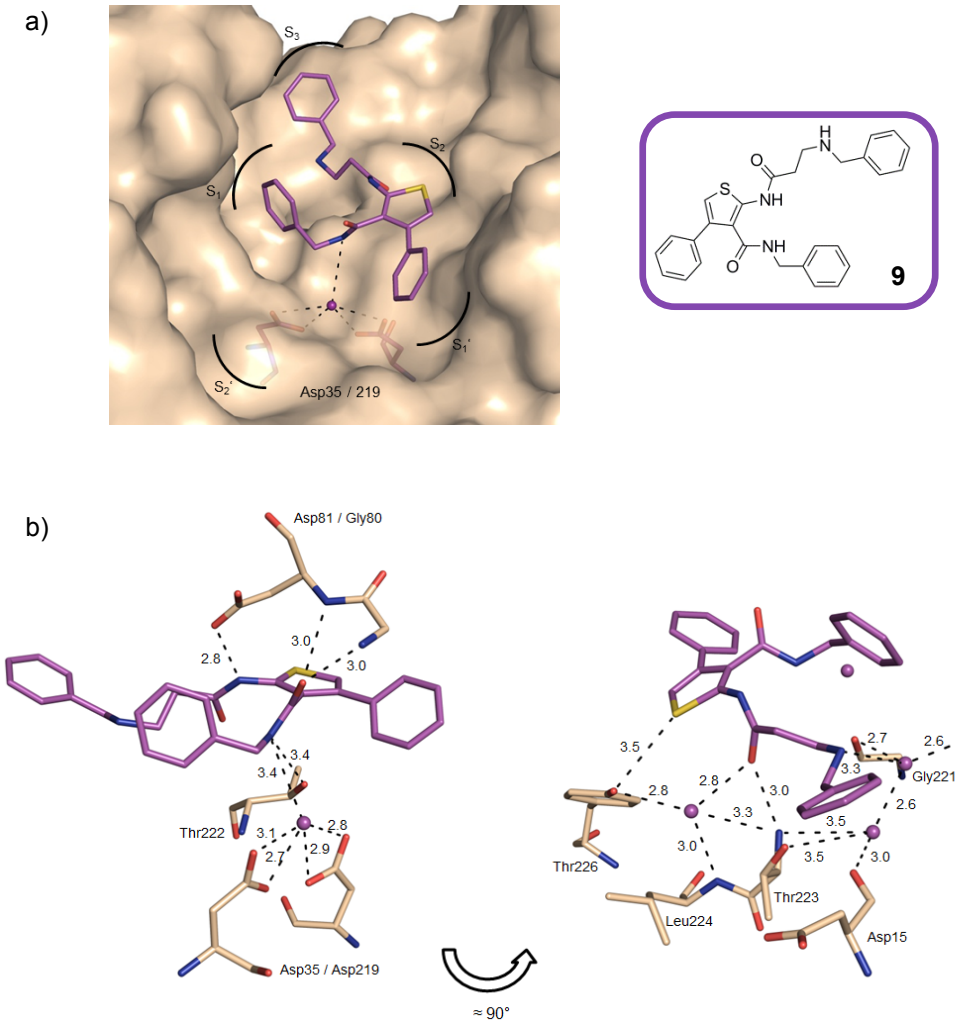


Figure 6.11. a) The schematic overview of the purple binding mode cluster is exemplarily shown for **9** (stick representation). Endothiapepsin is shown as surface, the *flap* amino acids 77-83 are excluded. b) Detailed H-bond network of compound **9**. Ligands and amino acids are shown in stick representation, color coded by atom type, distances in Å.

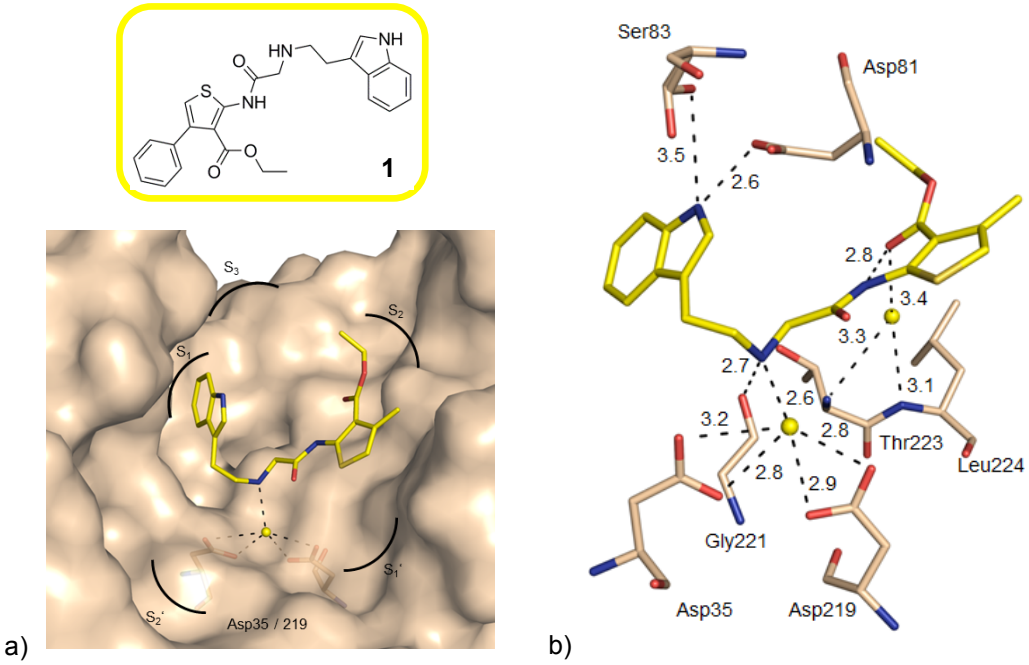


Figure 6.12. a) Schematic overview of **1** (stick representation). Endothiapepsin is shown as surface, the *flap* amino acids 77-83 are excluded. b) H-bond network of compound **1**. Ligands and amino acids are shown in stick representation, color coded by atom type, distances in Å.

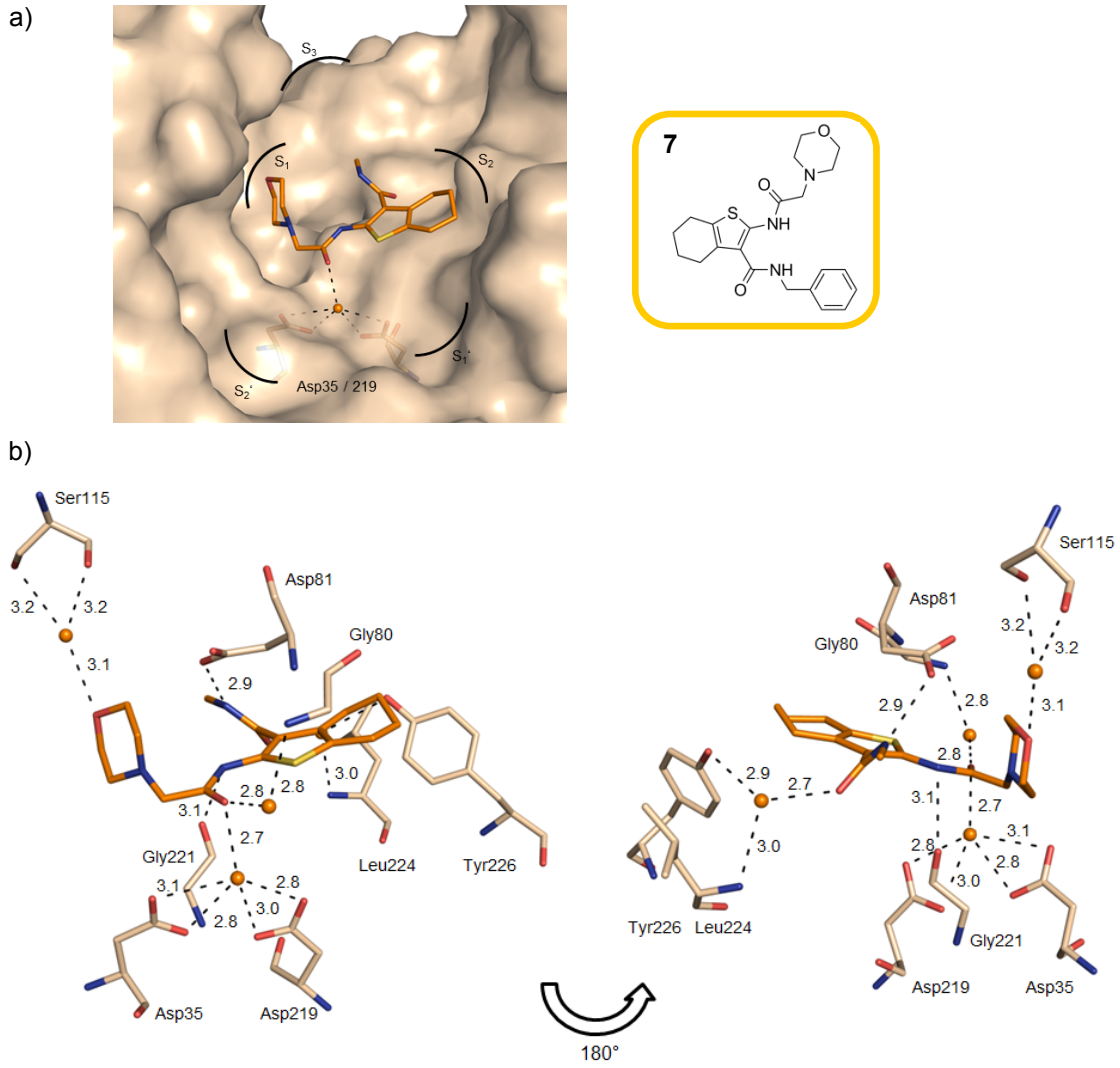


Figure 6.13. a) Schematic representation of the binding mode of **7** (orange cluster) in endothiapepsin (shown as surface), the *flap* amino acids 77-83 are excluded. b) H-bond network for **7**. Ligands and amino acids are shown in stick representation, color coded by atom type, distances in Å.

6.7 References

(1) Hubbard, R. E., Ed. *Structure-based drug discovery*; Royal Society of Chemistry: Cambridge, 2006.

(2) Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **2003,** *10*, 787-797.

(3) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998,** *27*, 249–284.

(4) Itzstein, M. von. The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discovery* **2007,** *6*, 967–974.

(5) Venkatraman, S. Discovery of boceprevir, a direct-acting NS3/4A protease inhibitor for treatment of chronic hepatitis C infections. *Trends Pharmacol. Sci.* **2012,** *33*, 289–294.

(6) Ghosh, A. K. *Structure-based design of drugs and other bioactive molecules: tools and strategies*; Wiley-VCH: Weinheim, 2014.

(7) Blundell, T. L.; Cooper, J.; Foundling, S. I.; Jones, D. M.; Atrash, B.; Szelke, M. On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition-state analogs. *Biochemistry* **1987,** *26*, 5585–5590.

(8) Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; Beer, T. de; Folmer, Rutger H. A. Discovery of a novel warhead against β-secretase through fragment-based lead generation. *J. Med. Chem.* **2007,** *50*, 5903–5911.

(9) Gewald, K.; Schinke, E.; Böttcher, H. 2-Amino-thiophene aus methyenaktiven Nitrilen, Carbonylverbindungen und Schwefel. *Chem. Ber.* **1966,** *99*, 94–100.

(10) Martin, A. *http://www.bioinf.org.uk/software/profit/*.

(11) McLachlan, A. D. Rapid comparison of protein structures. *Acta Crystallogr. Sect. A* **1982,** *38*, 871-873.

(12) Chi, G.; Manos-Turvey, A.; O'Connor, P. D.; Johnston, J. M.; Evans, G. L.; Baker, E. N.; Payne, R. J.; Lott, J. S.; Bulloch, Esther M. M. Implications of binding mode and active site flexibility for inhibitor potency against the salicylate synthase from mycobacterium tuberculosis. *Biochemistry* **2012,** *51*, 4868–4879.

(13) Reich, S.H.; Melnick, M.; Davies II, J. F.; Appelt, K.; Lewis, K. K.; Fuhry, M. A.; Pino, M.; Trippe, A. J.; Nguyen, D.; Dawson, H.; Wu, B.-W.; Musick, L.; Kosa, M.; Kahil, D.; Webber, S.; Gehlhaar, D. K.; Andrada, D.; Shetty, B. Protein structure-based design of potent orally bioavailable, nonpeptide inhibitors of human immunodeficiency virus protease. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92*, 3298–3302.

(14) Steuber, H.; Zentgraf, M.; La Motta, C.; Sartini, S.; Heine, A.; Klebe, G. Evidence for a novel binding site conformer of aldose reductase in ligand-bound state. *J. Mol. Biol.* **2007,** *369*, 186–197.

(15) Mpamhanga, C. P.; Spinks, D.; Tulloch, L. B.; Shanks, E. J.; Robinson, D. A.; Collie, I. T.; Fairlamb, A. H.; Wyatt, P. G.; Frearson, J. A.; Hunter, W. N.; Gilbert, I. H.; Brenk, R. One

scaffold, three binding modes: novel and selective pteridine reductase 1 inhibitors derived from fragment hits discovered by virtual screening. *J. Med. Chem.* **2009,** *52*, 4454–4465.

(16) Boström, J.; Hogner, A.; Schmitt, S. Do structurally similar ligands bind in a similar fashion? *J. Med. Chem.* **2006,** *49*, 6716–6725.

(17) Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screening* **2001,** *6*, 429–440.

(18) Zhang, R.; Monsma, F. Fluorescence-based thermal shift assays. *Curr. Opin. Drug Discovery Dev.* **2010,** *13*, 389–402.

(19) Bobkova, E. V.; Weber, M. J.; Xu, Z.; Zhang, Y.-L.; Jung, J.; Blume-Jensen, P.; Northrup, A.; Kunapuli, P.; Andersen, J. N.; Kariv, I. Discovery of PDK1 kinase inhibitors with a novel mechanism of action by ultrahigh throughput screening. *J. Biol. Chem.* **2010,** *285*, 18838-18846.

(20) Śledź, P.; Lang, S.; Stubbs, C. J.; Abell, C. High-throughput interrogation of ligand binding mode using a fluorescence-based assay. *Angew. Chem., Int. Ed.* **2012,** *51*, 7680–7683.

(21) Lo, M.-C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **2004,** *332*, 153–159.

(22) Sabnis, R. W.; Rangnekar, D. W.; Sonawane, N. D. 2-aminothiophenes by the gewald reaction. *J. Heterocycl. Chem.* **1999,** *36*, 333–345.

(23) Huang, Y.; Dömling, A. The Gewald multicomponent reaction. *Mol. Diversity* **2011,** *15*, 3-33.

(24) Chakrabarti *et al.* U.S. Patent 5,627,178, 1997.

(25) Bartholomäus, R. Synthese und Struktur-Wirkungsbeziehungen substituierter Thiophene als neuartige Leitstrukturen zur Inhibition von Plasmepsin II und IV. Diplomarbeit, Philipps-Universität Marburg, 2007.

(26) Varala, R.; Nuvula, S.; Adapa, S. R. Molecular iodine-catalyzed facile procedure for N-Boc protection of amines. *J. Org. Chem.* **2006,** *71*, 8283–8286.

(27) Jahani, F.; Tajbakhsh, M.; Golchoubian, H.; Khaksar, S. Guanidine hydrochloride as an organocatalyst for N-Boc protection of amino groups. *Tetrahedron Lett.* **2011,** *52*, 1260–1264.

(28) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010,** *53*, 2719–2740.

(29) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.* **1997,** *62*, 7512–7515.

(30) Aurelio, L.; Figler, H.; Flynn, B. L.; Linden, J.; Scammells, P. J. 5-Substituted 2-aminothiophenes as A1 adenosine receptor allosteric enhancers. *Bioorg. Med. Chem.* **2008,** *16*, 1319–1327.

(31) Kobayashi-Matsunaga, Y.; Ishii, T.; Hamaguchi, T.; Osada, H.; Sato, M. Synthesis and characterization of a novel protein tyrosine phosphatase inhibitor, 2-(cyclobutylamino)- *N*- (2-furylmethyl)-2-thioxoacetamide. *Lett. Drug Des. Discovery* **2005,** *2*, 224–227.

(32) Jung, B.; Englberger, W.; Wünsch, B. Molecular modeling directed synthesis of a bicyclic analogue of the δ opioid receptor agonist SNC 80. *Arch. Pharm.* **2005,** *338*, 281–290.

(33) Kuhnert, M.; Köster, H.; Bartholomäus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem., Int. Ed.* ***2015****, 54, 2849-53.*

(34) SHIONOGI & CO., LTD., WO 2003/070277.

(35) Al-Zaydi, K. M. A simplified green chemistry approaches to synthesis of 2-substituted 1,2,3-triazoles and 4-amino-5-cyanopyrazole derivatives conventional heating versus microwave and ultrasound as ecofriendly energy sources. *Ultrason. Sonochem.* **2009,** *16*, 805-809.

(36) Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A Small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011,** *54*, 7784–7796.

(37) Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973,** *22*, 3099–3108.

(38) Otwinowski, Z.; Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307–326.

(39) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 125–132.

(40) Erskine, P. T.; Coates, L.; Mall, S.; Gill, R. S.; Wood, S. P.; Myles, D. A.; Cooper, J. B. Atomic resolution analysis of the catalytic site of an aspartic proteinase and an unexpected mode of binding by short peptides. *Protein Sci.* **2003,** *12*, 1741–1749.

(41) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40*, 658–674.

(42) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 213–221.

(43) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 486–501.

(44) Sheldrick, G. M.; Schneider, T. R. [16] SHELXL: High-resolution refinement. *Methods Enzymol.* **1997,** *277*, 319–343.

(45) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993,** *26*, 283-291.

7. Fragment Screening by Thermal Shift Assay

7.1 Introduction

The concept of fragment-based drug discovery (FBDD) has successfully been exploited within the last 15 years and has led with vemurafenib (Zelboraf®) to one approved drug and more than 30 candidates that entered clinical trials so far.[1–3] Thus, FBDD has proven to be a valuable alternative method for hit identification. It is based on the idea that the chemical space is more efficiently probed with fragments than e.g. by screening a large HTS library.[2]

For successful FBDD the library composition plays an important role. In analogy to the Lipinki's "rule of five" which describes properties for a druglike and orally bioavailable molecule, the "rule of three" was defined as similar rule for fragments by researchers at Astex.[4,5] According to this rule, fragments should have a molecular weight below 300 Da, a ClogP ≤ 3 and three or less H-bond donors as well as H-bond acceptors. However, as the experience with the FBDD approach is continuously growing, also other library design strategies emerged in literature.[6]

As a consequence of their reduced size and complexity, fragments usually exhibit lower affinities (mM-30 µM) in comparison to HTS hits.[7] Thus, for the detection of these rather weakly interacting molecules, robust screening assays or sensitive biophysical techniques such as e.g. high-concentration screening, fragment screening by X-ray crystallography, STD-NMR experiments, or the thermal shift assay (TSA) among others are frequently applied.[8] However, little is known in literature about the hit-overlap utilizing different fragment screening methods. Are screening methods or rather the combination of different prescreening methods able to detect the hits to comparable extent as the time-consuming X-ray crystallography-based screening approach in which all fragments of interest are subjected to crystallization trials in absence of other prescreening options?

In a cooperation project with the research group of Prof. Klebe (Philipps-Universität Marburg), the in-house fragment library comprising 364 fragments[6] was evaluated against endothiapepsin, which represents a robust and well accessible protein to compare and analyze the hit-overlap of different fragment screening methods. In the context of this thesis, the TSA-based fragment screen was established and the results are discussed in the following.

7.2 Concept of the Thermal Shift Assay – an Overview

In drug discovery, the thermal shift assay is commonly used as screening method for hit identification, for affinity ranking of inhibitors as well as for protein stability studies to e.g. identify suitable buffer conditions for protein crystallization or assay conditions.[9–12] Recently, TSA was also used to evaluate drug candidates in a cellular context, e.g. to monitor the transport of drug molecules through cellular barriers to their targets, target specificity as well as dose-dependent target engagement in animals i.e. the analysis of ligand target interactions within cells from appropriate cell cultures or animal tissues.[13] Additionally, TSA has also been applied to obtain information about the binding mode of compounds utilizing wild-type and mutant enzymes.[14]

In a thermal shift assay, the thermal stability of a protein is determined. The thermal unfolding of the protein can be detected in the presence of an appropriate fluorescence dye leading to an increase in the fluorescence, which is measured in dependence of the temperature (Figure 7.1). The fluorescence dye used for TSA is quenched in aqueous solution like the assay buffer, but shows fluorescence in a hydrophobic environment such as the hydrophobic surface of an unfolded protein. Therefore, the increase in fluorescence indicates the unfolding of the protein and the beginning thermal denaturation. In general, binding of ligands to the target protein can either induce a stabilizing effect, or a destabilizing effect, or result in neutral behavior towards the protein of interest being subjected to a heating experiment.

Different dyes have been developed for the TSA measurements, with SYPRO Orange being the most famous one. Its excitation and emission wavelengths are 492 nm and 610 nm, respectively. Thus, it is well suited for the identification of small molecules as this high excitation wavelength generally decreases the likeliness for interfering with these. However, the quenching of the fluorescence of the dye by the respective ligand or other components might falsify the results, as all fluorescence-based assays can be affected by this effect if absorption and emission wavelength are in inappropriate proximity. Besides SYPRO Orange, 1.8-ANS (1-anilino-8-naphthalene sulfonate), bis-ANS, nile red, or dapoxyl sulfonic acid are used and represent an alternative if the measurements do not succeed with SYPRO Orange, or if another wavelength area is desired.[12]

The thermal unfolding of the protein is dependent on the protein stability and can be related to the temperature-dependent Gibbs free energy of unfolding (ΔG_u). The melting temperature (T_m) is the point where ΔG_u is zero and the concentrations of folded and unfolded protein are equal.

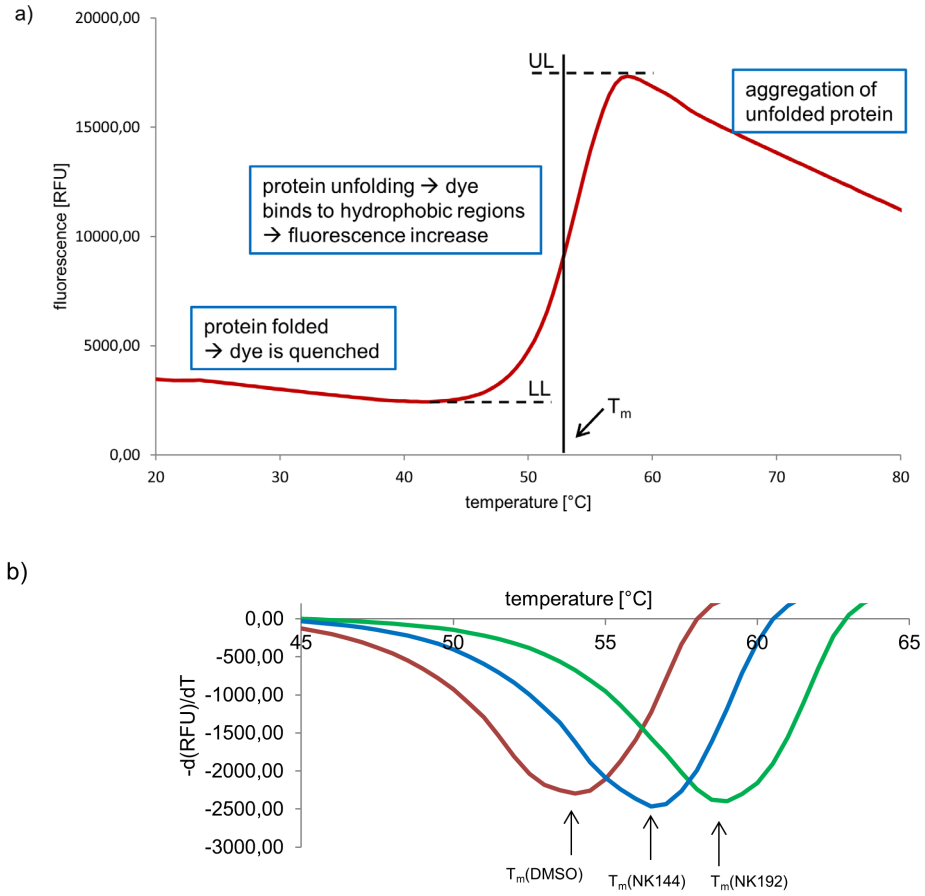


Figure 7.1. a) Typical TSA-curve (example: HIV-1 PR, determined for validation purpose in the context of this thesis). The increase in fluorescence indicates the protein unfolding. LL and UL are the lower and upper level in the fluorescence intensity; T_m is the melting temperature of the protein. b) Derivative of the HIV-1 PR TSA-curves with DMSO (red) and with two HIV-1 PR inhibitors (blue, green). The minimal turning point reflects the T_m-value. Clearly shifts of the melting temperature due to protein stabilization by the ligands are observed.

The melting temperature can be calculated from the first derivative (derivative method) or from the inflection point of the melting curve using the Boltzmann equation.[12,15,16]

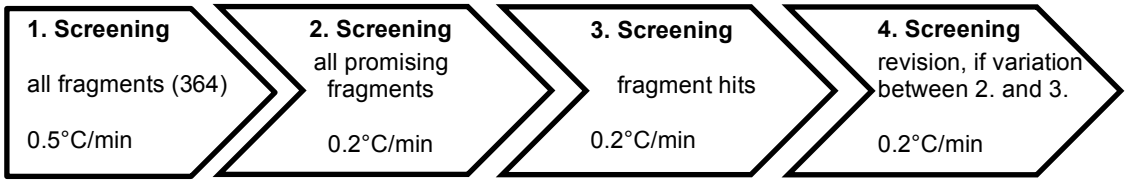
In general, the thermal shift assay is easy to handle, rapid, label-free, inexpensive, and only requires the target protein in low concentration, a fluorescence dye (commercially available), an assay buffer, and in case of an inhibitor screening also the respective inhibitors.[10,12] As no or only little prior knowledge of the protein or the protein function is required and no specific substrate or labeling is needed, the TSA is well-suited as incipient assay setup for the identification of inhibitors, especially for a new therapeutic target.[9]

A drawback of TSA is that this assay, using only native proteins, does not provide information about the way the ligand stabilized the protein: one cannot easily

discriminate whether a ligand stabilized a protein by addressing the active site or by binding to an allosteric site, which both could result in similar ΔT_m-values. Also a ligand can bind to more than one site of the protein simultaneously. The affinity ranking of inhibitors, not only if they comprise different scaffolds, should be interpreted with care as the T_m-shift is dependent on the enthalpy and entropy of binding.[10,12] An intriguing example is given in chapter 6 of this thesis.

7.3 Experimental Setup

Fragments in general exhibit a comparably low affinity to the target protein, which makes it challenging to develop an experimental setup being suitable to detect little changes in the protein stability, which, consequently, results in low T_m-shifts. Therefore, the TSA in this thesis was performed with a high fragment concentration of 2.5 mM to facilitate the detection of binding events. The in-house fragment library was screened according to the following procedure (the step size of the temperature increment per minute is given in the lower line):



7.4 Results, Discussion and Conclusion

In general, the thermal shift assay allows to distinguish between compounds which show a significant shift of the melting temperature in comparison to the DMSO-T_m-value (“hits”), compounds which do not provoke a T_m-shift, and compounds which destabilize the target protein (negative T_m-shift). Typically, for this classification the intervals of 2-3 fold the standard deviation above and below the mean of the recorded melting points are used as cutoff values.

Out of the 364 fragments of the in-house library, 16 fragments could not be measured due to insufficient solubility. Out of the remaining 348 fragments tested in total, the initial screening of the fragment library (Table 7.2, appendix) yielded 52 fragments which showed a shift of T_m higher than the 2-fold standard deviation, 260 fragments which had no significant influence on stabilization of the protein and 21 fragments which destabilized the protein. The data obtained from 15 fragments could not be analyzed accurately.

The fragments which destabilized the protein were not investigated in more detail as they were not considered as putative hits within the scope of this project.
Therefore, in a second screening the 52 fragment hits as well as some fragments which had shown a ΔT_m-value slightly under the 2-fold standard deviation cutoff were measured again, however, this time with a heating rate of 0.2°C/min (Table 7.4, appendix). The hits obtained by this second screening were evaluated again utilizing the same conditions (3. screening) and in case of discrepancies between these two screening results (2. and 3. screening), the fragment was evaluated a third time applying the above mentioned conditions.
Overall, 17 fragment hits which showed a shift of the melting temperature greater or equal than the 3-fold standard deviation (3σ) were identified (Table 7.1). The chemical structures of the ten screening hits with the highest shift of the melting temperature are shown in Figure 7.2.

Table 7.1. ΔT_m-values (mean of the 2.-4. screening) with respective standard deviation of the fragment hits. Fragments possessing a ΔT_m-value of ≥ 3σ (≥ 1.2°C) were defined as “hits”.

Nr.	ΔT_m [°C]	Nr.	ΔT_m [°C]
177*	3.4 ± 0.3	216	1.8 ± 0.1
149	2.8 ± 0.4	335	1.8 ± 0.0
284	2.0 ± 0.1	306	1.7 ± 0.2
017	1.9 ± 0.1	041	1.6 ±0.3
064	1.9 ± 0.1	236	1.5 ± 0.2
159	1.9 ± 0.1	127	1.3 ± 0.2
255	1.9 ± 0.1	337	1.3 ± 0.3
008	1.8 ± 0.2	290	1.2 ± 0.1
178	1.8 ± 0.1		

*fragment measured at 1.25 mM

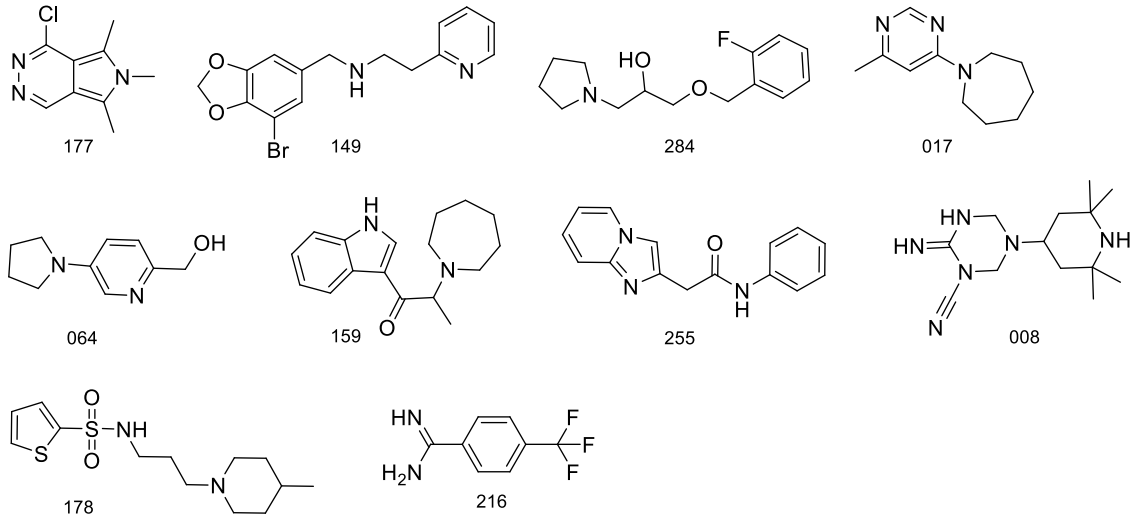


Figure 7.2. Chemical structures of the 10 TSA fragment hits with highest ΔT_m-value.

The screening of the fragment library by a fluorescence-based assay prior to that thesis,[6] performed by Dr. Helene Köster, yielded 55 Hits (> 40 % inhibition at 100 µM). These 55 hits were additionally subjected to a crystallographic screening by soaking of fragment mixtures (2 fragments), which resulted in 11 X-ray structures.[6]

The comparison of the TSA results with those obtained from the fluorescence-based assay and crystallographic screening reveals that most of the hits obtained by TSA could also be found by the fluorescence-based assay: out of the 17 TSA hits, 14 were also identified by the fluorescence assay and for 6 of them an X-ray structure could be determined (Figure 7.3). Considering also the TSA fragments which show a shift of the melting temperature ≥ the 2-fold standard deviation, the hit rate increases from 17 to 31 hits, however, the hit overlap in comparison to the fluorescence assay is only increased by 4 fragments (14 vs 18).

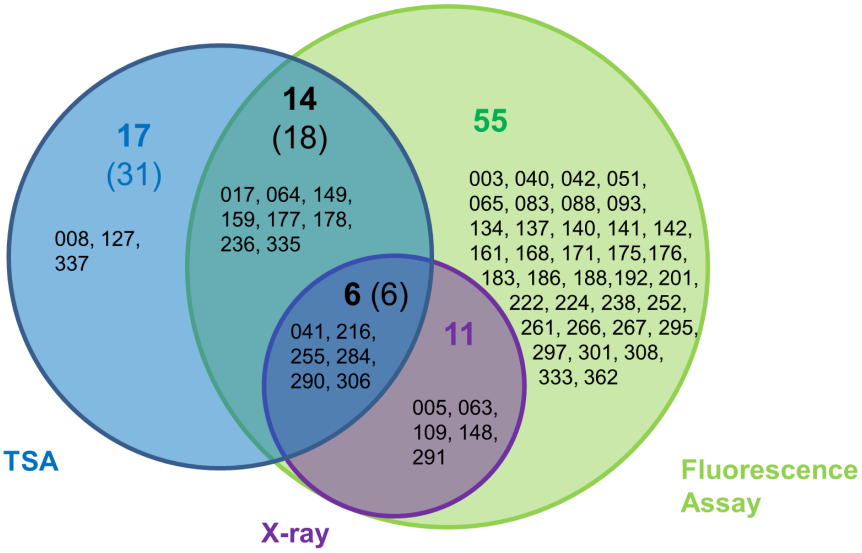


Figure 7.3. Hit overlap of the fragment hits found in the thermal shift assay and in the fluorescence assay. Hits of the TSA are defined as fragments which yield ΔT_m-value of ≥ the 3-fold standard deviation. In brackets the amount of TSA hits with ΔT_m-values ≥ 2σ are shown.

In summary, the TSA screening of the in-house library with in total 348 accessible fragments yielded 17 fragment hits, 14 of which were also identified in the fluorescence-based assay; additional 41 hits could only be detected in the fluorescence assay.

Depending on the outcome of a screen, in a typical setup the cutoff value is adapted to obtain a reasonable amount of hits, which are then evaluated by further screening methods or directly subjected to crystallization trials. By applying less stringent cutoff limits (ΔT_m-shift ≥ the 2-fold of the standard deviation instead of the previous analysis with a threshold of ≥ 3σ), the hit-rate increases in our case from 17 to 31 fragment hits, noteworthy, the hit overlap between both methods remaining nearly unchanged (14 vs 18).

In this part of the collaborative study, the suitability of the TSA-method in general and the established setup used for the identification of stabilizing fragments binding to endothiapepsin was confirmed. A final analysis, however, is not possible until the data of other screening methods as well as the X-ray screening results of the whole fragment library are available.

7.5 Experimental Section

7.5.1 Thermal Shift Assay

TSA measurements were performed in the iCycler IQ5 Real Time Detection System (Bio-Rad). For the assay, 96-well plates (Multiplate^{TM} Low-Profile 96-well unskirted PCR Plates, BioRad) were used with a total volume of 40 µL. The reaction mixture contained 36 µL assay buffer-SYPRO Orange-mix, 3 µL endothiapepsin (29 µM) and 1 µL DMSO/fragment. The assay buffer-SYPRO Orange-mix consisted of the endothiapepsin-assay buffer (0.1 M sodium acetate, pH 4.6, 0.01 % Tween20) and 0.2 % (V/V) SYPRO Orange (protein gel stain 5000x concentration in DMSO, Invitrogen).

The plate was heated from 20°C to 80°C applying a heating rate of 0.5°C/min (1. screening) or 0.2°C/min (2.-4. screening) to allow a more precise determination of the shift of the melting temperature. Unless specified otherwise, the final inhibitor concentration was 2.5 mM.

T_m-values were manually determined from the derivation of the resulting data (-d(RFU)/dT). A least three DMSO reference values ($T_{m(DMSO)}$) were determined per plate. The respective standard deviations of $T_{m(DMSO)}$ were used for defining the cutoff (2-fold or 3-fold of the standard deviation). To obtain the T_m-shifts of the fragments (ΔT_m), the unfolding temperature of endothiapepsin without a ligand ($T_{m(DMSO)}$) was subtracted from the unfolding temperature with ligand ($T_{m(lig)}$). Each T_m-value was at least measured in duplicat.

7.6 Appendix

Table 7.2. Results of the first screening run of all 364 fragments. Unless specified otherwise (column “conc.”), the fragments were measured at 2.5 mM.

x: fragment shows no significant shift of T_m (>2-fold standard deviation).
H: fragment shows a shift of T_m >2-fold standard deviation
d: fragment destabilizes the protein: $-\Delta T_m$ >2-fold standard deviation
n.a.: data could not be analyzed accurately
uk: concentration of the fragment unknown
-: fragment could not be measured (due to low solubility)

Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift
001		x	029		H	057		n.a.	085		x
002		x	030		x	058		x	086		x
003		x	031		d	059		n.a.	087		n.a.
004		x	032		d	060		x	088		n.a.
005		x	033		x	061		x	089		x
006		x	034	-	-	062		x	090		x
007		x	035		x	063		H	091		x
008		H	036	0.75	x	064		H	092		x
009		x	037		x	065		x	093		x
010	1.25	x	038	1.25	x	066		H	094		x
011		x	039		H	067		x	095		x
012		x	040		H	068		H	096		x
013		x	041		H	069		x	097	-	-
014		x	042		x	070		d	098		x
015		x	043	uk	n.a.	071		x	099		H
016		x	044		x	072		d	100		x
017		H	045		x	073		x	101	1.25	H
018		x	046		d	074		x	102		x
019	uk	x	047	1.25	H	075	-	-	103		x
020		x	048		x	076		n.a.	104		x
021		H	049		x	077		x	105		x
022		H	050	1.25	H	078		n.a.	106		x
023		d	051		x	079	uk	x	107	1.25	x
024		x	052		x	080		x	108	-	-
025		x	053		n.a.	081		x	109		x
026		d	054		x	082		x	110		H
027		x	055		d	083		x	111		x
028		x	056		H	084		d	112		x

Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift
113		x	153		x	193		d	233		x
114		x	154		x	194		x	234		x
115		x	155		x	195		x	235	1.25	x
116		x	156	-	-	196		x	236		H
117		x	157		x	197		d	237		x
118		x	158		H	198		x	238		x
119		n.a.	159		H	199		x	239		x
120		x	160		x	200		x	240		x
121		x	161		H	201		x	241		x
122		x	162		H	202		x	242		x
123		x	163		x	203		x	243		x
124		x	164		x	204		x	244	1.25	x
125	-	-	165		x	205		H	245		x
126		x	166		x	206		x	246		x
127		H	167		x	207		H	247		x
128		x	168		x	208		x	248		x
129	1.25	x	169		x	209		x	249		x
130		H	170		x	210	-	-	250		x
131		x	171		x	211		x	251		x
132		H	172		x	212		x	252		x
133		x	173		x	213		x	253		x
134		x	174		x	214		x	254		x
135		x	175		H	215		x	255		H
136		x	176		x	216		H	256		x
137		x	177	1.25	H	217		x	257		x
138		x	178		H	218	-	-	258	1.25	x
139		x	179		x	219		x	259	1.25	x
140		x	180		x	220		x	260		x
141		x	181	-	-	221		x	261		H
142		x	182	-	-	222		x	262		x
143		x	183		x	223		x	263		x
144		x	184		x	224		H	264		x
145		n.a.	185		H	225		x	265	-	-
146		x	186		x	226		x	266	1.25	x
147		x	187		x	227		H	267		x
148		x	188		x	228		x	268		x
149		H	189		x	229	-	-	269		x
150		x	190		x	230		x	270		x
151		x	191		d	231	1.25	x	271	-	-
152		x	192		x	232		x	272		x

Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift	Nr.	conc. [mM]	shift
273		x	296		x	319	0.83	x	342		d
274		x	297		x	320		x	343		x
275		x	298		x	321		H	344		H
276		x	299		x	322		x	345		d
277		x	300	uk	x	323		n.a.	346		x
278		x	301		x	324		x	347		x
279		x	302		x	325		x	348		x
280		x	303		x	326		n.a.	349		d
281		x	304		x	327		x	350		x
282		x	305		d	328		x	351		x
283		x	306		H	329		x	352		d
284		H	307		x	330		x	353		H
285		x	308		x	331		x	354		H
286		x	309		x	332		x	355		x
287		x	310	-	-	333		d	356	uk	x
288		x	311		n.a.	334		H	357		n.a.
289		x	312	-	-	335		H	358		d
290		H	313		d	336		x	359		n.a.
291		x	314		x	337		H	360		x
292		H	315		x	338		x	361		H
293		H	316		x	339		H	362		x
294		x	317		d	340		x	363		x
295		x	318	-	-	341		x	364		x

Table 7.3. Results of the 2.-4. screening: mean ΔT_m-values (at least in duplicate) of all fragment hits of the 1. screening, ordered by fragment numbers. Fragments with ΔT_m-values ≤ 0.4°C (standard deviation) are indicated with "x".

Nr.	ΔT_m	Nr.	ΔT_m	Nr.	ΔT_m
008	1.8 ± 0.2	127	1.3 ± 0.2	255	1.9 ± 0.1
017	1.9 ± 0.1	130	0.9 ± 0.2	261	0.7 ± 0.1
021	x	132	1.0 ± 0.1	284	2.0 ± 0.1
022	1.0 ± 0.4	149	2.8 ± 0.4	290	1.2 ± 0.1
029	0.8 ± 0.1	158	0.8 ± 0.4	292	0.6 ± 0.1
039	0.7 ± 0.3	159	1.9 ± 0.1	293	0.9 ± 0.1
040	0.9 ± 0.2	161	1.1 ± 0.1	306	1.7 ± 0.2
041	1.6 ± 0.3	162	0.7 ± 0.2	321	x
047*	x	175	1.0 ± 0.1	334	1.1 ± 0.2
050*	x	177*	3.4 ± 0.3	335	1.8 ±0.0
056	0.5 ± 0.3	178	1.8 ± 0.1	337	1.3 ± 0.3
063	0.6 ± 0.1	185	x	339	x
064	1.9 ± 0.1	205	0.8 ± 0.2	344	1.0 ± 0.1
066	1.0 ± 0.1	207	0.6 ± 0.0	353	x
068	x	216	1.8 ± 0.1	354	x
099	x	224	1.1 ± 0.1	361	0.6 ± 0.1
101*	0.7 ± 0.1	227	0.7 ± 0.1		
110	x	236	1.5 ± 0.2		

*) fragments were measured at 1.25 mM

Table 7.4. Fragments which yielded ΔT_m-values slightly under the 2-fold standard deviation in the first screening and were measured again, but did not show any significant shift of T_m in the second screening.

094	131	173	217	304	328
098	134	174	267	309	
111	138	176	272	316	
112	140	190	277	322	
121	142	196	289	327	

7.7 References

(1) Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discovery* **2013,** *12*, 5–7.

(2) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008,** *51*, 3661–3680.

(3) Fragments in the clinic: 2015 edition. Avaiable from: http://practicalfragments.blogspot.de/2015/01/fragments-in-clinic-2015-edition.html. [access: 16.04.2015].

(4) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001,** *46*, 3–26.

(5) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug discovery today* **2003,** *8*, 876–877.

(6) Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011,** *54*, 7784–7796.

(7) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004,** *3*, 660–672.

(8) Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012,** *51*, 4990–5003.

(9) Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screening* **2001,** *6*, 429–440.

(10) Lo, M.-C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **2004,** *332*, 153–159.

(11) Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **2006,** *357*, 289–298.

(12) Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007,** *2*, 2212–2221.

(13) Martinez Molina, D.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **2013,** *341*, 84–87.

(14) Śledź, P.; Lang, S.; Stubbs, C. J.; Abell, C. High-throughput interrogation of ligand binding mode using a fluorescence-based assay. *Angew. Chem., Int. Ed.* **2012,** *51*, 7680–7683.

(15) APPLICATION NOTE Protein Thermal Shift™ technology (Life technologies™). Available from:

www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_095306.pdf. [access: 26.02.2015].

(16) Vivoli, M.; Novak, H. R.; Littlechild, J. A.; Harmer, N. J. Determination of protein-ligand interactions using differential scanning fluorimetry. *J. Visualized Exp.* **2014**, 51809.

8. Summary

Aspartic proteases are involved in various physiological and pathophysiological processes. Famous examples highlighting the success of inhibition of aspartic proteases in drug discovery include, among others, e.g. aliskiren which represents a non-peptidic inhibitor of renin, and ten HIV-1 protease (HIV-1 PR) inhibitors which are or have been clinically used in the therapy of AIDS.
This thesis focuses on the identification and synthesis as well as kinetic and structural characterization of non-peptidic small molecule inhibitors of the two aspartic proteases HTLV-1 protease (HTLV-1 PR) and endothiapepsin.

The HTLV-1 PR, a promising target for the treatment of viral infections caused by the human T-cell leukemia virus type-1, is related to the well-known HIV-1 PR, however, it exhibits a substantially different substrate specificity and inhibition profile than the latter. However, due to the similarity in the active site, HIV-1 PR inhibitors provide a promising starting point for the identification and further optimization of novel small molecule inhibitors against HTLV-1 PR.
First, after successful establishment of an in-house HTLV-1 protease technology platform, the well-known HIV-1 PR inhibitor indinavir, which displays a K_i-value in the one-digit micromolar range (3.5 µM) against the HTLV-1 PR, was chosen as auspicious starting point although in comparison to the HIV-1 PR (540 pM) its affinity is strongly reduced. However, other highly potent HIV-1 protease inhibitors (saquinavir, ritonavir, nelfinavir, and amprenavir) do not show any relevant affinity against the HTLV-1 protease (K_i > 20 µM).
Within this thesis the X-ray structure of indinavir in complex with the HTLV-1 PR was determined at 2.40 Å resolution, representing, to the best of our knowledge, the first HTLV-1 PR crystal structure with a non-peptidic inhibitor. This structural information laid the foundation for rationalizing the rather moderate affinity of indinavir against the HTLV-1 PR and thus provided the basis for further structure-guided optimization strategies. Interestingly, indinavir binds at first glance similar to both enzymes. As the H-bond inventory between both proteases is quite conserved, one would not directly expect such a drastic difference in affinity. However, particularly the van der Waals contact inventory established by hydrophobic residues of indinavir seems to be less favorable in the HTLV-1 PR compared to the original HIV-1 PR complex thus presumably accounting for the observed affinity difference.
As a second approach for lead identification, the privileged structure concept was exploited as tool to identify novel small molecule scaffolds for HTLV-1 PR inhibition. A

screening of our in-house aspartic protease inhibitor library was performed and resulted in the identification of C_2-symmetric 3,4-bis-*N*-alkylsulfonamido-pyrrolidines and pyrrolidine-based bicyclic HIV-1 PR inhibitors as promising candidates for HTLV-1 PR inhibition. Both inhibitor classes were characterized in more detail regarding their kinetic as well as structural properties.

Out of a series of ten 3,4-bis-*N*-alkylsulfonamido-pyrrolidine inhibitors, AB84 exhibits an affinity of 15 nM (K_i-value) and represents, to the best of our knowledge, the most potent non-peptidic inhibitor of HTLV-1 PR described so far. The successfully determined crystal structures of AB84 and AB83, the latter being another representative of this inhibitor series, enabled structure-guided SAR interpretations thus laying the foundation for the deduction of design ideas for the further optimization of this inhibitor scaffold. In addition, with the determined crystal structures in hand the observed affinity differences within the members of this compound class towards HTLV-1 PR and HIV-1 PR could convincingly be rationalized.

The pyrrolidine-based bicyclic compounds exhibit affinities from the three-digit up to the one-digit micromolar range, with NK232, decorated with two benzhydryl moieties, being the most potent inhibitor of this series (K_i-value: 1.4 µM). The crystal structure of NK101 was determined at moderate resolution, nevertheless, this structure provides valuable information about the binding mode of this inhibitor scaffold.

Based on these fundamental insights and the deduced SAR described in this thesis, both scaffolds represent promising starting points for the further inhibitor optimization utilizing structure-based drug design.

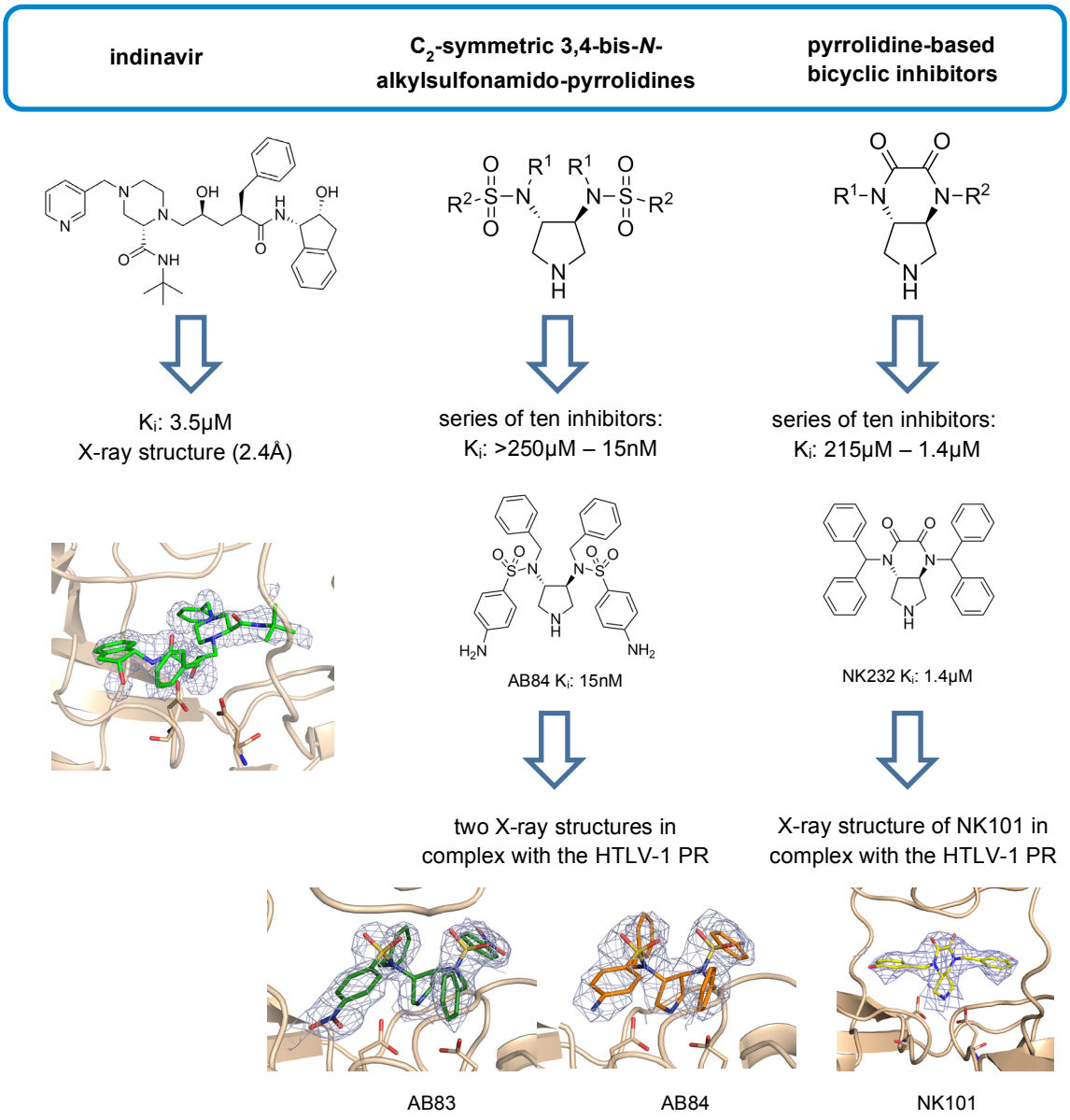


Figure 8.1. Novel scaffolds for HTLV-1 protease inhibition as promising starting points for structure-based lead optimization identified within this thesis.

The second part of this thesis deals with the aspartic protease endothiapepsin that serves as a model system for aspartic proteases in general. Various 2-aminothiophene compounds were synthesized as inhibitors of endothiapepsin utilizing the *Gewald* reaction. Surprisingly, the binding mode analysis of eight similar 2-aminothiophene inhibitors resulted in four completely different binding modes, hence, explaining retrospectively, why the initial deduction of the SAR based on the obtained affinity data had failed. The presented example highlights the complexity of binding events and their strong dependence on seemingly minor effects of the scaffold decoration. As structure-based drug discovery indispensably relies on the correct deduction and interpretation

of the underlying SAR, not only the incipient determination of the binding mode of solely a single lead representative or a set of identified hits at the beginning, but also the re-validation of the initially determined binding mode during the optimization process is of utmost importance. Moreover, we could demonstrate that any inconsistent affinity-thermal shift assay (TSA) correlation may serve as an easily available indicator to qualify candidates for structural revalidation, even if an interaction mode is believed to be known.

The detailed analysis presented herein highlights the necessity to continuously monitor binding modes during the hit-to-lead optimization process and questions the commonly accepted hypothesis that similar ligands usually bind in a similar fashion.

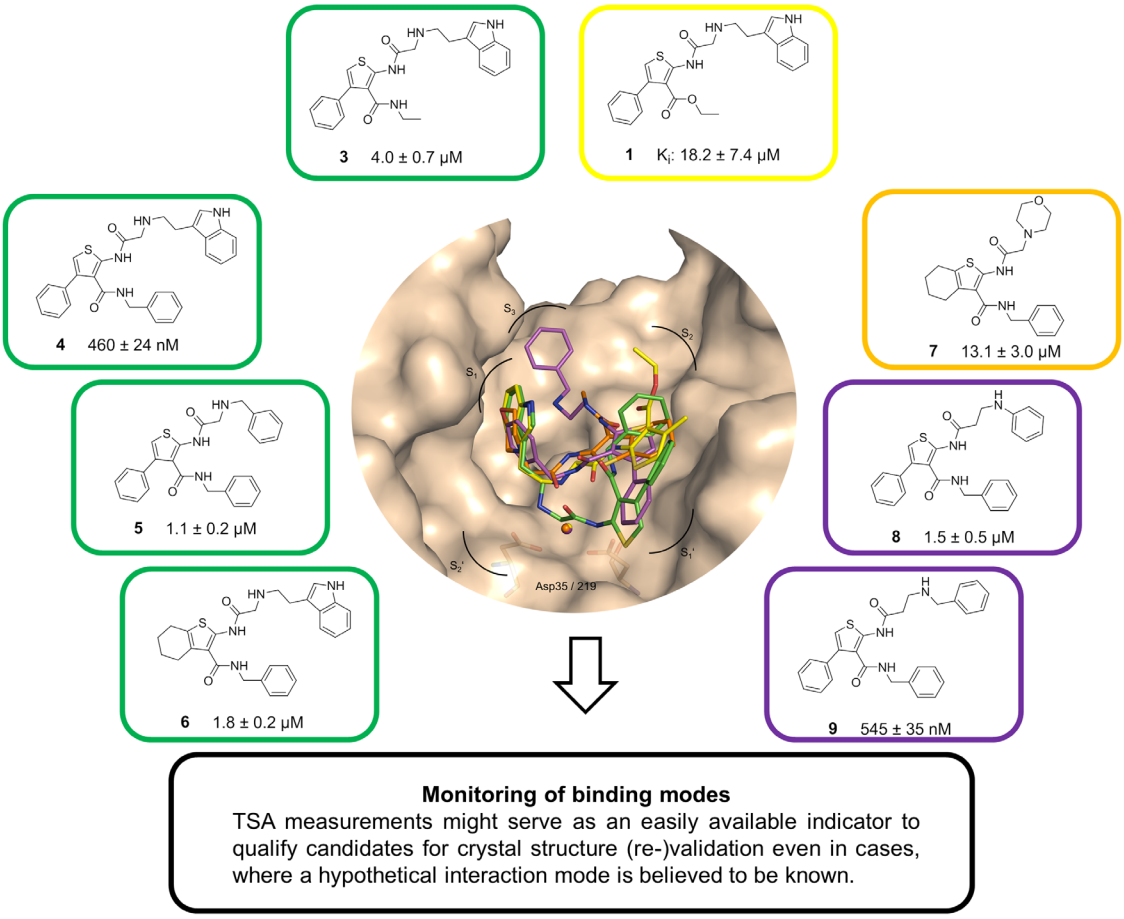


Figure 8.2. Multiple binding modes of *Gewald* reaction-based aspartic protease inhibitors: a representative case study for structure-based lead optimization projects challenging typical design paradigms in medicinal chemistry.

9. Zusammenfassung

Aspartylproteasen spielen in vielen physiologischen aber auch pathophysiologischen Abläufen eine bedeutende Rolle. Zahlreiche Beispiele belegen die erfolgreiche Wirkstoffentwicklung auf dem Gebiet der Aspartylprotease-Inhibitoren: z.B. ist der zur Behandlung von Bluthochdruck eingesetzte Wirkstoff Aliskiren ein nicht-peptidischer Inhibitor des Renin. Des Weiteren werden bzw. wurden zehn HIV-1-Protease-Inhibitoren erfolgreich in der Therapie der Immunschwächekrankheit AIDS eingesetzt.
Diese Arbeit umfasst die Identifizierung und Synthese sowie kinetische und strukturelle Charakterisierung von nicht-peptidischen Kleinmolekülen als Inhibitoren der beiden Aspartylproteasen HTLV-1 Protease und Endothiapepsin.

Die HTLV-1-Protease stellt ein vielversprechendes Zielenzym für die Behandlung von viralen Infektionen mit dem Humanen T-Zell-Leukämie-Virus dar und ist eng verwandt mit der bekannteren HIV-1-Protease. Trotz struktureller Ähnlichkeit beider Proteasen unterscheiden sich diese jedoch deutlich voneinander hinsichtlich ihrer Substrat-spezifität und dem Inhibitionsprofil gegenüber bekannten Inhibitoren. Jedoch stellen auf Grund der Ähnlichkeiten im aktiven Zentrum einige der bereits bekannten HIV-1-Protease-Inhibitoren einen geeigneten und vielversprechenden Startpunkt für die Identifizierung und Optimierung von neuartigen Kleinmolekülen zur Inhibition der HTLV-1-Protease dar, über deren Bindungsmodus vor Aufnahme dieser Arbeiten allerdings noch nichts bekannt war.
Nach erfolgreicher *in-house*-Etablierung der Expression, der Proteinreinigung, eines Affinitätsassays sowie geeigneter Kristallisationsbedingungen für die HTLV-1-Protease wurde zunächst der bekannte HIV-1-Protease-Inhibitor Indinavir, der gegenüber der HTLV-1-PR eine Affinität im einstellig mikromolaren Bereich aufweist, näher untersucht. Die Affinität (K_i-Wert: 3,5 µM) gegenüber der HTLV-1-Protease ist im Vergleich zur HIV-1-Protease (540 pM) zwar deutlich geringer, jedoch weisen andere hoch-affine HIV-1 PR-Inhibitoren (Saquinavir, Ritonavir, Nelfinavir und Amprenavir) keine nennenswerte Affinität gegenüber der HTLV-1-Protease auf (K_i > 20 µM). Im Rahmen dieser Arbeit konnte mit der Kristallstruktur von Indinavir im Komplex mit der HTLV-1-Protease, die nach unserem Wissen erste HTLV-1-Protease-Kristallstruktur mit einem nicht-peptidischen Inhibitor beschrieben werden. Die so erhaltenen struktu-rellen Informationen trugen maßgeblich zum Verständnis der moderaten Affinität von Indinavir gegenüber der HTLV-1-Protease bei und legten darüber hinaus den Grundstein für weitere strukturbasierte Optimierungsansätze. Interessanterweise

bindet Indinavir auf den ersten Blick sehr ähnlich im aktiven Zentrum beider Proteasen (HTLV-1- und HIV-1-Protease): das ausgebildete Wasserstoffbrücken-Netzwerk zwischen beiden Proteasen ist annähernd gleich und liefert zunächst keine direkte Erklärung für den so deutlichen Affinitätsunterschied. Dahingegen scheinen insbesondere die van der Waals-Interaktionen, die von den hydrophoben Gruppen des Indinavirs ausgebildet werden, im Fall der HTLV-1-PR weniger vorteilhaft im Vergleich zur HIV-1-Protease zu sein.

In einem zweiten Ansatz zur Leitstrukturidentifizierung wurde das Konzept der „privilegierten Strukturen“ zur Identifizierung neuartiger niedermolekularer Grundgerüste für die HTLV-1 PR-Inhibition verfolgt. Durch Screening der Arbeitskreis-internen Substanzbibliothek verschiedener Aspartylproteaseinhibitoren konnten C_2-symmetrische 3,4-bis-*N*-alkylsulfonamid-substituierte Pyrrolidine sowie bizyklische Pyrrolidine als vielversprechende Kandidaten zur Inhibition der HTLV-1-Protease identifiziert werden. Beide Inhibitorklassen wurden ursprünglich als HIV-1-Protease-Inhibitoren entwickelt und im Rahmen dieser Arbeit bezüglich ihrer enzymkinetischen und strukturellen Eigenschaften gegenüber der HTLV-1-Protease genauer untersucht.

Aus einer Serie mit zehn C_2-symmetrischen 3,4-bis-*N*-alkylsulfonamid-substituierten Pyrrolidinen stellt die Verbindung AB84 mit einer Affinität von 15 nM nicht nur den affinsten Vertreter dieser Serie, sondern, nach bestem Wissen, auch den bislang affinsten nicht-peptidischen HTLV-1-Protease-Inhibitor dar. Die erfolgreiche Bestimmung der Kristallstrukturen von AB84 sowie AB83, einem weiteren Vertreter dieser Inhibitorklasse, ermöglichte ebenfalls die strukturbasierte Interpretation der Struktur-Wirkungs-Beziehungen (SAR) und legt somit den Grundstein für die weitere strukturbasierte Leitstrukturoptimierung. Die Kristallstrukturen ermöglichten ebenfalls die Interpretation der Affinitätsunterschiede nicht nur innerhalb der Inhibitorserie, sondern darüber hinaus auch im Vergleich zur HIV-1-Protease.

Die bizyklischen Pyrrolidin-basierten Inhibitoren weisen Affinitäten im drei- bis einstellig micromolaren Bereich auf, wobei die mit zwei Benzhydrylgruppen substituierte Verbindung NK232 der bisher aktivste Vertreter dieser Inhibitorserie (K_i: 1,4 µM) ist. Eine Kristallstruktur dieser Verbindungsklasse wurde mit NK101 erfolgreich bestimmt. Obwohl der HTLV-1-Protease-NK101-Komplex nur eine moderate Auflösung aufwies, lieferte er dennoch erste wertvolle Informationen über den Bindungsmodus dieser Substanzklasse in der HTLV-1-Protease.

Basierend auf den erhaltenen, grundlegenden Erkenntnissen über den Bindungsmodus sowie der abgeleiteten SAR bieten beide Grundgerüste vielversprechende Optionen für die weitere strukturbasierte Leitstrukturoptimierung.

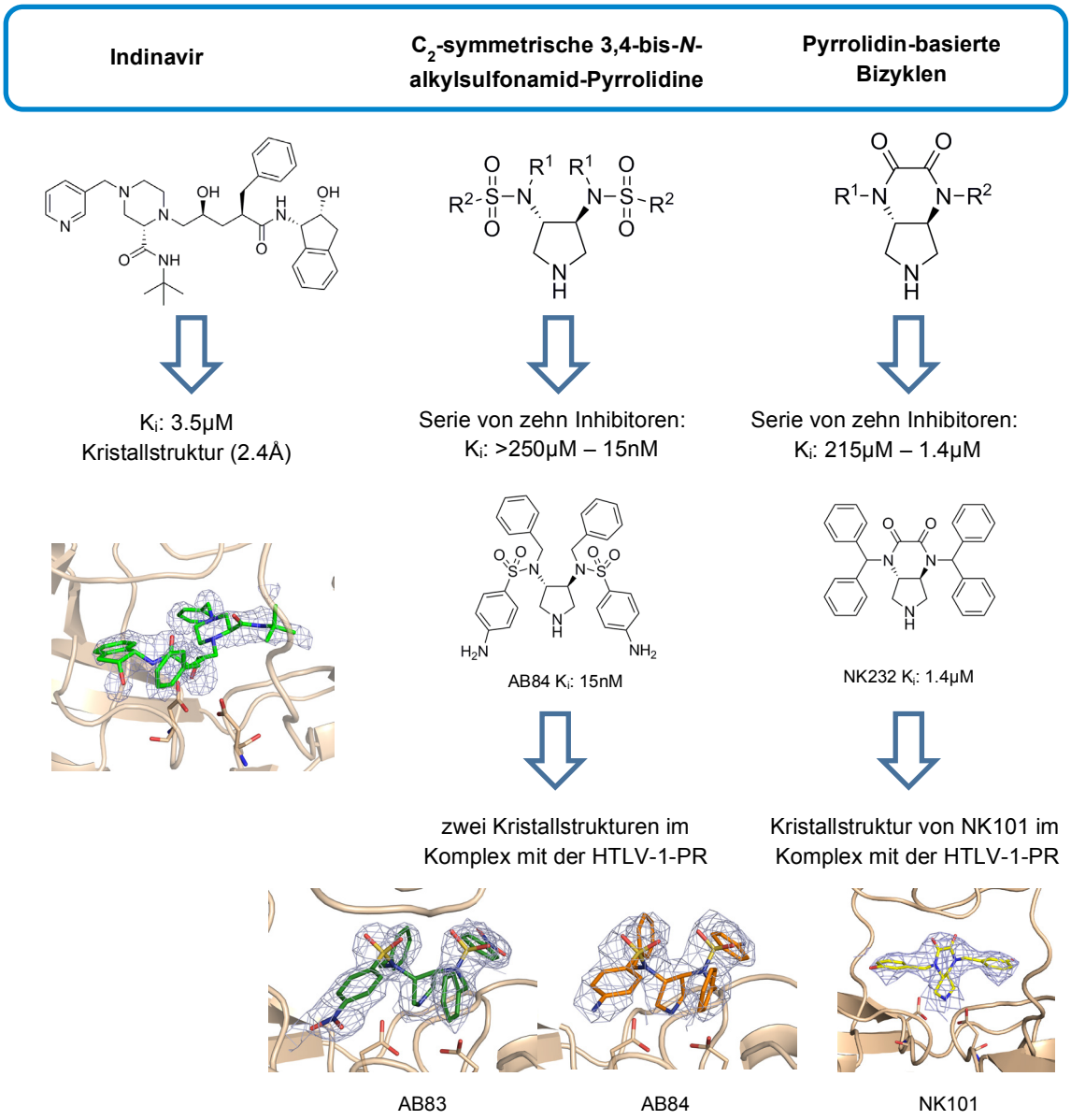


Abbildung 9.1. Neuartige Grundgerüste zur HTLV-1-Protease-Inhibition: aus dieser Arbeit hervorgegangene vielversprechende Startpunkte für eine strukturbasierte Leitstrukturoptimierung.

Der zweite Teil dieser Arbeit beschäftigt sich mit der Aspartylprotease Endothiapepsin, einem bekannten Modellenzym für Aspartylproteasen. Verschiedene Verbindungen mit einem 2-Amino-Thiophen Grundgerüst wurden basierend auf der *Gewald*-Reaktion synthetisiert. Überraschenderweise wurden innerhalb einer Serie von acht strukturell sehr ähnlichen Liganden vier verschiedene Bindungsmodi gefunden, was im Nachhinein erklärt, warum die auf den Affinitätsdaten und der Kristallstruktur der Leitstruktur basierte SAR nicht schlüssig interpretiert werden konnte. Dieses Beispiel unterstreicht eindrucksvoll die Komplexität des Bindungsgeschehens von Liganden und

dessen Abhängigkeit von kleinsten chemischen Veränderungen am Substitutionsmuster einer Leitstruktur. Strukturbasiertes Wirkstoffdesign beruht auf der korrekten Interpretation der zugrundeliegenden SAR, und somit ist nicht nur die anfängliche Bestimmung der Kristallstruktur(en) im Komplex mit einer/mehreren Leitstruktur(en) von größter Bedeutung, sondern ebenfalls die Überprüfung des angenommenen Bindungsmodus während des Leitstrukturoptimierungsprozesses. Darüber hinaus konnte gezeigt werden, dass Unstimmigkeiten bei der Korrelation von klassischen Affinitäts- mit TSA-Daten als Hinweis auf Kandidaten dienen können, für die eine Überprüfung des Bindungsmodus lohnenswert sein könnte, auch wenn ein vermeintlich zugrundeliegender Bindungsmodus bekannt zu sein scheint. Die hier präsentierte Fallstudie unterstreicht die Notwendigkeit der kontinuierlichen Überwachung der Bindungsmodi während eines Leitstrukturoptimierungsprozesses und stellt die im Allgemeinen angenommene Hypothese, dass strukturell ähnliche Liganden in ähnlicher Weise binden, in Frage.

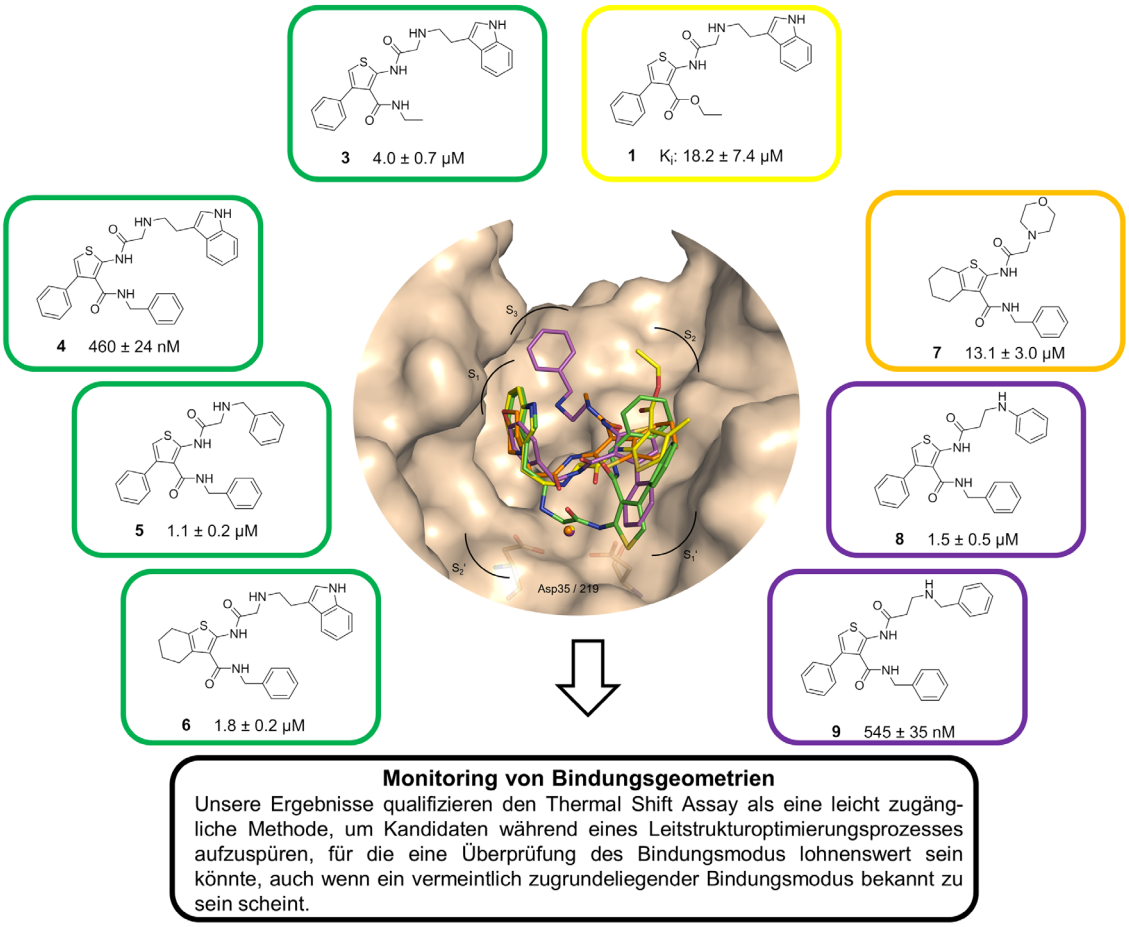


Abbildung 9.2. Wechselnde Bindungsmodi von *Gewald-R*eaktions-basierten Aspartylprotease-Inhibitoren: eine beeindruckende Fallstudie für die strukturbasierte Leitstrukturoptimierung stellt gängige Design-Paradigmen in der Medizinischen Chemie in Frage.

Acknowledgment

Mein herzlicher Dank gilt meiner Doktormutter Frau Prof. Dr. Wibke E. Diederich für die interessante Aufgabenstellung, die bei der Bearbeitung gelassenen Freiräume, die es mir erlaubt haben meine eigenen Ideen zu verfolgen, die konstruktiven Gespräche, das sorgfältige und geduldige Korrekturlesen der Papermanuskripte, sowie für die Unterstützung und stetige Hilfsbereitschaft während der gesamten Promotionszeit. Es war eine schöne, lehrreiche und produktive Zeit.

Ich danke Herrn Prof. Dr. Klaus Reuter für die Organisation und Leitung des S1-Labors, das offene Ohr für S1-bezogene Fragestellungen und für die freundliche Übernahme des Zweitgutachtens.

Herrn Prof. Dr. Carsten Culmsee und Herrn Prof. Dr. Jens Kockskämper danke ich für die bereitwillige Übernahme des Prüfungsvorsitzes und der Nebenfachprüfungen.

Herrn Prof. Dr. Gerhard Klebe möchte ich für die Nutzung des S1-Labors und der Computersoftware danken.

Mein herzlicher Dank gilt meinen Kooperationspartnern: Ich danke Dr. Holger Steuber für die Einführung in die Welt der Kristallographie und für die Verfeinerung der zahlreichen Kristallstrukturen. Ich danke Dr. Andreas Blum und Dr. Nina Klee für die Bereitstellung der synthetisierten HIV-1-Protease-Inhibitoren. Für die Synthese des Assaysubstrats danke ich Dr. Kornelia Hardes.

Meinen ehemaligen und aktiven Kollegen im AK Diederich danke ich für die schöne und ausgelassene Atmosphäre, die gute Zusammenarbeit und die tolle Zeit auch außerhalb des Laboralltags.

Ich danke Steffi Dörr für die Einarbeitung im S1-Labor, welche eine erfolgreiche *in-house* Etablierung des neuen Targets ermöglicht hat und stets für fröhliche Stimmung im S1 Labor gesorgt hat.

Für die weitgehenden fachlichen Diskussionen, Eure Hilfe und Unterstützung während meiner Promotionszeit und nicht zuletzt für die erholsamen Abende möchte ich mich

insbesondere bei Nina Klee, Frithjof Scheer, Helena Rimmer und Jessica Jüngel bedanken.

Bedanken möchte ich mich auch bei Hans-Dieter Gerber für das Korrekturlesen der Papermanuskripte und die Aufnahme des eiligen qHNMRs.

Dr. Alexander Wlodawer danke ich für die zur Verfügungstellung des HTLV-1 Plasmids, was uns die Initiierung des HTLV-1 Projekts ermöglicht hat.

Meinen Vertiefern danke ich für die hilfreiche Unterstützung, insbesondere möchte ich mich bei Markus Lakemeyer für sein großes Engagement bedanken.

Ich danke Lydia Hartleben für die Unterstützung und Geduld bei allen administrativen Angelegenheiten.

Ich möchte mich bei den Computer-Administratoren der AG Klebe insbesondere bei Felix Gut, Michael Betz, Timo Krotzky, Felix Terwesten und Phong Nguyen bedanken.

Reiner Müller danke ich für die hilfreichen Tipps im Fachbereichsdschungel, inkl. der jährlichen Suche nach den Waagen.

Ich danke den Wissenschaftlern der Beamline BESSY II (Berlin) für die Betreuung der Beamline und dem Helmholtz-Zentrum für Materialien und Energie (Berlin) für die freundliche Übernahme der Reisekosten.

Vielen Dank an die Mitarbeiter der Analytik-Abteilungen des Fachbereichs Pharmazie für die Durchführung der Serviceleistungen und die freundliche Hilfsbereitschaft bei Problemen und Fragestellungen.

Meinen Eltern, Geschwistern und Freunden danke ich für die bedingungslose und geduldige Unterstützung, die aufmunternden, kraftvollen Worte und die erholsamen Wochenenden. Vielen Dank für den festen Rückhalt, Euer Vertrauen und Verständnis.

Erklärung

Ich versichere, dass ich meine Dissertation

„**Synthesis, Identification, Kinetic, and Structural Characterization of Inhibitors of the Aspartic Proteases HTLV-1 Protease and Endothiapepsin”**

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den.....

.....

(Maren Kuhnert)

Curriculum Vitae

Diese Seite enthält persönliche Daten. Sie ist daher nicht Bestandteil der Online-Veröffentlichung.