Fragment based Drug Discovery; Design and Validation of a Fragment Library; Computer-based Fragment Screening and Fragment-to-Lead Expansion

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

Philipps Universität Marburg

dem Fachbereich Pharmazie

der PHILIPPS-UNIVERSITÄT MARBURG

vorgelegt von

Tobias Friedrich Craan

aus Korbach

Marburg/Lahn 2011

Vom Fachbereich Pharmazie der Philipps-Universität Marburg

als Dissertation angenommen am: 16.06.2011

Erstgutachter: Prof. Dr. Gerhard Klebe

Zweitgutachter: Dr. Andreas Heine

Tag der mündlichen Prüfung: 17.06.2011

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von März 2008 bis Juni 2011 durchgeführt.

Meinen Eltern

All truths are easy to understand once they are discovered; the point is to discover them.

Galileo Galilei

Abbreviations

AA	arachidonic acid
Å	Ångström (1Å = 10^{-10} m)
Acc	Hydrogen bond acceptor atom type probe
AEA	anandamide; N-arachidonylethanolamide
AKR	aldo-keto reductase
Ala	Alanine
AnD	Hydrogen bond doneptor atom type probe
Arg	Arginine
Aro	Aromatic atom type probe
Asn	Asparagine
Asp	Aspartic acid
BACE1	beta-site APP cleaving enzyme 1
Bcl	B-cell lymphoma
Bcr-Abl	Breakpoint cluster region- Abelson
B-raf	Serine/threonine-protein kinase B-Raf
CDK	Cyclin-dependent kinases
CSD	Cambridge Structural Database
CUDA	NVIDIA’s parallel computing architecture. It enables increases in computing performance by harnessing the power of the GPU.
Cys	Cysteine
Da	Dalton
DMSO	Dimethylsulfoxide
Don	Hydrogen bond donor atom type probe
DXI	D-xylose-isomerase
EctD	ectoine hydroxylase
ETP	Endothiapepsin
F_c	calculated structure factor amplitudes

F_o	observed structure factor amplitudes
FBDD	Fragment based drug discovery
FBLD	Fragment based lead discovery
FIH	factor inhibiting HIF (hypoxia-inducible factor)
FXa	Factor X(10) a
Gln	Glutamine
G-loop	Glycine rich loop
Glu	Glutamic acid
Gly	Glycine
H-bond	Hydrogen bond
His	Histidine
Hsp	heat shock protein
HTS	High throuput screening
Hyd	Hydropobic atom type probe
IC_{50}	Half maximal inhibitory concentration
Ile	Isoleucine
IspD	4-diphosphocytidyl-2C-methyl-D-erythritol synthase
K^+	Potassium ion
K_i	Competitive inhibition constant
K_d	Dissociation constant
K_m	Michaelis Menten constant
LE	Ligand efficiency
Leu	Leucine
LFA-1	Lymphocyte function-associated antigen 1
LTA4A	Leukotriene A4 Hydrolase
Lys	Lysine
M	Molarity (mol/L)
µM	Micromolar
MD	Molecular dynamics
Met	Methionine

MMP	Matrix metalloproteinase
nM	Nanomolar
NADP(H)	Nicotinamide adenine dinucleotide phosphate (reduced form)
NMR	Nuclear magnetic resonance
ns	nano second
o/w	Octanol/water partition
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDGFR	Platelet-derived growth factor receptor
PEG	Polyethylene glycol
pH	Potentialis hydrogenii
Phe	Phenylalanine
PKA	c-AMP dependent Protein Kinase A
PPKA	Peroxisome proliferator-activated receptor
Pro	Proline
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RO3	Rule of three
RO5	Rule of five
SAP2	Secretory aspartyl protease 2
Ser	Serine
SPR	Surface Plasmon Resonance
T_m	melting temperature
Thr	Threonine
TLN	Thermolysin
TPSA	Total polar surface area
Trp	Tryptophane
Tyr	Tyrosine
Val	Valine
VEGF	Vascular endothelial growth factor

Table of Contents

Abbreviations **9**

Table of Contents **13**

1 Motivation **17**

1.1 Aims of this Thesis 17

2 Introduction to Fragment Based Drug Design **19**

2.1 General Concept of Fragment Based Drug Design 19

2.2 Fragment Growing Approach 23

2.3 Fragment Library Design 24

3 A Small Non-rule-of-3-compatible Fragment Library Provides a High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes **27**

3.1 Introductory Remarks 27

3.2 Abstract 27

3.3 Introduction 28

3.4 Results 30

3.5 Discussion 44

3.6 Conclusions 45

3.7 Materials and Methods 46

4 HotSpot Analysis – a Promising Strategy for Lead Optimization; an Endothiapepsin Fragment Screen as Case Study **53**

4.1 Introductory Remarks 53

4.2 Introduction 53

4.3 Methods 54

4.4 Results 57

4.5 Conclusion and Outlook 65

5 Experimental Active Site Mapping as a Starting Point to Fragment Based Lead Discovery **67**

5.1 Introductory Remarks 67

5.2 Introduction 67

5.3 Results and Discussion 69

5.4 Conclusion 90

5.5 Materials and Methods 91

6 From Probe to Fragment and Lead: A Combined Approach of Experimental Fragment Screening and Computational De Novo Design 93

6.1 Introductory Remarks 93

6.2 Introduction 93

6.3 Results and Discussion 95

6.4 Conclusion 112

7 Early Steps of G-loop opening for Protein Kinase A (PKA) in complex with Phenol 115

7.1 Introductory Remarks 115

7.2 Introduction 115

7.3 Results and Discussion 120

7.4 Conclusions and Outlook 132

7.5 Methods 133

7.6 Supplementary Informations 134

8 New Scaffolds for Aldose Reductase: A Virtual Screening Study 141

8.1 Introductory Remarks 141

8.2 Drug Design 141

8.3 Target Family of AKRs 142

8.4 Ligand Database and Targeted Library Design 143

8.5 Conformational and Pharmacophore Search 145

8.6 Docking 145

9 Development of a Thermal Shift Assay 149

9.1 Introductory Remarks 149

9.2 Introduction 149

9.3 Concept of Thermal Shift Assay 149

9.4 Indrodustion to EctD 150

9.5 Validation of the Method Using Different EctD Variants 152

9.6 Residues Likely to be Involved in 2-Oxoglutarate Binding by EctD 156
9.7 Materials and Methods: Fluorescence-Based Thermal Shift Assay 157

10 RNA Editing Modulates the Binding of Drugs and Highly Unsaturated Fatty Acids to the Open Pore of Kv Potassium Channels. 159

10.1 Introductory Remarks 159
10.2 Abstract 159
10.3 Docking and Modelling Results 160

Conclusion and Outlook 163

Zusammenfassung und Ausblick 165

Literature 167

Danksagung 179

Erklärung 181

Curriculum Vitae 183

1 Motivation

1.1 Aims of this Thesis

In recent years fragment screening has become a popular approach to identify new lead structures. Fragments are usually defined by the Astex ‘rule of three’ (RO3). SPR, NMR spectroscopy, biochemical assays and X-ray crystallography are efficiently applied screening techniques to discover prospective fragments as binders. These methods require as a starting point a fragment library assembled by some predefined criteria. For example commercially available fragment libraries are designed using the ‘RO3’ as strict selection threshold.

Nevertheless, in Chapter 3 we critically ask whether these rules are too stringent for the design of an optimal fragment library, holding candidates that leave sufficient room for subsequent chemical modifications. With respect to strategies focusing on fragment growing and fragment merging, an appropriate number of functional groups for follow-up chemistry is required. Frequently, groups to enable a reasonable linking and growing chemistry display properties such as hydrogen bond donors and/or hydrogen bond acceptors. As such they can be used as favorable entry points to begin with the required optimization chemistry.

Therefore, we have designed an in-house fragment library. During library design we modified the ‘RO3’ from Astex, and we did no strict filtering by physico-chemical properties during fragment enumeration. We could assemble a rather small fragment library of 364 members which we validated on endothiapepsin - a well established in-house target.

The development of novel tools for computational approaches in the field of fragment-based drug design and their validation with respect to experimental data are important tasks to estimate the scope and relevance of computational methods. In Chapter 4 and 5 HotspotsX analyses using eleven determined endothiapepsin-fragment structures and using different molecular probes for diverse targets were performed.

Based on the molecular probe phenol, we performed a case study on protein Kinase A (PKA) in collaboration with Merck Serono which is presented in Chapter 6. Starting from a phenol molecule, we performed a virtual screening and discovered a prospective hit which could be structurally characterized in a second crystal structure. This initial seed was grown into the binding pocket and lead to a 70 µM lead compound. The affinity of this compound could be increased to 110 nM in three design cycles.

The PKA-phenol complex structure displays surprisingly a rather closed G-loop conformation, which must be induced by the phenol molecules. MD simulations described in Chapter 7, were performed to get some first insights into the G-loop opening mechanism leading to the uncomplexed apo protein.

In Chapter 8 a virtual screening approach was described which resulted in new scaffolds for the inhibition of aldose reductase.

Biophysical assays are highly desired to detect fragment hits for a novel protein. In Chapter 9 a thermal shift assay is described which has been applied to detect fragment hits and which has also been used to characterize variant mutations of EctD.

Nils Decher et al. experimentally observed the binding of arachidonic acid in a K^+ ion channel. To rationalize these findings, a docking study is described in Chapter 10 that was performed to get first ideas about the possible binding mode and structural explanations of the experimentally observed data.

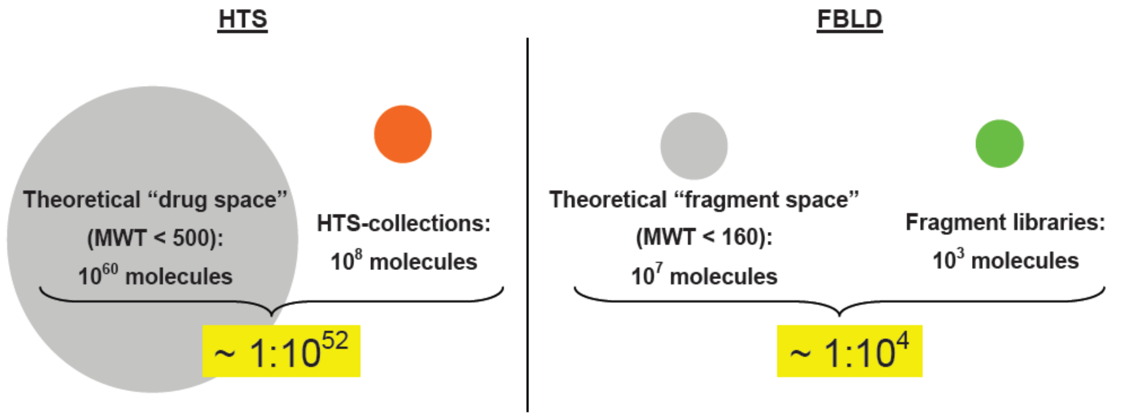
2 Introduction to Fragment Based Drug Design

2.1 General Concept of Fragment Based Drug Design

The more novel and interesting target structures for a therapeutic treatment are discovered and validated via proteomics and structural genomics, the higher will be the demand for efficient strategies to discover leads that interfere with protein function. In the nineties, high-throughput screening and combinatorial chemistry have been established to resolve the bottleneck for an efficient lead discovery. Subsequently, virtual computer screening has been added as an alternative to complement these approaches. However, success rates were not as expected and the size of the usually discovered hits was in the range of common drug molecules not leaving much space for optimization without significantly exceeding the molecular weight limit of approx. 500 Da.[1] To better rank the size of the discovered hits with their actual potency, the concept of ligand efficiency[2] was introduced. Highly efficient leads exhibit good potency combined with low-molecular weight.

Fragments are much smaller than usual HTS compounds. Due to the size difference between fragments and HTS candidate molecules the library to explore a certain chemical space will be much smaller for fragments.

Figure 2.1 illustrates the difference in the theoretical chemical space. An HTS screen using a collection of 10^8 compounds of a size up to 500 Da matches a rather small portion of the theoretically possible chemical space. However, a fragment library of approximately 1000 fragments matches a significantly larger portion of chemical space covering molecular weights up to 160 Da. Thus, in FBLD it appears simpler to cover a larger portion of chemical space compared to the required compound depository in HTS.



Source: Fragment-Based Approaches in Drug Discovery, (2006), Wiley VCH

Figure 2.1 Chemical space used in HTS compared to FBLD approaches.[3]

Advances in biophysical techniques to detect protein-ligand binding and increasing success to acquire structural information about protein-ligand complexes by crystallography[4-6] or NMR spectroscopy[7-9] allowed to push the limits of the compounds to be screened to lower molecular weight. Particularly NMR and Surface Plasmon Resonance (SPR) are nowadays reliable enough to detect very small and weak binders which still exceed sufficient and convincing potency ("high" ligand efficiency).[2] Other approaches intend to screen directly with proteins in crystalline state.[6, 10] Fragment-type[11] leads appear as special challenge as they provide – once characterized in terms of a crystal structure – wide opportunities for optimization into prospective drug candidates.

Figure 2.2 illustrates the different starting points for drug development. The goal is the area indicated by the red square which represents the Lipinski-like drug molecules. The bottom right corner represents a drug candidate with ~ 500 Da/ ~ 35 heavy atoms and an affinity of ~ 10 nM. This compound would have a LE of 0.3 kcal/mol*heavy atom. Considering the different starting points it can be seen that a fragment as entry point with low affinity but high LE should be more convenient to develop a drug-like molecule, as affinity can be gained by directly growing the fragment. An HTS hit on the other hand is a starting point based on higher affinity and higher molecular weight but lower LE. First step in the development of such hits should be the increase of LE. Therefore, the molecular weight has to be reduced first without significant loss of affinity.

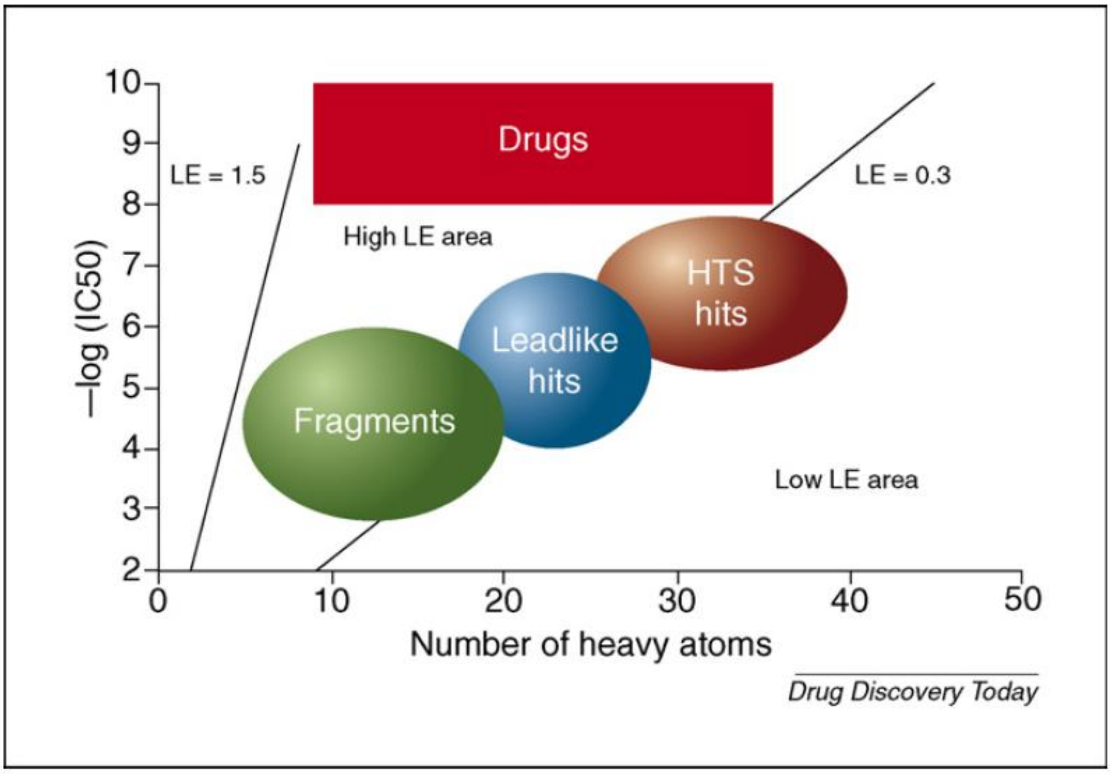


Figure 2.2 The starting points for drug development taken from Siegal et al.[12]

Meanwhile, many examples have been described in literature and an impressive number of reviews has been written.[13-24] Mostly, this method has been applied in industry or specialized small biotech companies. Primarily these studies considered the following targets: HSP90[25-30], different kinases[31-38], phosphatases[39-41], antibacterial and anti-infective targets[42-46] , BCL-2 [47] and BACE.[47-50] Table 2.1 lists clinical candidates originally discovered and developed by fragment-based drug design approaches. Nowadays a significant number of pharmaceutical companies uses this approach to develop novel drug candidates.

Table 2.1 Fragments in the clinic Sep. 2010[51]

Phase 3		
PLX-4032	Plexxikon	B-RafV600E

Phase 2		
ABT 263	Abbott	Bcl-2/Bcl-xL inhibitor
ABT 869	Abbott	VEGF & PDGFR inhibitor
AT9283	Astex	Aurora inhibitor
LY-517717	Lilly/Protherics	FXa inhibitor
Indeglitazar	Plexxikon	PPAR agonist
VER-52296/	NVP-AUY-922	Hsp90 inhibitor

Phase 1		
ABT-518	Abbott	MMP-2 & 9 inhibitor
ABT-737	Abbott	Bcl-2/Bcl-xL inhibitor
AT13387	Astex	Hsp90 inhibitor
AT-7519	Astex	CDK1,2,4,5 inhibitor
DG-051	deCODE	LTA4H inhibitor
IC-776	Lilly/ICOS	LFA-1 inhibitor
LP-261	Locus	Tubulin inhibitor
PLX-5568	Plexxikon	Kinase inhibitor
SGX-393	SGX	Bcr-Abl inhibitor
SGX-523	SGX	Met inhibitor
SNS-314	Sunesis	Aurora inhibitor

Some of the compounds from industry have been published, mostly those of Astex Pharmaceuticals. Astex is a company with a business model that is entirely attributed to fragment-based approaches. Figure 2.3 illustrates the development of two kinase[38, 52]and one Hsp90 clinical candidates.[30]

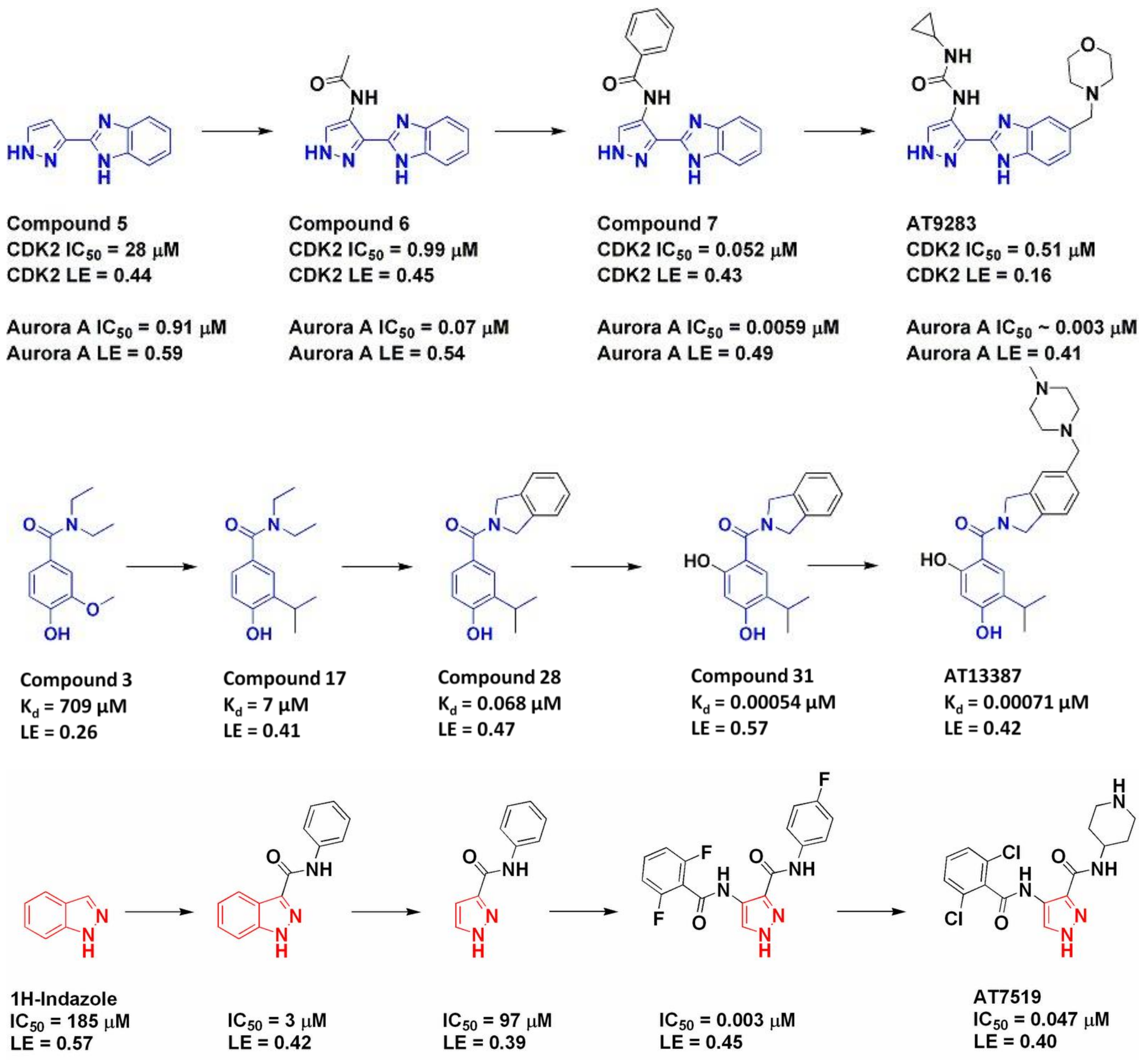

Compound 5
CDK2 IC$_{50}$ = 28 μM
CDK2 LE = 0.44
Aurora A IC$_{50}$ = 0.91 μM
Aurora A LE = 0.59
Compound 6
CDK2 IC$_{50}$ = 0.99 μM
CDK2 LE = 0.45
Aurora A IC$_{50}$ = 0.07 μM
Aurora A LE = 0.54
Compound 7
CDK2 IC$_{50}$ = 0.052 μM
CDK2 LE = 0.43
Aurora A IC$_{50}$ = 0.0059 μM
Aurora A LE = 0.49
AT9283
CDK2 IC$_{50}$ = 0.51 μM
CDK2 LE = 0.16
Aurora A IC$_{50}$ ~ 0.003 μM
Aurora A LE = 0.41
Compound 3
K$_d$ = 709 μM
LE = 0.26
Compound 17
K$_d$ = 7 μM
LE = 0.41
Compound 28
K$_d$ = 0.068 μM
LE = 0.47
Compound 31
K$_d$ = 0.00054 μM
LE = 0.57
AT13387
K$_d$ = 0.00071 μM
LE = 0.42
1H-Indazole
IC$_{50}$ = 185 μM
LE = 0.57
IC$_{50}$ = 3 μM
LE = 0.42
IC$_{50}$ = 97 μM
LE = 0.39
IC$_{50}$ = 0.003 μM
LE = 0.45
AT7519
IC$_{50}$ = 0.047 μM
LE = 0.40

Figure 2.3 Development scheme of some Astex clinical candidates. Top: the development of AT9283, an Aurora kinase inhibitor.[52] Middle: the development of AT13387, an Hsp90 inhibitor[30] and bottom: AT7519, a CDK inhibitor.[38]

2.2 Fragment Growing Approach

The idea to move towards smaller and smaller initial “leads” reminds of an old concept developed in the early phase of structure-based drug design. In the late eighties and early nineties computational *de novo* design was developed with much enthusiasm. Starting with a very small “seed”, initial ligands are gradually grown into the binding pocket of the target protein. Even though received with much sympathy in the beginning, the approach rapidly got out of fashion. This was either attributed to its high complexity, too low success rate also due to limited crystallographic access and experience to determine complex structures or it was simply overrun by the upcoming high-throughput technologies; and at present we clearly witness that fragment screening tends to move into the high-throughput area again.

Nevertheless, the appealing aspect of *de novo* design was its rigorous rational concept, only once ligand binding is fully understood this kind of design based on first principles can work. However, two decades later we have to confess that the binding process is still so little understood and therefore a purely rational approach remains as such still rather ambiguous.

On the other hand, meanwhile our experimental techniques have been much better developed and allow much faster access to structural information. Therefore, it may be asked whether a combination of *de novo* design supported by multiple crystal structure analyses will now allow for an alternative strategy in lead discovery, reconsidering much of the early concepts of *de novo* design.

Therefore, we picked a representative example from the heavily studied family of protein kinases to perform a feasibility study. Ongoing from very small, promiscuously binding probes to potent leads using the concepts of *de novo* design. However, strongly supported by iterative crystal structure analysis particularly in the early phase of drug design where the detection of ligand binding would be impossible applying routine screening techniques.

2.3 Fragment Library Design

Drug-like molecules are usually defined by the Lipinski rules (‘RO5’).[1] This ‘rule of five’ tries to summarize properties to be met to achieve oral availability which is essential for an active substance as prospective candidate for clinical trials. Similar rules have been defined for fragments and became popular as the Astex ‘rules of three’ (‘RO3’).[11] These rules reduce the various thresholds from five to three. In particular, the crucial molecular weight is < 300 g/mol, the number of hydrogen bond donors and acceptors is ≤ 3, XlogP is ≤ 3,[11] the number of freely rotatable bonds is ≤ 3 and the polar surface area is ≤ 60 Å^2. As a consequence fragment libraries assembled and offered commercially by many providers over the last years try to stick to these rules and select the fragments largely in agreement with Astex ‘RO3’.

In this context, we wanted to critically ask whether these rules are too strict for a fragment library holding candidates that leave sufficient room for subsequent chemical modifications. In terms of strategies focusing on fragment growing and fragment merging, an appropriate number of functional groups is required to still perform reasonable linking chemistry. Frequently, the groups used for the follow-up chemistry display properties as hydrogen donors and/or hydrogen acceptors. Therefore in the following they can be used also as synthetic handles to start the required optimization chemistry. As a consequence, we designed a small fragment library without adhering strictly to the RO3 criteria for fragments. It should be well suited either for crystallographic screening and follow-up chemistry. Subsequently, we screened endothiapepsin with this library. Besides finding new fragment binders for this protease and their crystallographic characterization, we were interested to find out how well the discovered fragments agree with the Astex ‘RO3’.

Endothiapepsin serves as a model system for the large group of pepsin-like aspartic proteases. In this family, there are many proteins that are involved in serious diseases such as malaria (plasmepsins), fungal infections (secreted aspartic proteinases), Alzheimer's disease (β-secretase) and hypertension (renin). In the eighties this protein was the working horse for the development of blood pressure depressions before the crystal structure of renin became available. This underlines the relevance of endothiapepsin as suitable reference protein. In the past it has already been successfully used as a model system by others. Here, it provided important information to understand the details of the catalytic mechanism[53] as well as developing renin inhibitors.[54] Endothiapepsin was also a successful model in discovering novel fragments binding to β-secretase.[55]

3 A Small Non-rule-of-3-compatible Fragment Library Provides a High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes

3.1 Introductory Remarks

The present study has been accomplished together with Helene Köster and the industry partners Boehringer Ingelheim, Proteros Biostructures and Merck Serono in the framework of a common BioChancePlus BMBF project. The following chapter has been prepared as a publication in a scientific journal. Therefore, some redundancies with the text in the previous chapter are present. My contribution to this work has been the data mining of the fragment library and the selection of fragments at the Marburg site.

3.2 Abstract

Drug-like molecules are defined by the Lipinski rules of five. Similar rules have been defined to characterize fragments. Reducing the various thresholds from five to three and became popular as Astex rule of three. They have been heavily applied during assembly of fragment libraries and providers frequently use these rules to select fragments for commercial offer.

In this contribution, we ask whether these rules are too stringent to compose a fragment library with candidates leaving sufficient room for subsequent chemical modifications with respect to fragment growing and merging, where an appropriate number of functional groups for chemical transformations is required. Usually these groups exhibit properties as hydrogen bond donors and/or acceptors and provide favorable entry points to start the required optimization chemistry. Therefore, we designed a small fragment library without strictly applying criteria set by the rule of three on the physico-chemical properties. In consequence, our 364-membered fragment library is not consistent with the rule of three.

As a case study for fragment binding, we studied the aspartic protease endothiapepsin with our fragment library. As initial screening we employed a biochemical assay based on the cleavage of a fluorogenic substrate. All compounds were screened at 1 millimolar concentration. In this pre-screen we defined 'hits' to inhibit the target enzyme by at least 40 %. Fifty-five hits were suggested out of our 364-membered library; which were subsequently subjected to a crystallographic study by soaking them into native endothiapepsin crystals. Eleven complex crystal structures were determined covering fragments which exhibit diverse binding modes. They can be divided into three categories: direct binding to the aspartates of the catalytic dyad; binding to the aspartates mediated by a water molecule; and no direct interaction with the catalytic dyad. They cover binding to different specificity pockets of the protease. Only four of the eleven fragments are consistent with the Astex rule of three. Restriction of our library to this rule would have limited our list of fragment hits with respect to the variety of chemotypes.

3.3 Introduction

The accelerated discovery of novel and interesting target structures for therapeutic intervention, often stimulated by proteomics and structural genomics, increasingly demands for efficient strategies to provide first leads that interfere with protein function. In the nineties high-throughput screening and combinatorial chemistry have been established to resolve the bottleneck for an efficient lead discovery. Subsequently, virtual computer screening has been added as an alternative to complement these approaches. However, success rates were not as expected and the size of the discovered hits was in the range of common drug molecules, not leaving much space for optimization without significantly exceeding the molecular weight limit of about 500 g/mol.[1] To better evaluate the size of a discovered hit with respect to its actual potency, the concept of ligand efficiency[2] was introduced. Highly efficient leads exhibit good potency combined with low molecular weight.

Advances in biophysical techniques to detect protein-ligand binding and increasing success to acquire structural information about protein-ligand complexes by crystallography[4-6] or NMR spectroscopy[7-9] allowed to push the limits of the compounds to be screened to lower molecular weight. Particularly NMR and Surface Plasmon Resonance (SPR) are nowadays supposed to be reliable enough to detect very small and weak binders which exhibit sufficient and convincing potency (“high” ligand efficiency).[2] Other approaches intend to screen directly with proteins in crystalline state.[6, 10] Fragment-type[11] leads appear as special challenge as they provide – once characterized in terms of a crystal structure – wide opportunities for optimization into prospective drug candidates. Many examples have meanwhile been described in literature and an impressive number of reviews has been written[13-24], mostly developed in industry or specialized small biotech companies along the following targets: HSP90[25-30], different kinases[31-38], phosphatases[39-41], antibacterial and anti-infective targets[42-46] , BCL-2[47] and BACE.[47-50]

Drug-like molecules are usually defined by the Lipinski rules (‘RO5’).[1] This ‘rule of five’ tries to summarize properties to be met to achieve oral availability which is essential for an active substance as prospective candidate for clinical trials. Similar rules have been defined for fragments and became popular as the Astex ‘rules of three’ (‘RO3’).[11] These rules reduce the various thresholds from five to three. In particular, the crucial molecular weight is < 300 g/mol, the number of hydrogen bond donors and acceptors is ≤ 3, XlogP is ≤ 3,[11] the number of freely rotatable bonds is ≤ 3 and the polar surface area is ≤ 60 $Å^2$. As a consequence fragment libraries assembled and offered commercially by many providers over the last years try to stick to these rules and select the fragments largely conform with Astex ‘RO3’.

In this contribution we want to critically ask whether these rules are too strict for a fragment library holding candidates that leave sufficient room for subsequent chemical modifications. In terms of strategies focusing on fragment growing and fragment merging an appropriate number of functional groups is required. Frequently these groups display properties as hydrogen donors and/or hydrogen acceptors. They can be used also as synthetic handles to start the required optimization chemistry. Therefore, we designed a small fragment library without adhering strictly to the RO3 criteria for fragments. They should be well-suited either for crystallographic screening and follow-up chemistry. Subsequently, we screened endothiapepsin with this library. Besides finding new fragment binders for this protease and their crystallographic characterization we were interested to find out how well the discovered fragments agree with the Astex ‘RO3’.

Endothiapepsin serves as a model system for the large group of pepsin-like aspartic proteases. In this family there are many proteins that are involved in serious diseases such as malaria (plasmepsins), fungal infections (secreted aspartic proteinases), Alzheimer's disease (β-secreatase) and hypertension (renin). In the eighties, endothiapepsin was the working horse for the development of blood pressure depressions before the crystal structure of renin became available. This underlines the relevance of endothiapepsin as suitable reference protein. In the past, it has already been successfully used as a model system by others. Here, it provided important information to understand the details of the catalytic mechanism[53] as well as developing renin inhibitors.[54] Endothiapepsin was also a successful model in discovering novel fragments binding to β-secretase.[55]

3.4 Results

Library Design

Initially, we inquired several chemical suppliers (ASINEX, ChemBridge, MayBridge, InterBio Screen, LifeChemicals, Enamine, Specs, Vitas M Laboratory) to name their commercially offered compounds with ≤ 20 non-hydrogen atoms, including only C, N, O, F, Cl, Br, P and an availability of at least 100 mg. In spring 2009, this resulted in 238,224 compounds. Since the library was intended to be screened crystallographically, fragments containing strong X-ray scatterers such as bromine were considered advantageous. Due to the high molecular weight of such electron rich atoms we restricted the size of our fragments by the number of non-hydrogen atoms instead of molecular weight. Thus, we defined the fragments to have between 8 and 20 non-hydrogen atoms. The range of the molecular weight in the final library is between 122 g/mol and 359 g/mol with an average at 224 g/mol (Figure 3.1, Table 3.1).

Next, we filtered for particular functional groups to discard potentially toxic, unwanted or chemically unstable moieties. The applied filters are similar to those developed by Baurin and colleagues at Vernalis.[56]

All physico-chemical properties have been calculated under the assumption of standard protonation states at physiological pH conditions using MOE.[57] Rotatable bonds were not restrained (Figure 3.1). Therefore, the number of rotatable bonds in the final database varies from 0 to 7 with an overall average of 1.7 (Table 3.1). The logP value was calculated within the MOE software (clogP(o/w))[57] and was filtered to be ≤ 3. For particular chemical motifs the threshold was expanded to 5.4 (Figure 3.1). Thus, the calculated lipophilicity spreads from -1.25 to 5.39 with a mean of 1.58 (Table 3.1).

For hydrogen-bond acceptors we filtered differently from the ‘RO3’ to be more comparable to the ‘RO5’ criteria.[1] There the number of hydrogen-bond acceptors is multiplied by a factor of two considering 10 hydrogen-bond acceptors as appropriate. Transferring this factor to the ‘RO3’ for fragments, six hydrogen-bond acceptors are acceptable. Only in special cases we allowed fragments to exceed this number of hydrogen-bond acceptors as upper limit (Figure 3.1). In the final fragment library the hydrogen-bond acceptor range falls between 1 and 7 with an average of 3.7. The hydrogen-bond donors have been selected in agreement with the original ‘RO3’. Only a few chemical scaffolds had a larger number of hydrogen-bond donors, so that this property finally varies from 0 to 4 with an average of 1.3. For the total polar surface area (TPSA) we increased the threshold from 60 Å^2 to 80 Å^2. Again, some chemotypes were allowed to deviate from this upper threshold. In the final library the TPSA ranges between 15 Å^2 and 126 Å^2 with a mean of 52 Å^2.

The filtered fragments were clustered, manually prioritized, and selected by visual inspection to avoid strong accumulation of similar chemotypes and to cover a sufficient range of differing chemical scaffolds. Every entry was requested to contain at least one ring system. The over-whelming majority shows one or two ring systems (frequently fused) and about 20 % contain three ring systems.

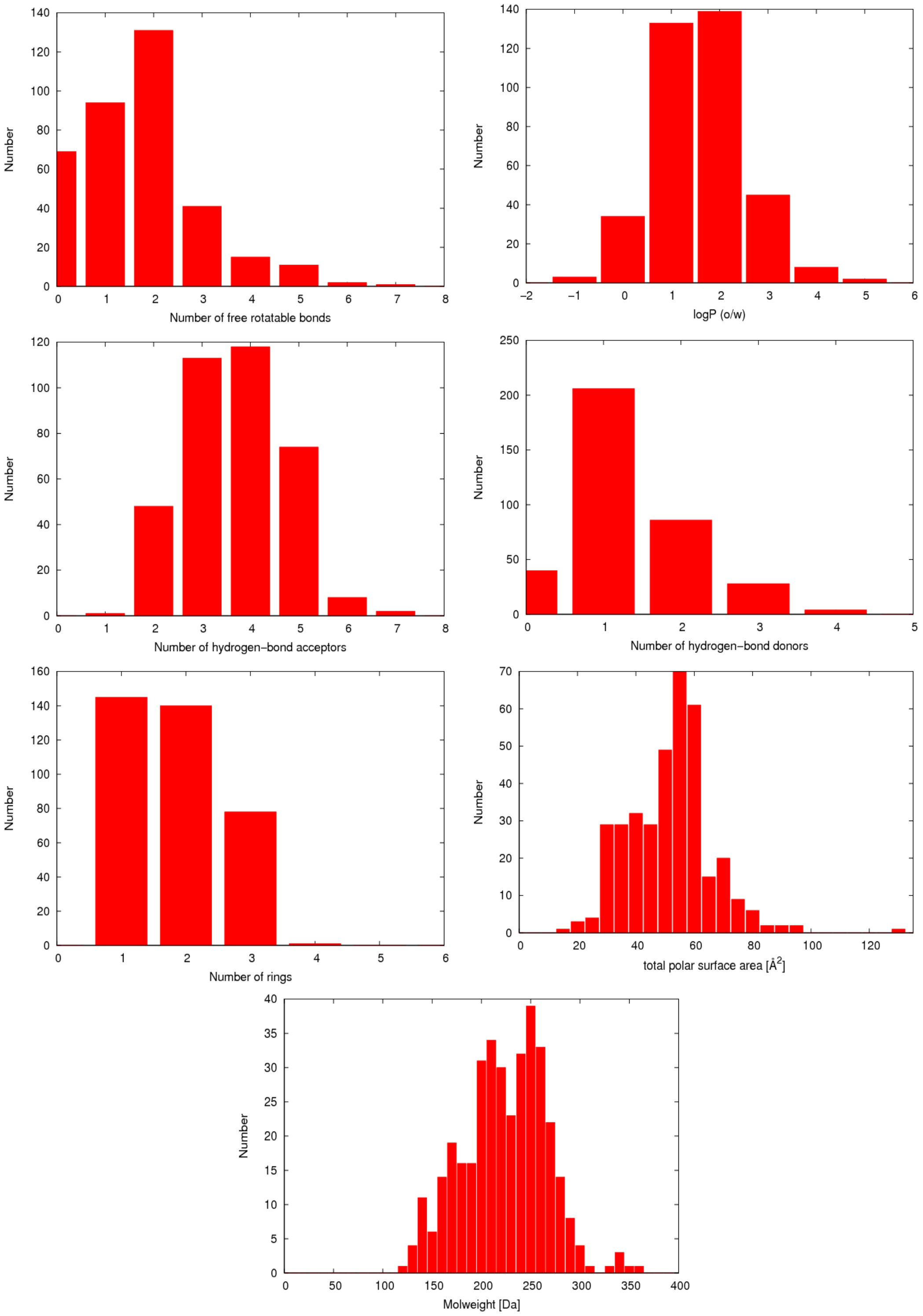


Figure 3.1: The distribution of physico-chemical and structural properties in the library.

Table 3.1: Physico-chemical parameters of the library.

	min	max	average
Number of heavy atoms	8	20	15
Molweight	122	359	224
Lipinski donor	0	4	1.3
Lipinski acceptor	1	7	3,7
clogP(o/w)	-1.3	5.4	1.6
Free rotabtable bonds	0	7	1.7
TPSA	15	126	52

Considering in detail the finally selected 364 compounds of our library, 141 fragments conform to the 'RO3' whereas 223 fragments do not agree with the respective criteria. Finally, experimental solubility was determined to be > 1 mM for 76 % of the compounds. Considering upper and lower threshold limits of the 'RO3', we exceed the ranges of all considered rules, but with respect to the average values our selection falls well into the limitation of the rules. Only the average number of acceptors is clearly beyond the range set by the original 'RO3'.

Validating the library on Endothiapepsin

To validate the suitability of our library for fragment-screening purposes, all entries were tested against endothiapepsin in a fluorescence-based competition assay. All compounds were screened at 1 mmol/L concentration. Due to insufficient solubility under the applied assay conditions 14 fragments were excluded and another 18 compounds could only be screened at 500 µmol/L. Finally further 17 entries were excluded from the assay screen, due to self-fluorescence.

For the screen, the hit criterion was set to at least 40 % inhibition of protein function, leading to 55 entries which entered into a subsequent crystallographic follow up screen. The fragments were soaked into native endothiapepsin crystals in mixtures of two compounds. In case one compound was identified in a crystal structure, the other one from the mixture was subjected to individual soaking. Simultaneous binding of two fragments was not observed.

In total, eleven complex crystal structures were obtained. The hit rate of the fluorescence-based assay with 55 hits out of a 364 membered fragment library is remarkably high. The observed eleven complex crystal structures out of the initially detected 55 hits suggests that either the composition of our assembled fragment library is efficiently done and the complete initial screening based on the fluorescence-based assay was a good filter. Table 3.2 shows the assay results. The inhibition of endothiapepsin ranges from 42 % to 100 %. Forty percent inhibition was set as threshold to accept a test candidate as ‘hit’. The eleven fragments of which we could determine a crystal structure are highlighted. The percent inhibition does not correlate with the probability to penetrate into the crystals and reveal successfully a crystal structure. Interestingly enough we were not able to successfully determine crystal structures of the eight fragments suggesting full inhibition in the fluorescence-based assay, however, we were able to crystallize fragments with the enzyme that displayed less than 50 % inhibition.

Table 3.2: Summary of the screening results. Hits leading to a crystal structure are highlighted. Values marked with an asterisk are measured at a 500 µmol/L inhibitor concentration. All other were measured at 1 mmol/L.

ID	% Inhibition	ID	% Inhibition	ID	% Inhibition
149	100	297	78	201	51
177	100 *	186	76	065	50
178	100	333	76	**291**	**50**
236	100	175	75	**041**	**49**
238	100	301	75	088	47
042	99	040	73	137	47
064	99	**063**	**66**	188	47
017	97	252	64	261	47
306	**93**	224	63	051	46
003	92	159	60	141	46
168	92	183	60	140	45
005	**89**	335	60	**216**	**45**
083	89	134	57	295	44
109	**89**	176	55	362	43
284	**87**	192	55 *	171	42
255	**84**	266	54 *		
290	**84**	267	54		
142	80	308	53		
161	80	222	52		
093	79	**148**	**51**		

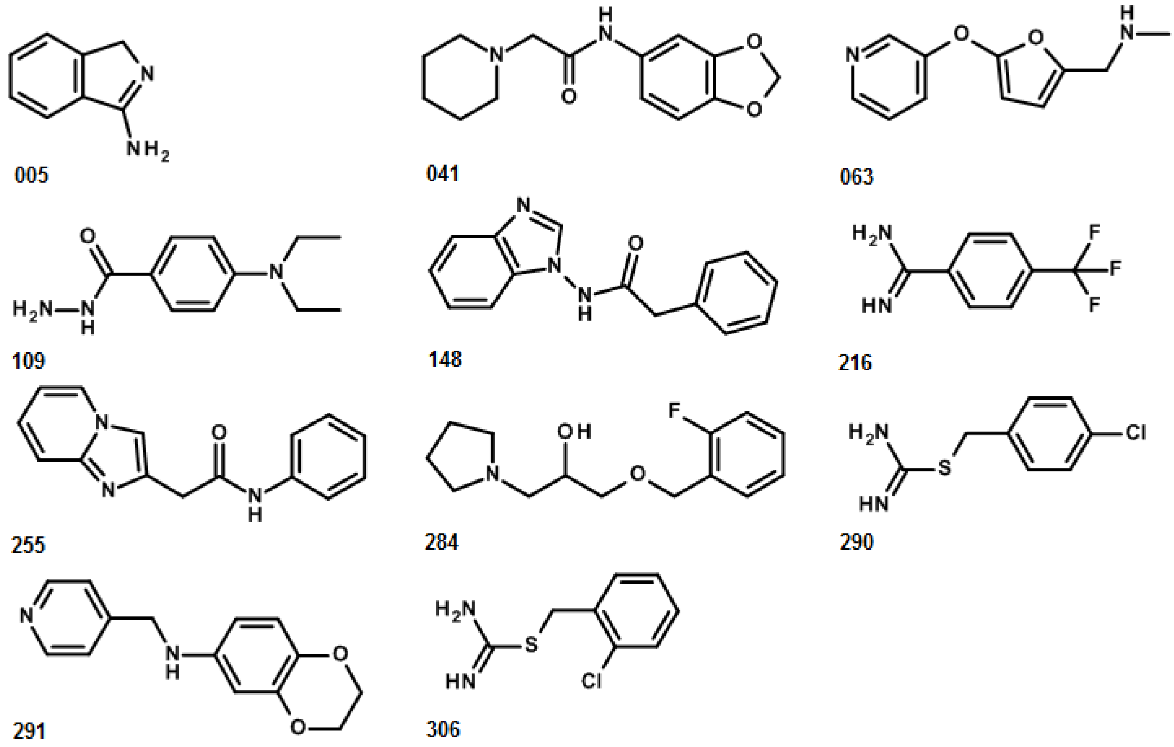


Figure 3.2: Chemical formulae of the eleven hits that successfully provided complexes in crystal structure analysis.

To find meaningful parameters that correlate with successful crystal soaking and structure determination we closely examined the crystallographically characterized fragments. Table 3.3 lists the physico-chemical parameters of these fragments. The molecular weight ranges from 168 g/mol to 262 g/mol which is in agreement with the Astex rules. All eleven fragments have between 1 and 3 hydrogen bond donors. Remarkably the hydrogen-bond acceptors show a different picture. Six out of the eleven hits possess four or more acceptors. Only five compounds exhibit between 1 and 3 acceptors, the number principally allowed by the 'RO3'. Therefore, with respect to this criterion the majority of our discovered fragments do not match the Astex rules. The clogP is in most cases below 3 and agrees with the Astex 'RO3'. Only one fragment displays a clogP value of 3.3 and departs somewhat from these limits. The number of freely rotatable bonds range from 0 up to 6. The majority has less than three freely rotatable bonds, only two fragments differ and show 4 and 6 rotatable bonds. Thus, the last two fragments infringe the Astex rules. The total polar surface area is in all cases below 60 and thus fully agrees with the limitation set by the rules.

In summary, seven of the eleven fragments which lead to a crystal structure violate at least one parameter of the Astex ‘RO3’.

Table 3.3 Physico chemical parameter of the fragments with a crystal structure

Id	Heavy atoms	Weight[57] [g/mol]	Lip don[57]	Lip acc[57]	clogP[57]	Free rotatable bond	TPSA[57] [Å^2]
005	10	169	2	2	2.2	0	38.4
041	19	262	1	5	1.6	3	50.8
063	15	204	1	4	1.0	4	47.3
109	15	207	3	4	1.0	2	58.4
148	19	251	1	4	2.2	2	46.9
216	13	261	3	2	3.3	1	49.9
255	19	251	1	4	1.8	3	46.9
284	18	253	1	3	2.0	6	32.7
290	12	201	3	2	2.9	2	49.9
291	18	242	1	4	1.7	2	43.4
306	12	201	3	2	2.9	2	49.9

Solubility has been discussed to be a crucial property.[30] To investigate the impact of solubility, we measured the aqueous buffer solubility of our fragments at pH 7.4. During library design we applied computer tools[58, 59] and property calculations to select fragments with expected high solubility. The most important parameter to predict this property was clogP which varied from -1.25 to 5.39 with an average of 1.58 (Table 3.1). The measured solubilities are presented in Table 3.4. A total of 36 fragments could not be evaluated due to self-fluorescence. 76 % of the fragments show solubility greater than 1 mmol/L indicating a quite successful computational selection procedure. To further characterize the discovered crystallographic hits, we compared the solubility of these compounds with the assay hits and with all entries of the library. As can be seen in Table 3.4 the solubility of the assay hits varies over the entire solubility range with a slight tendency towards better solubility. All fragments which also penetrate into the crystals show a solubility of greater than 1.5 mmol/L. This observation suggests that better soluble fragments are more likely to show up as an assay hit. With regard to the probability to penetrate into the crystals apart from reasonable inhibition properties pronounced solubility seems to be an even more important aspect.

Table 3.4: Solubility data for the entire library: first column, for the 55 initial hits second column and for the eleven crystal structure fragments. ND = not detected due to self-fluorescence.

	all (364)	**Hits (55)**	**Structure (11)**
N.D.	10 %	2 %	0 %
0	1 %	0 %	0 %
0 - 0.5 mM	6 %	2 %	0 %
0.5 - 1 mM	7 %	2 %	0 %
1 - 1.5 mM	10 %	14 %	0 %
1.5 - 2 mM	38 %	53 %	64 %
> 2 mM	28 %	27 %	36 %

Analysis of the Crystal Structures

Endothiapepsin is secreted by the *fungus Cryphonectriaparasitica*. Like most aspartic proteinases the protein is activated in acid media. After cleavage of the N-terminal propeptide, the active protein consists of 330 amino acids with a molecular weight of 34 kDa. The catalytic site consists of two aspartates. While Asp35 is deprotonated Asp219 is supposed to carry one proton.[53, 60] This protonation state is stabilized by the surrounding amino acids.[61] In the uncomplexed state a water molecule is bound between the two aspartates. During the first step of proteolysis this water molecule is able to attack the scissile bond of a peptide leading to its cleavage. The active site of the protein is covered by the flap, a highly flexible ß-hairpin loop. The catalytic dyad as well as the most important pockets are shown in Figure 3.3.

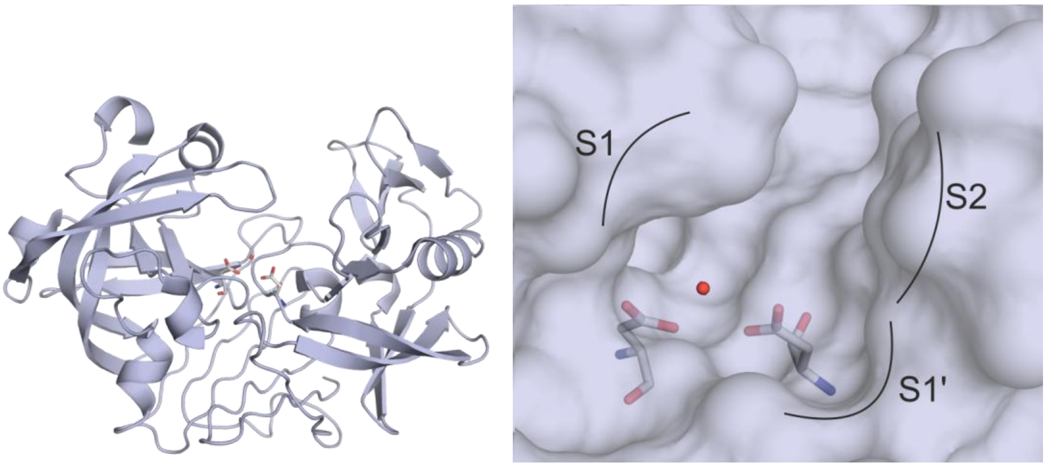


Figure 3.3: Left: overall folding structure of endothiapepsin with the catalytic dyad in the center. Right: Blow-up of the uncomplexed binding pocket with the catalytic water and the two catalytic aspartates. The adjacent specificity pockets S1’, S1 and S2 are indicated on the solvent accessible surface.

Eleven crystal structures of a fragment in complex with endothiapepsin were obtained with resolutions ranging from 1.25 to 1.90 Å. The soaked fragments disclose their binding mode in all complexes by a well-resolved difference electron density and suggest population of at least 90 %. The compounds show diverse binding modes which can be divided into three categories: Direct binding to the two catalytic aspartates, binding to the two catalytic aspartates mediated by a water molecule and no direct interaction with the catalytic dyad.

Direct binding to the two catalytic aspartates could be observed in seven cases making this the most prominent interaction motif (Figure 3.4). Fragment 109 addresses the two catalytic aspartates through its terminal nitrogen of the hydrazinocarbonyl moiety. The terminal nitrogen interacts with both aspartates and with a water molecule bridging an interaction to Gly37 in the front part of the S1’ pocket. The major part of the ligand skeleton is oriented towards the S1 pocket. The fragment is populated to 90 % and shows some disorder concerning the two terminal ethyl substituents (Figure 3.4). Additionally, a DMSO molecule is bound in the front part of the S4 pocket interacting with Thr223.

Fragment 005 binds to the catalytic dyad with its exocyclic amino group placed between the two aspartates. The endocyclic nitrogen atom of the five-membered ring forms a hydrogen bond to Thr222 of the S2 pocket. The fused aromatic benzene ring is located in the S1 pocket performing a π-stacking with Tyr79 (Figure 3.4).

Although fragments 148 and 255 are structurally similar their binding modes show interesting differences. Fragment 148 interacts only with one of the catalytic aspartates (Asp35). In addition, a hydrogen bond to Thr222 and to Gly221 is formed. Furthermore, this fragment forms via the carbonyl oxygen of the amide group an H-bond to the NH group of Gly80. Its benzimidazole moiety occupies the S1 pocket while the phenyl ring at the opposite end is orientated towards the S2 pocket. 255 binds above the catalytic dyad forming hydrogen bonds to both aspartates. Another hydrogen bond is formed via its carbonyl amide group towards the flap addressing Gly80. The larger fused 5-6-ring system occupies the S1 pocket similar to the binding mode of 148. While 148 is addressing the S2 pocket, in this case the phenyl ring is pointing towards the S1’ pocket (Figure 3.4).

Fragments 216, 306 and 290 are forming three most likely salt-type hydrogen bonds towards the catalytic dyad by placing an amidine group next to the two aspartates (Figure 3.4). All prefer a double paired hydrogen bond with Asp35 whereas Asp219 is only addressed via a single H-bond contact. It is difficult to speculate on the most likely protonation state of the dyad. Usually Asp35 is assumed to be fully deprotonated whereas Asp219 should carry one proton.[62] Regarding the observed binding mode we suggest that this protonation is given here as well.

Interestingly in β-secretase a complex with a benzamidine derivative has been reported, however with reverse orientation (PDB code: 3KMX).[50] Nevertheless, the distances across the catalytic dyad in β-secretase are somewhat larger than in endothiapepsin which might - with some care - explain the deviating binding properties observed in the former protease. The aromatic portions of the inhibitor are positioned in front of the S1’ pocket. 306 and 290 vary only in the position of the attached chlorine. While the *para*-attached chlorine in 290 is pointing into the surrounding solvent environment, the *ortho*-substitution in 306 enables the chlorine to interact with Phe194. Here, the chlorine atom is pointing with about 4 Å distance towards the ring plane of Phe194 ring. This geometry departs somewhat from the ideal interaction pattern described for chlorine-aromatic ring system contacts.[63] In the structure of 216, an additional DMSO molecule is found exactly at the same position as in the complex with 109.

Binding mediated via a water molecule can be observed for 063 and 291 (Figure 3.5). In both cases, the catalytic water molecule bridges an interaction between a pyridine-type nitrogen and the catalytic dyad.

Besides the interaction to the catalytic aspartates, 63 forms two hydrogen bonds with its secondary amino group at the opposite end of the fragment. The functional group is positioned between the flap aspartate (Asp81) and Ser115. The central part of the fragment binds to the S1 pocket.

Fragment 291 also addresses Asp81 with a nitrogen. The pyridine ring is able to perform a π-stacking with Tyr79. The benzdioxane moiety reaches into the S3 pocket.

Fragments 41 and 284 show no direct or possibly a very weak and extended interaction with the two aspartates of the catalytic dyad (Figure 3.6). Fragment 41 interacts with the flap residues by forming two hydrogen bonds, one through the piperidine nitrogen to the terminal carboxylate group of Asp81 and one via its amide carbonyl to the backbone NH group of the same residue. The adjacent amide nitrogen forms a hydrogen bond towards Thr222 in front of the S2 pocket. The aromatic piperonyl ring system is pointing towards the S1’ pocket, whereas the hydrophobic portion of the piperidine ring occupies the S1 pocket.

Fragment 284 forms hydrogen bonds via its hydroxyl function to the backbone nitrogen of Gly80 of the flap and to Gly221 at the bottom of the binding pocket. As this pyrrolidine nitrogen is most likely protonated under the acidic buffer conditions applied for soaking a second rather long H-bond contact (3.5 Å) is formed to the water molecule hosted at the pivot between both aspartates of the catalytic dyad. The hydrophobic portion of the pyrrolidine moiety occupies the S1 pocket while the opposing aromatic ring is oriented towards the S2 pocket.

The difference electron densities of 41 and 284 are less well defined compared to the other nine fragments, indicating a rather large residual mobility of the fragments in the binding pocket. Possibly this observation is due to the fact that these fragments are not in direct contact with the strongly fixed, most likely negatively charged catalytic dyad.

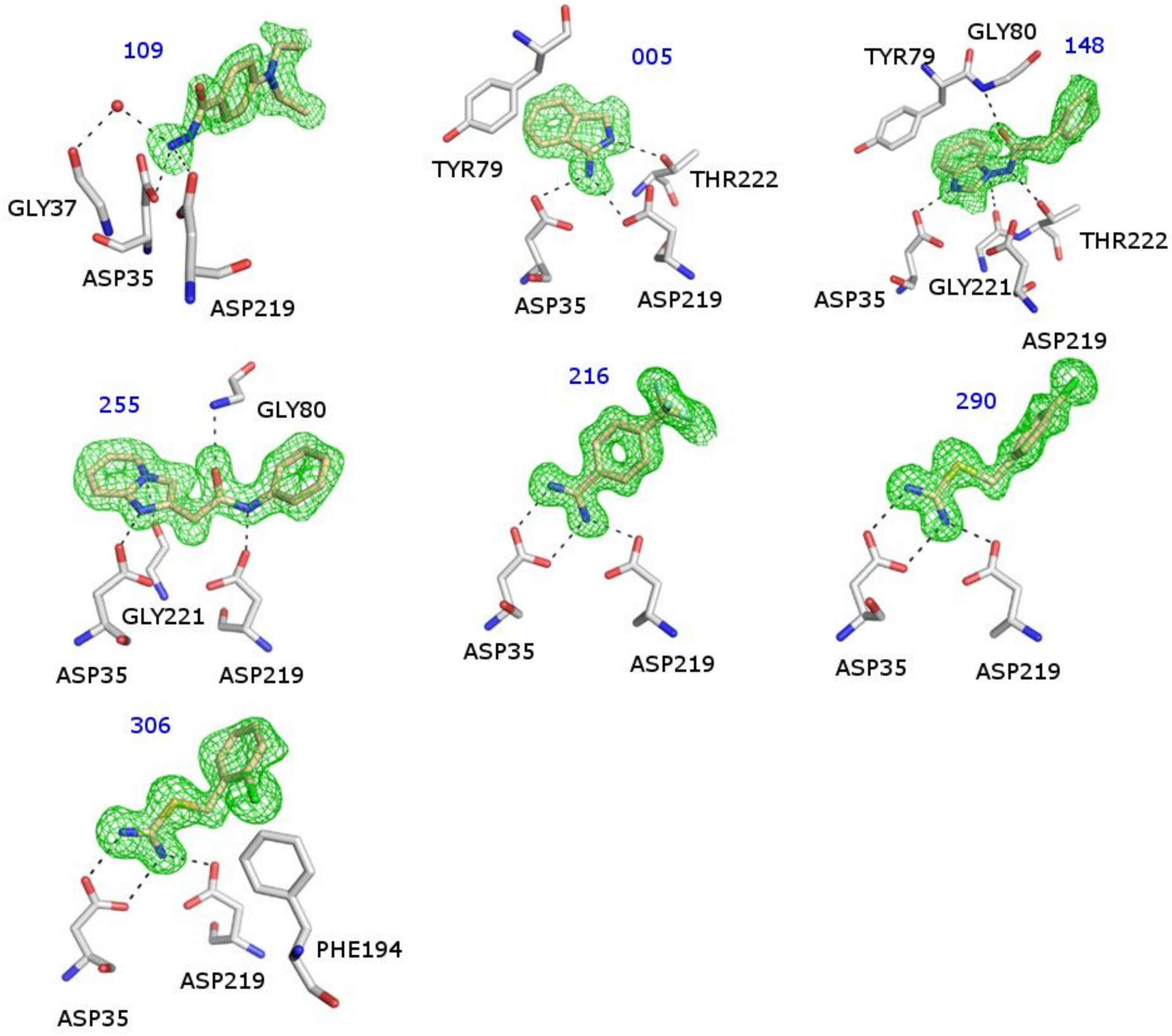

109
GLY37
ASP35
ASP219
005
TYR79
THR222
ASP35
ASP219
148
GLY80
TYR79
THR222
ASP35
GLY221
ASP219
255
GLY80
GLY221
ASP35
ASP219
216
ASP35
ASP219
290
ASP35
ASP219
306
PHE194
ASP219
ASP35

Figure 3.4: The seven crystal structures which directly interact with the two catalytic aspartates. Nitrogen atoms are colored blue, oxygen atoms red, ligand carbon atoms in salmon and protein carbon atoms in white. The green mesh shows the F_o-F_c difference electron density at a σ-level of 2.0.

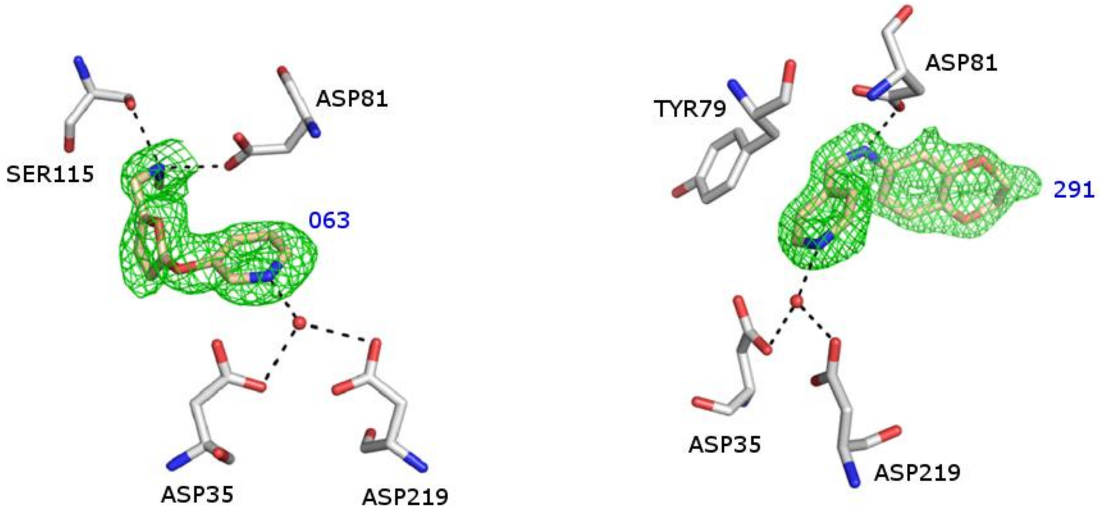

SER115
ASP81
063
ASP35
ASP219
TYR79
ASP81
291
ASP35
ASP219

Figure 3.5: The two fragments interacting with the catalytic aspartates by an interstitial water molecule. Nitrogen atoms are colored blue, oxygen atoms red, ligand carbon atoms in salmon and protein carbon atoms in white. F_o-F_c different electron density at σ-level of 2.0.

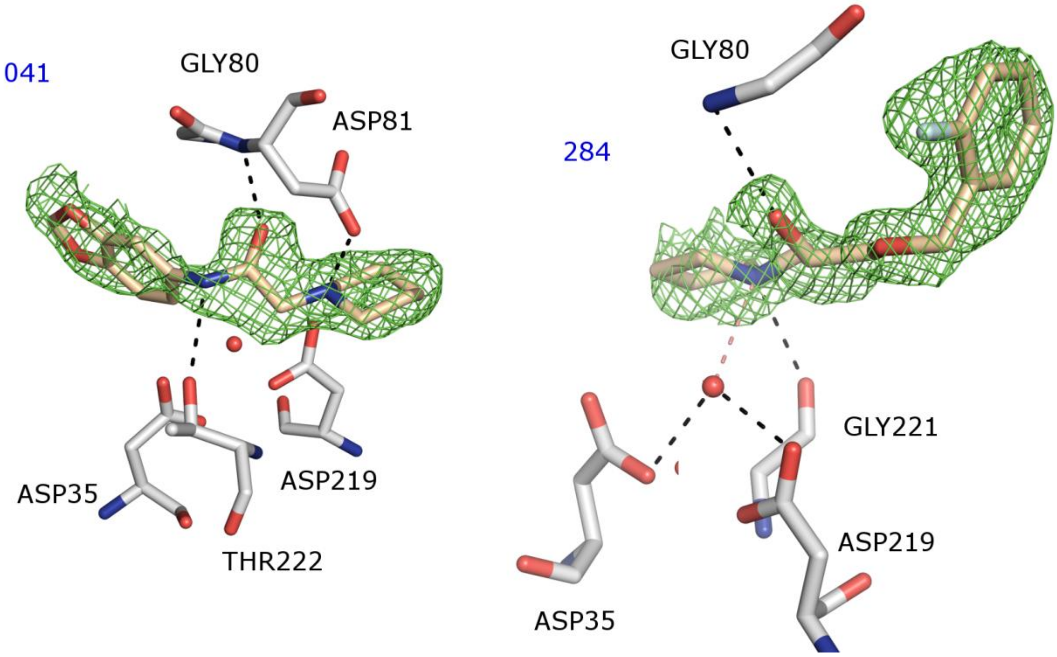

041
GLY80
ASP81
ASP35
THR222
ASP219
GLY80
284
GLY221
ASP219
ASP35

Figure 3.6 Fragments showing no direct interaction with the two aspartates. Nitrogen atoms are colored blue, oxygen atoms red, ligand carbon atoms in salmon and protein carbon atoms in white. F_o-F_c different electron density at σ-level of 2.0.

An overlay of all crystal structures is shown in Figure 3.7. Almost the entire volume of the various sub-pockets of this protease are occupied by the different fragments. Besides binding to the catalytic dyad several preferred interactions can be observed. The following interactions highlight some of these contacts. Seven compounds (005, 041, 148, 255, 284, 291 and 109) orient a hydrophobic ligand portion towards the S1 pocket. A ligand carbonyl group addressing the backbone nitrogen of the flap aspartate is observed in three cases (041, 148, 284). Thr222 at the bottom of the binding pocket is addressed by an NH functionality of 005 and 148. Remarkably, all these interactions are also observed for pepstatin, a well-known highly potent inhibitor of most aspartic proteases not designed by a medicinal chemistry program but optimized by microorganisms using principles of evolution.

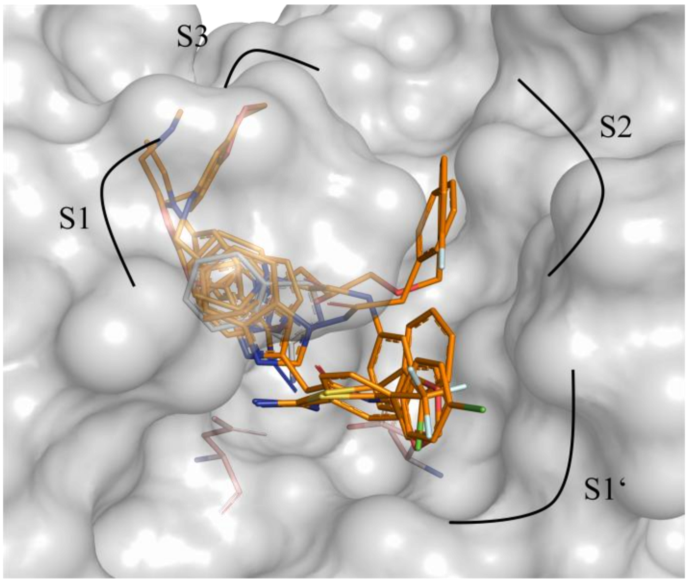


Figure 3.7: Overlay of all eleven fragment structures. Binding pocket is in surface representation, specificity pockets are indicated. Carbon atoms are colored in salmon, nitrogen in blue, oxygen in red, chlorine in green and fluorine in cyan.

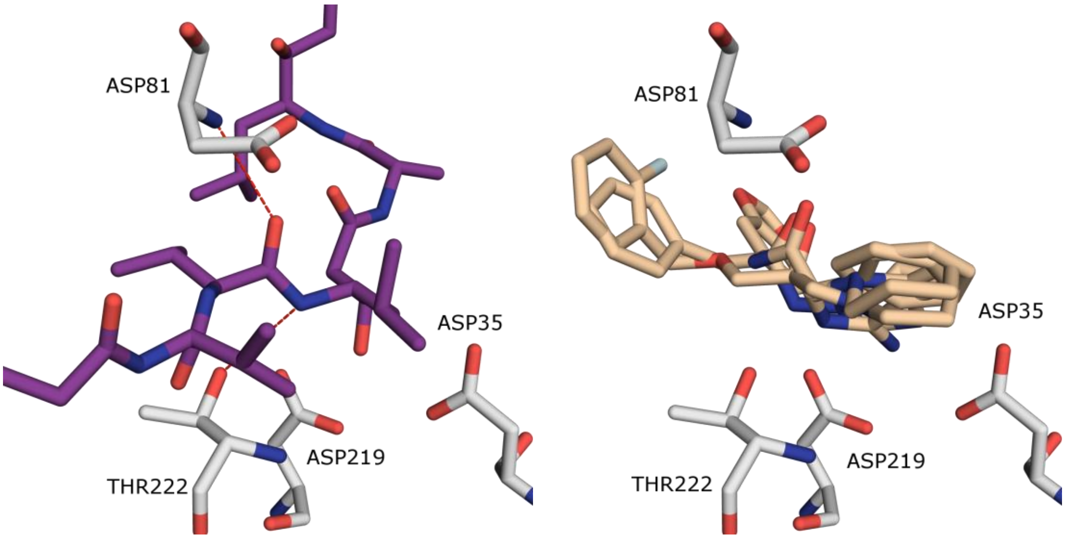


Figure 3.8: Left: Pepstatin bound to EP (PDB code 4ER2). Right: The flap aspartate and Thr222 are shown in white. Two interactions performed by pepstatin are marked with dashed lines.

3.5 Discussion

The fluorescence-based assay screening results and the obtained crystal structures with endothiapepsin suggest that the library has been convincingly designed for crystallographic fragment screening; bearing in mind that this library was assembled for all kinds of targets and not specifically for aspartyl proteases. Nevertheless, this conclusion is preliminary and based on the result with one single target, for which we obtained eleven complex crystal structures out of the 364-membered library. In this context it is worth mentioning that confidential results obtained by one industry partner in this project suggested also remarkable hit rates for an additional target.

The solubility of the fragments in the assembled library is very promising and underlines the importance of this property, as 66 % of the compounds showed at least solubility of 1.5 mmol/L and 76 % are at least 1 mmol/L soluble. This criterion is critical for success, because fragments are weak binders and therefore have to be applied in rather high concentrations for crystallization and assay experiments.

As indicated earlier, the library design was not consistent with the ‘RO3’ of Astex. The assembled compounds are in agreement with these rules with respect to molecular weight, logP and the TPSA. However, we cover a larger range concerning the number of hydrogen-bond donors and in particular acceptors and in the number of free rotatable bonds. If we had designed a library strictly within the limitations of the ‘RO3’, we would have missed seven fragments of our detected hits for endothiapepsin. Particularly remarkable is the fact that all four remaining RO3-conform crystallographically successful fragments have amidine-like moieties. However, with respect to drug design and medicinal chemistry follow-up programs they would provide a rather narrow range of chemotypes. Clearly a research study applying fragment-based lead discovery expects a much broader range of chemical diversity. Therefore, the strict threshold of less than three donors and acceptors and the number of rotatable bonds is questionable.

With respect to fragment growing and fragment merging, a sufficient number of synthetically accessible functional groups is of utmost importance for subsequent chemical synthesis. Widening up such criteria will enlarge the available fragment pool and might lead to higher hit rates of binding fragments available for further ligand design. As a matter of fact, usually the functional groups are motifs that show hydrogen-bond donor or hydrogen-bond acceptor properties.

3.6 Conclusions

The suitability of a fragment library for successful crystallization and follow-up optimization is not only determined by the library design. The correct choice of the experimental screening conditions is equally essential to discover reasonable hits. The first step is usually a screening assay followed by a crystallographic hit validation. In our case we have chosen a functional assay at 1 mmol/L compound concentration and demanded at least 40 % protein inhibition. On the one hand, this criterion was quite successful, as we were able to determine eleven complex crystal structures out of 55 assay hits. On the other hand it is remarkable that according to our crystallographic screen results no obvious correlation between potency observed in the assay and probability to obtain a crystal structure can be established. It might be even possible that there are still compounds with an even lower inhibition rate which would penetrate into the crystals and indicate their successful binding by the determination of their binding poses in the crystallographic screens. Ranking fragments by assay results seems indicative but by no means reliable. In literature, several protocols have been reported applying alternative screening protocols based on techniques such as SPR, NMR, biochemical functional assay, mass spectrometry or replacement titration experiments to name a few. Interestingly, hit lists were generated showing sometimes minimal overlap. The physical basis for these observations is still rather unclear. However, structure-based fragment lead discovery needs a well resolved structure as an entry point into a medicinal chemistry follow-up program. Usually this is a crystal structure. Thus final consequence of these observations might be to perform fragment screening primarily on protein crystals, a perspective presently not followed as the X-ray facilities would have to be expanded and adapted to this strategy. Most setups are not yet suitable for this concept and would require further automation and more frequent access to synchrotron beam time. Up to now, we also refrained from such attempts as it appears only feasible with the required automation and approved access to synchrotron beam time to include all library fragments into the screening process.

Nevertheless, in this project eleven crystal structures with a fragment bound to endothiapepsin could be determined. These fragments show diverse binding modes filling up almost the entire volume of the various specificity pockets. They can be divided into three different categories which include direct contacts to the catalytic aspartates, binding mediated by a water molecule, and no direct binding to the catalytic dyad. The various binding modes provide novel ideas to address the active site of aspartic proteases. The fragment structures are also suitable to elucidate experimentally the hot spot of binding. An alignment of all structures enables us to “map out” possible interaction patterns and binding motifs with the protease. Remarkably, pepstatin, a potent non-selective aspartyl protease inhibitor uses all the prominent interactions highlighted by our fragments to achieve potent binding (Figure 3.8).

3.7 Materials and Methods

Inhibition Assay

Endothiapepsin was purified from Suparen® (provided by DSM food specialties) by exchanging the buffer to 0.1 M acetate buffer pH 4.6 using a Vivaspin 20 with a molecular weight cut off at 10,000 Da. The protein concentration was measured by absorbance at 280 nM assuming an extinction coefficient of 1.15 for 1 mg/mL solutions.[64]

A 100 mM stock solution in DMSO was prepared for all compounds of the fragment library. Due to solubility reasons, in 18 cases only 50 mM could be achieved. Fourteen compounds were not soluble in DMSO and therefore excluded from the screen. As substrate we used Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH_2 (purchased from Bachem). The assay was performed with a Tecan Safire^2 microplate reader at an excitation wavelength of 337 nM and an emission wavelength of 414 nM. The K_m of the substrate towards endothiapepsin was determined to be 1.6 µM. The assay buffer (0.1 M acetate buffer pH 4.6 containing 0.01 % Tween 20) was premixed with the substrate and the screening compound whereas the protein was added directly before measurement. The final reaction volume was 200 µM containing 4 nM endothiapepsin, 1.8 µM substrate and 1 mM test compound (or 500 µM in the cases where a 50 mM stock solution was used). Blanks were prepared in the same way using DMSO instead of the compound stock solution. During measurement the fluorescence increased due to substrate cleavage. For data analysis the initial slope of the fluorescence in the compound containing wells were compared to the initial slope of the blanks. Each compound was measured twice. The final result represents the average of both measurements.

Crystallization and Structure Determination

Crystals of unbound endothiapepsin were grown similarly as described previously.[55] We used the sitting drop vapor diffusion method and a crystallization temperature of 16° C. The drops contained 2 µL of protein solution (5 mg/mL) and 2 µL of mother liquor. The reservoir solution consisted of 1 mL 0.1 M NH_4Ac, 0.1 M acetate buffer pH 4.6 and 26 % PEG 4000. Crystals were ready for soaking after about two weeks. For each compound, a 1 M stock solution in DMSO was prepared. The crystals were soaked 1-2 days in reservoir buffer containing 25 % glycerol and a mixture of two compounds, each at a final concentration of 50 mM. To ensure a clear identification of the bound fragment the two compounds were chosen by maximal chemical shape diversity. In case the compound precipitated the precipitate was ignored. After soaking, the crystals were flash-frozen in liquid nitrogen. In case one compound was binding, the other one was subsequently soaked again in a different mixture. Fragment 109 was soaked alone for practical reasons. In two cases (255, 148) the bound compound was soaked again because the difference density was not sufficient for placement of the fragment. These compounds were soaked without an additional fragment. In both cases the difference electron density improved significantly.

In-house datasets were collected on a sealed Cu fine focus X-ray device using a MAR345 image plate detector. Synchrotron datasets were collected at BESSY beamline 14.2 or 14.3 in Berlin. Bessy beamline 14.2 provides a Rayonix MX-255 CCD detector, beamline 14.3 a Rayonix SX-165mm CCD detector. All datasets were collected at 100 K and processed using HKL2000.[65]

The structures were determined by molecular replacement using the program Phaser.[66] Search model was the 0.90 Å structure of endothiapepsin bound to a short peptide (PDB code: 1OEW). For cross-validation of the refinement 5 % of the reflections were chosen at random for inclusion in the R_{free} set. After initial simulated annealing with CNS, refinement was done using SHELXL.[67] During the last cycles of refinement of the structures containing 109, 148, 216, and 290 anisotropic refinement was applied. The structure containing fragment 284 was refined using PHENIX.[68] After each cycle the models were inspected and subsequently improved using Coot.[69] In most structures some unexplained difference density remained after refinement. This additional density which also occurred in the binding pocket is probably due to the high amount of soaked compound leading to unspecific binding. It is also possible that this indicates some DMSO molecules showing limited occupancy as only in two structures fully occupied DMSO molecules could be assigned.

Solubility Assay

A 500 µM buffered compound solution was prepared using a 10 mM compound stock solution in DMSO. This solution was transferred into three wells (100 µL /well, triplicates) of a 96 well filtration plate (Millipore; 0.2 µm, Hydrophilic PVDF, Durapore-MSGVN2250) and incubated at room temperature for 90 min while agitating at 100 rpm (sample wells). In parallel, the same solution was transferred into three wells (100 µL /well, triplicates) of a 96-well, UV transparent plate (Greiner, #655801) prefilled with 100 µL/well acetonitrile thereby assuring complete dissolution of the compounds (control wells). In order to separate precipitate from dissolved compound the filtration plate was centrifuged (2000 rpm, 5 min, RT). The filtrates were collected directly in the wells of the UV transparent plate. After adding 100 µL/well acetonitrile to the filtrate wells, absorbance spectra were taken (250-500 nM). Spectra of the sample and control wells are integrated and averaged over the triplicates. Kinetic solubility was calculated by 500 µM x (integral sample wells / integral control / wells).

Crystallographic Table 1:

Bound fragment	**109**	**005**	**148**	**255**	**216**	**290**
PDB code	3PBZ	3PBD	3PMY	3PM4	3PCW	3PLD
Data collection and processing						
Collection site	BL 14.2	In-house	BL 14.2	In-house	BL 14.2	BL 14.3
λ [Å]	0.91841	1.54178	0.91841	1.54178	0.91841	0.89440
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters						
a [Å]	45.1	45.3	45.4	45.2	45.3	45.3
b [Å]	73.3	73.1	73.1	73.5	73.1	73.0
c [Å]	52.4	52.7	52.8	52.5	52.5	52.8
β [°]	109.2	109.8	109.6	109.2	109.5	109.6
Resolution [Å]	30 - 1.48	30 – 1.70	30 – 1.38	40 - 1.68	20 – 1.25	40 – 1.40
Highest resolution shell	1.51-1.48	1.73–1.70	1.40-1.38	1.71-1.68	1.27-1.25	1.42-1.40
Unique reflections	51357	35536	65186	37129	86999	63118
R_{sym}[%][a]	5.8 (25.5)	4.6 (20.8)	4.6 (21.9)	4.1 (21.2)	3.6 (18.9)	5.9 (33.0)
Completeness [%] [a]	95.5(86.0)	99.5(99.3)	97.7(87.5)	100(100)	98.0(84.5)	99.3(97.6)
Redundancy[a]	2.7 (2.0)	2.8 (2.6)	3.0 (2.0)	3.9 (3.6)	3.6 (2.1)	3.1 (2.6)
I/σ[a]	15.9 (2.7)	29.4 (4.7)	22.1 (3.8)	31.4 (6.0)	27.9 (3.9)	18.5 (2.4)
Refinement						
Final R_{free}	17.9	19.6	16.2	19.3	15.0	16.6
Final R_{work}	12.8	15.3	12.2	15.3	11.6	12.9
No. of water molecules	260	323	292	303	304	300
Ramachandran plot						
Most favored regions (%)	94.2	92.8	94.6	93.9	93.5	94.6
Additional allowed regions (%)	5.8	7.2	5.4	6.1	6.5	5.4
Mean B-factors (Å^2)						
Protein atoms	14.2	16.3	11.4	16.5	12.0	13.0
Water molecules	27.7	28.7	25.4	28.4	26.3	26.7
Ligand (fragment)	28.3	28.6	31.0	20.5	21.5	23.7
Ligand (other)[b]	27.6	24.4	17.5	25.0	24.3	19.9
RMSD bond length [Å]	0.009	0.007	0.011	0.008	0.013	0.010
RMSD bond angles	2.7	2.4	2.7	2.5	2.8	2.7

Crystallographic Table 2

Bound fragment	**306**	**63**	**291**	**41**	**284**
PDB code	3PLL	3PB5	3PI0	3PGI	3PMU
Data collection and processing					
Collection site	BL 14.3	In-house	BL 14.3	In-house	BL 14.2
λ [Å]	0.89440	1.54178	0.89440	1.54178	0.91841
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters					
a [Å]	45.2	45.3	45.3	45.3	45.2
b [Å]	73.1	73.0	73.0	72.8	73.7
c [Å]	52.7	52.5	52.7	52.8	52.5
β [°]	109.6	109.5	109.4	109.6	109.3
Resolution [Å]	40 – 1.73	30 – 1.90	40 – 1.64	30 – 1.90	20 – 1.43
Highest resolution shell	1.76-1.73	1.93-1.90	1.67-1.64	1.93-1.90	1.45-1.43
Unique reflections	32328	25453	38684	25348	59547
R_{sym}[%][a]	7.2 (34.7)	7.2 (30.0)	5.3 (33.3)	6.5 (28.7)	7.3 (42.9)
Completeness [%] [a]	96.4(94.7)	100(99.6)	98.0(97.2)	99.2(96.0)	99.4(98.2)
Redundancy [a]	3.1 (2.8)	4.0 (3.6)	3.0 (2.6)	3.6 (3.2)	3.1 (2.7)
I/σ [a]	13.2 (2.6)	19.0 (4.1)	20.4 (3.0)	26.7 (4.3)	14.3 (2.3)
Refinement					
Final R_{free}	22.0	22.2	18.4	21.5	18.9
Final R_{work}	16.3	16.1	15.1	16.2	16.5
No. of water molecules	236	237	254	239	264
Ramachandran plot					
Most favored regions (%)	93.1	92.8	94.2	93.1	93.9
Additional allowed regions (%)	6.9	7.2	5.8	6.9	6.1
Mean B-factors ($Å^2$)					
Protein atoms	14.8	17.1	13.0	21.4	13.7
Water molecules	24.9	27.5	25.4	30.6	24.5
Ligand (fragment)	18.2	30.8	25.5	46.0	32.9
Ligand (other) [b]	22.3	32.2	24.8	40.4	-
RMSD bond length [Å]	0.007	0.006	0.008	0.006	0.005*
RMSD bond angles	2.4	2.2	2.5	2.2	1.1*

[a] values in parentheses are for the highest resolution shell

[b] other ligands are Glycerol and/or DMSO

* low values are due to stronger geometrical restrains used in PHENIX compared to SHELXL

4 HotSpot Analysis – a Promising Strategy for Lead Optimization; an Endothiapepsin Fragment Screen as Case Study

4.1 Introductory Remarks

The present study is based on eleven crystal structures of endothiapepsin with bound fragments. The structures were determined by Helene Köster (Univ. Marburg) and are described in the previous chapter 3 (page 27). This study was accomplished together with the master student Rajathees Rajaratnam and Gerd Neudert.

4.2 Introduction

Figure 4.1 shows the eleven fragments for which we were able to determine crystal structures in complex with endothiapepsin. Fragments are rather small molecules and therefore, a commonly agreed concept is the idea that fragments are bound due to a specific interaction with the target. The diverse chemical constitution of fragments will lead to different physico-chemical properties and thus to deviating binding profiles. In the present example, fortunately the entire pocket is occupied by the ensemble of all detected fragments. Therefore, we are able to map out the physico-chemical profile of the binding pocket of endothiapepsin by an experimental approach.

As computational alternative, programs such as Grid[70], SuperStar[71] and our in-house program HotSpotsX are available to predict the physico-chemical properties of binding pockets. Currently, HotSpotsX is developed by Gerd Neudert and considering the experimentally derived physico-chemical fingerprint of the binding site of endothiapepsin, we wanted to validate HotSpotsX.

Therefore, we asked the question whether the HotSpotsX predicted interaction maps are in agreement with the localization of corresponding atom types in the experimentally determined crystal structures.

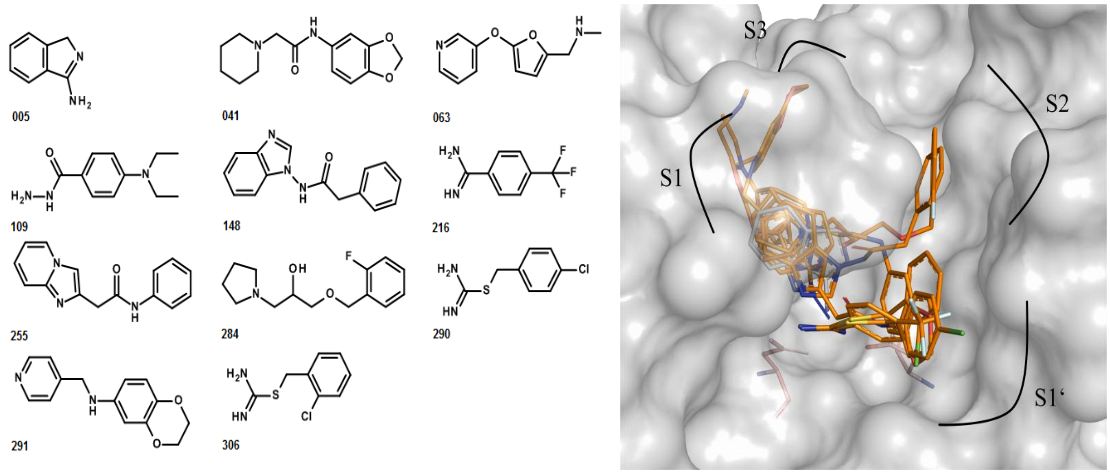


Figure 4.1: Eleven fragments are shown as chemical formula (left) and a composed picture of their positions in the binding site of endothiapepsin (right).

The eleven crystal structures show no significant backbone movements and no side chain rotation could be observed. Thus, the protein geometry as found in the complex structure with 005 has been used as a template.

4.3 Methods

In this study two programs, developed by Gerd Neudert, have been used (personal communication).

4.3.1 fconv[72]

The first program is named fconv and can be used for molecule data handling and data parsing problems. This program defines internal atom types to describe the local chemical environment, hybridization and bonding state for each atom in a molecule.

In total, 157 atom types were defined and classified into five different generic physico-chemical properties: H-bond donor, acceptor, doneptor, aromatic and hydrophobic portion. The atom types that did not match any of the properties have been defined X. Figure 4.2 shows one example for each property.

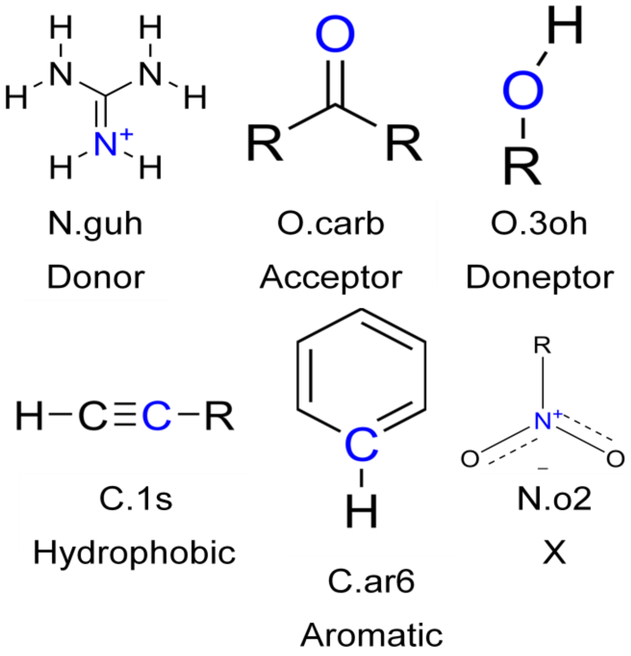

N.guh
Donor
O.carb
Acceptor
O.3oh
Doneptor
C.1s
Hydrophobic
C.ar6
Aromatic
N.o2
X

Figure 4.2: Examples for the defined physic-chemical properties

Table 4.1 shows the final classification. 29 atom types have been defined as acceptors, 15 atom types as doneptors; 9 atom types as aromatic, 11 atom types as donor and 18 atom types as hydrophobic.

Table 4.1 Internal atom types of fconv sorted by physico-chemical properties. Acceptor (Acc), doneptor (AnD), aromatic (Aro), doneptor (Don) and hydrophobic (Hyd) properties

Acc		**AnD**	**Aro**	**Don**	**Hyd**
O.carb	O.co2	O.3oh	C.ar6	N.guh	C.1s
N.ar2	O.2po	N.r3	C.ar6x	N.ar6p	C.2r3
N.1	O.2so	N.gu1	C.arp	N.arp	C.3r3
N.oh	O.2p	N.gu2	C.arx	N.ar3h	C.1p
N.aas3	O.2s	N.mi1	C.ar	N.ohac	C.2p
N.aat3	O.3po	N.mi2	N.ar6	N.ims	C.2s
N.2n	O.3so	N.aap	N.ar3	N.amp	C.2t
N.2s	O.o	N.2p	O.ar	N.ams	C.et
N.3t	O.3es	N.3n	S.ar	N.samp	C.ohp
O.r3	O.3eta	N.3p		N.sams	C.ohs
O.n	O.3et	N.3s		N.mih	C.oht
O.2co2	S.r3	O.h2o		N.4H	C.3p
O.2es	S.thi	O.noh			C.3s
O.2hal	S.2	O.3ac			C.3t
O.am		O.ph			C.3q
					S.sh
					S.s
					S.3
29		**15**	**9**	**12**	**18**

4.3.2 HotSpotsX

The second program used in this context is HotSpotsX. Based on different knowledge-based potentials, this program can predict interaction fields for different atom types in the binding pocket. Areas with highly favorable interaction values are defined as hotspots. They result from specified contour maps which encompass the sum of the corresponding atom types. Negative values are favorable values. Due to the classification of the atom types, the physico-chemical properties of binding pockets can be predicted. The classification of the pyhsico-chemical properties is described in Table 4.1, this gerneralized maps are the sum of the individual atom types and therefore, the absolute values are higher. Currently two different sets of potentials are available. One is based on stored structural data in the Cambridge Structural Database (CSD)[73], the data depository of small molecule crystal structures. The second is derived from structural information in the Protein Data Bank (PDB)[74]; in this database protein crystal structures have been deposited.

4.3.3 Pymol

The PyMOL Molecular Graphics System, Version 1.1 was used to visualize the molecules and the images are rendered by this program.[75]

4.4 Results

Each of the eleven fragments was analyzed individually in terms of physico-chemical properties.

4.4.1 Analysis of acceptor functionalities

The ether functionality of 063 is a nice example for a hydrogen-bond acceptor. Figure 4.3 displays on the left hand side the contour map for acceptors. This map was produced condidering 29 different atom types to describe acceptor probe atoms (Table 4.1).

The oxygen of the ether motif coincides well with the acceptor density. On the right hand side the contour map of the atom type of O.3eta is shown, this atom type is defined for oxygens in ethers bound to an aromatic carbon. Also for this more specified atom type O.3eta the contour map reflects the actually found position of the acceptor hotspot convincingly well.

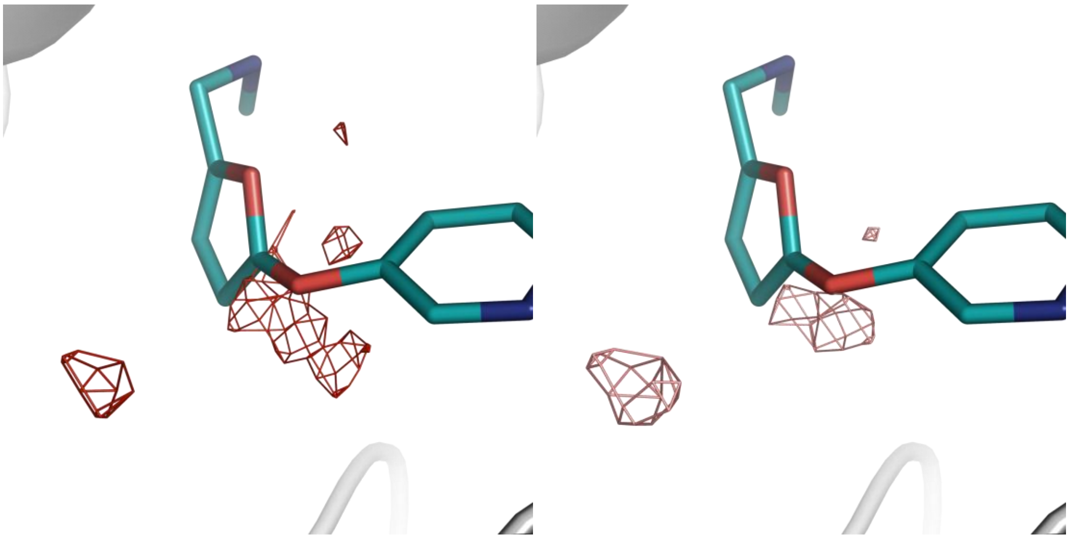


Figure 4.3: Fragment 063 in the binding site of endothiapepsin is shown. Left: The contour map of acceptors is shown at a map level of 59 % above the minimal map level (-293102/ -498904). Right: The contour map for the atom type O.3eta is shown at a map level of 56 % above the minimal map level (-15947/-28467).

4.4.2 Analysis of doneptor functionalities

The exocyclic nitrogen of 005 is a good example for a doneptor functionality as it can act either as a hydrogen-bond donor or hydrogen-bond acceptor simultaneously. Figure 4.4 displays the contour map for this type of doneptor functionality, and is therefore based on 15 different atom types representing donor probe atoms (Table 4.1). Figure 4.4 displays the fragment in the binding pocket. On the left, the contour map of the doneptor property and on the right the contour map for the specific atom type N.aap is displayed which is defined as primary aromatic amine. The exocyclic nitrogen fits perfectly well into the calculated density. In the course of the catalytic mechanism, a water molecule is placed at exactly this pivotal position between the two aspartates of the dyad. Water molecules are well known to act as doneptors.

It is remarkable that the computational approach predicts a doneptor functionality at this position and underlines that the results calculated by HotSpotsX_CSD are very useful indicators for hotspot analysis in ligand design.

The nitrogen of 005 does not fit comparably well into the contour map for N.aap. The map is less significant as the position of this atom only becomes apparent at a map level of 38 %. This fact suggests that the contour maps based on the merged atom-type properties are better suited to identify hotspots in a certain area of the binding pocket.

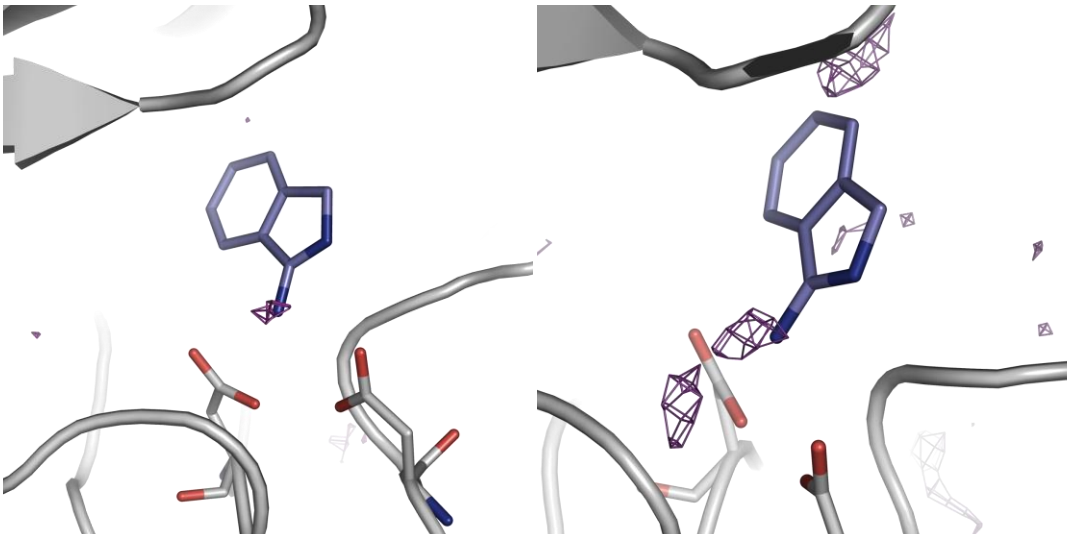


Figure 4.4: Fragment 005 in the binding site of endothiapepsin is shown. Left: The contour map of a doneptor property is shown at a map level of 79 % above the minimal map level (-189653/ -240154). Right: The contour map for the atom type N.aap is shown at a map level of 38 % above the minimal map level (-9050/ -23841).

The terminal nitrogen of the hydrazino-carbonyl functionality of fragment 109 also acts as a doneptor. In Figure 4.5 the fragment in the binding site with the corresponding contour map is shown. On the left, the map for the merged doneptor property is shown and on the right the map for the specific atom type N.3n is indicated. This atom type is defined as sp^3 nitrogen bound to another nitrogen. In both cases, the calculated maps are in agreement with the experimentally determined position.

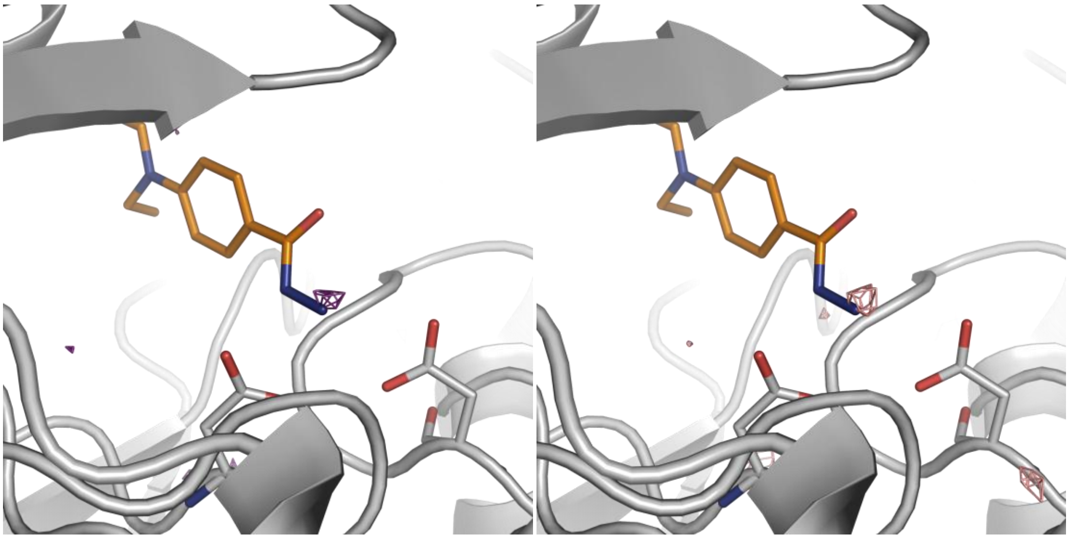


Figure 4.5: Fragment 109 in the binding site of endothiapepsin is shown. Left the contour map of a doneptor is shown at a map level of 79 % from a minimal map level (-189653/ -240154). Right the contour map for the atom type N.3n is shown at a map level of 75 % from a minimal map level (-14500/ -19349).

4.4.3 Analysis of donor functionalities

The benzamidine functionality of fragment 216 is another example for a donor functionality. Figure 4.6 displays the fragment in the binding site. On the left hand side the contour map for donor functionalities is shown. To generate this map an atom type is used which has been merged from eleven different atom types. On the right hand side the contour map produced with the specific atom type N.mih is presented. This atom type is defined by nitrogens in protonated amidino groups (also set if protonation state is unknown). The hot spot is similar to the doneptor hot spot and it is placed next to the catalytic dyad. In both contour maps the amidino group coincides well with the extremum in the calculated map.

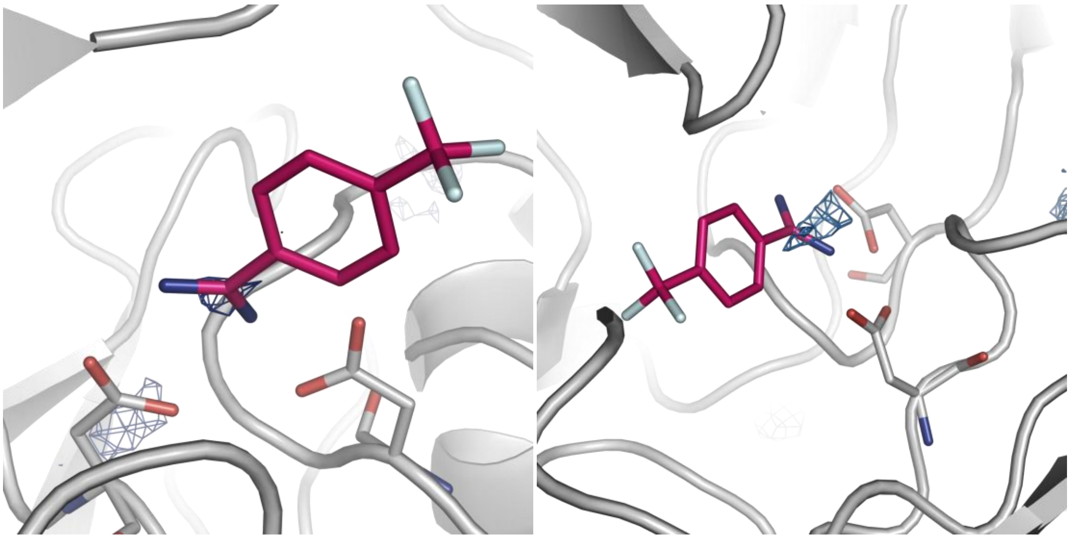


Figure 4.6: Fragment 216 is shown in the binding site of endothiapepsin. Left: The contour map of the donor is shown at a map level of 68 % above the minimal map level (-176722/ -260327). Right: The contour map for the specific atom type N.mih is shown at a map level of 65 % above the minimal map level (-17671/ -27017).

The nitrogen of the carbamido-sulfanyl functionality of fragment 290 and fragment 306 are also examples exhibiting donor functionalities. Figure 4.7 and Figure 4.8 show on the left hand side the contour map for general donor functionalities. On the right hand side the contour map for the specific atom type N.mih is presented. In both cases the nitrogen fits well into the calculated contour map. This result is not surprising, as the amidino motif of the fragments is exactly in the same orientation for all three fragments.

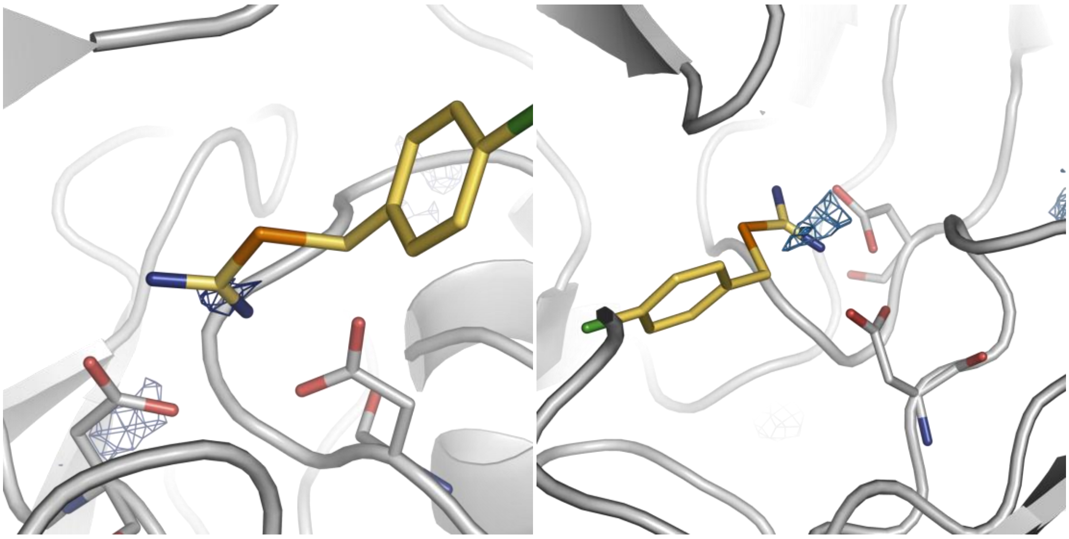


Figure 4.7: Fragment 290 is shown in the binding site of endothiapepsin. Left: The contour map of a donor is shown at a map level of 68 % above the minimal map level (-176722/ -260327). Right: The contour map for the specific atom type N.mih is shown at a map level of 65 % above the minimal map level (-17671/ -27017).

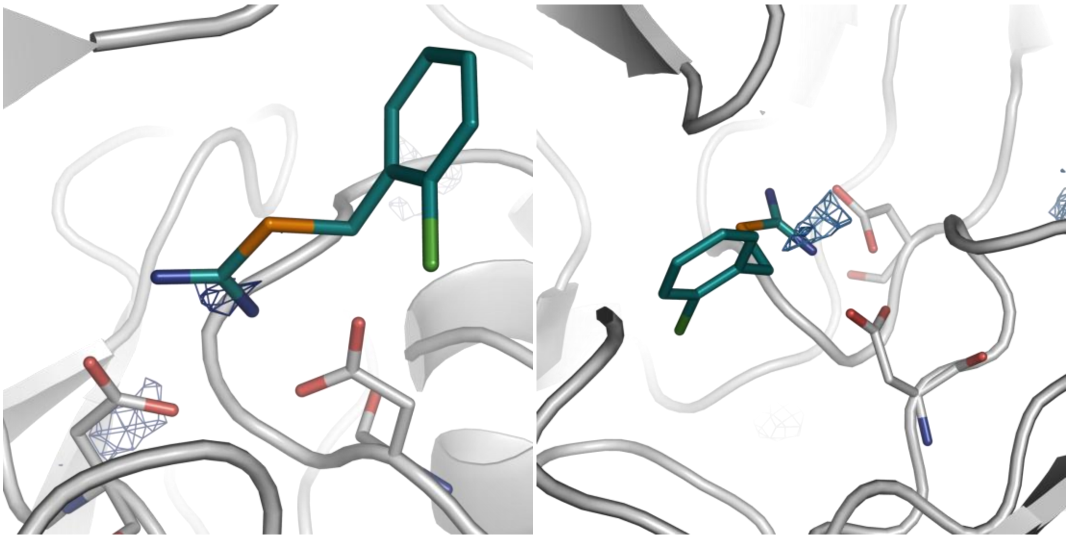


Figure 4.8: Fragment 306 is shown in the binding site of endothiapepsin. Left: The contour map of a donor is shown at a map level of 68 % above the minimimal map level (-176722/ -260327). Right: The contour map for the specific atom type N.mih is shown at a map level of 65 % above the minimal map level (-17671/ -27017).

4.4.4 Analysis of aromatic functionalities

Fragment 148 is a good example to study aromatic properties in the binding site of endothiapepsin. The fragment has two aromatic ring moieties linked by a spacer. The catalytic dyad is addressed by one ring nitrogen and one nitrogen found in the spacer. Figure 4.9 shows on the left hand side the contour map for an aromatic probe atom. This map was produced considering nine different atom types to describe an aromatic probe atom. On the right hand side the map for the specific C.ar6 is shown, this atom type is defined as a sp^2 carbon in a benzene ring. The larger region preferred by aromatic properties coincides well with the aromatic moiety found in the fragment.

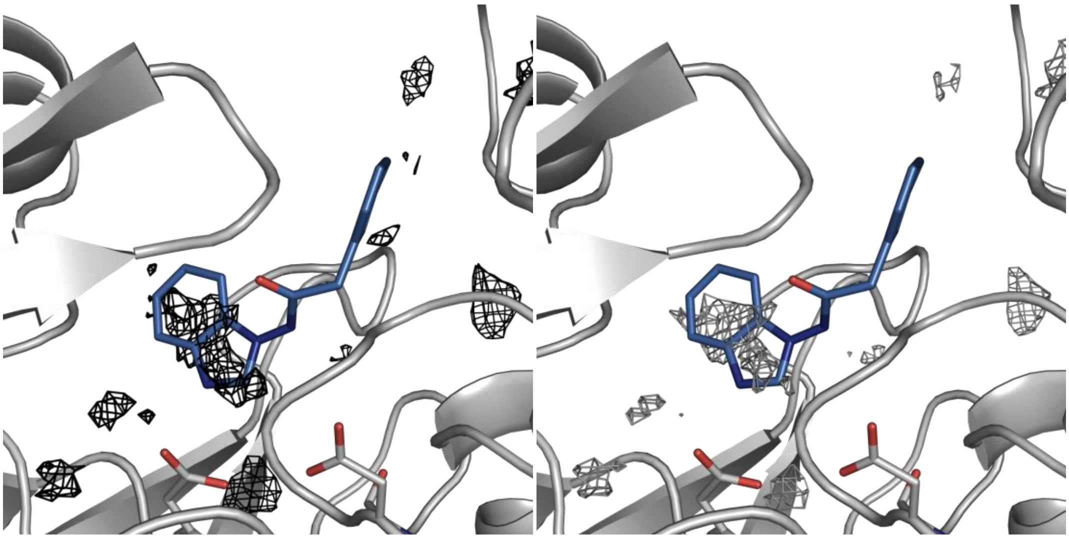


Figure 4.9 Fragment 148 is shown in the binding site of endothiapepsin. Left: The contour map for aromatic probe is shown at a map level of 56 % above the minimal map level (-112067/ -199231). Right: The contour map for the specific atom type C.ar6 is shown at a map level of 63 % above the minimal map level (-18102/ -28799).

Figure 4.10 shows fragment 255. This compound contains two aromatic moieties as well. The bicyclic system is well placed into the most favorable area of the calculated contour map. The catalytic dyad is addressed by two nitrogens, one is part of the larger bicyclic system and the other one is localized in the amide bond. Thus, the aromatic aminophenyl portion is in a different position compared to the benzyl moiety in fragment 148. In the environment of this aromatic ring system no highly favorable aromatic interaction field is computed. This may be an indication that the hydrogen bond of the exocyclic nitrogen is orienting the attached aromatic ring in this area of the binding pocket.

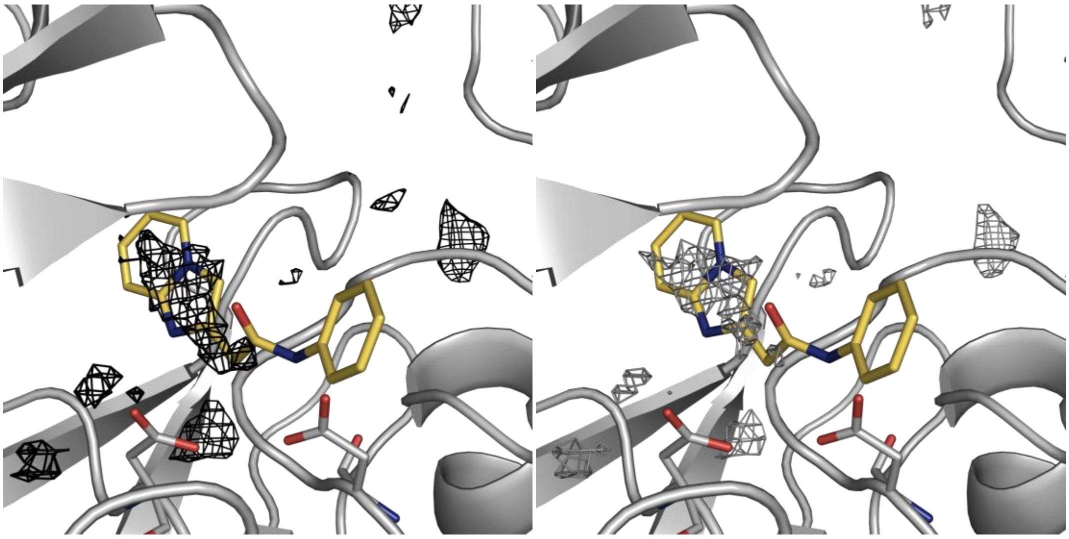


Figure 4.10: Fragment 255 is shown in the binding site of endothiapepsin. Left the contour map for aromatic probe is shown at a map level of 56 % above the minimal map level (-112067/ -199231). Right the contour map for the atom type C.ar6 is shown at a map level of 63 % above the minimal map level (-18102/ -28799).

4.4.5 Analysis of hydrophobic functionalities

Fragment 284 will be used to study hydrophobic interactions. Figure 4.11 displays the contour map for the hydrophobic functionalities. This map is from a generalized atom type based on 18 different atom types (Table 4.1). The pyrrolidine motif of this fragment is placed in the calculated map. In the figure on the right the contour map is computed for the specific atom type for C.3s. This atom type is defined as sp^3 carbon connected to two non-hydrogen atoms. Here, the calculated contour map fits well.

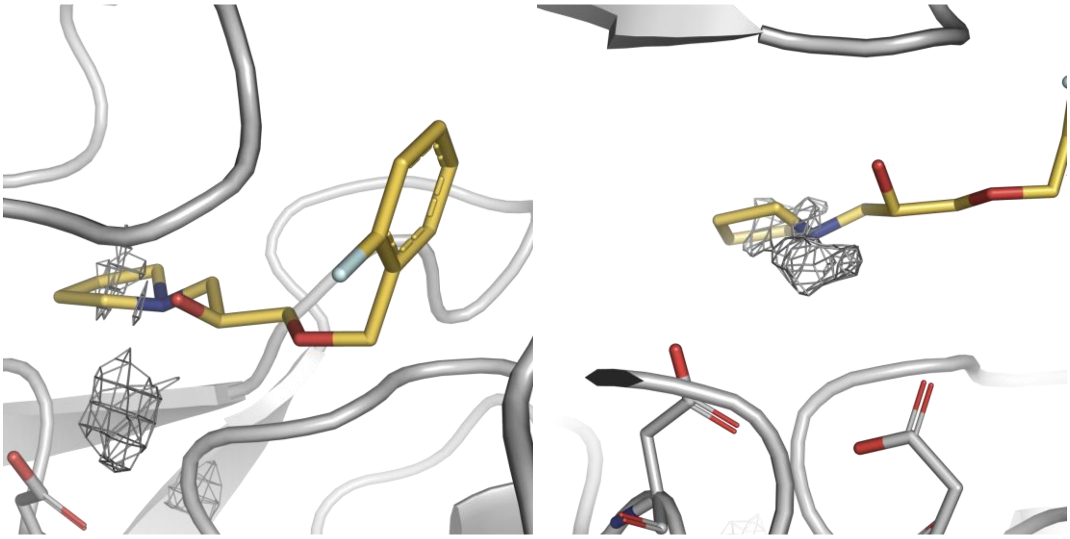


Figure 4.11 shows fragment 284 in the binding site of endothiapepsin. Left: The contour map of a hydrophob probe is shown at a map level of 60 % above the minimal map level (-280171/ -463165). Right: The contour map for the specific atom type C.3s is shown at a map level of 62 % above the minimal map level (-21981/ -35206).

4.5 Conclusion and Outlook

Overall the calculated maps for endothiapepsin are predicting the positions of corresponding atom types actually found in small fragment crystal structures in the binding pocket remarkably well.

For the acceptor, doneptor and donor probes the merged atom types performed slightly better then the single atom types. Only one case for the doneptor probe atoms was observed with a large difference, here the map level differ from 79 % to 38 %. For the hydrophobic and the aromatic probe atoms the single atom types had a slightly better predictive power compared to the merged atom types 62 % to 60 % and 63 % to 56 %, respectively.

A reason for this phenomenon can be the different population of the defind atom types in the CSD database.[73] The knowledge based potentials are driven form a subset of the CSD database. Here, the population of hydrophobic and aromatic atoms is higher compared to the hydrophilic atom types. Thus, the potentials for the hydrophobic and aromatic atom types are based on a larger data pool compared to the hydrophilic atoms. This could be an explanation for the better predictive power for the merging hydrophilic atom types.

On the basis of the predicted hot spots for the different atom types a design of putative ligands will be performed. The synthetic realization of these designed ligands is planned in the groups of Prof. Diederich in Marburg, Germany and Prof. Anna Hirsch in Groningen, Netherlands.

5 Experimental Active Site Mapping as a Starting Point to Fragment Based Lead Discovery

5.1 Introductory Remarks

This chapter is prepared to be published in a scientific journal. The present study is based on experimental HotSpot mapping performed by Dr. Jürgen Behnen[76] and Helene Köster (Univ. Marburg). The experimental data were used to validate the computational HotSpot calculations with DrugScore. My contribution to this work was the calculation and interpretation of the DrugScore HotSpot maps.

5.2 Introduction

High-throughput-screening (HTS) is still considered as the major source for lead discovery in pharmaceutical drug research.[77] However, the success rate of finding promising leads by HTS is still rather unsatisfactory after 20 years of experience, even though well-tailored screening libraries are scanned.[78] Enormous HTS libraries have been assembled with up to several million compounds, often compiled by candidates from late-phase drug discovery projects. Therefore, more drug-like than lead-like molecules end-up in these depositories[79] and leave little room for optimization concomitantly keeping their existing drug-likeness.[80] Lead optimization involves attachment of novel and additional substituents which will for drug-size molecules easily exceed their molecular weight and hydrophobicity beyond the limits acceptable for drug-like molecules.[81]

Contrary to HTS, fragment-based lead discovery (FBLD) starts with molecules of low molecular weight, typically in a range of 120-250 Da, however, binding affinities will only attain milli- to micro-molar range.[82] To better estimate the actual perspective of a hit in FBLD “ligand efficiency” is consulted as a descriptor. It ranks the actually achieved potency of a discovered hit with its potential to be improved by fragment growing or fragment linking.[83-85] Meanwhile, more than ten drug candidates developed from initial fragment hits are in clinical trials, which underlines the growing relevance and acceptance of fragment-based approaches.[14, 86-89]

The present study merges the concepts of multiple-solvent-crystal-structure determination (MSCS) and FBLD. MSCS was developed in the mid-nineties as a new crystallographic approach in the group of Dagmar Ringe.[90-92] The idea of this concept was to soak small organic solvent molecules into protein crystals to map out their binding sites experimentally. The method was first applied to porcine pancreatic elastase.[93] Here, we want to extend the MSCS approach from solvent to small highly soluble probe molecules still well beyond the molecular weight limits applied to standard-sized fragments, but with rather general applicability to a broad range of proteins. They help to map out active sites for their hot spots of binding with respect to hydrophilic and hydrophobic properties. Aside of an efficient active-site mapping, the detected probe poses are valuable indicators to establish a protein-based pharmacophore. Furthermore, they can serve as an entry point to embark into a fragment growing project (Chapter 6).

Fragment growing projects are usually performed as an iterative study of experimental structural biology, computational design and synthesis.[14] Starting with seminal approaches such as GRID by Peter Goodford[94] and MCSS by Martin Karplus[95] a plethora of tools have been developed to map out computationally the hot spots of binding in protein pockets.[96-100] These methods are either based on sophisticated force fields and apply concepts of molecular dynamics simulations or they follow empirical approaches that are based on statistical potentials or occurrence propensities observed in experimental structures. Early on, we suggested the knowledge-based scoring function DrugScore[100] to be used for such hot spots analyses. To further optimize DrugScore with respect to the retrieved data source and the applied atom-type definition[101] we compared the experimentally determined poses of solvents up to small probe molecules with respect to the predicted hot spots. To estimate the potential whether the determined poses of our probe molecules can be used as a starting point for a fragment growing strategy, we compared the observed poses with larger, already characterized protein-ligand complexes in the PDB.

An important criterion for the selection of appropriate probe molecules is their very high solubility. As we have to anticipate very low binding affinity, this fact can only be compensated by high ligand concentrations applied during soaking. In addition, such weak binding is difficult to detect reliably. Therefore, we performed our screening of small highly soluble molecules directly on protein crystals by X-ray crystallography.[102, 103]

To establish our experimental setup, we selected the zinc protease thermolysin (TLN) as target protein. This enzyme has already been used by others for extensive MSCS studies.[104, 105] English et al.[105] succeeded to soak isopropanol, acetonitrile, acetone, DMSO and phenol into thermolysin crystals. With our attempts we were able to expand this list by a set of additional molecules. Subsequently, we transferred these probe molecules to other enzymes such as protein kinase A (PKA), D-xylose-isomerase from *Streptomyces rubiginosus* (DXI), 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (IspD) of the non-mevalonate pathway, and the aspartyl proteases endothiapepsin (ETP) and secretory aspartyl protease 2 (SAP2).

5.3 Results and Discussion

5.3.1 Probe molecule characterization by X-ray crystallography

Thermolysin: As described in the study by English et al., phenol binds to the S1’ pocket of TLN (PDB code 1FJW).[105] For reasons of better crystallographic detection, we decided to use the phenol derivative 3-bromo-phenol containing the strongly diffracting bromine. As expected, 3-bromo-phenol binds in virtually the same way as previously found by English et al. As depicted in Figure 5.1 A and B, both phenol and 3-bromo-phenol, bind to the S1’ pocket of TLN forming hydrogen bonds to Glu143 and through long distance an electrostatic interaction (4.1 or 4.4 Å, resp.) to the catalytic zinc ion. Even though this fragment was successfully soaked, we decided to use in the following unsubstituted phenol. Due to its higher logP value [2.63 vs. 1.48 (http://www.emolecules.com)], 3-bromo-phenol exhibits inferior solubility and puts some limitations on this molecular probe. Then, we decided to soak other small soluble molecules with varying properties (for details see experimental section) into TLN crystals and revealed successfully X-ray structures with aniline, urea, *N*-methylurea, 2-bromoacetate, 1,2-propanediol and N_2O. The latter 1,2-propanediol was used as racemate, therefore we checked whether the R- or S-enantiomer could be assigned to the difference electron density map. Similar to phenol, aniline also binds in the S1´-pocket of TLN (Figure 5.1 C). Surprisingly, aniline assumes a different binding mode compared to phenol. The aromatic ring system coincides with the binding pose of phenol, but the amino group binds via hydrogen bonds to Glu143, Ala113 and Asn112 and it does not form the long-range electrostatic interaction with the zinc ion. The deviating orientation of the amino group results in an induced-fit adaption of Glu143 moving this residue 1.8 Å towards Ala113 and forcing Asn112 into a slightly twisted orientation (Figure 5.1 D).

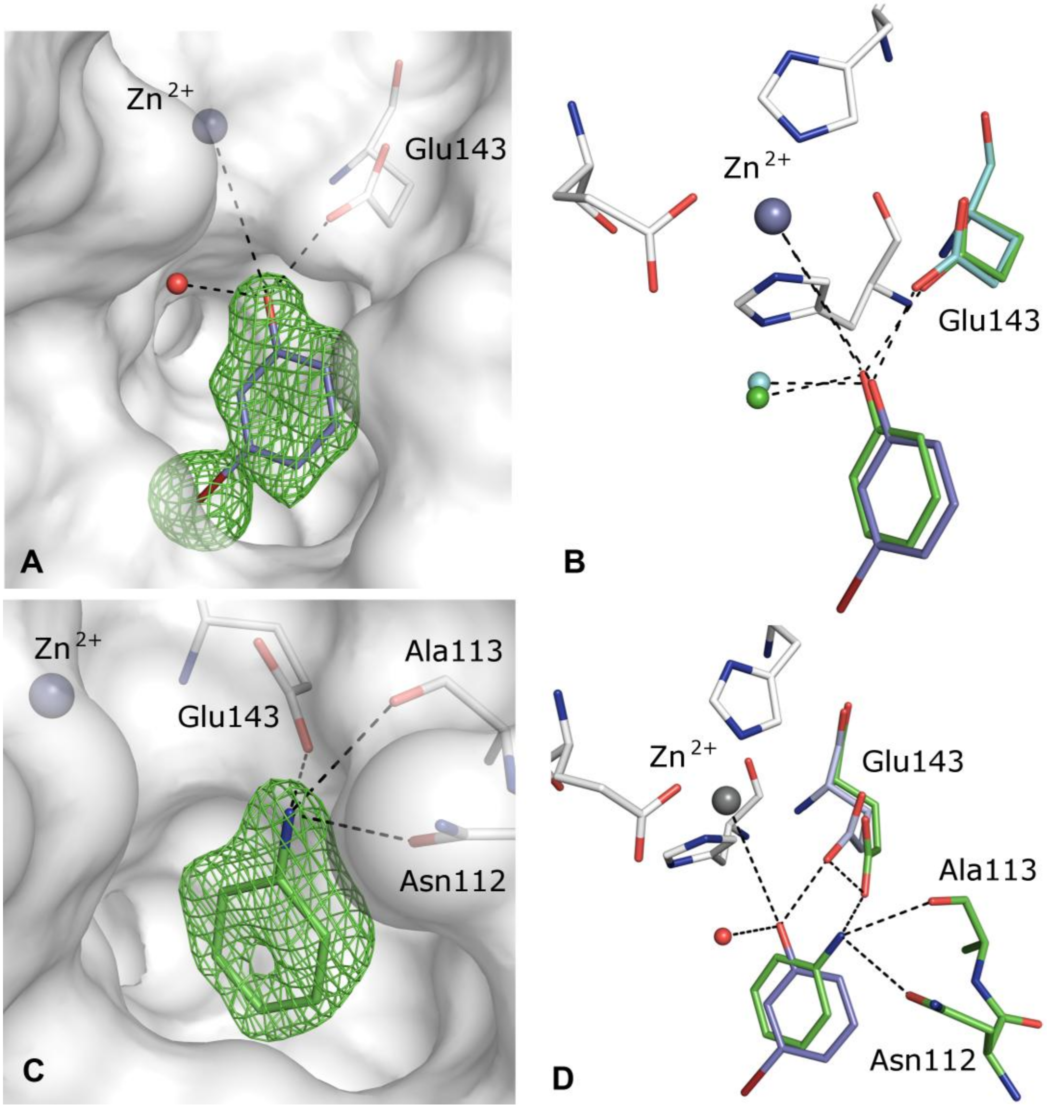


Figure 5.1 A: Surface representation of the S1' pocket of TLN. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with 3-bromo-phenol (blue) are shown in white stick models. In all cases oxygen atoms are depicted in red, nitrogen atoms in blue, bromine in brown and the catalytic zinc ion is shown as gray sphere, water as red sphere. The difference electron density (F_o-F_c) for the 3-bromo-phenol molecule is shown in green at a level of 3.0 σ. **B:** Close up view of the phenol/3-bromo-phenol binding geometry in the TLN complex. Phenol is shown in green and 3-bromo-phenol in blue. The interacting side chains of amino acid residues are displayed as green stick models for phenol, in light-blue stick models for 3-bromo-phenol and as white stick models for the zinc coordinating residues (Glu166, His146 and His142). Dashes symbolize the interaction between one molecule and the respective partner. **C:** Surface representation of the S1' pocket of TLN. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with aniline (green) are shown as white stick models. The difference electron density (F_o-F_c) for the aniline molecule is shown in green at a level of 3.0 σ. **D:** Close up view of the aniline/3-bromo-phenol binding geometry in the TLN complex. Aniline is shown in green and 3-bromo-phenol in blue. The interacting side chains of amino acid residues are displayed as green stick models for aniline, in light-blue stick models for 3-bromo-phenol and as white stick models for the zinc coordinating residues (Glu166, His146 and His142).

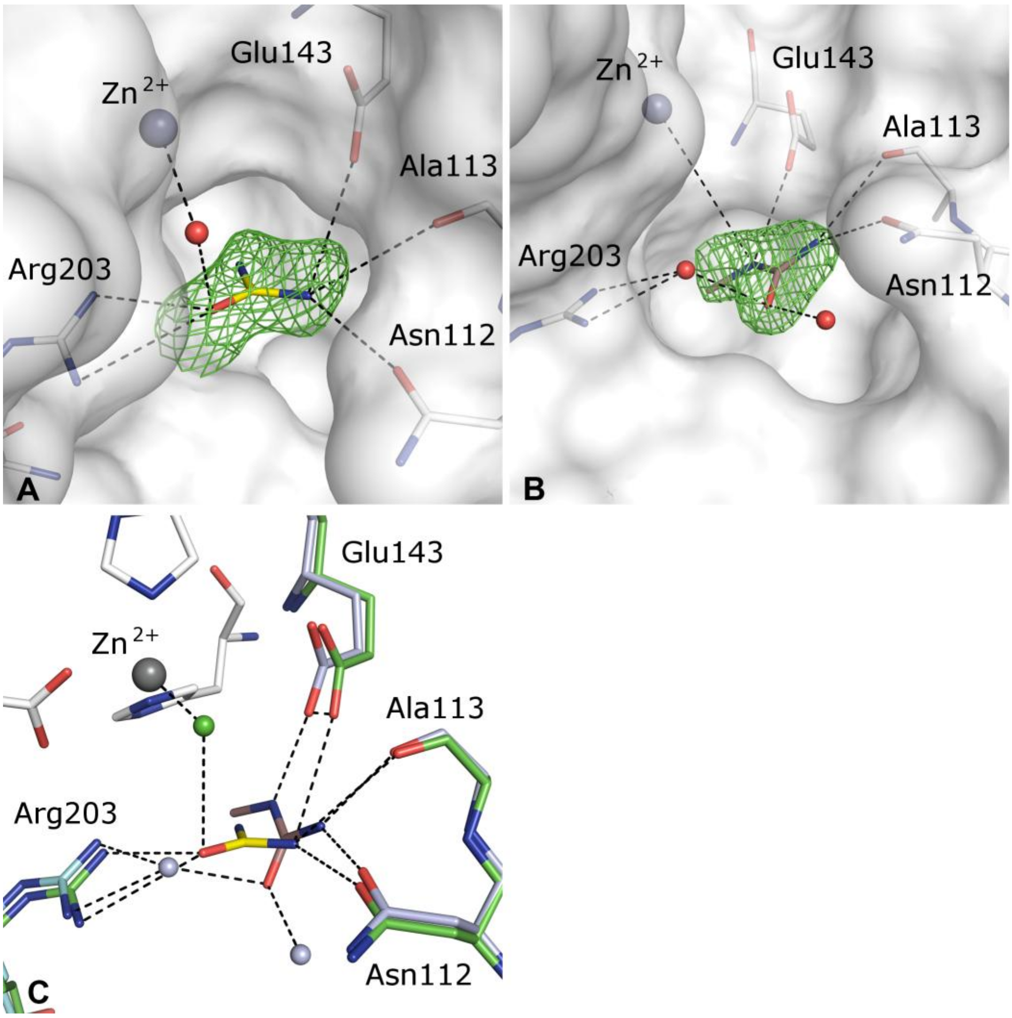


Figure 5.2 A: Surface representation of the S1’ pocket of TLN. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with urea (yellow) are shown as white stick models. In all cases oxygen atoms are depicted in red, nitrogen atoms in blue, the catalytic zinc ion is shown as gray sphere and water molecules as small spheres. The difference electron density (F_o-F_c) for the urea molecule is shown in green at a level of 3.0 σ. **B:** Surface representation of the S1’ pocket of TLN. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with *N*-methylurea (violet) are shown in white stick models. The difference electron density (F_o-F_c) for the *N*-methylurea molecule is shown in green at a level of 3.0 σ. **C:** Close up view of the urea/*N*-methylurea binding geometry in the TLN complex. Urea is shown in yellow and *N*-methylurea in violet. The interacting side chains of amino acid residues are displayed as green stick models for urea, including the respective water molecule, in light-blue stick models for *N*-methylurea, including the respective water and as white stick models the zinc coordinating residues (Glu166, His146 and His142). Dashes symbolize the interaction between one molecule and the respective partner.

Soaking experiments with urea as a probe involves the complication that its overall shape does not allow to distinguish between carbonyl and amino groups in the difference electron density. This complication stimulated us to also use *N*-methylurea as a lower symmetric analogue of urea. This molecule allows to distinguish carbonyl and amino groups in the electron density map. In both structures one molecule is found in the S1’ pocket (Figure 5.2 A, B). Interestingly, as indicated in Figure 5.2 C, *N*-methylurea binds somewhat differently compared to urea and must therefore be regarded as a probe with deviating properties. In the S1´subsite *N*-methylurea adopts a binding pose that is rotated by 90° with respect to urea. The position of the urea carbonyl oxygen is occupied in the *N*-methylurea complex by a picked-up water molecule. The latter mediates a hydrogen bond network between Arg203 and the carbonyl group of *N*-methylurea. In the case of urea, the carbonyl oxygen is directly H-bonded to the guanidinium group of Arg203 and forms an additional H-bond to the zinc ion transmitted by a water molecule. In both cases, additional hydrogen bonds are formed to Glu143, Ala113 and Asn112.

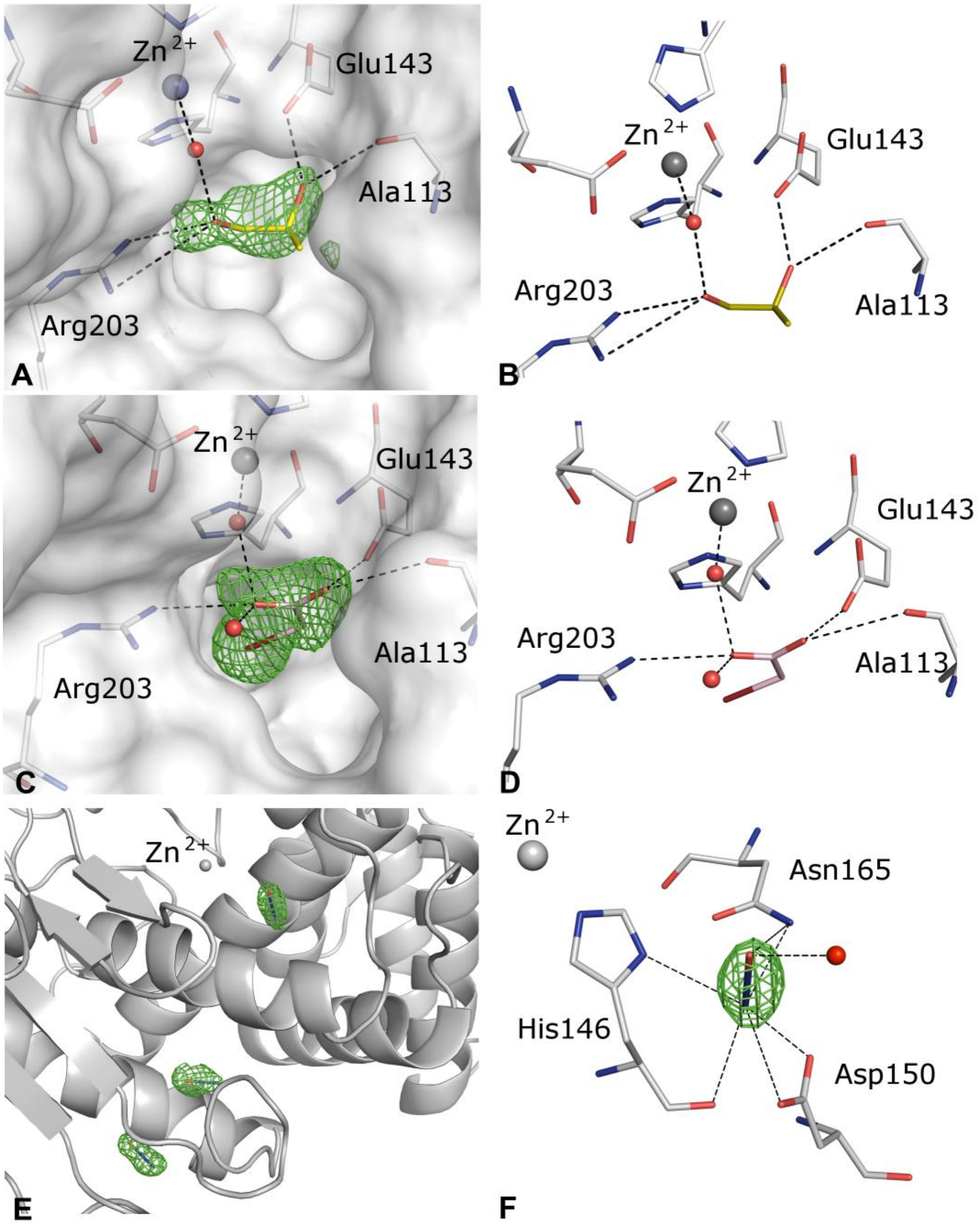


Figure 5.3 A: Surface representation of the S1' pocket of TLN. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with S-1,2-propanediol (yellow) are shown as white stick models. In all cases oxygen atoms are depicted in red, nitrogen atoms in blue and the catalytic zinc ion is shown as gray sphere. The difference electron density (F_o-F_c) for the S-1,2-propanediol molecule is shown in green at a level of 3.0 σ. **B:** Close up view of the S-1,2-propanediol binding geometry in the TLN complex. S-1,2-propanediol is shown in yellow. The interacting side chains of amino acid residues are displayed as white stick models for S-1,2-propanediol and for the zinc coordinating residues (Glu166, His146 and His142). **C:** Surface representation of the S1'pocket of TLN. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with 2-bromoacetate (pink) are shown in white stick models. The difference electron density (F_o-F_c) for the 2-bromoacetate molecule is shown in green at a level of 3.0 σ. **D:** Close up view of the 2-bromoacetate binding geometry in the TLN complex. 2-bromoacetate is shown in pink. The interacting side chains of

amino acid residues are displayed as white stick models for 2-bromoacetate and for the zinc coordinating residues (Glu166, His146 and His142). **E:** Cartoon model of TLN and the zinc displayed in gray. Three nitrous oxide molecules are shown as stick models with their corresponding difference electron density (F_o-F_c) shown at a level of 3.0 σ. **F:** Close up view of the nitrous oxide binding geometry within the TLN active site. Here, nitrous oxide interacts with Asp150, Asn165, His146 and a water molecule. The difference electron density (F_o-F_c) is displayed at a level of 3 σ.

The binding pose of S-1,2-propanediol in the S1'pocket of TLN is displayed in Figure 5.3 A and B. Interestingly, S-1,2-propanediol combines the previously described binding modes of 3-bromophenol, aniline, urea and *N*-methylurea by forming hydrogen bonds on one end with the 2-hydroxyl group to Glu143 and the backbone oxygen of Ala113 and on the other end with the 1-hydroxyl group to the guanidinium group of Arg203. We also succeeded to discover 2-bromoacetate as a probe molecule for TLN (Figure 5.3 C). It binds similarly to urea and 1,2-propanediol in the S1' pocket of TLN. The first carboxyl oxygen forms hydrogen bonds to Glu143 and Ala113 and the second to Arg203. In addition to Arg203, bromoacetate interacts with two water molecules (Figure 5.3 D). Finally, we attempted to pressure gases into TLN crystals. With N_2O we were able to obtain a high resolution structure that allowed placement of three N_2O molecules into the difference electron density (Figure 5.3 E). One N_2O molecule is found at a deeply buried position where previous studies could detect xenon and phenol as binders.[105] Nonetheless, this site is very narrow and clearly of no relevance for any drug binding. In the active site one molecule could be detected. The linear N_2O mediates somewhat similar to interstitial water molecules long-range contacts in the unoccupied active site of thermolysin. For example, Figure 5.3 F shows an N_2O molecule contacting via both terminal ends either Asp150, the backbone carbonyl of His146, Asn165, and His146Nδ. Similarly to urea, unambiguous assignment of nitrogen and oxygen to the N_2O density remains ambiguous. Furthermore, the volatile molecule is not easy to fully populate in crystals. Due to its low diffraction power N_2O is only of limited use and may be mistaken for a disordered water molecule. However, for highly resolved structures a binding pose might be proposed. Next, a gas of high diffraction power, xenon, was used as probe molecule to map out hydrophobic sites in the binding pocket of TLN. Here, assignment of the xenon atom to the electron density was straightforward. Unfortunately, xenon binding was not observed within the binding pocket of TLN but in a remote hydrophobic cavity. While unsuitable for mapping of binding pockets, further xenon experiments with endothiapepsin and acetylcholinesterase revealed its potential for in-house SAD phasing (Behnen et al., unpublished results).

Aside of the described probes a large number of additional substances were tested but failed. Among them the polar aromatic probe benzamidine that has been applied successfully to elucidate preferred binding sites next to aspartates. Possibly we failed with this probe in TLN due to the cationic character of the zinc-ion contained in the binding site.

The results with thermolysin stimulated us to screen particularly the probe molecules phenol, *N*-methylurea and 1,2-propandiol against other proteins. To test benzamidine we included two aspartyl proteases in our test panel.

Protein Kinase A: A crystal structure of PKA in complex with phenol shows three bound phenol molecules. Two address the active site of the protein (Figure 5.4 A, B), while a third is found in a remote position on the protein surface. The binding mode of the latter will not be further discussed as it does not coincide with the usually addressed ATP binding pocket. The first phenol probe addresses the hinge region adopting a binding mode which allows its hydroxyl group to act simultaneously as H-bond donor to the backbone carbonyls of Glu121 and Val123 (2.9 or 4.3 Å, resp.) and as acceptor to the backbone amino group of Val123 (2.8 Å). This interaction pattern is commonly found for many potent kinase inhibitors. The ring plane of the second phenol is twisted by approximately 90° and occupies the ribose sugar pocket. Its hydroxyl group interacts with Glu127, the backbone carbonyl of Leu49 and one water molecule. The phenyl moiety is located directly below the glycine rich loop (Figure 5.4 B). This usually very flexible loop is fixed in the present structure by two hydrogen bonds formed between the backbone NH and the hydroxyl group of Ser53 and the carboxylate of Glu184.

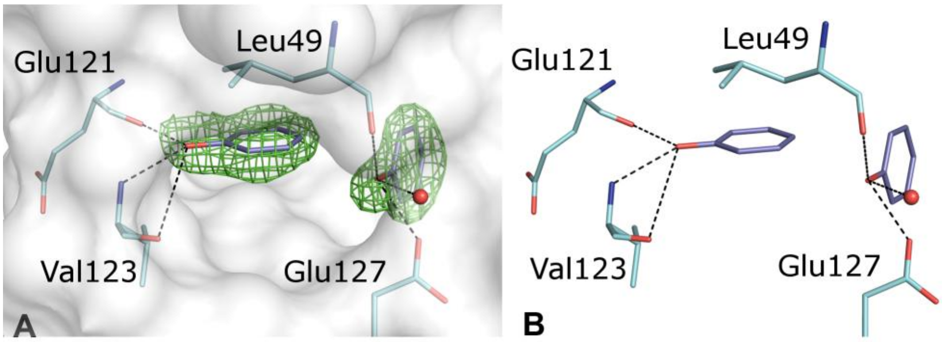


Figure 5.4 A: Surface representation of the active site of PKA. The solvent-accessible surface of the protein is shown in gray. Amino acids involved in interactions with phenol (blue) are shown in light-blue stick models. In all cases oxygen atoms are depicted in red and nitrogen atoms in blue. The difference electron density (F_o-F_c) for the phenol molecule is shown in green at a level of 3.0 σ. **B:** Close up view of the phenol binding geometry in the PKA complex. Phenol is shown as stick model in blue. The interacting side chains of amino acid residues are displayed as light-blue stick models. Dashes symbolize the interaction between one molecule and the respective partner.

D-xylose-isomerase (DXI) and 4-Diphosphocytidyl-2C-methyl-D-erythritol synthase (IspD): We could detect an S-1,2-propanediol molecule in the active site of DXI (Figure 5.5 A). The 2-hydroxyl group coordinates one catalytic manganese ion, the carboxylate group of Glu180 and interacts with an additional water molecule. The 1-hydroxyl function is involved in H-bonds to Nε of His53 and fixes an incorporated water molecule (Figure 5.5 B). A R-1,2-propanediol molecule is found in the active site of both crystallographically independent monomers of IspD (Figure 5.5 C). Their binding modes are very similar, thus only one complex will be discussed in detail. The 1-hydroxyl group of R-1,2-propanediol binds via hydrogen bonds to the backbone NH and the side chain hydroxyl group of ThrA140. Furthermore, a contact to the backbone carbonyl of AlaB163 is observed. The 2-hydroxyl group forms an H-bond to the carboxylate group of AspB106, the guanidinium group of ArgB109 and the side chain ammonium group of LysB213 (Figure 5.5 D).

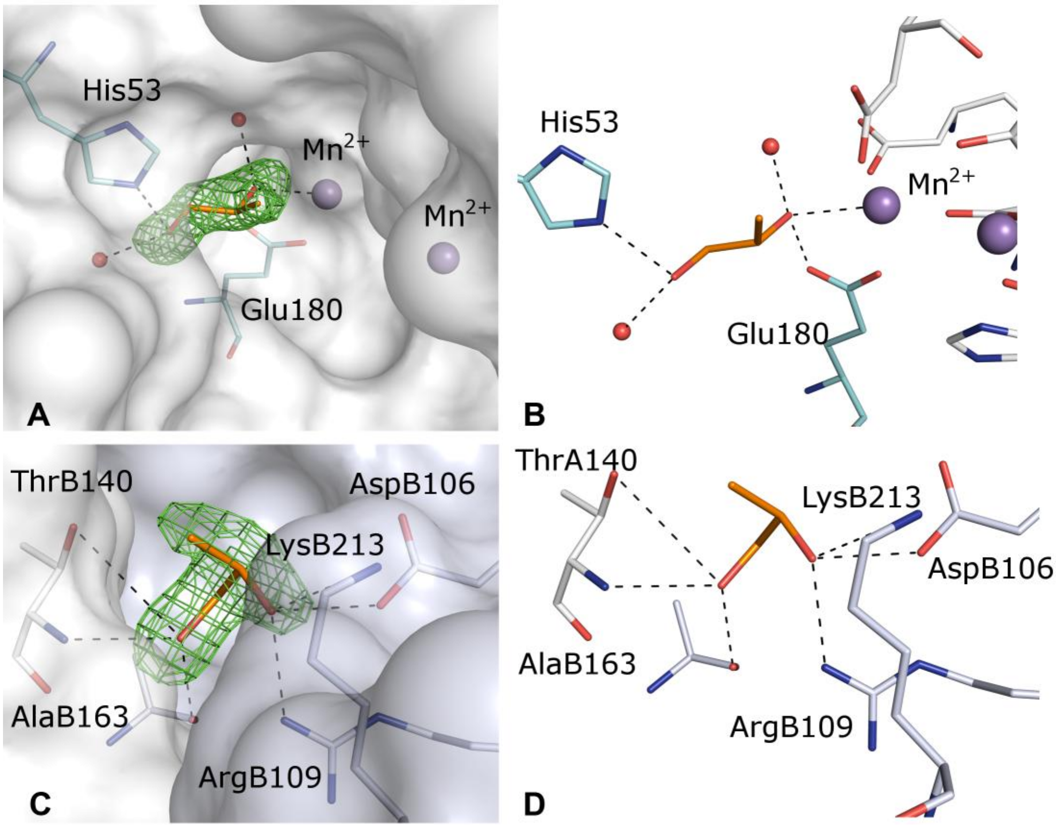


Figure 5.5 A: Surface representation of the active site of DXI. The solvent-accessible surface of DXI is shown in gray. Amino acids involved in interactions with S-1,2-propanediol (orange) are shown in light-blue stick models and the two manganese ions as purple spheres. In all cases oxygen atoms are depicted in red and nitrogen atoms in blue. The difference electron density (F_o-F_c) for the S-1,2-propanediol molecule is shown in green at a level of 3.0 σ. **B:** Close-up view of the S-1,2-propanediol binding geometry in the DXI complex. S-1,2-propanediol is shown in orange as stick model. The interacting side chains of amino acid residues are displayed as light-blue stick models for S-1,2-propanediol and as white stick models for the two manganese coordinating residues (Asp286, Asp244, Glu216, Asp254 and His219). Dashes symbolize the interaction between one molecule and the respective partner. **C:** Surface representation of the active site of IspD. The solvent-accessible surface of monomer1 is shown in gray and for monomer two in light-blue. Amino acids involved in interactions with R-1,2-propanediol (orange) are shown in light-blue stick models for monomer two and in white stick models for monomer one. The difference electron density (F_o-F_c) for the R-1,2-propanediol molecule is shown in green at a level of 3.0 σ. **D:** Close up view of the R-1,2-propanediol binding geometry in the IspD complex. R-1,2-propanediol is shown in orange as stick model. The interacting side chains of amino acid residues are displayed as light-blue stick models for monomer two and as white stick models for monomer one.

Secretory Aspartyl Protease 2 (SAP2) and Endothiapepsin (ETP): The active sites of both aspartyl proteases were probed with benzamidine, a popular molecule used to address Asp189 in the S1-pocket of trypsin-like serine proteases.[106, 107] In SAP2 two benzamidine probes could be detected, while ETP even contained three benzamidine molecules. The first amidine addresses in both enzymes with very similar geometry one aspartate (Asp32 SAP2, Asp35 ETP) of the catalytic dyad with twinned H-bonds, a second Asp is involved in hydrogen-bond contacts (Asp218 SAP2, Asp219 ETP, Figure 5.6 A, B). The aromatic moiety is oriented towards the S1’ pocket. In SAP, the second benzamidine is located in a completely different environment in the S2 pocket. It virtually bridges across the active site forming H-bonds with one NH_2 group to the backbone carbonyl of Gly220 and the side chain OH of Thr221 (3.0, 2.8 Å). The opposing amino group involves the carboxylate side chain of Asp86 (2.9 Å) and the Tyr84 backbone carbonyl oxygen (3.4 Å) in an H-bonding network and forms a 2.7 Å hydrogen bond to the hydroxyl group of an additionally observed MPD (2-methyl-2,4-pentanediol) molecule. The phenyl ring planes of both adjacent benzamidines are at about 4.5 Å in somewhat larger distance than a van der Waals contact (Figure 5.6 C). The second benzamidine is found in ETP to bind into the S1 pocket next to Asp81 and Ser115. It forms with one of its NH_2 groups H-bonds to the carboxylate group of Asp81 (2.7 Å) and to the hydroxyl group of Ser115 (3.0 Å), while the other NH_2 group forms H-bonds to the backbone carbonyl oxygen of Ser115 (3.2 Å) and a water mediated contact to Asp119 (Figure 5.6 D). Apart from two benzamidine molecules observed in ETP (Figure 5.6 D), a DMSO molecule is found in the active site somewhat below the position where in SAP the second benzamidine is located. The third benzamidine in ETP is located at the protein surface, forming twinned H-bonds to the carboxylate of Asp279 and a carbonyl oxygen interaction of Cys255.

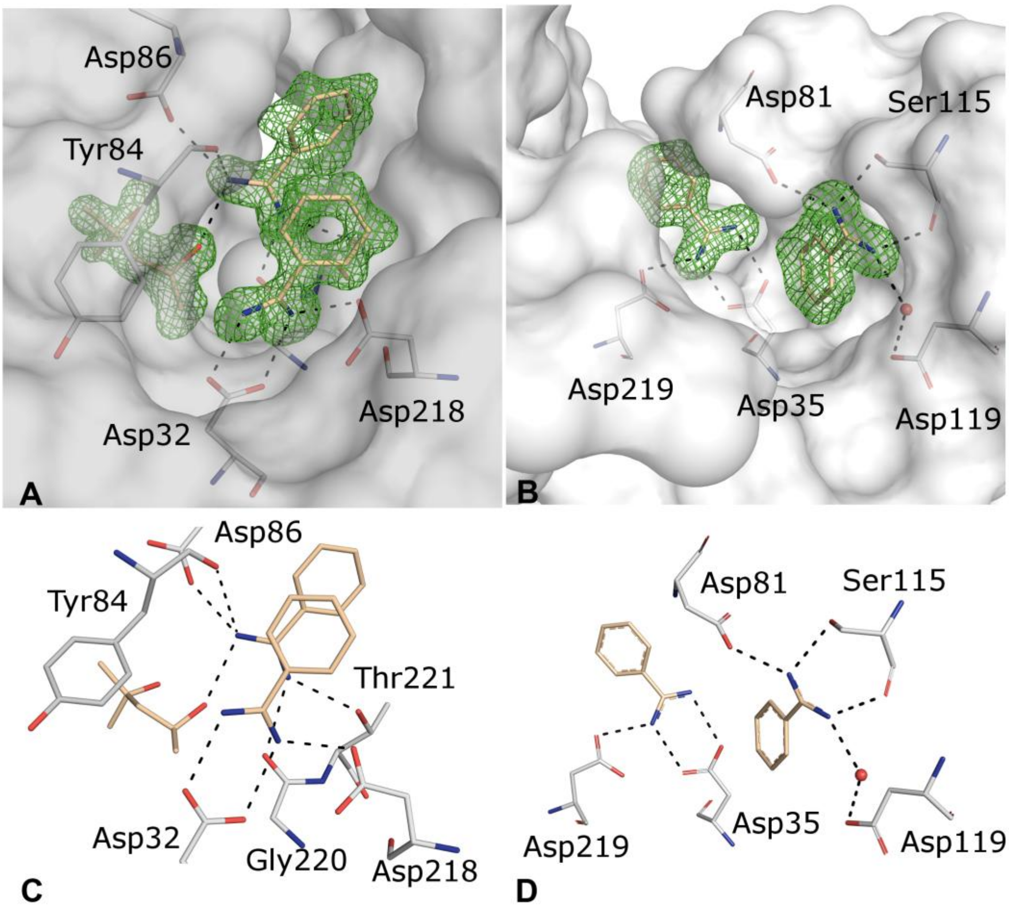


Figure 5.6 A: Surface representation of the active site of SAP. The solvent-accessible surface of SAP is shown in gray. Amino acids involved in interactions with benzamidine (light brown) are shown as white stick models. Furthermore a 2-methyl-2,4-pentanediol molecule is observed in close proximity to the benzamidine molecules. In all cases oxygen atoms are depicted in red and nitrogen atoms in blue. The difference electron density (F_o-F_c) for the shown molecules is displayed in green at a level of 3.0 σ. **B:** Surface representation of the active site of ETP. The solvent-accessible surface of ETP is shown in gray. Amino acids involved in interactions with benzamidine (light brown) are shown as white stick models. The difference electron density (F_o-F_c) for the two benzamidine molecules is displayed in green at a level of 3.0 σ. **C:** Close up view of the benzamidine binding geometry in the SAP complex. Two benzamidine and a 2-methyl-2,4-pentanediol molecule are shown in light brown. The interacting side chains of amino acid residues are displayed as white stick models. Dashes symbolize the interaction between one molecule and the respective partner. **D:** Close up view of the benzamidine binding geometry in the ETP complex. Two benzamidine molecules are shown in light brown. The interacting side chains of amino acid residues are displayed as white stick models.

5.3.2 DrugScore hot spot calculations

Our approach of experimental binding-site mapping with several rather generally applicable probes can be used as a starting point to construct a protein-based pharmacophore. Furthermore, as mentioned in the introduction, several computational tools have been developed to underscore the most favorable binding regions in an active site, the so-called hot spots of binding, among them the programs GRID, DrugScore or SuperStar.[98-100] In this study we applied DrugScore based on our most recent atom type setting. This scheme is based on 157 distinct types and data collected from small molecule crystal structures in the CSD to characterize our crystallographically studied probes.[101] The computational probes used in the different cases are listed in Table 5.1.

For visualization purposes the contour level of carbon or aromatic atoms (black), oxygen or acceptor atoms (red), nitrogen or donor atoms (blue) and doneptor atoms (an atom type that can serve either as H-bond donor or H-bond acceptor, e.g. hydroxyl functionality, yellow) is shown at a level showing best clarity in the maps, but in most cases the contour level has been set to approximately 60 % above the value of the next local minimum in the studied active site.

Table 5.1 Probe characterization and DrugScore contour levels to the respective protein

Probe characterization	**Enzym/Contour level**
Phenol: aromatic probe (Aro); doneptor probe (AnD); Br atom C.ar6 = sp2 carbon in an aromatic ring system; O.ph = phenolic hydroxyl group	TLN: Aro: 54 % (-86000/ -159875), AnD: 64 % (-265000/ -410239), Br: 63 % (-10200/ -16207) PKA: C.ar6: 56% (-14000/ -25180), O.ph: 30 % (-2800/ -9326)
Aniline: aromatic probe (Aro), donor probe (Don)	TLN: Aro 63 % (-95000/ -149644), Don 63 % (-115000/ -182022)
Urea: donor probe (Don), acceptor probe (Acc)	TLN: Don 53 % (-100000/ -188767), Acc 67 % (-170000/ -252746)
N-methylurea: donor probe (Don), acceptor probe (Acc)	TLN: Don 59 % (-110162/ -185202), Acc 62 % (-160000/ -255317)
1,2-Propanediol: doneptor probe (AnD), hydrophobic probe (Hyd)	TLN: AnD 45 % (-62163/ -138674) Hyd 71 % (-265873/ -376115) IspD: AnD 37 % (-44800/ -120867) Hyd 83 % (-319000/ -385520) DXI: AnD 50 % (-78500/ -157084), Hyd 69 % (-300000/ -434688)
Benzamidine: aromatic probe (Aro); donor probe (Don)	ETP: Aro 64 % (-90000/ -140728), Don 68 % (-110000/ -162225) SAP: Aro 52 % (-72000/ -137181), 70 % Don (-108000/ -153264)
2-Bromoacetate: C.3s = sp3 carbon; Br atom	TLN: C.3s: 70 % (-20576/ -29468), Br 70 % (-10934/ -15660)

Phenol/Aniline: Comparing the crystallographically determined binding poses of 3-bromo-phenol and aniline in TLN, DrugScore predictions indicate that the hot spots are in good agreement with the experimentally determined positions of the aromatic ring moiety in aniline and phenol as well as the amino functionality of aniline (Figure 5.7 A, B). Only the calculated favorable sites for a doneptor function and bromine atoms differs slightly from the position found in the X-ray structures as DrugScore does not predict the observed and electrostatically favored orientation towards the zinc ion. The DrugScore hot spot analysis of aromatic carbons in PKA agrees well with the observed binding modes of the two phenols in the active site of the protein. The polar OH group fits only for the phenol molecule binding to the hinge region. Figure 5.7 C shows the hot spot contouring for C.ar (black) and O.ph (red) together with the experimentally determined binding mode.

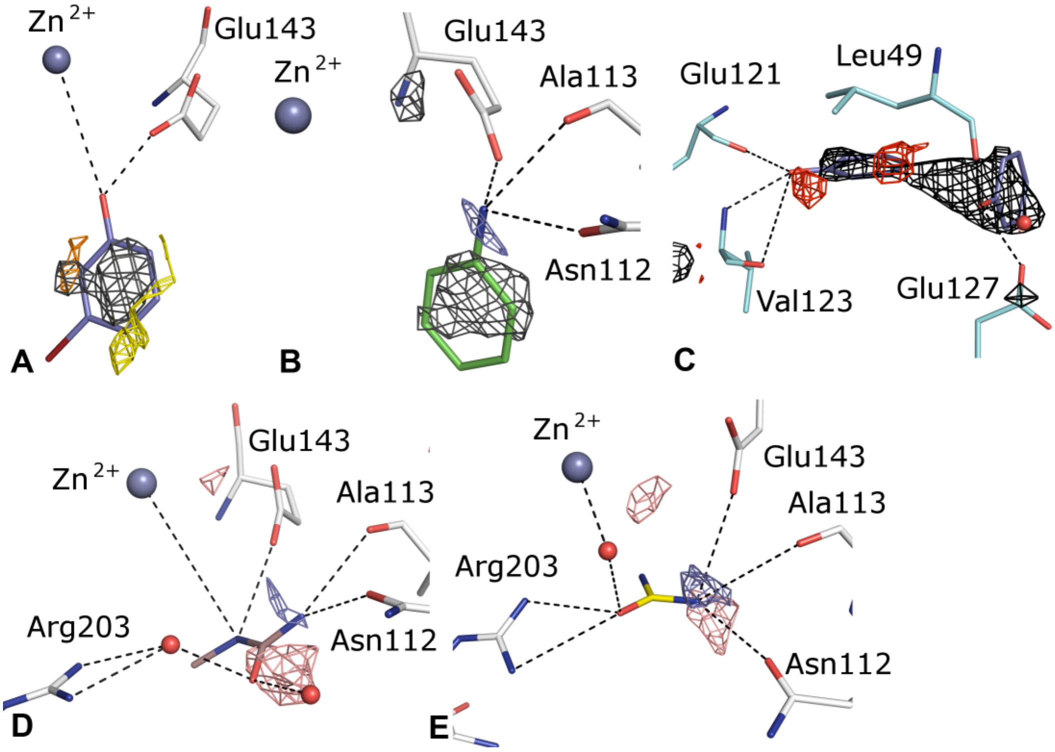


Figure 5.7 A: DrugScore^{HotSpot} analysis of the TLN-3-bromo-phenol complex; bromo-phenol is displayed in blue and Glu143 as white stick model and the catalytic zinc ion as gray sphere. In all cases oxygen atoms are depicted in red, nitrogen atoms in blue, bromine atoms in brown. Interaction distances are displayed as dashed lines. Hotspots are displayed in color-coded mesh using the respective probes for aromatic atoms (black), doneptor atoms (yellow) and bromine atoms (brown) at a map level of 54 % (-86000/ -159875), 64 % (-265000/ -410239) and 63 % (-10200/ -16207) above the minimal map level, respectively. **B:** DrugScore^{HotSpot} analysis of the TLN-aniline complex; aniline is displayed in green and Glu143, Ala113 and Asn112 as white stick model and the catalytic zinc ion as gray sphere. Hotspots are displayed in mesh, aromatic (black) and donor (blue) at a map level of 63 % (-95000/ -149644), (-115000/ -182022) in both cases. **C:** DrugScore^{HotSpot} analysis of the PKA-phenol complex; phenol is displayed in blue and the interacting residues Glu121, Val123, Leu49 and Glu127 as light-blue stick model. Hotspots are displayed in mesh; C.ar6 (black) and O.ph (red) at a map level of 56 % (- 14000/ -25180) and 30 % (- 2800/ -9326), respectively. **D:** DrugScore^{HotSpot} analysis of the TLN-*N*-methylurea

complex; *N*-methylurea is displayed in violet, the interacting residues Glu143, Ala113,Asn112 and Arg203 as white stick model and the catalytic zinc ion as gray sphere. Hotspots are displayed in mesh; donor atoms (blue) and acceptor atoms (red) at a map level of 59 % (-110162/ -185202) and 62 % (-160000/ -255317), respectively. **E:** DrugScore^{HotSpot} analysis of the TLN-urea complex; urea is displayed in yellow, the interacting residues Glu143, Ala113, Asn112 and Arg203 as white stick model and the catalytic zinc ion as gray sphere. Hotspots are displayed in mesh; donor atoms (blue) and acceptor atoms (red) at a map level of 59 % (-100000/ -188767) and 62 % (-170000/ -252746).

Urea/*N*-Methylurea: As described above, one urea and one *N*-methylurea molecule bind to the S1'-pocket of TLN. In both cases, the atom types for donor atoms (blue) and acceptor atoms (red) were used as probes for the DrugScore calculations. For *N*-methylurea, the positions of the carbonyl and one amino function are correctly predicted in TLN (Figure 5.7 D). Unfortunately, only the position of the amino function of urea is in agreement with the crystallographically determined binding in TLN. Possibly the presence of the neighbouring highly charged zinc ion reduces the predictive power of our DrugScore maps again (Figure 5.7 E).

2-Bromoacetate: For 2-bromoacetate as probe molecule, maps were calculated for aliphatic carbon (C.3s; black) and for bromine atoms (brown). As displayed in Figure 5.8 A, the predicted favorable regions for an aliphatic carbon (black) and for the bromine generally agrees with the experimentally observed poses in TLN.

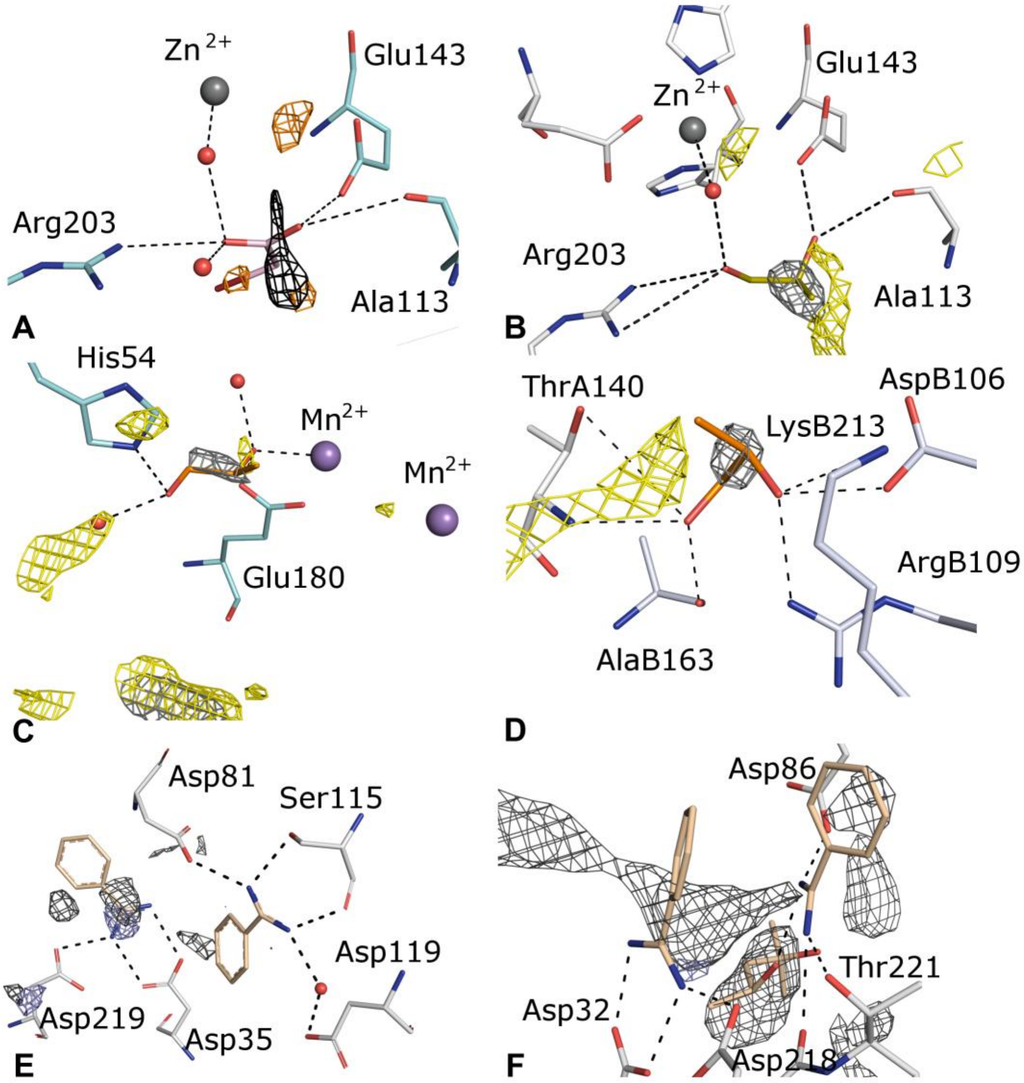


Figure 5.8 A: DrugScore$^{\text{HotSpot}}$ analysis of the TLN-2-bromoacetate complex; 2-bromoacetate is displayed in pink, the interacting residues Glu143, Ala113 and Arg203 as light-blue stick model and the catalytic zinc ion as gray sphere. In all cases oxygen atoms are depicted in red, nitrogen atoms in blue and bromine in brown. Interaction distances are displayed as dashed lines. Hotspots are displayed in color-coded mesh using respective probes C.3s (black) and Br (brown) at a map level of 70 % (-20576/ -29468), (-10934/ -15660). **B:** DrugScore$^{\text{HotSpot}}$ analysis of the TLN-S-1,2-propanediol complex; S-1,2-propanediol and Glu143, Ala113 and Arg203 are displayed as white stick model and the catalytic zinc ion as gray sphere. Hotspots are displayed in mesh; doneptor atoms (yellow) and hydrophobic atoms (gray) at a map level of 45 % (-62163/ -138674) and 71 % (-265873/ -376115), respectively. **C:** DrugScore$^{\text{HotSpot}}$ analysis of the DXI-S-1,2-propanediol complex; S-1,2-propanediol is displayed in orange and His54 and Glu180 as light-blue stick model and the catalytic manganese ion as purple spheres. Hotspots are displayed in mesh; doneptor atoms (yellow) and hydrophobic atoms (gray) at a map level of 50 % (-78500/ -157084) and 69 % (-300000/ -434688), respectively. **D:** DrugScore$^{\text{HotSpot}}$ analysis of the IspD-R-1,2-propanediol complex; R-1,2-propanediol is displayed in orange and the interacting residues ThrA140 as white stick model, AlaB163, ArgB109, AspB106 and LysB213 as light blue stick model. Hotspots are displayed in mesh; doneptor atoms (yellow) and hydrophobic atoms (gray) at a map level of 37 % (-44800/ -120867) and 83 % (-319000/ -385520), respectively. **E:** DrugScore$^{\text{HotSpot}}$ analysis of the

ETP-benzamidine complex; benzamidine is displayed in light brown and the interacting residues Asp219, Asp35, Asp81, Ser115 and Asp119 as white stick model. Hotspots are displayed in mesh; aromatic atoms (black) and donor atoms (blue) at a map level of 64 % (-90000/ -140728) and 68 % (-110000/ -162225), respectively. **F:** DrugScore^{HotSpot} analysis of the SAP-benzamidine complex; benzamidine is displayed in light brown and the interacting residues Asp32, Asp218, Asp86 and Thr221 as white stick model. Hotspots are displayed in mesh; aromatic atoms at aromatic atoms (black) and donor atoms (blue) at a map level of 52 % (-72000/ -139181) and 70 % (-108000/ -153264), respectively.

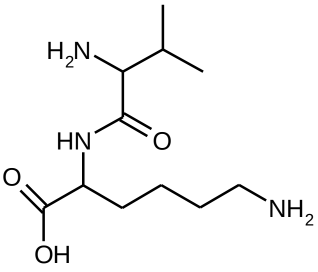
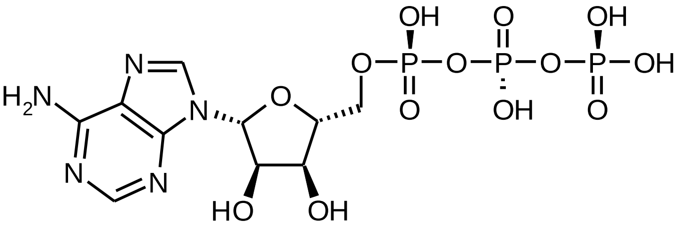
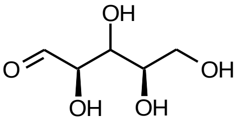
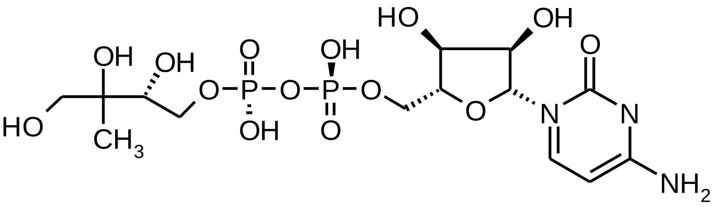
1,2-Propanediol: In the case of 1,2-propanediol as probe molecule maps were calculated for hydrophobic (gray) and for doneptor atoms (yellow). As indicated in Figure 5.8 B – D, in all cases the predicted favorable region for a hydrophobic atom (gray) agrees with the experimentally observed poses and some of the calculated doneptor maps (yellow) coincide with the observed positions of hydroxyl groups on our probe molecule.

Benzamidine: In the case of benzamidine as probe molecule the contour maps for aromatic carbon atoms (black) and donor groups (blue) have been calculated for ETP and SAP2. In both cases, two benzamidine molecules are present in the binding pocket. One of the molecules interacts with the catalytic dyad. For both proteins the calculated donor map fits remarkably well to the position of the polar nitrogens. The binding area of the aromatic moiety is reasonably well indicated in the maps (Figure 5.8 E, F). Interestingly enough the map of the aromatic atom probe predicts the hydrophobic portion of the bound MPD molecule very well.

5.3.3 Superposition with larger ligands in related crystal structures

To estimate how well the detected probe molecules actually reflect the binding of similar portions in much larger ligands, we compared the observed probe poses with crystallographically studied ligands in the respective enzymes. We elucidated whether position and orientation of the found probes match with functionalities embedded into larger scaffolds. The considered ligands are listed in Table 5.2.

Table 5.2 Schematic structures of reference ligands

Abbreviation	**Structure**
Val-Lys (TLN)	
ATP (PKA)	
D-Xylose (DXI)	
CDPME (IspD)	

Thermolysin: We superimposed the probe molecules aniline, *N*-methylurea and 1,2-propanediol in TLN with the dipeptide Val-Lys bound to the enzyme (PDB code 8TLN).[108] The aromatic ring system of aniline and the hydrophobic Val-side chain are located in similar position in the hydrophobic S1' pocket (Figure 5.9 A). Occupation of this pocket by hydrophobic ligand portions is quite essential for potent TLN binding.[109] More interestingly, the binding patterns of the *N*-terminus of Val-Lys (light-blue) as well as the NH_2 groups of aniline (green) and *N*-methylurea (violet) agree well in space and deviate by only 1.1 Å. They all adopt the same orientation and form hydrogen bonds to Glu143, Ala113 and Asn112 (Figure 5.9 A). Moreover, the position of the first hydroxyl group of 1,2-propanediol falls next to the carbonyl group of the Val residue.

Both form H-bonds to the guanidinium function of Arg203. The second hydroxyl oxygen superimposes with the *N*-terminus and binds as this group to Glu143 and Ala113 (Figure 5.9 B).

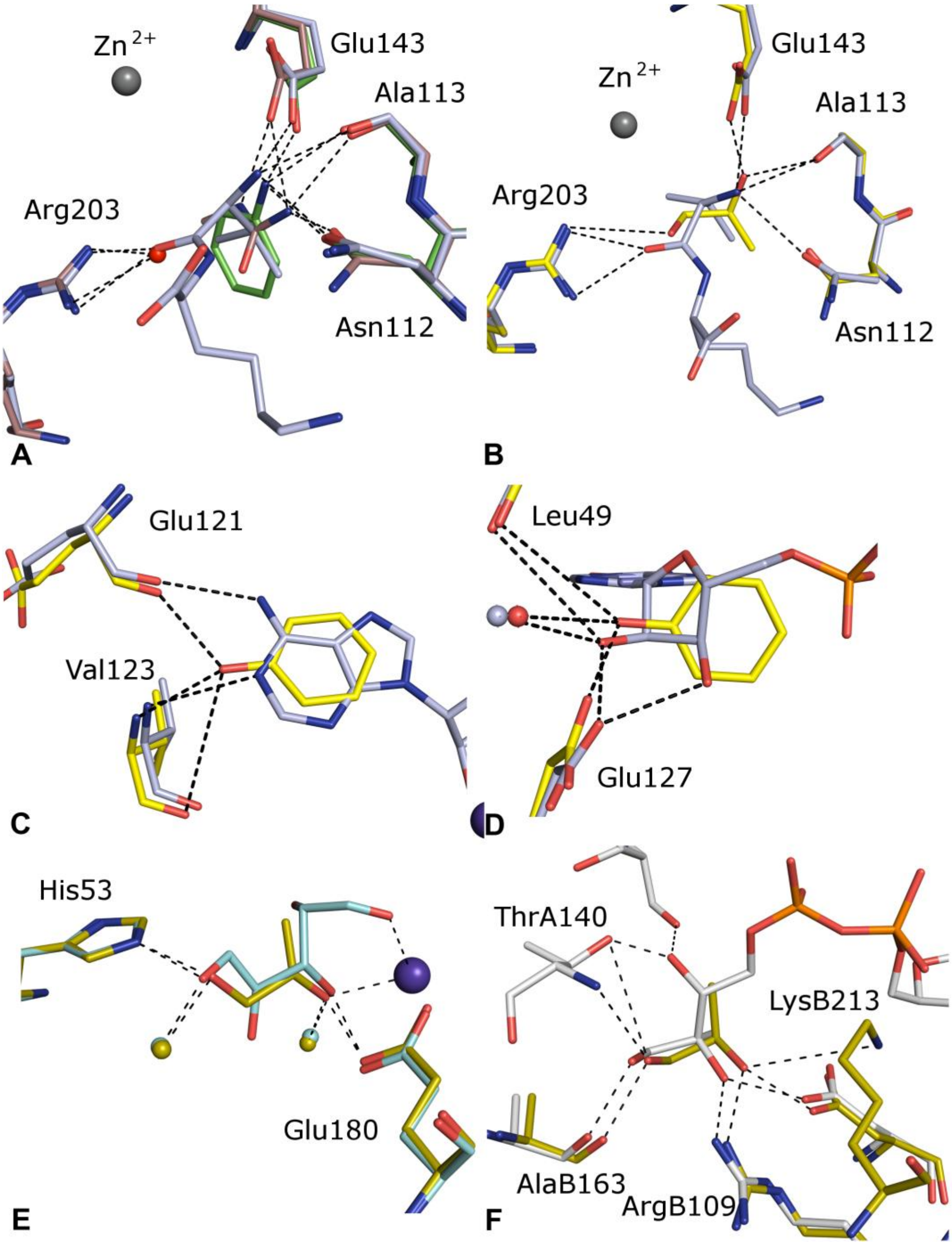

Zn 2+
Glu143
Ala113
Arg203
Asn112
A
Zn 2+
Glu143
Ala113
Arg203
Asn112
B
Glu121
Val123
C
Leu49
Glu127
D
His53
Glu180
E
ThrA140
LysB213
AlaB163
ArgB109
F

Figure 5.9 Superposition of **A:** aniline (green) and *N*-methylurea (violet) and **B:** S-1,2-propanediol (yellow) with Val-Lys (light-blue) in TLN. Aniline, *N*-methylurea, S-1,2-propanediol as well as the side chain amino acid residues are displayed as stick models. Amino acid residues from the aniline complex are displayed in green, from the *N*-methylurea complex in violet, the S-1,2-propanediol complex in yellow and the Val-Lys complex in light-blue. In all cases, oxygen atoms are depicted in red and nitrogen atoms in blue. Dashes symbolize the interaction between one molecule and the respective partner.

Superposition of **C:** phenol (yellow) and ATP (light-blue) at the hinge region of PKA and **D:** in the sugar pocket of PKA. ATP and phenol as well as the side chain amino acid residues are displayed as stick models in corresponding colors. **E:** Superposition of S-1,2-propanediol (yellow) and the linear form of D-xylose (light-blue) in the active site of DXI. Both the linear form of D-xylose and S-1,2-propanediol as well as the side chain amino acid residues are displayed as stick models in corresponding colors. **F:** Superposition of R-1,2-propanediol (yellow) and CDPME (white) in the active site of IspD. Both CDPME and R-1,2-propanediol as well as the side chain amino acid residues are displayed as stick models in corresponding colors.

Protein Kinase A: In case of PKA two phenol molecules were detected in the active site. In Figure 5.9 C these fragment structures have been superimposed with a complex of the substrate ATP (PDB code 1ATP).[110] N1 and the exocyclic 6-NH_2 group of the purine moiety of ATP (light-blue) and the hydroxyl group of phenol (yellow) address the hinge region Val123 and Glu121 in similar fashion (Figure 5.9 C). The phenyl portion of phenol also matches the hydrophobic centre of the purine ring system.

The position of the second phenol coincides with the ribose sugar moiety of ATP (Figure 5.9 D). The hydroxyl groups at the 2-position of the five-membered ribose ring and of phenol form H-bonds to Glu127 at very similar sites. Additionally, the hydroxyl functionalities of the ribose and phenol interact with a water molecule found in both complexes.

4-Diphosphocytidyl-2C-methyl-D-erythritol synthase and D-xylose-isomerase: Both in DXI and IspD, 1,2-propanediol molecules were detected. Their binding poses were compared with the substrate complexes of DXI and IspD. In case of DXI, the complex hosting the open-chain form of D-xylose (PDB code: 1XIC) was superimposed with our S-1,2-propanediol complex.[111] The positions of both OH groups of the latter match perfectly well in position and orientation with the first and third hydroxyl group of open-chain D-xylose (Figure 5.9 E). They show very similar binding patterns. The first OH group of S-1,2-propanediol (yellow) and D-xylose (light-blue) form hydrogen bonds to His53 and one water molecule. The second OH group of the probe molecule and third OH group of D-xylose built-up a hydrogen-bond network to Glu180 and an additional water molecule mediates a contact to one of the catalytic manganese ions in the active site.

For IspD, the R-1,2-propanediol complex was compared with that of the substrate CDP-methylerythritol (PDB code 1INI).[112] The R-1,2-propanediol molecule (yellow) matches perfectly well with the terminal OH group of methylerythritole (white) (Figure 5.9 F). Both hydroxyl functionalities and even the methyl group coincide and both hydroxyl groups form hydrogen bonds to the carboxylate of AspB106, one guanidinium NH of ArgB109, the backbone carbonyl of AlaB136, the terminal NH_2 of LysB213, and the backbone NH of ThrA140.

5.4 Conclusion

In this study, we demonstrated that different highly soluble molecular probes with molecular weight much lower than standard fragment size can bind to proteins originating from different classes: for TLN, PKA, DXI, IspD, ETA, and SAP2 probe fragment complexes with phenol, aniline, urea, *N*-methylurea, 2-bromoacetate, 1,2-propanediol and benzamidine could be determined. Furthermore, we could find evidence that the positions of the polar H-bond forming functional groups and hydrophobic portions of these molecules are in good agreement with *in silico* hot spot predictions using DrugScore. Remarkably, the observed poses of the probe molecules coincide well in space and interaction pattern with similar molecular portions embedded in much larger ligands for which complex crystal structures have been determined. This observation confirms the fundamental hypothesis of fragment-based lead discovery that binding poses even of very small molecular probes do not strongly deviate or move once a ligand is grown further into the binding site. This underscores that these probes actually populate at a given hotspot of binding. In that respect the small probes can be regarded as relevant seeds for further design. This opens a very promising prospective to also use the binding poses of probes such as phenol, aniline, urea, *N*-methylurea, 2-bromoacetate, 1,2-propanediol or benzamidine as convenient starting points of fragment-like de novo design.

The hot spot prediction by DrugScore has significantly improved with the new atom-type setting. Very fast to calculate, DrugScore can be consulted for further design to subsequently grow small initial seeds into the binding pocket by attaching further substituents in agreement with the predicted hot spots.

5.5 Materials and Methods

DrugScore^{HotSpots}

For the calculation of hotspots for all protein-fragment-complexes the respective enzyme structure was used. The fragment and water molecules were not considered during hotspot analyses in order to derive sites of favorable interactions with the non-complexed binding pocket. The atom-type assignment is based on the classification in DrugScore^{CSD}.[113]

6 From Probe to Fragment and Lead: A Combined Approach of Experimental Fragment Screening and Computational De Novo Design

6.1 Introductory Remarks

This chapter is prepared to be submitted to a scientific journal. Accordingly, some redundancies with the Introduction to this thesis might occur. The study was accomplished together with Helene Köster, Dr. Sascha Brass, Hans-Dieter Gerber and Barbara Wienen (Univ. Marburg) in cooperation with Merck Serono in Darmstadt in a joint collaboration funded by the BioChancePlus BMBF-project FragScreen. My contribution to this study was the design and development of the compounds in the three subsequent cycles.

6.2 Introduction

The more novel and interesting target structures for a therapeutic treatment are discovered and validated via proteomics and structural genomics, the higher will be the demand for efficient strategies to discover leads that interfere with protein function. In the nineties high-throughput screening and combinatorial chemistry have been established to resolve the bottleneck for an efficient lead discovery. Subsequently, virtual computer screening has been added as an alternative to complement these approaches. However, success rates were not as expected and the size of the usually discovered hits was in the range of common drug molecules not leaving much space for optimization without significantly exceeding the molecular weight limit of perhaps 500 Da.[1] To better rank the size of the discovered hits with their actual potency, the concept of ligand efficiency[2] was introduced. Highly efficient leads exhibit good potency combined with low molecular weight.

Improvements in biophysical techniques to record protein-ligand binding and increasing success in the acquisition of structural information about protein-ligand complexes by crystallography or NMR spectroscopy allowed to push the limits of the compounds to be screened to lower and lower molecular weight. Particularly NMR and SPR are nowadays powerful enough to detect very small and weak binders which still exceed good and convincing potency ("high" ligand efficiency). Such fragment-type[11] leads appear as special challenge as they provide – once characterized in terms of a crystal structure – wide opportunities for optimization into prospective drug candidates. Meanwhile, many examples have been described in literature and an impressive number of reviews have been written.[13-24] These examples, mostly performed in industry or specialized small biotech companies, focused on the following targets: HSP90[25-30], different kinases[31-38], phosphatases[39-41], antibacterial and anti-infective targets[42-46] , BCL-2[47] and BACE.[47-50]

The wish to push towards smaller and smaller initial "leads" reminds of an old concept developed in the early phase of structure-based drug design. In the late eighties and early nineties computational de novo design was developed with much enthusiasm. Starting with a very small "seed", a ligand was gradually grown into the binding pocket of the target protein. Even though received with much sympathy in the beginning, the approach rapidly got out of fashion. This was either attributed to its high complexity and too low success rate or it was simply overrun by the upcoming high throughput technologies; and we clearly witness at present that also fragment screening moves into the high throughput domain.

Nevertheless, the appealing aspect of de novo design was its rigorous rational concept, only once ligand binding is fully understood this kind of design based on first principles can work. However, two decades later we have to confess that the binding process is still little understood and therefore a purely rational approach remains as such much too ambiguous.

On the other hand, meanwhile our experimental techniques are much better developed and give fast access to structural information. Therefore, it may be asked whether a combination of de novo design supported by multiple crystal structure analyses on small molecule fragments or molecular probes will allow for an alternative strategy in lead discovery, reconsidering much of the early concepts of de novo design.

Therefore, we picked a representative example from the heavily studied family of protein kinases to perform a feasibility study. Ongoing from small promiscuously binding probes, potent nanomolar leads were developed using the concepts of *de novo* design. The process was strongly supported by crystal structure analysis particularly in the early phase where the detection of ligand binding would be impossible applying routine screening techniques.

6.3 Results and Discussion

6.3.1 Probing the active site

We selected PKA as a well established model kinase as it is easily accessible, well to crystallize, and stable enough for many biophysical investigations. Also in the early nineties, Dagmar Ringe and Greg Petsko brought up a concept of soaking small molecule probes, mostly solvent molecules, into protein binding sites. Structures determined with these probes can be understood in the sense of a de novo design approach as initial "seeds" to grow from there to putative leads. We picked up this concept and searched for molecular probes that can be applied rather generally to many protein binding sites to probe their binding properties. This approach can be compared with a kind of "experimental active-site mapping" as usually performed computationally by programs such as GRID[70], MCSS[114], DrugScore[115] or Super Star[71].

Among the most prospective probes we found phenol to access at least in our screens a fairly large number of proteins (Chapter 5). Apart from these, also water can be seen as such a "promiscuous" probe. Stimulated by these results we also tried to crystallize PKA with phenol. A screening using an SPR-based binding assay did not succeed in detecting any binding of this probe.

However, the crystal structure of PKA with phenol shows three phenol molecules to be bound (Figure 6.1). Aside of two phenol molecules which access the ATP binding site of the protein, a third one is found in a remote site on the protein surface. The first phenol probe addresses the hinge region adopting a binding mode which allows the hydroxyl group to operate as H-bond donor towards the backbone carbonyl of Glu121 and as an acceptor towards the backbone NH of Val123. This type of interaction pattern is common for many potent kinase inhibitors. The second phenol is twisted by 90° and occupies the sugar subpocket. Its hydroxyl group interacts with Glu127. The benzyl moiety occupied an area directly under the glycine rich loop (G-loop). This usually very flexible loop is fixed by two hydrogen bonds built between the backbone NH and the hydroxyl group of Ser53 and the carboxylate of Asp184. In literature, three major conformational families and thus overall different positions of the glycine rich loop have been described. The first is represented by the APO structure (pdbcode:1J3H[116]); here the loop is in totally open geometry. The second is observed with some hinge binders (e.g. pdb code: 1JLU[117]); here the loop is in a medium open conformation. The third one is observed in the PKA-ATP (pdb code 1L3R[118],1ATP[119]) complex structure, here the glycine rich loop interacts with the phosphate groups of ATP and is in closed state.[120]

The third phenol is found on top of the G-loop outside the catalytic center and could possibly describe a potential new allosteric site (Figure 6.1).

To complement this experimental active site mapping by computational means, we applied our recently extended DrugScore^{CSD} mapping approach (Neudert & Klebe, to be published) to the PKA active site. Using an aromatic carbon and a phenolic hydroxyl group as probes, the method quite nicely highlights those areas actually found to be accommodated by the phenol molecules (Figure 6.1).

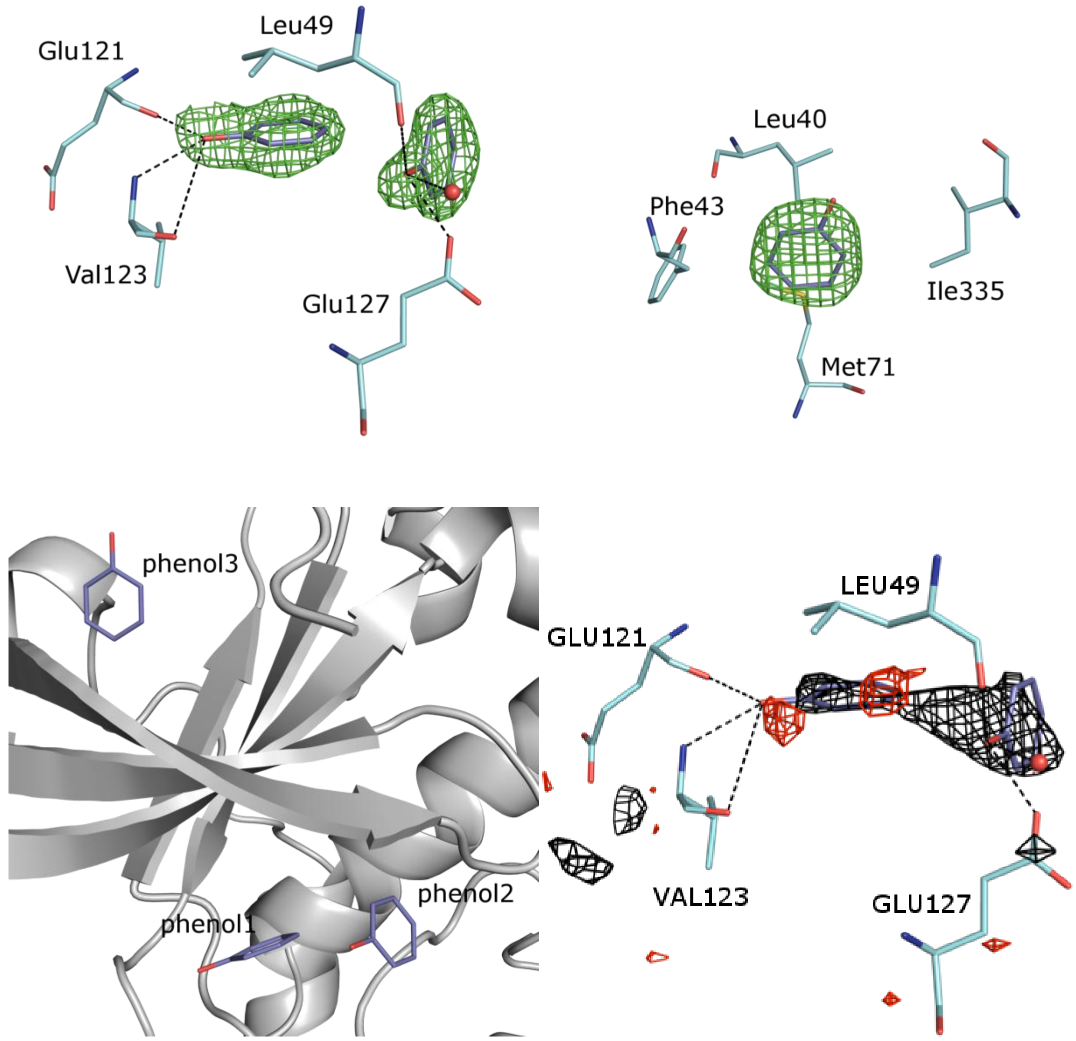


Figure 6.1 Top left: The F_o-F_c difference electron density map for the two phenols in the ATP binding; Top right: The single phenol molecule on top of the G-loop; contouring at σ-level 2.0; Bottom left: An overview of the phenol molecules with the kinase folding pattern; Bottom right: The calculated DrugScore contour maps of C.ar and O.ph of the PKA-phenol complex. Hotspots are displayed in mesh; C.ar6 (black) and O.ph (red) at a map level of 56 % (- 14000/ -25180) and 30 % (- 2800/ -9326), respectively.

Based on our initial phenol seed we wanted to consult the computer to detect better suited and appropriately decorated phenol derivatives as putative fragment leads in order to keep the furcated H-bonded binding mode of the phenol molecule towards the hinge region.

To obtain a broad scope of possible candidates we performed a targeted virtual screening. The test library was compiled from about 4,000 commercially available fragment-like candidates that exhibited one ring portion with several decorations. All fragments satisfied the ‘rule of three’ criteria as defined by Congreve et al..[11]

FlexX[121] was used for docking using the binding mode observed for the hinge binding phenol and further guided by a pharmacophore derived from the DrugScore maps. Among the best scored hits were several phenol derivatives. Confirming their binding to PKA by SPR failed, indicating that the affinity was clearly beyond 2 mM. However, with 2-methyl-4-acetylphenol we succeeded in obtaining a PKA complex crystal structure. In this complex only one phenol moiety is bound. It addresses, similarly to the unsubstituted phenol, the hinge region, places, as expected, its methyl group in a hydrophobic niche, and picks up via its carbonyl group of the acetyl substituent a water network in-between Thr183 and Lys72. Two water molecules remain as interstitial bridges in this region and mediate contact between the fragment and the protein. For this complex an open G-loop position is obtained (Figure 6.2).

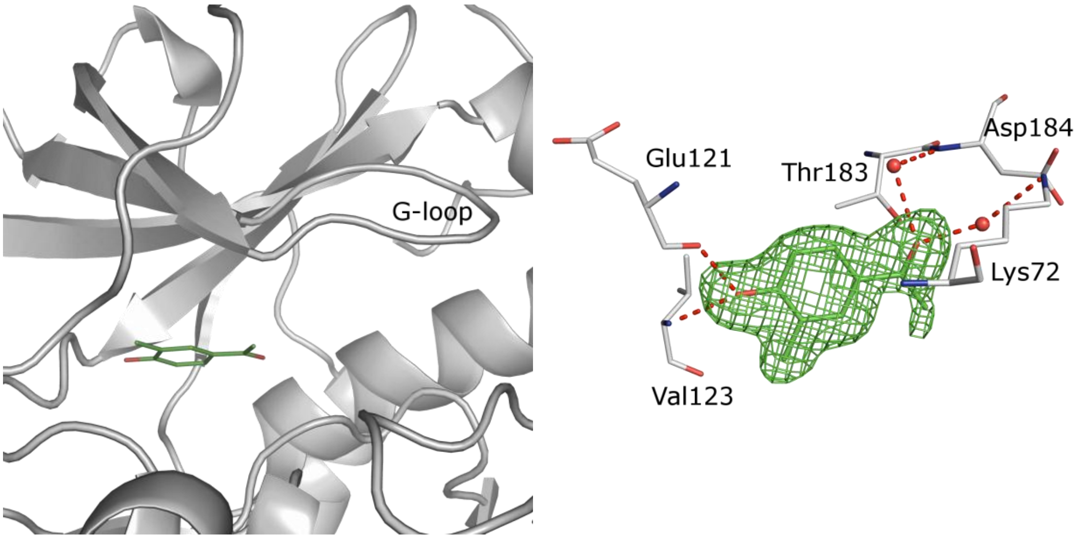


Figure 6.2 The virtual screen hit is shown; Left: The binding pocket of PKA in complex with the virtual screen hit; Right: The F_o-F_c electron density map of the virtual screening hit is presented in green mesh at a σ-level of 2.0.

Structural expansion of the initial seed

Next, we consulted DrugScore[CSD] [115] again to suggest additional interactions to be picked up in our further design. Interestingly enough, an area remote from the water cluster, clamped from the top by the hydrophobic faces of the peptide backbone at the tip of the G-loop (residues 51-53), from the far end of a small pocket by the aromatic phenyl portion of Phe54, a favorable binding region for an aromatic ligand moiety is suggested as well as an H-bond acceptor functionality is proposed to be placed under the G-loop. Another favorable position for an acceptor motif is suggested near Lys72.

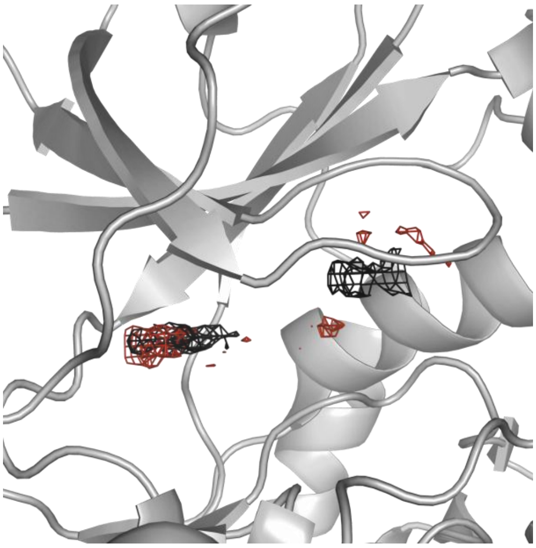


Figure 6.3 The HotSpotsX calculated maps for aromatic interactions (black contours) at a map level of 52 % (-100000/ -193546); and O.co2 as an acceptor representative (red contours) at a map level of 32 % (-7500/ -23475) above the minimum of the map is shown.

6.3.2 Design of a putative lead skeleton and synthesis strategy

The composite picture of the different seed and fragment structures along with our experimental and computational active-site mapping can be summarized in the following design hypothesis. A phenol moiety decorated *ortho* to the OH group by a hydrophobic expansion (methyl, chlorine, methoxy, fused phenyl ring) has to be expanded in *para* position by a chain of about four members to bridge the gap towards the indicated aromatic moiety. The terminus of the linking chain should be decorated with a carbonyl group to interact with Lys72 in a similar fashion as the water molecule in the virtual screening hit and to coincide with the acceptor hot spot suggested by the DrugScore calculations. Furthermore, this linker should be the attachment point for an aromatic substituent to fill the cavity clamped by the peptide stretch and next to Phe54 as indicated by DrugScore. The linker itself should be selected in a way to pick up the water network observed towards the side opposing the crevice formed by Thr183.

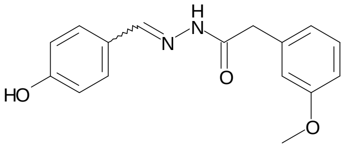
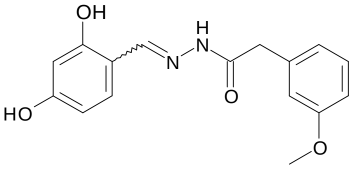
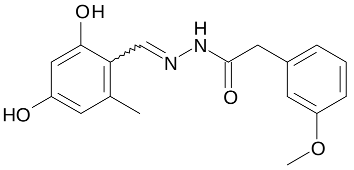
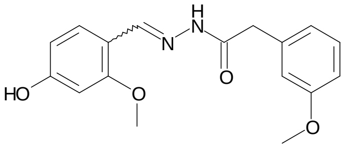
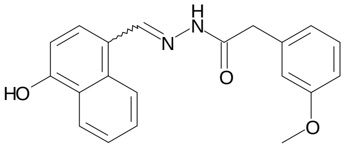
The described binding pattern has to be matched with an appropriate chemistry keeping synthetic accessibility and availability of a broad range of building blocks in mind. As central moiety in the linker, we went for substituted hydrazones as these derivatives can easily be made and the formed central group provides the correct spacer length and the required patterns of H-bond donor and acceptor facilities. The portion addressing the remote hydrophobic site and Thr183/Lys72 can then be represented by a broad range of phenyl acetic acid derivatives, as the hydrazone unit is introduced via hydrazine replacing the OH functionality at the carboxylate group of the latter derivatives. The produced mono-substituted hydrazide can then be reacted to the desired hydrazones with various 4-hydroxybenzaldehydes to introduce, with appropriate linker-geometry, the phenolic portion competent to address the hinge region.

The selection and assemble of candidates for the first round of synthesis followed our recently described KNOBLE[122] approach. About 200 putative candidates for synthesis were assembled in the computer and docked with FlexX against the PKA active site. Scoring the docking results suggested the following five molecules (Table 6.1) as first candidates for synthesis.

6.3.3 Validation of the first leads

Subsequent to synthesis, the binding potency of our first lead candidates was tested by an SPR assay. Binding in the range of 6.5 to 120 µM was recorded.

Table 6.1 The structure of the five synthesized compounds of the first synthesis cycle with the K_d and the ligand efficiency is shown.

Name	**Structure**	**K_d** [µM]	**LE** [kcal/mol*heavy atom]
1		70	0.27
2		35	0.28
3		6.5	0.31
4		>100	
5		15	0.26

We succeeded to determine the crystal structures with two ligands (compound 1, 3) from the first synthesis campaign. As only difference they experience a deviating substitution pattern at the aromatic ring addressing the hinge region. The derivative lacking any substituent in *ortho* position next to the hydroxyl group (70 µM) binds very similarly to the initial phenyl probe and the virtual screening fragment hit. An additional *meta* substitution to the 1-hydroxyl group seems to increase the affinity by a factor of two. Compound 3 forms, via the second OH group, an additional contact to a water molecule also interacting with Thr183. The methyl group at the other *meta* position seems ideal to fill the hydrophobic niche next to the hinge region. This additional substitution leads to a more potent derivative (6.5 µM). A methoxy substituent at this position (compound 4) seems detrimental to binding (>100 µM). Computer modeling suggests that this group cannot adopt the preferred coplanar arrangement with the phenyl portion. Most likely this is due to its more demanding steric bulk in the small hydrophobic niche and deviation from planarity costs some price in affinity. Modeling also suggests that the naphthyl derivative (compound 5, 15 µM) is to big to perfectly occupy the hydrophobic niche in the lower part of the hinge region.

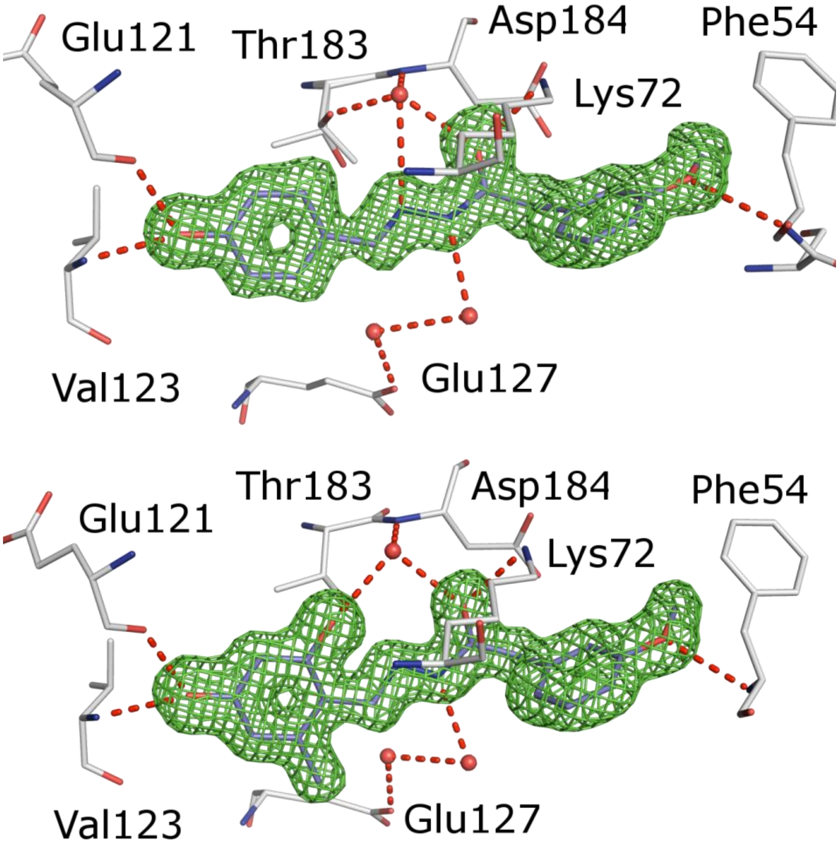


Figure 6.4 Top: The crystal structure of our initial lead compound 1 and bottom: Crystal structures of compound 3 from the fist synthesis cycle. Protein carbons are colored white, ligand carbons violet; oxygen in red; nitrogen in blue. The difference electron density is shown as green mesh at a σ level of 3.0.

Apart from these deviating interactions of the structurally varying phenolic portion, the unchanged remaining part of the two ligands binds to the protein in a very similar way and conforms quite closely to our design hypothesis. The NH donor group of the hydrazone forms, as expected, an H-bond to the conserved water molecule and the carbonyl oxygen picks up an interaction pattern with Thr183 and Lys72. The second aromatic portion binding next to the glycine-rich loop and nicely fills the area indicated as preferred for aromatic carbons by DrugScore^{CSD} (Figure 6.5). The convincing predictive power of the novel CSD-based DrugScore potentials can be demonstrated by superimposing the previously calculated DrugScrore maps with subsequently determind crytal structure of compound 1. The contours encompass the ligand with nearly ideal shape (Figure 6.5).

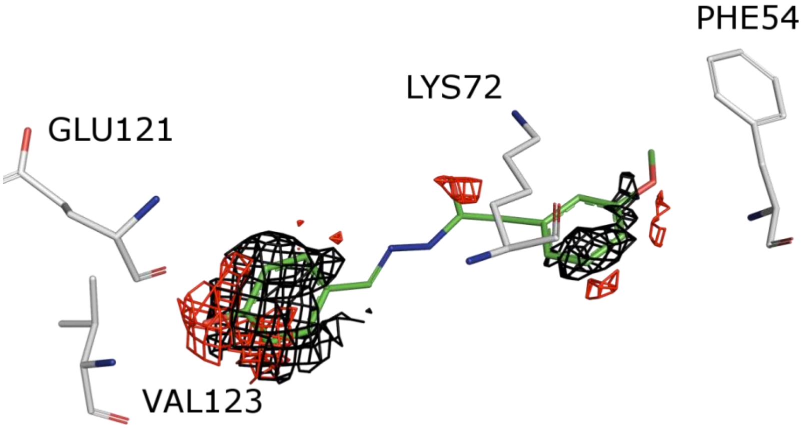


Figure 6.5 Calculated hotspots and the initial lead compound are shown. The HotSpotsX calculated maps for aromatic interactions (black contours) at a map level of 52 % (-100000/ -193546); and O.co2 as an acceptor representative (red contours) at a map level of 32 % (-7500/ -23475) above the minimum of the map is shown.

In this first synthesis cycle the affinity could be increased towards single-digit micromolar range and further more the ligand efficiency could be increased from initially 0.27 to 0.33 kcal/mol*heavy atom.

6.3.4 Extension of the first leads

To increase potency we anticipated to introduce a newly formed salt bridge with the protein. To achieve this goal the scaffold would have to be expanded by a primary amino group. This basic motif can form a salt bridge towards the carboxylate of the adjacent Asp184 residue. Figure 6.6 displays the intended salt bridge.

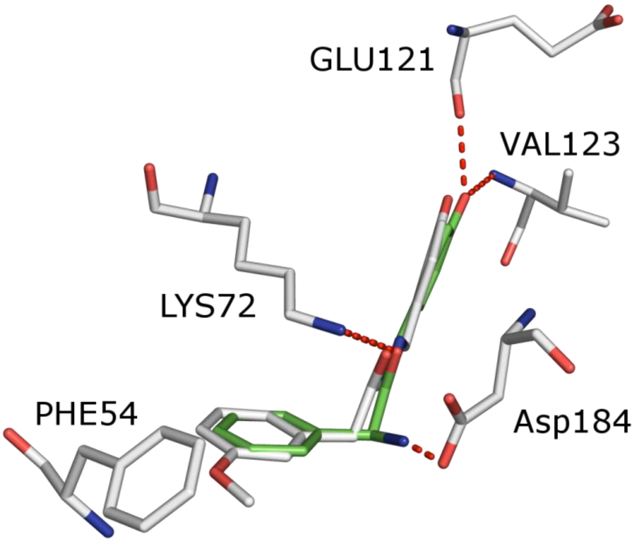
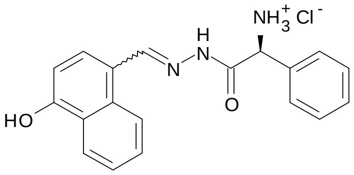
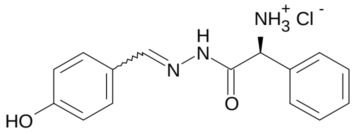
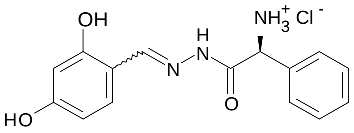
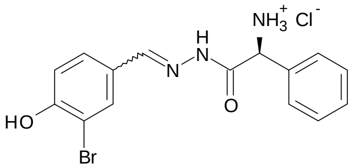
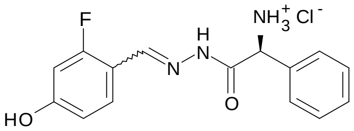
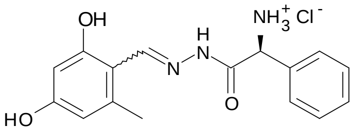


Figure 6.6 The new design hypothesis of gaining affinity through the formation of a salt bridge between the newly attached amino group of the ligand and Asp184 is shown.

For the hinge binder with the most promising substitution pattern the amino derivatives have been synthesized. Introducing the additional amino function to the lead structure creates a novel stereo center. The docking simulation with FlexX was consulted to predict the required stereochemistry. The docking tool suggests the S-phenylglycine derivatives as the active ones. Since synthesis required introduction of an α-amino acid, availability of chiral starting materials was crucial. Therefore, both stereo isomers of the unsubstituted phenylglycine were considered. In the next synthesis round we omitted the *meta*-methoxy group at the phenyl acetic acid compounds. Thus, enantiomerically pure compounds could be used as commercially available starting material.

Table 6.2 displays the affinity data determined by SPR for the second synthesis cycle. Unfortunately, the affinity could not be enhanced as expected. The affinity of the synthesized candidates remained in the one-digit micromolar range. As a direct comparison of similarly substituted derivatives shows, the affinity of, e.g. compound 1 with respect to compound 7 could be increased from 70 µM to 3.4 µM and simultaneously the LE improved from 0.27 to 0.38. Equally the affinity of compound 2 and compound 8 could be increased from 35 µM to 2.4 µM and the LE enhanced from 0.28 to 0.37. For the compounds pair 3/11 potency is nearly similar 6.5 µM and >10 µM. As LE could be increased in the second synthesis cycle in almost all cases, introduction of an amino group appeared quite promising even though affinity could not be enhanced as expected.

Table 6.2 Affinity data achieved for the derivatives of the second synthesis cycle.

Name	**Structure**	$\mathbf{K_d}$ [µM]	**LE** [kcal/mol*heavy atom]
6		3.8	0.31
7		3.4	0.38
8		2.4	0.37
9		1.8	0.38
10		>10*	
11		>10*	

* did not reach saturation state

Fortunately, crystal structure determination of two examples of the second synthesis cycle was successful. Figure 6.7 displays the crystal structures with compounds 9 and 11. Both ligands address similarly to the initial seed the hinge region. The introduced amino group builds, as expected, the newly formed salt bridge with Asp184 (Figure 6.7).

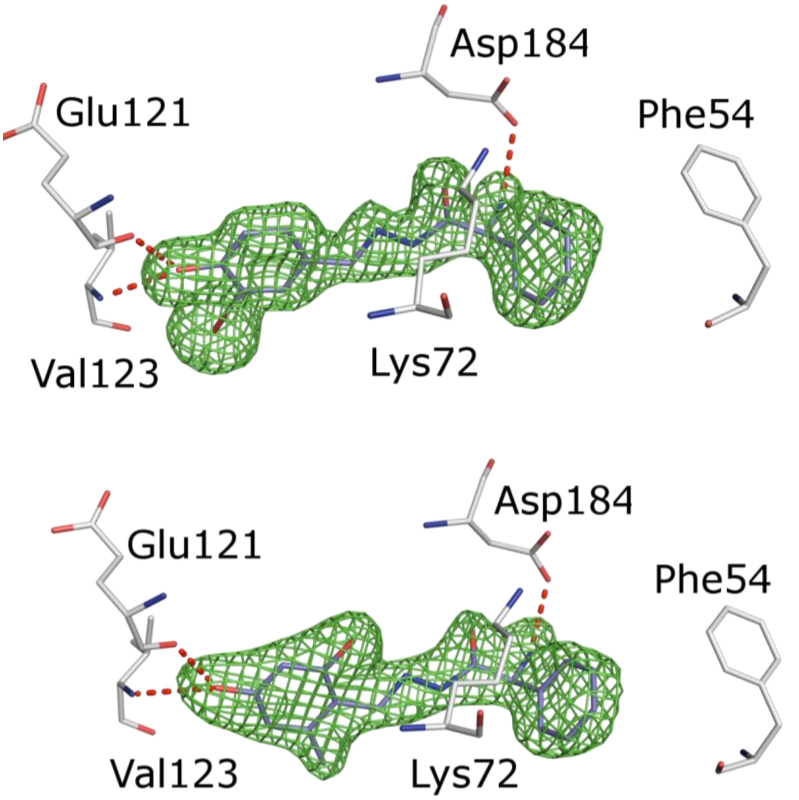


Figure 6.7 Top: crystal structure of compound 9 of the second synthesis cycle. Bottom: Crystal structure of compound 11.

With respect to the G-loop region of the binding pocket, the ligands of the first and the second synthesis round differ slightly in orientation (Figure 6.8). As a result of the newly formed salt bridge, the ligands of the second round are dragged towards Asp184. Thus, the initially designed H-bond between the carbonyl functionality and Lys72 is ruptured. The distance in the first series amounts to 2.8 Å (Figure 6.8 top, dashed red line) and the distance in the second series increases to 4.2 Å (Figure 6.8 bottom, dashed blue line).

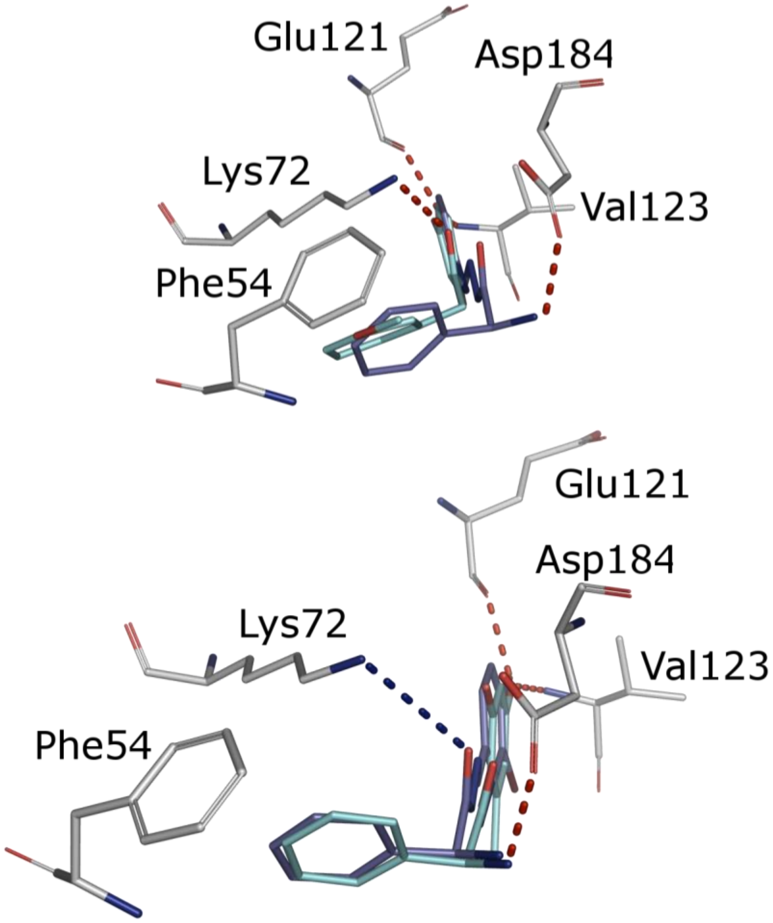


Figure 6.8 Top: Crystal structure of compounds 1 (carbons light blue) and 9 (carbons violet), Bottom: Crystal structure of compounds 9 and 11 (light blue). Protein carbons are colored white; oxygen red and nitrogen in blue.

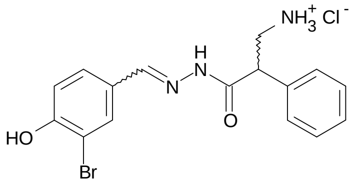
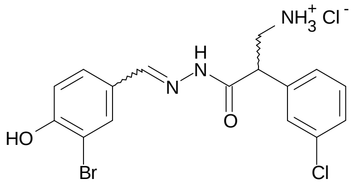
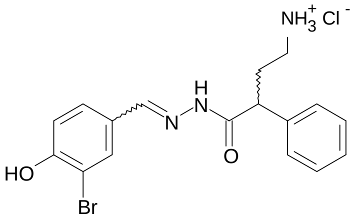
For the compounds of the first cycle we expected some of the potency to result from the H-bond formed by their carbonyl functionality with Lys72 and from the aromatic interactions of the benzene moiety of the former phenyl acetic acid component. For the derivatives of the second cycle the H-bond between Lys72 and this carbonyl group is lost and the aromatic group is partly pushed out of its position as a price to form the new salt bridge. Anyhow, we could gain LE in the second design cycle, suggesting a more favorable contribution of the salt bridge compared to the C=O --- Lys72 H-bond and the geometrically slightly better contacts of the aromatic portion.

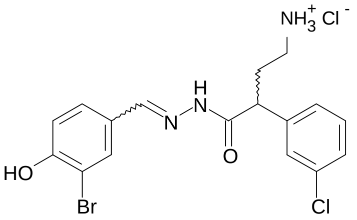
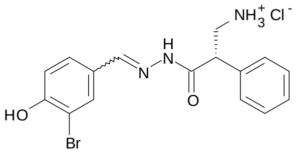
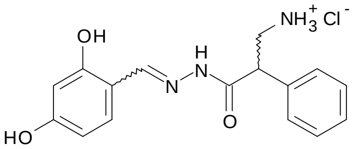
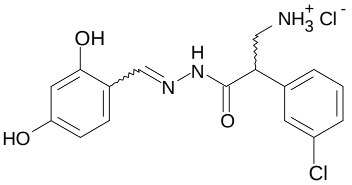
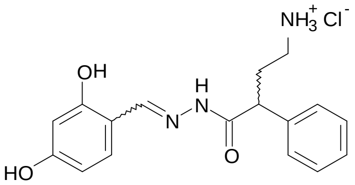
To combine all interactions in a more favorable geometry, we initiated a third design cycle. We planned to introduce a spacer of one and two CH_2 groups to place the amino group in a more optimal position, still allowing the favorable geometry of the remaining skeleton as observed for our leads of the first design cycle.

6.3.5 Introduction of a spacer

In our third synthetic cycle we concentrated on the most promising hinge portions, the *ortho* Br-phenol or *meta* hydroxyphenol scaffolds. The phenyl acetic acid portion of our lead series, anticipated to occupy the back part of the pocket, has not been further optimized. The contribution of a methoxy substituent as investigated in our first series appeared not very promising. Therefore, we now considered a *meta*-chloro substituent as it appears for structural reasons ideally placed to form a chlorine-π interaction with Phe54. Table 6.3 summarizes the affinity data found for the derivatives of the third cycle.

Table 6.3 Affinity data achieved for the derivatives of the third synthetic cycle.

Name	**Structure**	**Ki** [µM]	**LE** [kcal/mol*heavy atom]
12		0.23	0.41
13		0.11	0.41
14		0.36	0.38

15		0.20	0.38
16		10.89	0.31
17		0.85	0.36
18		0.33	0.39
19		0.87	0.38

The affinity increases for both, the compounds with the one- and two-membered methylene spacer. In the second cycle we introduced α-amino acids which are well accessible as enantiopure starting materials. Thus only E/Z isomers have to be regarded which were always produced at the hydrazone linker. In the third cycle we could not start as easily with similar enantiopure material to introduce the β- and γ- amino acids as they are not commercially available. The synthesis created therefore both stereo isomers at the introduced chiral center along with E/Z isomerism. Due to the mixture of four compounds being measured in the assay, some unspecific binding might be overlaid and therefore, it is important to be careful with the interpretation of the assay values.

Nevertheless, for compound 12 with a single CH_2 spacer the affinity could be enhanced by a factor of 7 to 230 nM. The additional *meta*-chloro substituent in compound 13 further improved potency to 111 nM.

The derivatives with a two-membered CH_2 spacer exhibit affinities of 359 nM for compound 14 and 200 nM for the corresponding *meta*-chloro derivative compound 15. Obviously this design cycle provided the affinity boost as expected for the introduction of the amino group. As there was an enantiomerically pure β-amino acid commercially available lacking the chloro substituent, however showing the wrong stereo chemistry at the chiral center, we also synthesized this derivative for validation purposes. It loses affinity (compound 16 (11 µM) vs. compound 12 (0.23 µM)) and underlines the importance of correct stereochemistry. Apart from our R-/S- chirality differences we face in all derivatives E/Z isomerism giving rise to a compound ratio between 20-80 % to 50-50 % according to NMR assignment. Considering that the sample subjected to our assay is actually a mixture of four diastereomers and keeping in mind that inversion of the stereochemistry could result in an affinity loss of nearly two orders of magnitude, most likely the active stereoisomer from our diastereomeric mixture is a rather potent one- to two-digit nanomolar inhibitor.

We could determine structures with compounds 14 and 15 which also gave clarity about the stereochemistry of the most potent isomer. Figure 6.9 illustrates the binding mode of both compounds. The phenol moiety of compound 15 addresses, as in all examples, the hinge region and the carbonyl function interacts via an H-bond with Lys72; the amino function forms a salt bridge to Asp184 and an H-bond to Asn171; and the meta-chloro substituent interacts via a Cl-π stacking with Phe54. For compound 14, the phenol moiety addresses, similar to compound 15, the hinge region. The H-bond to Lys72 is also present. The amino function of the ligand shows in this case, however, two different conformations. One is similar to the conformation observed for compound 15. In the crystal structure this conformation is populated to 55 %. In the other conformation, the ligand amino function interacts with Thr51 instead of Asp184 and Asn171 (Figure 6.9 bottom) which shows the ligand amino side chain in particially eclipsed instead of gauche geometry. The second conformation and the competitive interaction with uncharged Asn171 might give an explanation why the compounds with one spacer atom show a slightly better affinity. The previously described water mediated interaction to Glu127 is also present in this case.

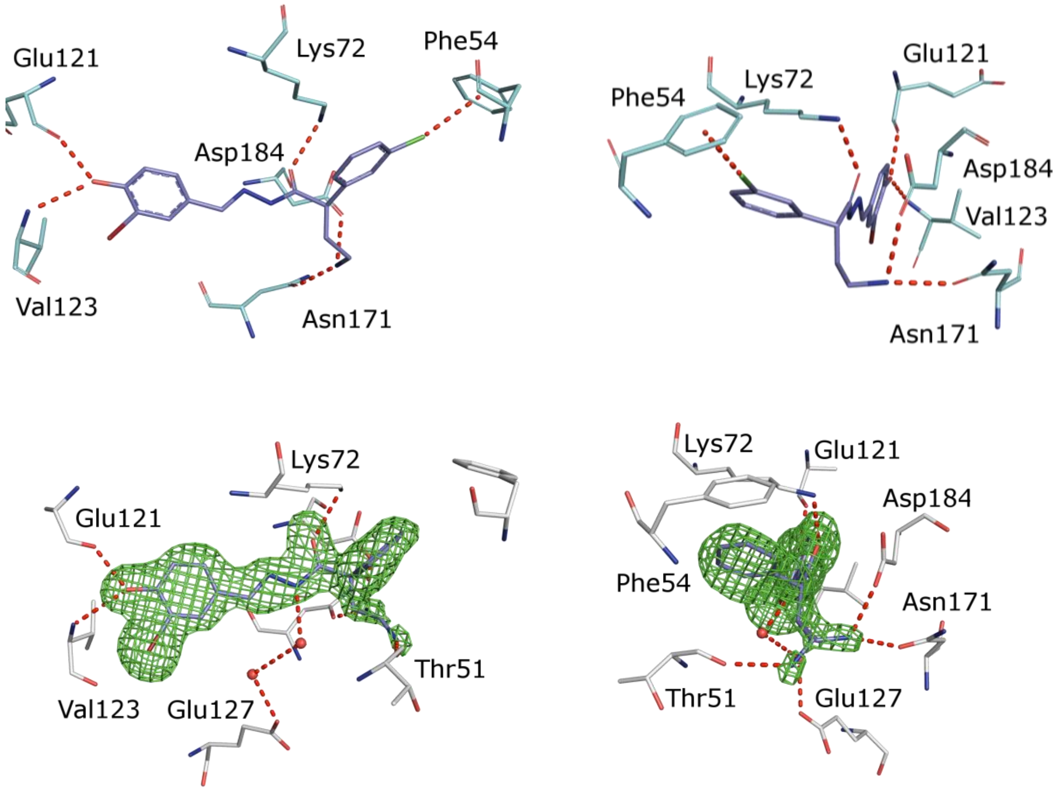


Figure 6.9 Top: Crystal structure of compound 15 in complex with PKA. This is a preliminary structure. Left and right picture are perpendicular. Bottom: Crystal structure of compound 14. Ligand carbon atoms are colored in violet; protein carbon in white; oxygen in red and nitrogen in blue. Electron density map is shown in green mesh at a σ-level of 3.0. Left and right picture are perpendicular.

Figure 6.10 illustrates a superposition of compounds 3, 9, 15 from the different design cycles. In all cases the phenol moiety interacts with the hinge region. For derivatives of the first and third cycles interactions between the ligand's carbonyl functions and Lys72 are well established. In the second cycle, this H-bond is largely expanded due to the newly formed salt bridge introduced here. The terminal aromatic portion is dragged out of the optimal position. In the third cycle, this aromatic portion returns back and superimposes remarkably well with the geometry observed for the first cycle. The amino functions of the molecules 9 and 15 resulting from the second and third cycle are not superimposed, as their amino groups are separated by different spacer length. They both interact with Asp184. For 15 we observed additionally an H-bond with Asn171.

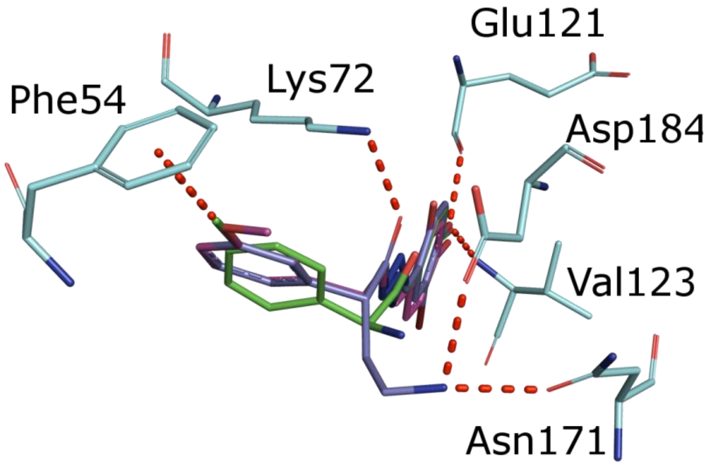


Figure 6.10 Superposition of compound 3 from the first cycle (carbon pink), compound 9 from the second cycle (carbons green) and compound 15 from the third cycle (carbon violet). The protein carbons are colored in light blue; oxygen in red and nitrogen in blue.

6.4 Conclusion

The experimentally detected probe molecule phenol has been used as a seed in combination with a computational HotSpots analysis as a starting point for a fragment-type *de novo* design study. We generated an initial lead with 70 µM potency. The affinity of this starting structure could be increased to nanomolar potency through three iterative design cycles. In the first cycle the affinity could be increased to 6.5 µM and LE from 0.27 to 0.31. In the second cycle the affinity could be enhanced only marginally, but LE could be improved to 0.38. The minor improvement resulted from a non-ideal geometry of the groups introduced to the lead skeleton. One previously introduced polar interaction was sacrificed to form a newly introduced second one. The redesigned compounds from the last synthesis round are now in the low nanomolar range. Definite affinity assignment is difficult, as the final products yielded by the applied synthesis scheme resulted in a mixture of four diasteroisomers. However, crystal structure analysis shows which stereoisomer is the most potent binder. Targeted synthesis of the expected less potent stereoisomer indicates an affinity loss of about two orders of magnitude and suggests strong dependence of binding affinity on the correct stereochemistry of the ligands.

7 Early Steps of G-loop opening for Protein Kinase A (PKA) in complex with Phenol

7.1 Introductory Remarks

The present study is based on a crystal structure of Protein Kinase A (PKA) in complex with phenol (PDB code: 3NX8). The analyzed structure is identical to the one already used as starting point in the design study presented in chapter 6 (page 93). This study was carried out in collaboration with Dr. Rafi Ahmad (The Norwegian Structural Biology Centre, University of Tromsø, Norway).

7.2 Introduction

Kinases, alternatively known as phosphotransferases, are enzymes which can phosphorylate target enzymes. Through this phosphorylation, the target enzyme is activated. The adversaries of kinases are the phosphatases which can remove phosphate groups from target enzymes. This dephosphorylation deactivates the target enzyme.

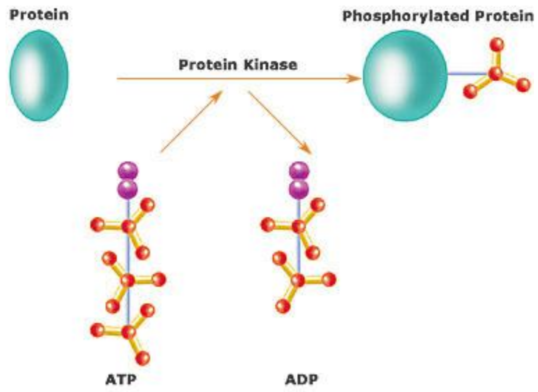


Figure 7.1 A schematic overview of the function of protein kinases is displayed.[123]

Usually the terminal phosphate group of ATP is transferred to the hydroxyl group of a serine, threonine or tyrosine of the target enzyme (Figure 7.1). Due to the universal applicability of this activation mechanism kinases are involved in many different physiological processes e.g. transcription[124] or apoptosis.[125] One kinase can activate different targets. Therefore, the three-dimensional structures of many kinases show strong adaptive properties. Figure 7.2 displays the overall structure of PKA colored by

B-factors. The three-dimensional structure consists of different loop regions. These regions are colored in red as they have large B-factors due to high residual mobility. The glycine rich loop (G-loop) forms the top of the ATP binding site and is known to be structurally rather mobile.

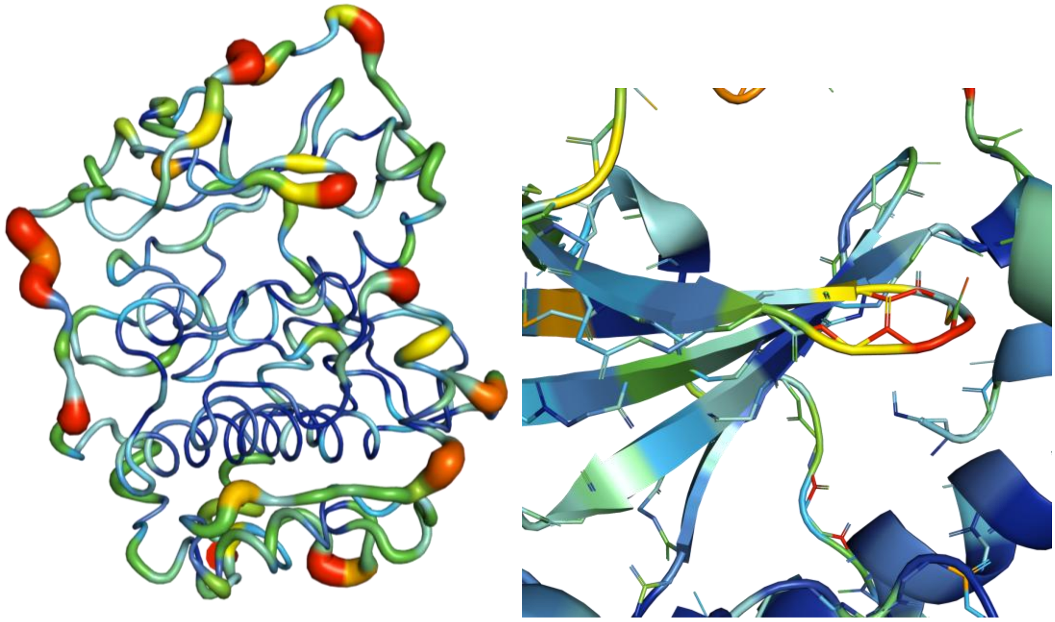


Figure 7.2 Left: The overall backbone structure of PKA is presented color-coded by B-factor. Regions colored in red are highly flexible and regions in blue are less flexible. Right: The G-loop region of PKA colored by B-factor (min 1.0Å^2, mean 30.8 Å^2, max 99.9 Å^2). (pdb code: 1JLU)

According to Taylor et al.[120] three different conformational orientations of the rather flexible G-loop can be described. The structure of uncomplexed apo-enzyme of PKA has been determined in an open form of the G-loop. A kind of "intermediate form" is the most frequently found state in many PKA examples in the PDB database, as it adopts its geometry when a ligand binds to the hinge region. In complex with ATP, the closed form of the G-loop is found. This can be explained by the interactions experienced between the phosphate groups of ATP and the G-loop. Figure 7.3 schematically shows the three possible conformations giving the trace of the backbone chain in three representative examples (pdb code: 1L3R, 1JLU, 1J3H).

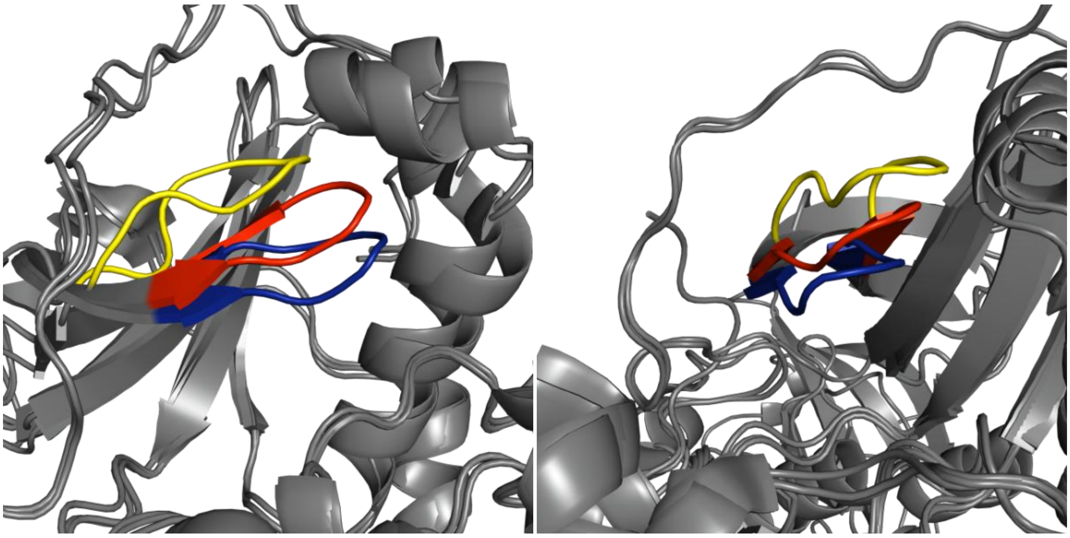


Figure 7.3: Three known conformations of the G-loop are shown. In blue PKA is displayed in complex with ATP (1L3R) (closed form), in red (1JLU) the loop is found in a medium position and in yellow the apo PKA (1J3H) (open conformation) is given. Left and right images show the binding site in perpendicular orientation.

Starting point of this study is a crystal structure of PKA in complex with phenol (PDB code: 3NX8) which interestingly shows the G-loop in a geometry closing up the ATP site even further as in the complex with the natural substrate. Figure 7.4 shows this complex. Three phenol molecules are bound to the protein. Two are placed in the ATP binding site of the kinase. One is bound to the hinge region and the other one is located below the glycine-rich loop (G-loop). The third phenol is found on top of this loop outside the catalytic center and could describe a potential new allosteric site.

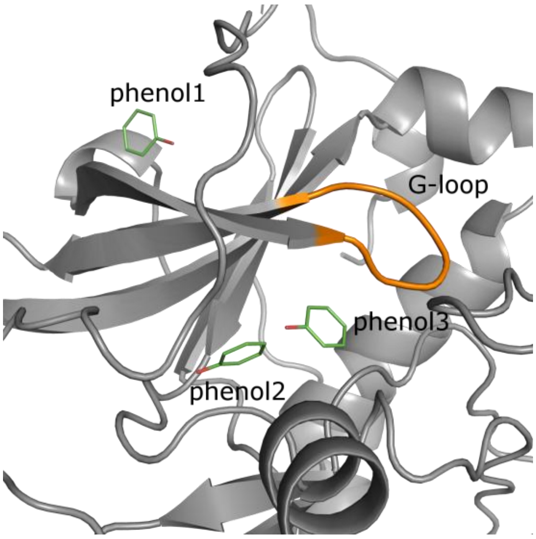


Figure 7.4 Crystal structure of PKA in complex with three phenol molecules. The protein is shown in cartoon and the three phenols are presented as green sticks, red oxygens. Phenol1 is sitting on top of the G-loop (orange), phenol2 is bound to the hinge region and phenol3 is bound underneath the G-loop.

The three orientations of the G-loop, all of which are described in literature [120], were compared with the observed position of the G-loop conformation in the PKA-phenol complex (orange). Interestingly, the G-loop is folded into the ATP binding site by an additional 4 Å (based on Cα Ser53) compared to the PKA-ATP complex (blue) (Figure 7.5). Due to a relibase search (Jan. 2010) for PKA's in the PDB this observed G-loop conformation corresponds to the most inward-folded geometry of the 85 deposite PKA structures, the G-loop is fixed by two newly formed hydrogen bonds.

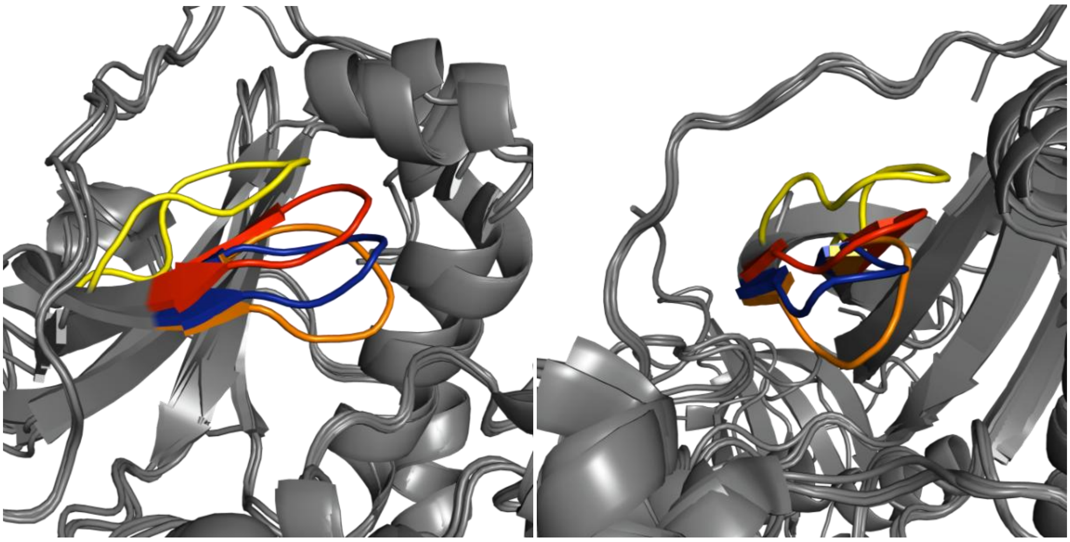


Figure 7.5 The three known G-loop orientations together with our PKA-phenol complex (orange) are shown. In blue, PKA in complex with ATP (1L3R), in red (1JLU) in a medium position, in yellow the apo PKA (1J3H) and in orange PKA in complex with phenol are shown. The images on the left and right are perpendicular.

Two hydrogen bonds are formed by the backbone nitrogen and the hydroxyl group of Ser53, which resides on top of the G-loop, with the terminal carboxylate function of Asp184. The distances of the hydrogen bonds are 2.9 Å and 3.0 Å, respectively (Figure 7.6).

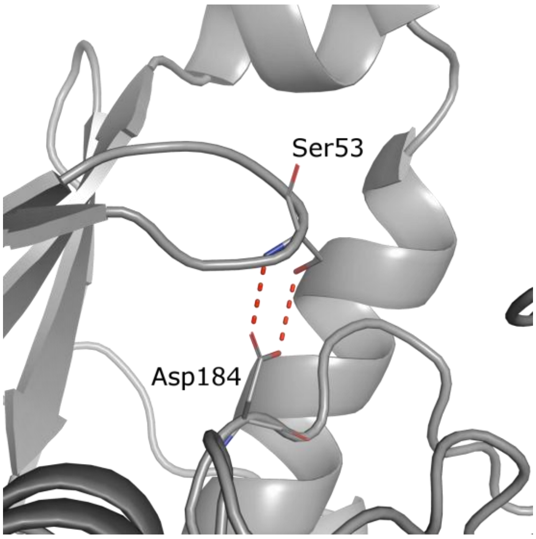


Figure 7.6 The PKA-phenol complex is shown. The newly formed hydrogen bonds are represented by red dashed lines. The distances of the hydrogen bonds are 2.9 Å and 3 Å, respectively.

As this complex represents the most inward-folded complex, the question arisis whether the third non-active-site phenol molecule found on top of the loop induces this G-loop folding and whether this could indicate a possible new allosteric binding site. If the G-loop is in its closed state, ATP cannot enter the ATP-binding site and this might represent an alternative way to inhibit the kinase. Accordingly, we tried to obtain first insights of small structural movements of secondary structure elements by molecular dynamics (MD) simulation using the program AMBER.[126]

7.3 Results and Discussion

7.3.1 MD overview

A series of six MD simulations was performed to investigate first steps of the G-loop opening. MD 1 considered all three phenol molecules to be part of the complex. We would expect for this structure that the G-loop would stay in a closed state.

MD 2 was started with the two phenol molecules in the ATP binding site and the phenol molecule sitting on top of the G-loop was removed. If the phenol molecule occupies a putative allosteric pocket, we would anticipate that the G-loop would start to open up.

MD 3 was performed considering the phenol molecule binding to the hinge region of the ATP binding site and the phenol molecule sitting on top of the G-loop. The phenol molecule occupying the hydrophobic niche under the G-loop was removed. We would expect that during this simulation, the G-loop stays in a closed conformation as the proposed allosteric phenol molecule is present.

MD 4 was performed representing the apo-enzyme starting with the geometry of the PKA-phenol complex. In this case all three phenol molecules were removed. The starting structure has an in-folded G-loop conformation. In this case we anticipate the G-loop to start the opening cascade.

MD 5 was also performed representing the apo-enzyme starting with the geometry of the PKA-phenol complex. This expanded calculation was performed on a GPU server. Also in this case we anticipate the first opening movements of the G-loop.

MD 6 was performed representing the apo-enzyme of PKA this time starting with the geometry of PKA with our virtual screening hit complex (chapter 6). In this case the G-loop is in an intermediate position and we suggest the G-loop to close up or to oscillate.

7.3.2 MD 1 considering all three phenols

We first analyzed whether the observed crystal structure complexed by the three phenol molecules remains stable or whether an intermediate opening during the simulation is observed. The presence of the above-mentioned hydrogen bonds was used as an indicator to study the behavior of the G-loop along the MD trajectory.

The hydrogen bonds between the backbone NH of Ser53 and the serine hydroxyl OH group and the carboxylate oxygens of Asp184 have been recorded in terms of their mutual distances. We defined as occurrence limit for the formation of a hydrogen bond a distance ≤ 2.2 Å between the hydrogen and the corresponding heavy atom. If a hydrogen-bond was present in a pico second time frame under consideration, the value for this frame was set to 1, else it was set to 0. For visualization the data points had to be condensed.

For each picosecond we checked for the occurrence of a hydrogen bond; if a hydrogen bond was present, the value was set to 1; and if no hydrogen bond was present the value was defined to be 0. For the binned 100 ps timeframe we could easily sum the defined values and calculate a % value. The % values are presented on the y-axis.

Figure 7.7 displays the formation of a hydrogen bond between the hydroxyl group of Ser53 and the two carboxylate oxygens of Asp184. The statistics show that the hydroxyl group of Ser53 interacts with both oxygens of the acidic group, intermediately swapping the two oxygens as interaction partners. During the first approximately 11 ns it is mostly interacting with the OD1 oxygen. For the next 3 ns, it changes its partner to the OD2 oxygen. Obviously the acid group performs an intermediate jump rotation of about 180°. However, overall the hydrogen bond is present during the entire MD simulation.

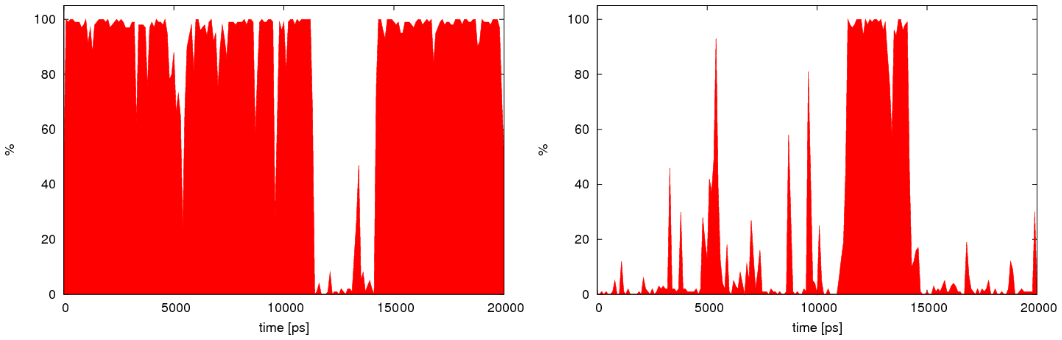


Figure 7.7 The occurrence of hydrogen bonds between the hydroxyl group of Ser53 and the two oxygens of the acid function of Asp184 is shown. Left picture is representing the OD1oxygen and the right picture presents the OD2 oxygen of Asp184.

Table 7.1 summarizes the H-bond networks occuring during the MD simulations.

The second putative hydrogen bond between the backbone NH and the carboxylate function of Asp184 is constantly present along the entire MD trajectory. Also here the actual interaction partner of NH found in the carboxylate group is swapped (Supplementary Informations).

In total, an occurrence of more than two H-bonds between the functional groups of Ser53 and the carboxylate group of Asp184 is observed. This is likely due to the classification according to distance cutoffs. Nevertheless, the analysis clearly suggests that the short hydrogen bonded contact between Ser53 and Asp184 remains fully intact along the entire MD simulation.

Table 7.1 The setup of the six MD simulations and resulting H-bond networks is summerized. In the last five colums the defined H-bonds occurring during the simulation are listed with their corresponding time frames of existence.

MD	**Phenol**			**Starting structure**	**Simulation time [ns]**	**H-bond Network**				
	1	2	3			Asp184 -Ser OH	Asp166 -Ser OH	Asp184 - SerNH	Asp184 -Phe54 NH	Asp184 -Thr51 OH
1	yes	yes	yes	close	20	0-20 ns	-	0-20 ns	-	-
2	no	yes	yes	close	20	0-12.5 ns	12.5-20 ns	0-20 ns	-	-
3	yes	yes	no	close	20	0-20 ns	-	0-20 ns	-	-
4	no	no	no	close	32	6-12 ns	0-6 ns and 12-32ns	0-20 ns	6-12 ns	-
5	no	no	no	close	90	0-55 ns	-	0-39 and 40-55 ns		39-40ns and 55-90 ns
6	no	no	no	intermediate	90	-	-	-	-	-

In summary we can conclude that over the simulation time of 20 ns, the G-loop does not open for the complex containing to all three phenols.

7.3.3 MD 2 lacking the phenol molecule in the assumed allosteric site

It was even more interesting to investigate whether the structure would open in case the phenol molecule bound on top of the G-loop is absent during a molecular dynamics run. The other two phenol molecules were considered to be present in the active site during the MD simulation. To answer this question, we performed a 20 ns second MD simulation. The occurrence of the Ser53-Asp184 hydrogen bonds was again used as an indicator for the conformational properties of the G-loop during the MD simulation.

The hydrogen bond between the hydroxyl group of Ser53 and either one of the two oxygens of the acidic function of Asp184 remains intact for the first approximately 12.5 ns. At this point the hydrogen-bond ruptured and never re-forms. This could be a first indication that the G-loop opens up along an irreversible pathway.

To obtain a more detailed insight into the trajectory, an additional hydrogen bond must be analyzed. Once the hydroxyl group of Ser53 stops its interaction to Asp184 it rotates its hydroxyl group and builds another hydrogen bond to Asp166 (2.7 Å). Figure 7.8 shows the two different orientations of the Ser53 OH side chain. It can obviously interact with both carboxylate groups of either Asp184 or Asp166 (Table 7.1).

After 12.5 ns a new hydrogen bond between the hydroxyl group of Ser53 and the carboxylate function Asp166 is formed. Overall the hydroxyl group of Ser53 is involved in hydrogen bonding along the entire 20 ns simulation time. The simulation time of 20 ns shows the change from Asp184 (2/3 of the time) to Asp166 (1/3 of the time) as one singular transition. Therefore, no conclusions of the occurrence frequency and the reversibility of this event can be drawn. However, the simulation underlines that both geometries can be adopted, likely with rather similar energy content (Table 7.1).

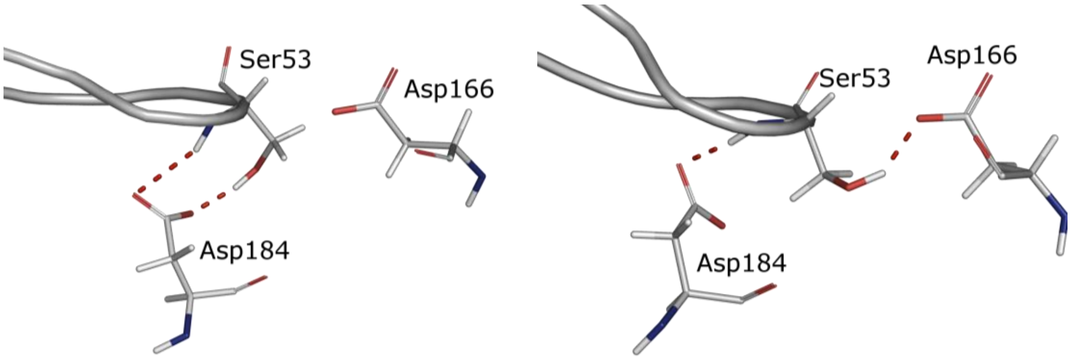


Figure 7.8 The different orientations of the side chain of Ser53 are shown. Left: the hydrogen-bond is formed with Asp184. Right: The hydrogen-bond is formed with Asp166.

A second hydrogen bond between the backbone NH of Ser53 and the terminal carboxylate of Asp184 can be observed. This contact is present over the entire 20 ns. During the time span of the MD simulation, the acidic function of Asp184 frequently performs jump rotations, swapping the interaction partner of the neighboring Ser53 NH. Obviously the movement of the side chain of Ser53 resulting in the H-bond formation to Asp166 has no influence on the hydrogen-bond contact of Ser53 NH and Asp184 carboxylate.

In summary, the G-loop does not open in a 20 ns MD simulation. When the phenol molecule which binds on top of the loop is removed, one of the initial hydrogen bonds present in the complex with the three phenol molecules is lost due to the rotation of Ser53 OH. This side chain forms a new hydrogen bond to the carboxylate group of Asp166. Thus, over the whole simulation time the inventory of locally formed hydrogen bonds remains unchanged. The hydrogen bond build by the backbone NH to the carboxylate group of Asp184 remains unruptured during the 20 ns simulation period. Removal of the assumed allosteric phenol molecules shows some impact on the local H-bonding network as the swapping of the Asp184 to Asp166 contact of Ser53 OH is only observed, once this phenol is not present. However, overall the G-loop remains in virtually the same geometry as observed in the crystal structure with the three phenol molecules. The change of the H-bond network suggests a pathway for the early step of the G-loop opening.

7.3.4 MD 3 simulation lacking the phenol molecule below the G-loop

In a further simulation run it was investigated whether the loop would open up in case the phenol molecule below the G-loop is removed. To answer this question another 20 ns MD simulation was computed. The phenol molecule occupying the hinge region and the phenol molecule occupying the possible allosteric site were considered to be present during the MD simulation. The occurrence of the crucial hydrogen bonds has again been used to record the properties of the G-loop.

The hydroxyl side chain of Ser53 interacts with the carboxylate function of Asp184 over the entire MD trajectory. The rotation of this side chain, which was observed in the former simulations, cannot be seen in this case. Also the hydrogen-bond between the backbone NH group and the carboxylate function of Asp184 is constantly present along the entire MD simulation (Table 7.1).

Summarizing the results of this simulation, the G-loop does not open within the 20 ns.

7.3.5 MD 4 simulations using the apo enzyme of PKA

In the following we addressed the situation that results from the total removal of all three phenol molecules during the MD simulation. In principle, this should lead to the uncomplexed apo enzyme. It has been described in literature that the G-loop adopts an open conformation in this situation.[120] Thus, we would expect the G-loop to open up during our simulation. As major movements are required, we expanded the computing time to 32 ns.

In this simulation the rotation of Ser53-OH towards Asp166 occurs in the first frames of the trajectory. However, the hydrogen bond between the hydroxyl side chain of Ser53 and the carboxylate group of Asp184 is reestablished during the time interval of ~ 6 ns – ~ 12 ns. During the remaining time, the hydroxyl group is rotated towards Asp166 (Table 7.1). Accordingly the Ser53 OH group is hydrogen-bonded during the entire MD simulation. Also the hydrogen bond between the backbone NH group and the carboxylate function of Asp184 is constantly present along the entire MD simulation.

Further analysis of the trajectory indicates that another hydrogen bond is formed in the G-loop region. Figure 7.9 displays two snapshots of the MD simulation. The left part of the figure shows the hydroxyl group of Ser53 pointing toward Asp184. This snapshot is taken at about 10 ns. On the right hand side of the figure the hydroxyl group is pointing towards Asp166. Thereby, Phe54 is also shifted, leading to a new hydrogen-bond between the backbone nitrogen of Phe54 and the oxygen of Asp184.

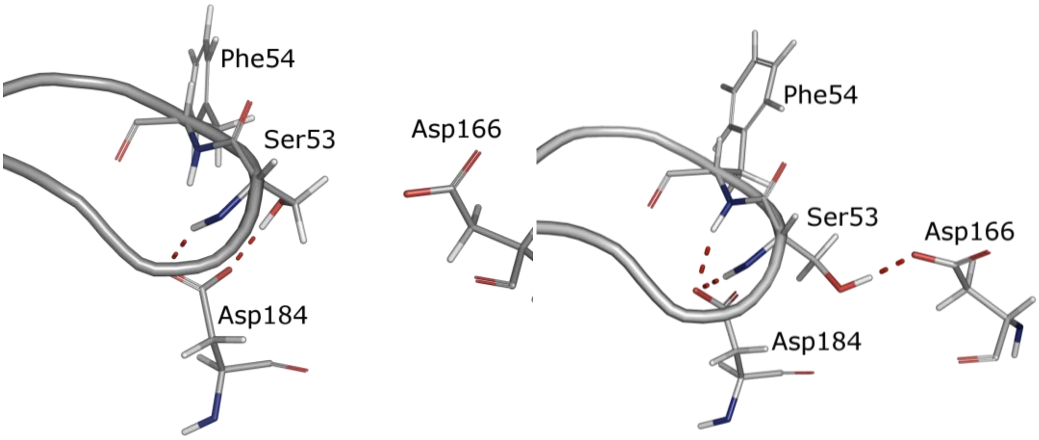


Figure 7.9 The different orientations of the side chain of Ser53 are shown. Left: The hydrogen bond is formed between the hydroxyl group of Ser53 and a carboxylate oxygen of Asp184. Right: The hydrogen bond is formed between the hydroxyl group of Ser53 and one of the carboxylate oxygens of Asp166, an additional hydrogen bond between the backbone nitrogen of Phe54 and the oxygen of Asp184 can be observed.

The newly formed hydrogen bond is present while the side chain of Ser53 is pointing towards Asp166. In the time slot between 6 ns – 12 ns the new hydrogen bond is hardly present. During this period the side chain of Ser53 points toward Asp184 most of the time.

Obviously also during this 32 ns simulation the G-loop does not open. One of the initial hydrogen bonds found in the triphenol complex disappears upon rotation of Ser53. The side chain then forms a new hydrogen bond contact to Asp166 and simultaneously the backbone nitrogen of Phe54 experiences also a new hydrogen bond to Asp184. Thus, over the whole simulation period at least two hydrogen bonds are present. Mostly even three hydrogen bond connections can be recorded. Opening of the G-loop is not achieved in the 32 ns simulation.

7.3.6 MD 5 simulation using the apo enzyme of PKA

This MD simulation was started under very similar conditions as MD4, however to collect the trajectory over a longer time span (90 ns) we moved to AMBER11. AMBER11 provides the new feature to calculate on graphic cards (GPU), which are highly parallelized and therefore the calculation time can be reduced dramatically. For this simulation we expected the G-loop to open up.

The H-bond between the hydroxyl group of Ser53 and the acid function of Asp184 is mostly present during the first 55 ns; only for a short time interval (~ 2.5 ns) around 40 ns this H-bond is not present. Surprisingly, after the 55 ns simulation period the hydroxyl group does not form any H-bond. In the MD simulations presented previously the hydroxyl group rotated towards Asp166 and formed an H-bond with the acid functionality of this adjacent residue.

The H-bond between the backbone NH and the acid function of Asp184 is present for the first 55 ns with a short period of a complete loss in the frames next to ~ 40 ns. After the time period of the first 55 ns this H-bond breaks irreversibly.

As the H-bonds previously observed in all other simulations are broken after 55 ns simulation time; we suggest that the G-loop further opens after that point. Nevertheless considering the overall geometry, the loop does not move in the remaining time span of 35 ns MD simulation to a geometry which can be described as an open conformation. Figure 7.10 superimposes ten representative loop geometries.

It can be seen that the G-loop remains inward-folded over the entire simulation time. These observations suggest that another H-bond network is formed.

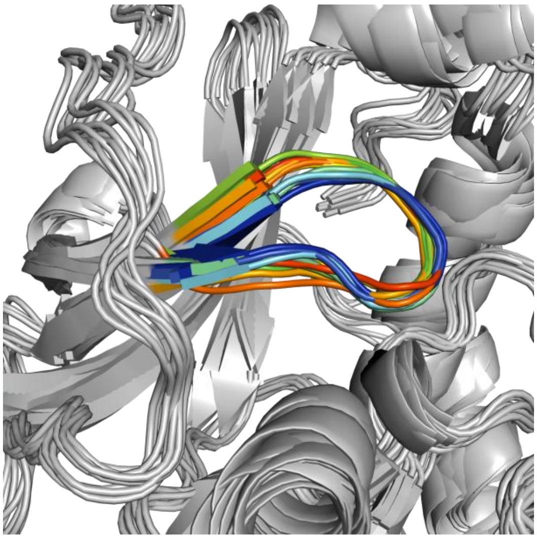


Figure 7.10 Ten representative orientations of the G-loop along a 90 ns MD simulation of the apo form of PKA are shown.

Detailed analysis shows that a new H-bond contact is formed between the hydroxyl function of Thr51 and the acid function of Asp184. Here, the acid function of Asp184 rotates by approximately 90° (Figure 7.11).

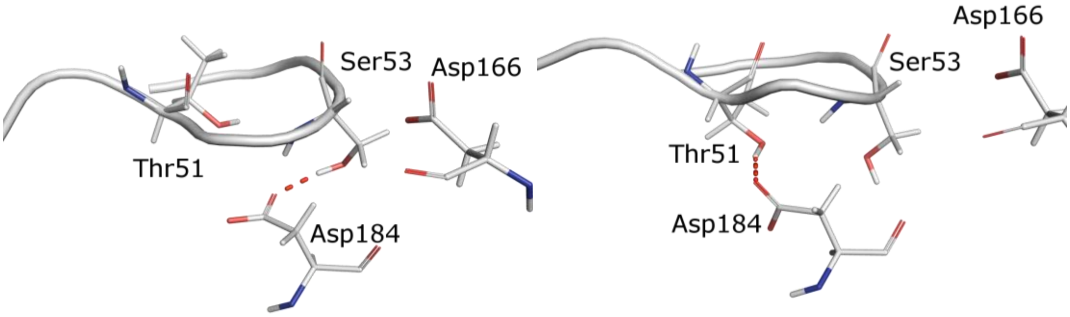


Figure 7.11 Left: The H-bond between Asp184 and Ser53. Right: The H-bond between Asp184 and Thr51 formed after 90° rotation of Asp184.

This newly formed H-bond has already been transiently present at ~ 40 ns and it remains stable from 55 ns until 90 ns. Thus, Asp184 remains involved in at least one H-bond over the entire simulation. One H-bond seems to be strong enough to hold the loop in a closed state.

7.3.7 MD 6 simulation using an intermediate G-loop conformation as a starting point

A second expanded simulation (90 ns) with AMBER11 on GPU’s started with an intermediate G-loop geometry. In this case we expect the loop to close up or to oscillate. The question is whether it can oscillate to a conformation which is able to form the above described H-bonds.

Figure 7.12 shows some representative cluster means over a time period of 90 ns. The G-loop is moving over a rather large range of 9 Å (based on Cα of Ser53). Figure 7.13 also includes the PKA-phenol complex (gray). The G-loop conformation of the PKA-phenol complex is folded in by another 7 Å (based on Cα of Ser53) compared to the G-loop conformations of the MD simulation.

The G-loop of this simulation moves between the conformations known from literature. The loop oscillates between the apo enzyme, which is assigned to an open state and the ATP bound conformation which is known to be in a closed state. The simulation departing from the intermediate state does not indicate any movements of the G-loop towards a closed conformation. The trajectory does not pass through a conformation similar to the one present in our triphenol complex.

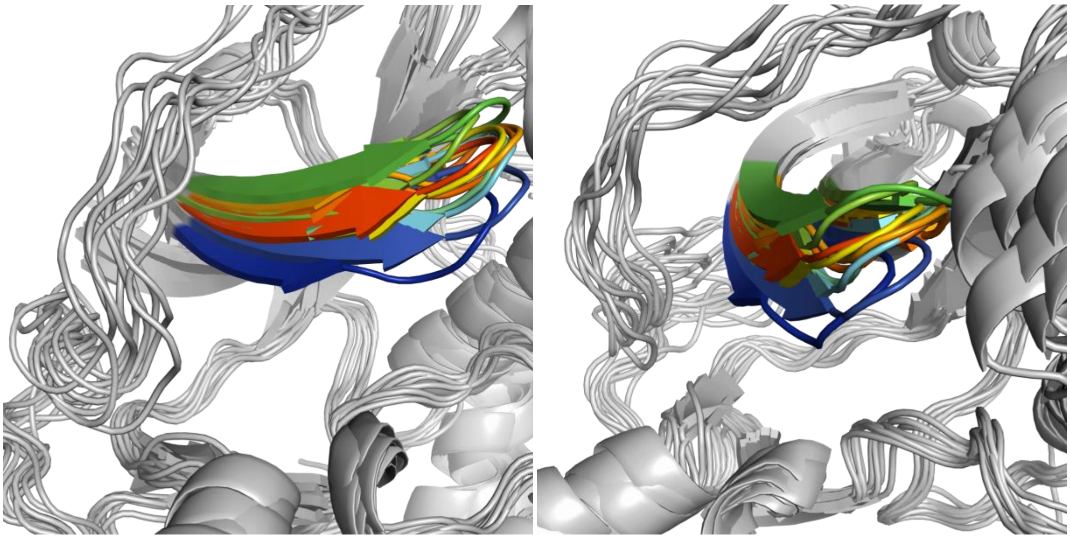


Figure 7.12 Ten representative clusters of a 90 ns MD simulation are shown. Starting point was an intermediate open structure, observed with the virtual screening hit (PDB code: 3OOG) presented in Chapter 6. Left and right picture are perpendicular.

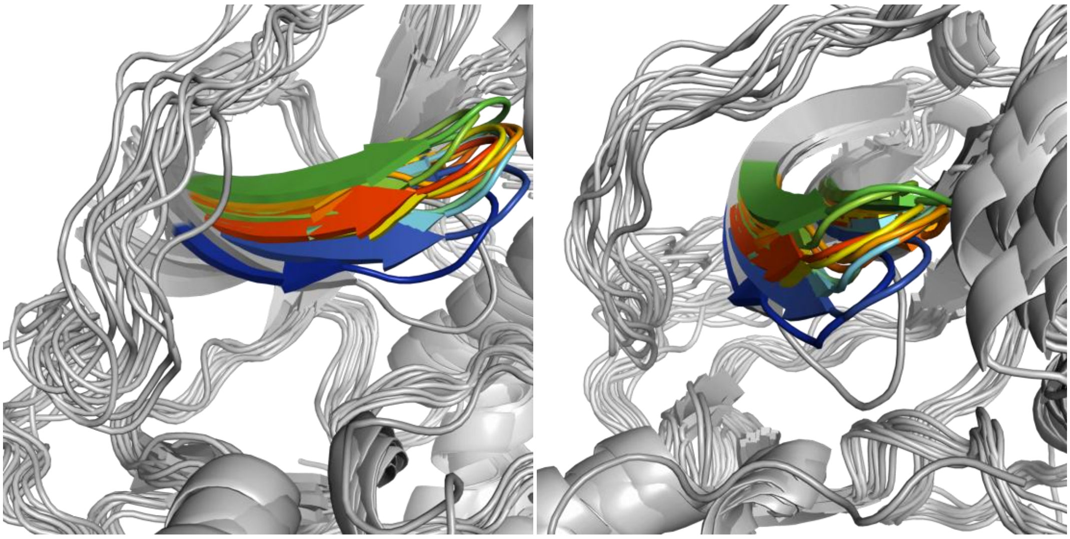


Figure 7.13 Ten representative clusters of a 90 ns MD simulation are shown. Starting point was an intermediate open structure observed with the virtual screening hit presented in chapter 6. In gray the G-loop conformation found in the PKA-triphenol complex is displayed. Left and right picture are perpendicular.

7.4 Conclusions and Outlook

Apparently, a time window of 90 ns is not large enough to observe major conformational transitions of the G-loop.

Vogtherr[127] and colleagues observed in their NMR studies that movements of the DFG-loop in p38 kinases occurs on a rather long time scale clearly beyond the 90 ns considered in this study. This result leads to the conclusion that we would have to expand our simulation time by at least another order of magnitude. Due to limitations in computer power, the required calculations are currently hardly possible.

The movements of single amino acids observed at the beginning of the simulation and leading to different H-bond networks indicate that small initial movements are the start for larger secondary structure movements supposedly initiating the overall irreversible opening of the G-loop.

A 20 ns simulation considering all three bound phenol molecules remains stable over the entire simulation time. Removal of the anticipated ‘allosteric phenol’ seems to allow for some larger movements also involving the swapping of H-bonds formed by Ser53. Moving to a simulation with all three phenol molecules removed from their crystal structure should relax into the geometry observed for the apo protein. However, as described, full relaxation to this state cannot be observed in the considered simulation time span. Instead some incipient changes initiated by rotations of Ser53 and Asp184 seem to be the starting points for the irreversible opening of the G-loop. Nevertheless, the simulations also show that the presence of at least a single H-bond across these residues is sufficient to hold the G-loop in closed state. We interpret the observed movements as early steps towards an opening of the G-loop as found in the crystal structure of the apo protein.

7.5 Methods

7.5.1 MD setups

For the MD 1-5 the starting geometry was the PKA-phenol complex (PDB code: 3NX9). For the different MD simulations the various phenol molecules were removed as described. For MD 6 the PKA-virtual screening hit complex structure (Chapter 6, PDB code: 3OOG) was used after removing the ligand.

All preparing steps for the protein were performed using AMBER Tools 1.4[128] using the ff99SB force field. For the phosphoserine and the phosphothreonine the contributed parameters form the University of Manchester were used (phosphoserine with unprotonated phosphate group, SER-PO3 and phosphothreonine with unprotonated phosphate group, THR-PO3).[129]

The parameters for the phenol molecules were computed with ANTECHAMBER[130] using the GAFF force field.[131]

Hydrogen bond atoms were edited using the AMBER templates and two sodium counter ions were added. The TIP3P water box was added as an octahedron around the protein using a boxsize of 10 Å.

The initial minimization of the solvent and ions was calculated in 3000 steps. The MD simulations were started by heating the system to 300 K over a time period of 50 ps and were equilibrated over a time span of 2.5 ns keeping the pressure and the volume constant.

The simulations MD 1-4 have been calculated using the parallel pmemd of AMBER10[126] on a supercomputer based in Tromsø, Norway. The expanded simulations (MD 5 and MD 6) have been calculated in-house on a CUDA based GPU server using AMBER11.[126]

The analysis of the hydrogen bonds and the clustering was performed by ptraj, a program of AMBER Tools 1.4.[128]

The hydrogen-bond plots were drawn with gnuplot.[132]

The PyMOL Molecular Graphics System, Version 1.1 was used to visualize the molecules and the pictures were rendered with this program.[75]

7.6 Supplementary Informations

7.6.1 MD 1 plots

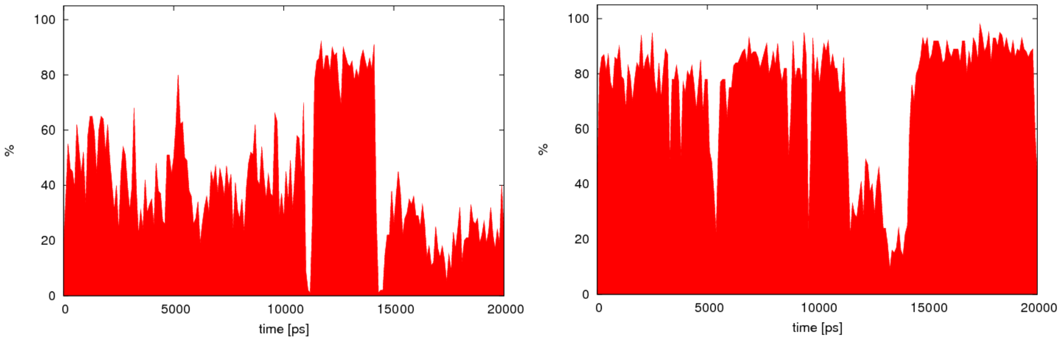


Figure 7.14 The occurrence of hydrogen bonds between the backbone NH of Ser53 and the two oxygens of the acid function of Asp184 are shown. In all following picture, the left picture represents the OD1 oxygen and the right picture presents the OD2 oxygen of the carboxylate function of Asp184.

7.6.2 MD 2 plots

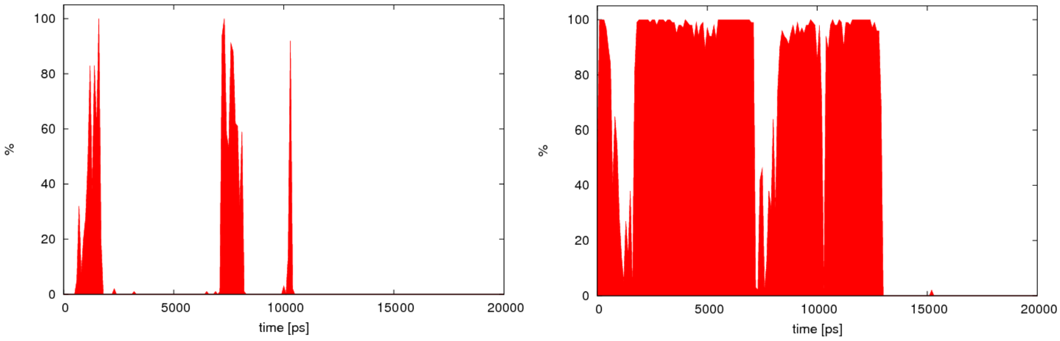


Figure 7.15 The occurrence of hydrogen bonds between the hydroxyl group of Ser53 and the two oxygens of the acid function of Asp184 are presented. Beyond about 12.5 ns the contact to Ser53 moves irreversibly into an H-bond towards Asp166 (s. Figure 7.16)

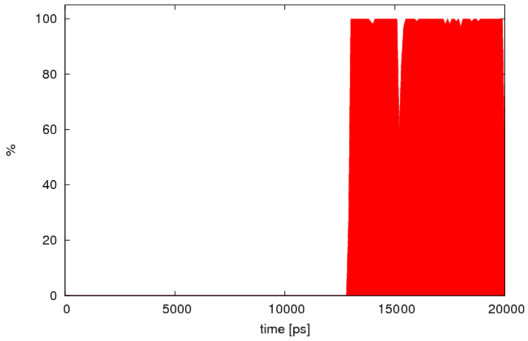


Figure 7.16 The occurrence of hydrogen bonds between the hydroxyl group of Ser53 and the oxygens of the acid function of Asp166 is shown. These contacts are only formed after 12.5 ns of simulation time.

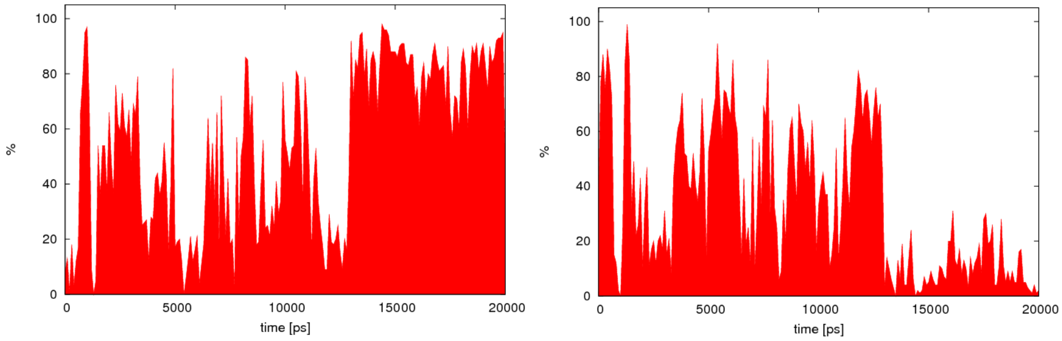


Figure 7.17 The occurrence of hydrogen bonds between the backbone nitrogen of Ser53 and the two oxygen of the acid function of Asp184 is shown.

7.6.3 MD3 plots

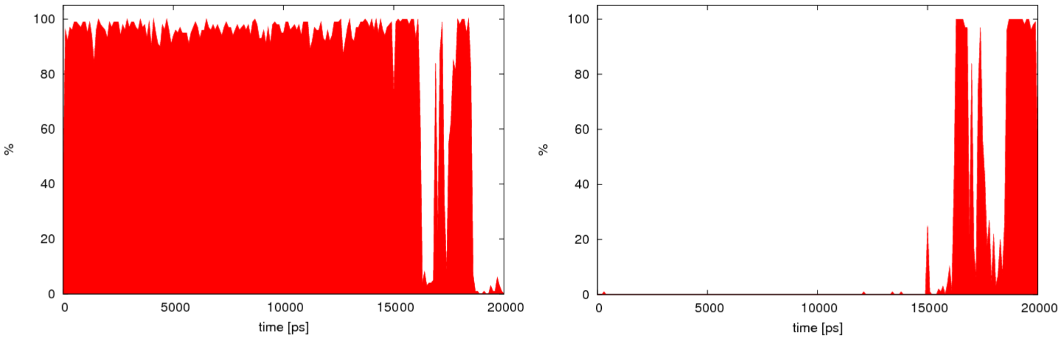


Figure 7.18 The occurrence of hydrogen bonds between the hydroxyl group of Ser53 and one of the two oxygen of the carboxylate group of Asp184 is presented.

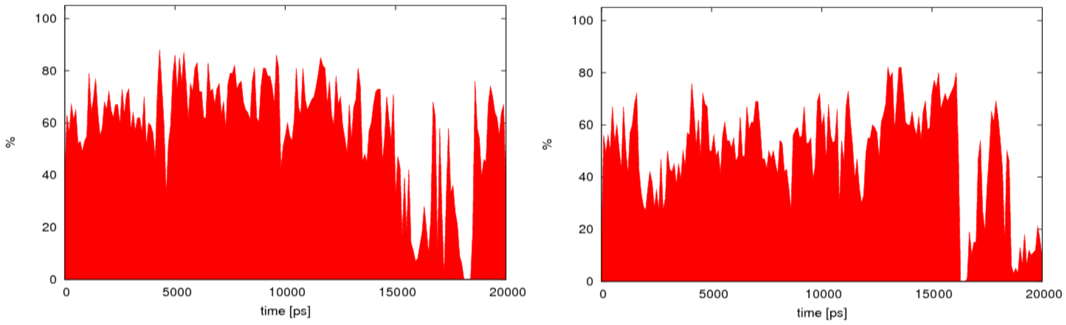


Figure 7.19 The occurrence of hydrogen bonds between the backbone nitrogen of Ser53 and the two oxygens of the acid function of Asp184 is shown.

7.6.4 MD 4 plots

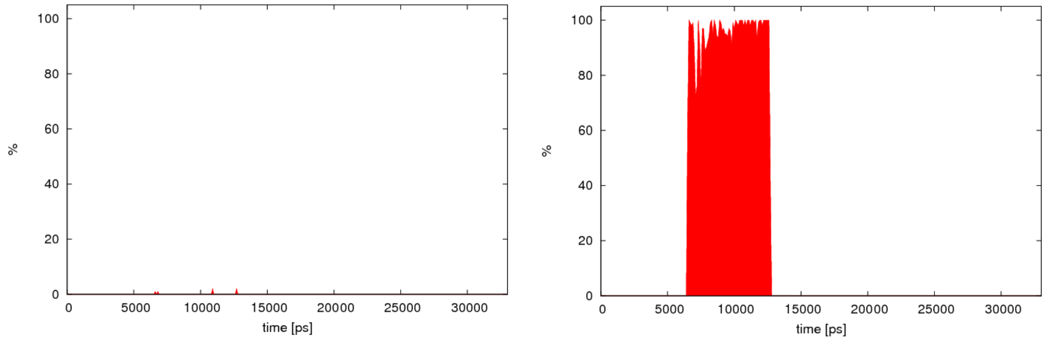


Figure 7.20 The occurrence of hydrogen bonds between the hydroxyl group of Ser53 and the two oxygen of the acid function of Asp184 is shown.

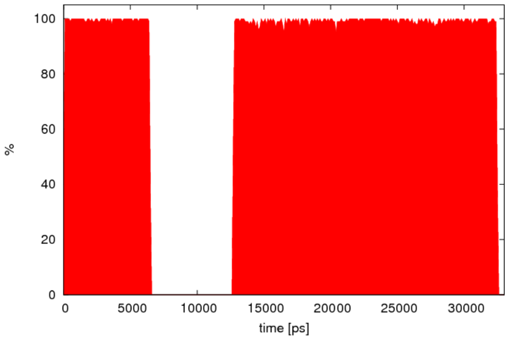


Figure 7.21 The occurrence of hydrogen-bond between the hydroxyl group of Ser53 and the oxygen of the acid function of Asp166 is shown.

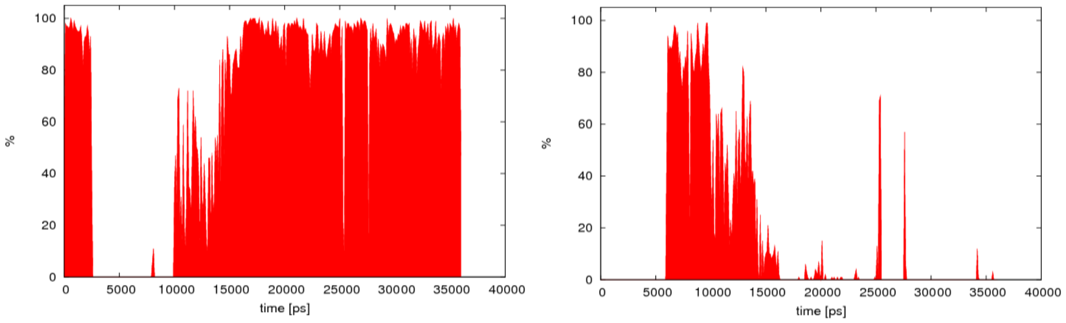


Figure 7.22 The occurrence of hydrogen bonds between the backbone NH group of Ser53 and the two oxygens of the acidic function of Asp184 is shown.

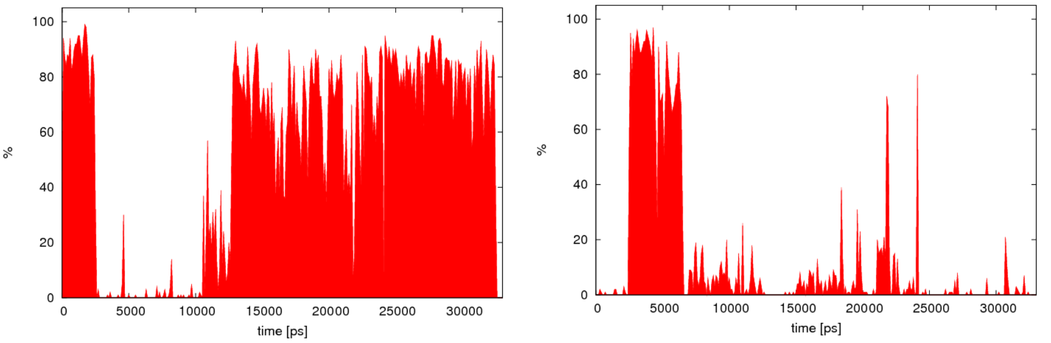


Figure 7.23 The occurrence of hydrogen bonds between the backbone nitrogen of Phe54 and the two oxygens of the acid function of Asp184 is shown.

7.6.5 MD 5 plots

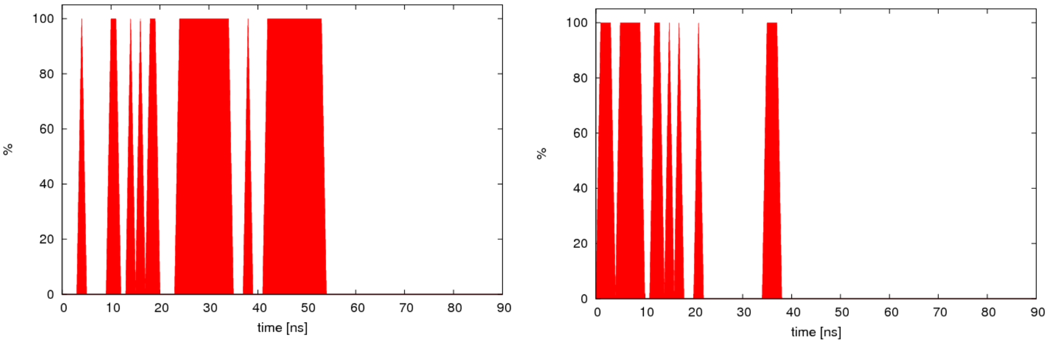


Figure 7.24 The occurrence of hydrogen bonds between the hydroxyl group of Ser53 and the two oxygen of the acid function of Asp184 is shown.

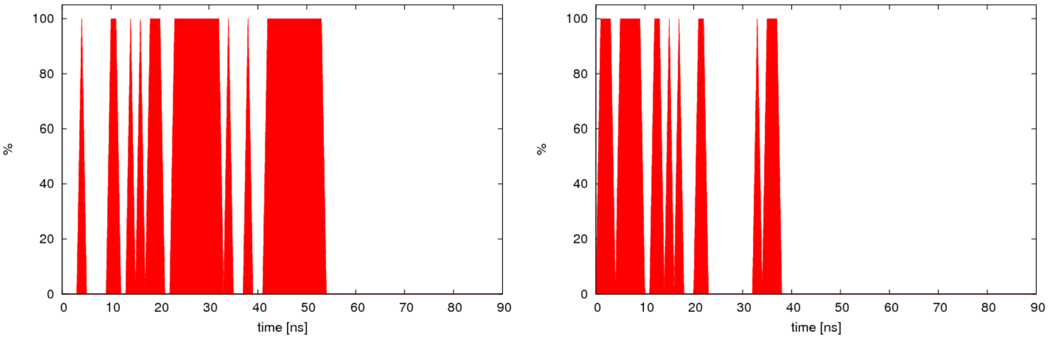


Figure 7.25 The occurrence of hydrogen-bond between the backbone NH of Ser53 and the oxygen of the acid function of Asp184 is shown.

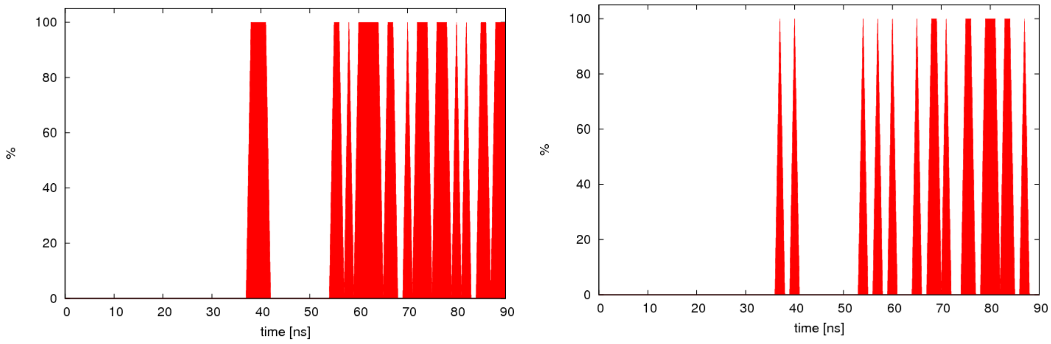


Figure 7.26 The occurrence of hydrogen bonds between the hydroxyl group of Thr51 and the two oxygens of the acid function of Asp184 is shown.

8 New Scaffolds for Aldose Reductase: A Virtual Screening Study

8.1 Introductory Remarks

The present study was accomplished in cooperation with Jakub Gunera who performed his bachelor thesis[133] and Cornelia Koch, who accomplished a PhD thesis on aldose reductase. Dr. Sascha Brass synthesized some of the suggested virtual screening hits.

8.2 Drug Design

One of the greatest challenges in the early stages of drug design is the search for novel lead structures. These structures are starting points for the development of potential drugs and templates for further optimization. Firstly, the affinity of these scaffolds has to be increased towards low nanomolar range, however simultaneously considering membrane permeability, lipophilicity and bioavailability. An experimental way to find lead structures is the so-called High Throughput Screening (HTS). This very costly screening method requires huge compound libraries (~ 1,000,000 compounds), an available robust assay and an elaborate experimental setup to screen fully automated the library. In consequence, only big companies can afford this method. Due to these high costs but also as methodological alternatives, other approaches such as virtual screening have been developed to find new leads. This computational method requires structural information about the target protein and computational resources. Fortunately, the number of determined protein-ligand complex structures has steadily increased over the last years (> 72000 structures deposited in the PDB[134] as of April 2011) and the computational resources also became available, so that these approaches have gained popularity. By virtual screening even larger ligand libraries (~ 10,000,000 compounds) can be screened easily, usually based on a predefined pharmacophore, which comprises the physico-chemical properties within the binding site. Another approach is the development of new compounds by incremental construction of ligands within the binding site. This is called *de-novo* design.[135]

8.3 Target Family of AKRs

The aldo-keto reductase (AKR) superfamily contains more than 140 proteins. They accomplish many different physiological roles. Most of them are NADP(H)-dependent oxidoreductases, which are responsible for metabolization of carbohydrates, steroids and prostaglandins. In addition, many other endogenous aldehydes and ketones, but also xenobiotics, are metabolized by AKRs. In the following we want to focus on an in-house target, the human aldose reductase.

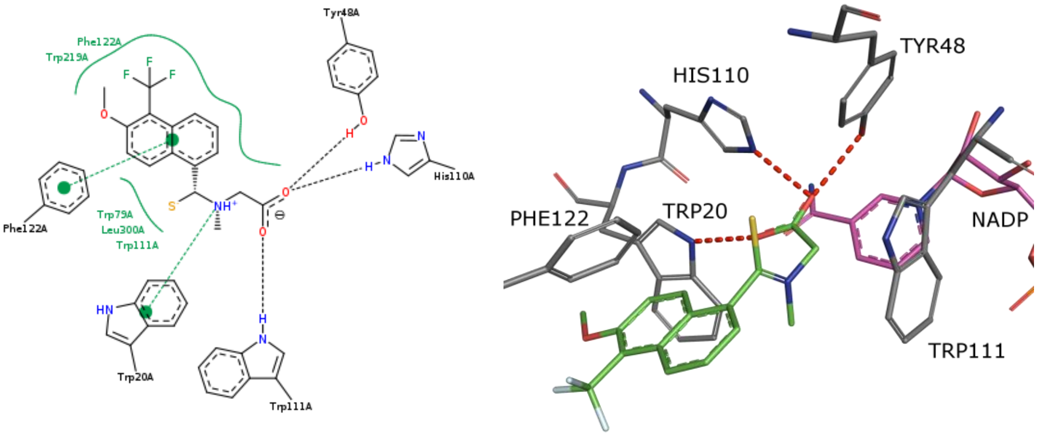


Figure 8.1 Aldose reductase in complex with tolrestat (2FZD). Left: A schematic overview of the binding pocket and the binding mode of tolrestat. The H-bonds are presented in black as dashed lines, hydrophobic interactions are shown by green dashed lines. Right: The crystal structure of tolrestat. The key residues are colored gray, tolrestat in green and the co-factor in purple, H-bonds as dashed red lines.

Figure 8.1 displays the binding pocket of Aldose Reductase. All known inhibitors of aldose reductase address the so-called ‘anion binding pocket’ which is built by His110, Tyr48, Trp111 and the cofactor NADP$^+$. Most of the known inhibitors are addressing this area of the pocket by a carboxylic acid or by a hydantoin moiety.

Figure 8.2 displays known inhibitors of aldose reductase. It can be seen that all of these compounds exhibit a carboxylic acid or a hydantoin motif.

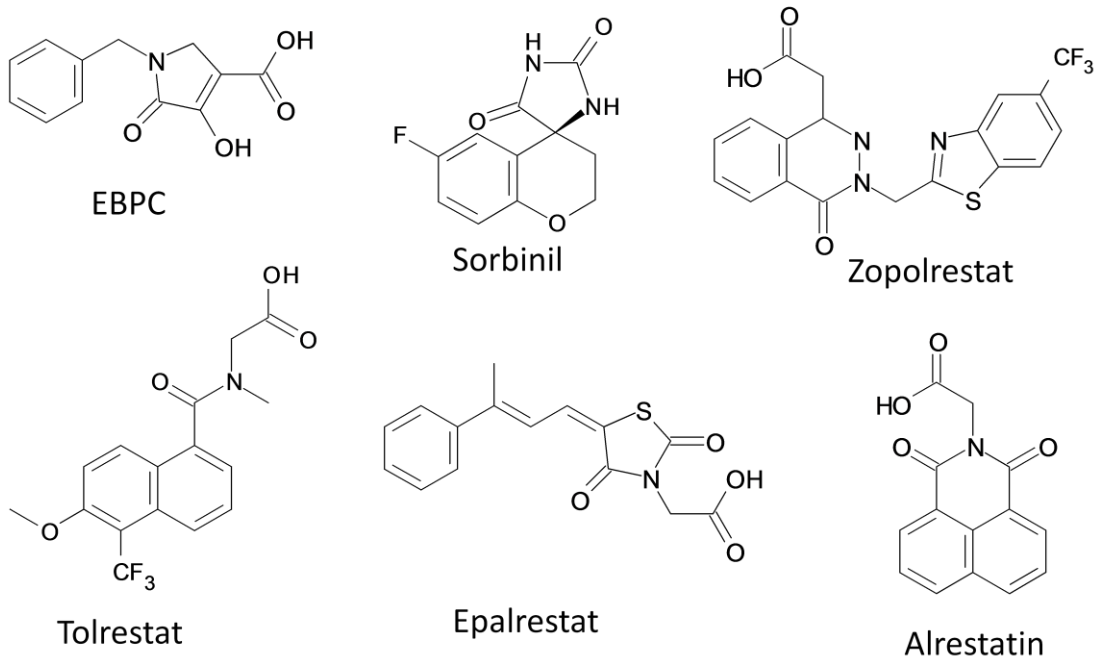


Figure 8.2 Known inhibitors for AKRs.

8.4 Ligand Database and Targeted Library Design

To embark onto Structure Based Drug Discovery (SBDD) by virtual screening more than crystallographic data of the target protein and its receptor-ligand structures are necessary. Easily accessible databases of commercially available and by some predefined criteria suitable compounds are of great importance to obtain reliable screening results. Therefore, the ZINC[136] database was selected as a matter of choice for the ligand database. Out of more than 13 million 3D structures deposited in ZINC, the lead-like subset of 1,900,000 entries was selected. The lead-likeness was proposed as a further refinement of the concept of drug-likeness[1], which is a collection of defined properties and constraints. These properties have been defined by Lipinski et al. [1]and became recognized as the “rule-of-five”. The rule suggests that low oral availability is more probable whenever:

- there are more than five hydrogen-bond donors (cumulation of oxygen and nitrogen atoms with one or more hydrogens bound)

- there are more than ten hydrogen-bond acceptors (cumulation of oxygen and nitrogen)

- the molecular weight is above 500 Da

- the octanol-water partition coefficient logP > 5

ZINC defines lead-like molecules more strictly. They show lower molecular weight
(< 350 Da) and lipophilicity (logP < 3).

Derivatives of carboxylic acids and hydantoins are suitable functional groups to form interactions with key residues of the target protein aldose reductase. Figure 8.2 shows known potent inhibitors. As can be seen, all of them are derivatives of carboxylic acids or hydantoin. Furthermore, they have aromatic or hydrophobic moieties. Therefore, we decided to build a ligand library based on structures including the most prominent features found in these inhibitors.

In the first step, the lead-like data subset of the ZINC[136] database, containing 1,900,000 entries, was filtered for carboxylic acid and hydantoin motifs. This substructure search was performed with our in-house tool fconv.[72] The ligand set was reduced to 23,307 compounds. To further decrease the number of entries, MOE[57] was used to extract molecules with a maximum number of four rotatable bonds. This second filtering step reduced the dataset to 8,409 compounds.

The considered ligand set was complemented by 45 compounds synthesized in the group of Prof. Schlitzer, Marburg. These compounds were particularly designed as putative inhibitors for aldose reductase.

8.5 Conformational and Pharmacophore Search

As described in section 8.2, large ligand libraries up to 10,000,000 compounds can be scanned by virtual screening. As it is impracticable to perform a docking run with several million compounds, one of the major steps in the applied work flow of the screening process is the successive reduction of the large number of structures to those which actually satisfy some predefined crucial requirements. These can be a given molecular weight threshold, the number of H-bond donor/acceptor functionalities, the number of rotatable bonds or formal charges, etc. The program MOE provides the functionality to filter large datasets in terms of sterical and physico-chemical properties. Conformational search combined with a pharmacophore hypothesis reduces the initial dataset to a set of suitable compounds, which are then subjected to docking. In order to check conformance with the predefined pharmacophore hypothesis, an exhaustive generation of all biologically relevant conformations of the ligands is highly desirable.

Therefore, ten conformations for each ligand have been generated leading to ~ 80 000 conformations. The resulting database was used to perform the virtual screening. A pharmacophore constraint was constructed based on the tolrestat complex with Aldose Reductase. However, to allow for larger diversity, only two H-bond acceptor constraints with respect to His110 and Tyr48 were taken into account. Furthermore, the shape and volume of the active site was rebuilt by excluded volume constraints and added as a third constraint to prevent steric clashes of ligands with the active site.

The pharmacophore query revealed 2,957 unique compounds, most of them derivatives of acetic acids. These virtual screening hits have been used for the subsequent docking approach.

8.6 Docking

The docking runs were performed with the program FlexX. Furthermore, we used two different scoring functions: The parent scoring functions implemented in FlexX and dsx (drugscoreX 0.46), an in-house developed scoring function.

All docking solutions were rescored with dsx (drugscoreX) and the overall best 250 poses were visually inspected in order to confirm a reliable orientation of ligand functional groups with respect to interactions with the key residues Tyr48, His110 and Trp111.

Among the most promising solutions we found sulfonamide derivatives which interact with the key residues via their carbonic acid portion (Figure 8.3).

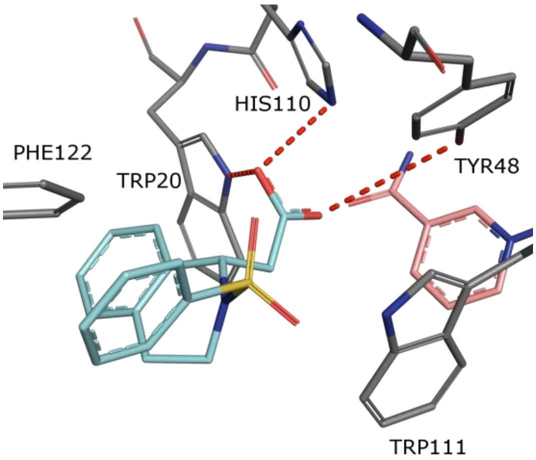


Figure 8.3 The docking solution of SB320 (s. Figure 8.4) is shown; The key residues are presented in stick representation; the cofactor NADP$^+$ is shown in stick modus; carbon atoms of the cofactor are colored in pink; carbon atoms of the protein side chains are colored in gray; carbon atoms of the ligand are colored in cyan; oxygens in red; nitrogens in blue and sulphur atoms in yellow.

We used commercially available compounds for this virtual screening approach, the selected sulfonamides were also easily accessible synthetically. Hence, a series of compounds was synthesized in-house. Figure 8.4 shows the chemical structures of these compounds and the measured K_i values.

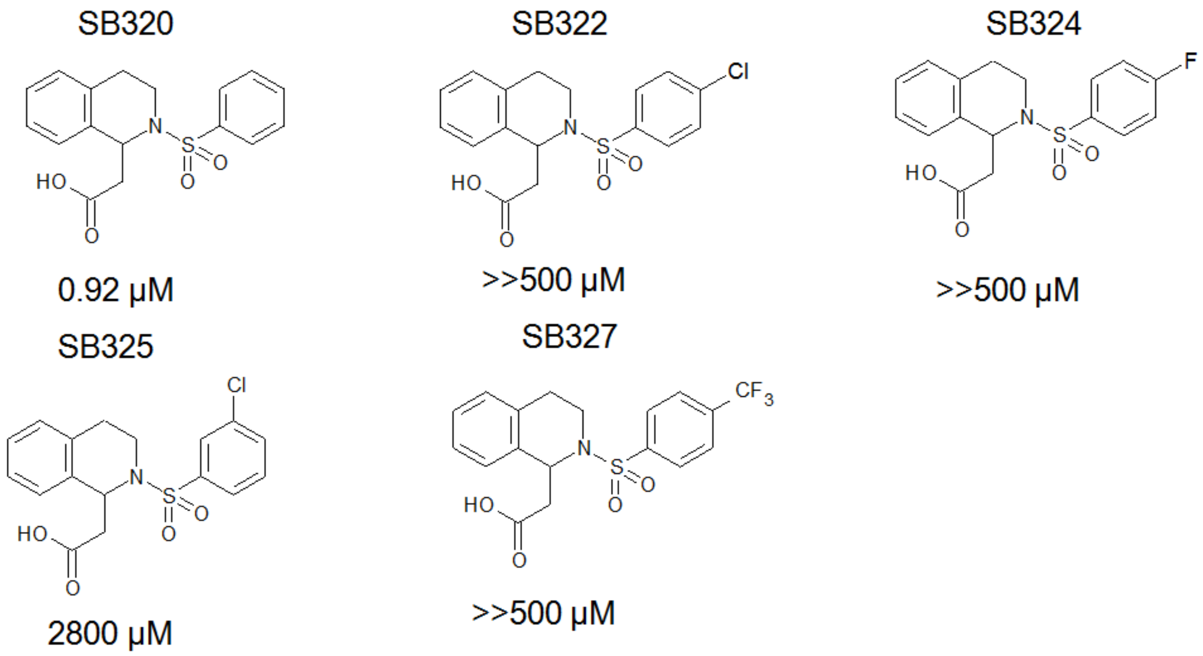


Figure 8.4 The synthezised compounds with affinity data against aldose reductase

SB320 is with 920 nM a rather promising starting point for subsequent lead optimization. Surprisingly, all other compounds from this narrow compound series show almost no inhibition. The predicted docking poses cannot explain the observed discriminations in the measured assay data. Therefore, before embarking onto lead optimization, a crystal structure determination of the SB320 in complex with aldose reductase would be necessary.

9 Development of a Thermal Shift Assay

9.1 Introductory Remarks

The present study was accomplished in cooperation with Prof. Klaus Reuter, Prof. E. Bremer and Marco Pittelkow.

9.2 Introduction

The experimental characterization of fragment binding to a target protein is still an area currently heavily explored. Presently, biophysical methods such as SPR, NMR and X-ray crystallography or biochemical assays are used to determine fragment hits. In recent years, a further assay referred to as thermal unfolding or thermal stability assay has become popular to discover fragment hits. Here, protein stabilization due to ligand binding is measured as a function of increasing temperature.

The ability to measure protein stability is a useful tool with regard to many different practical applications. For example, it can be used to determine optimal buffer conditions for protein storage, or to identify and characterize protein-ligand or protein-fragment binding. In addition, measurements of protein stability can be used to screen for mutations that can lead to changes in protein stability.[137]

We decided to validate the thermal stability method by performing a case study with mutant variants of the enzyme EctD.

9.3 Concept of Thermal Shift Assay

Cummings et al.[138] introduced the concept of the thermal unfolding assay in 2006. Protein stability is measured as the ability of a protein to remain in its native folded state depending on temperature. This method is based on the real-time measurement of the fluorescence of a dye, such as SYPRO Orange, which has affinity for hydrophobic parts of a protein. SYPRO Orange shows weak fluorescence in its unbound state in water and hydrophilic environments. However, its fluorescence strongly increases when the dye is bound to hydrophobic patches of a protein molecule which become better accessible upon protein unfolding. The measured fluorescence is linear to the amount of attached dye to the protein. In this method, the protein of interest is subjected to stepwise increases in temperature using a real-time PCR machine, in the presence of SYPRO Orange.

Fluorescence is measured and plotted against temperature, enabling determination of the melting temperature (T_m). The T_m is usually defined as the inflection point of the sigmodial-like curve of fluorescence plotted against temperature. A high melting temperature indicates high protein stability.[137-140]

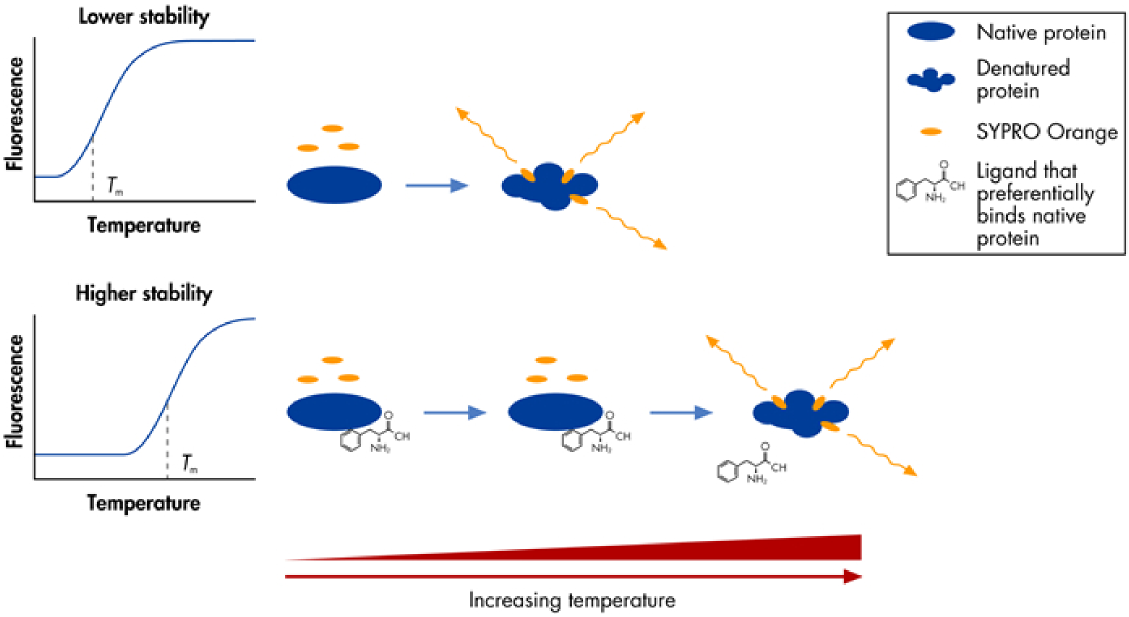

Lower stability
Fluorescence
Tm
Temperature
Higher stability
Fluorescence
Tm
Temperature
Native protein
Denatured protein
SYPRO Orange
Ligand that preferentially binds native protein
Increasing temperature

Figure 9.1: Fluorescence of SYPRO Orange is dramatically increased when it is bound to hydrophobic patches exposed upon protein denaturation. Protein stability (e.g., T_m) can be estimated by analyzing the temperature dependence of fluorescence intensity. Addition of ligands that preferentially bind to either the native or denatured protein will affect T_m.[137]

9.4 Indroduction to EctD

The biochemical characterization of the ectoine hydroxylase EctD from the moderate halophile *Virgibacillus salexigens* (formerly *Salibacillus salexigens*)[141] and from the soil bacterium *Streptomyces coelicolor*[142] has revealed that the EctD enzyme is a member of the non-heme containing Fe^{2+} and 2-oxoglutarate dependent dioxygenases superfamily (EC 1.14.11) (Figure 9.2). Members of this enzyme superfamily are found wide-spread in nature and catalyze a broad spectrum of oxidative reactions including cyclizations, ring fragmentations, C-C bond cleavages, epimerizations, desaturations, halogenations and hydroxylations of widely varying organic compounds.[143-145] This group of enzymes typically couples the decarboxylation of 2-oxoglutarate with the formation of a high-energy ferryl-oxo intermediate that acts as a hydrogen-abstracting species. The formed Fe(IV)=O species is directly responsible for the oxidation of the organic substrate bound by the enzyme. Fe^{2+} and 2-oxoglutarate dependent dioxygenases probably constitute the most versatile group of all oxidizing biological catalysts.[146]

Structural studies of a number of Fe^{2+} and 2-oxoglutarate dependent dioxygenases revealed a common protein fold that contains a highly conserved iron-binding motif, the so-called 2-His-1-carboxylate facial triad.[143-146] The amino acid sequence of EctD possesses this iron-binding motif[141, 142, 147] and the EctD enzyme catalyzes a type of reaction that is common among Fe^{2+} and 2-oxoglutarate dependent dioxygenases.[148] In this reaction, O_2-dependent hydroxylation of the substrate ectoine is accompanied by the oxidative decarboxylation of 2-oxoglutarate to form succinate and CO_2.

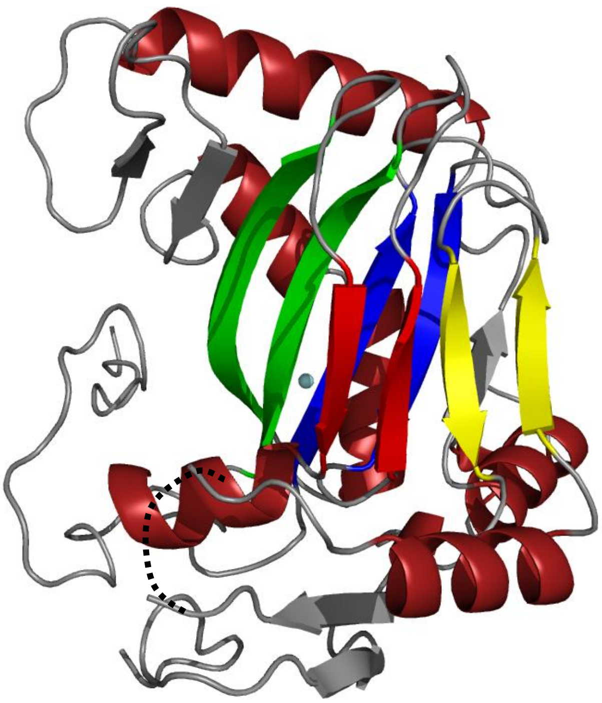


Figure 9.2: Ribbon representation of ectoine hydroxylase, EctD, with the successive segments of the double-stranded β-helix (DSBH) colored according to the scheme of Branden & Tooze.[149] The bound Fe^{3+} is shown as a bright blue sphere. A disordered putative loop region connecting DSBH β-strands IV and V is indicated by a dashed line.

9.5 Validation of the Method Using Different EctD Variants

To characterize wild-type EctD enzyme and different variants of EctD with respect to bivalent cation binding, we used the above-described fluorescence-based thermal shift assay.[139] The method exploits the energetic coupling between ligand binding and protein unfolding, typically resulting in increased thermal protein stability in the presence of a specific ligand. Accordingly, the protein "melting point" (T_m), defined as midpoint temperature of the protein-unfolding transition, increases as a function of concentration and affinity of the ligand. In agreement with observations made by Ericsson et al.[150] in thermofluor-based studies, no proper melting curves of EctD could be obtained in the presence of Fe(II) salts, since Fe^{2+} apparently quenches the fluoroprobe. Therefore, we tested if one of the bivalent cations, Mn^{2+}, Co^{2+}, Ni^{2+} and Zn^{2+}, was able to substitute for Fe^{2+} within the active centre of EctD and thereby increase the thermal stability of the protein. A number of crystal structures of Fe^{2+} and 2-oxoglutarate dependent dioxygenases have shown that Fe^{2+} can often be replaced by these cations although the proteins complexed with these non-physiological cations are enzymatically inactive.[151-155] Surprisingly, the presence of Co^{2+}, Ni^{2+} or Zn^{2+} led, even at low concentrations (less than 1 mmol·L^{-1}), to the immediate denaturation and precipitation of the protein. Precipitation of EctD by Co^{2+} and Ni^{2+} could be prevented by the addition of ectoine at a concentration of 100 mmol·L^{-1}. The underlying mechanisms for this observation are unclear at present but it seems possible that ectoine serves as a chelator for various bivalent cations. In contrast, Mn^{2+} could be added to EctD at concentrations of up to 100 mmol·L^{-1} without precipitating the protein. The T_m of EctD (~ 32 °C in the absence of any ligand) was significantly increased when the Mn^{2+} concentration was raised (Figure 9.3 a). At a concentration of 10 to 20 mmol·L^{-1} of Mn^{2+}, maximal thermostabilization of EctD was reached resulting in an increase in T_m by ~ 6 °C. This Mn^{2+}-dependent thermostabilization was not observed for the EctD(Asp148Ala) and, compared to wild-type EctD, drastically reduced for the EctD(His146Ala), EctD(Asp148Glu) and EctD(His248Ala) variants (Figure 9.3 a). These observations indicate that a substantial loss in affinity of EctD to the Fe^{2+} substitute Mn^{2+} occurs when a single amino acid residue of the Fe^{2+}-complexing 2-His-1-carboxylate facial triad is altered. This view is consistent with the lack of enzymatic activity of the mutated EctD enzymes under standard assay conditions and our inability to detect any iron in the enzyme preparations of these variant EctD proteins. Pavel et al.[156] have shown, that binding of Fe^{2+} to the active centre of the Fe^{2+} and 2-oxoglutarate dependent dioxygenase, clavaminate synthase, increases the affinity of the co-substrate 2-oxoglutarate to the enzyme by an order of magnitude. To investigate if, conversely, binding of the Fe^{2+} substitute Mn^{2+} to the EctD protein is influenced by the co-substrate, we determined the T_m of EctD at various Mn^{2+}-concentrations in the presence of saturating concentrations of 2-oxoglutarate. In the presence of excess co-substrate, a Mn^{2+} concentration as low as 0.5 mmol·L^{-1} induced a raise in T_m by ~ 9 °C (Figure 9.3 b).

The highest applied Mn^{2+} concentration (50 mmol·L^{-1}) caused a raise in T_m by ~ 10.5 °C. This observation suggests that the presence of the co-substrate 2-oxoglutarate causes an enhanced affinity of the EctD enzyme for Mn^{2+} and indicates - by inference - that binding of the natural ligand Fe^{2+} by EctD might be affected in a similar fashion as well.

We also investigated if Mn^{2+} was, in the presence of saturating concentrations of 2-oxoglutarate, able to cause thermostabilization of the mutated EctD variants defective in iron binding, EctD(His146Ala), EctD(Asp148Ala), EctD(His248Ala) and EctD(Asp148Glu). At a Mn^{2+} concentration of 50 mmol·L^{-1} an increase in T_m by ~ 7.5 °C was observed for EctD(His146Ala) and by ~ 5.5 °C for EctD(Asp148Glu) indicating that also for these variants excess amounts of 2-oxoglutarate cause a gain of affinity to Mn^{2+}. An effect of the co-substrate 2-oxoglutarate on Mn^{2+} affinity was noted for EctD(His248Ala) as well. For this variant, T_m was increased by ~ 3.5 °C at a Mn^{2+} concentration of 50 mmol·L^{-1}. A T_m shift of less than 1 °C, which was observed under these conditions for EctD(Asp148Ala) was not significant considering the error range of the used thermofluor assay (Figure 9.3 b).

In summary, the data documented above provide evidence that binding of the bivalent cation to the EctD enzyme is facilitated by the presence of the co-substrate 2-oxoglutarate. Since Fe^{2+} will certainly exhibit a considerably higher affinity to EctD than Mn^{2+}, this may be of limited relevance for enzymes with an intact 2-His-1-carboxylate facial triad. For SyrB2-like halogenases, however, in which the acidic residue of this triad is naturally replaced by alanine,[157] binding of 2-oxoglutarate may be a prerequisite for efficient iron binding and enzymatic activity. In this context, it is noteworthy that all available crystal structures of SyrB2 and of FIH variants with mutated Asp201 which contain a bound Fe^{2+}, contain the co-substrate 2-oxoglutarate as well.[157, 158]

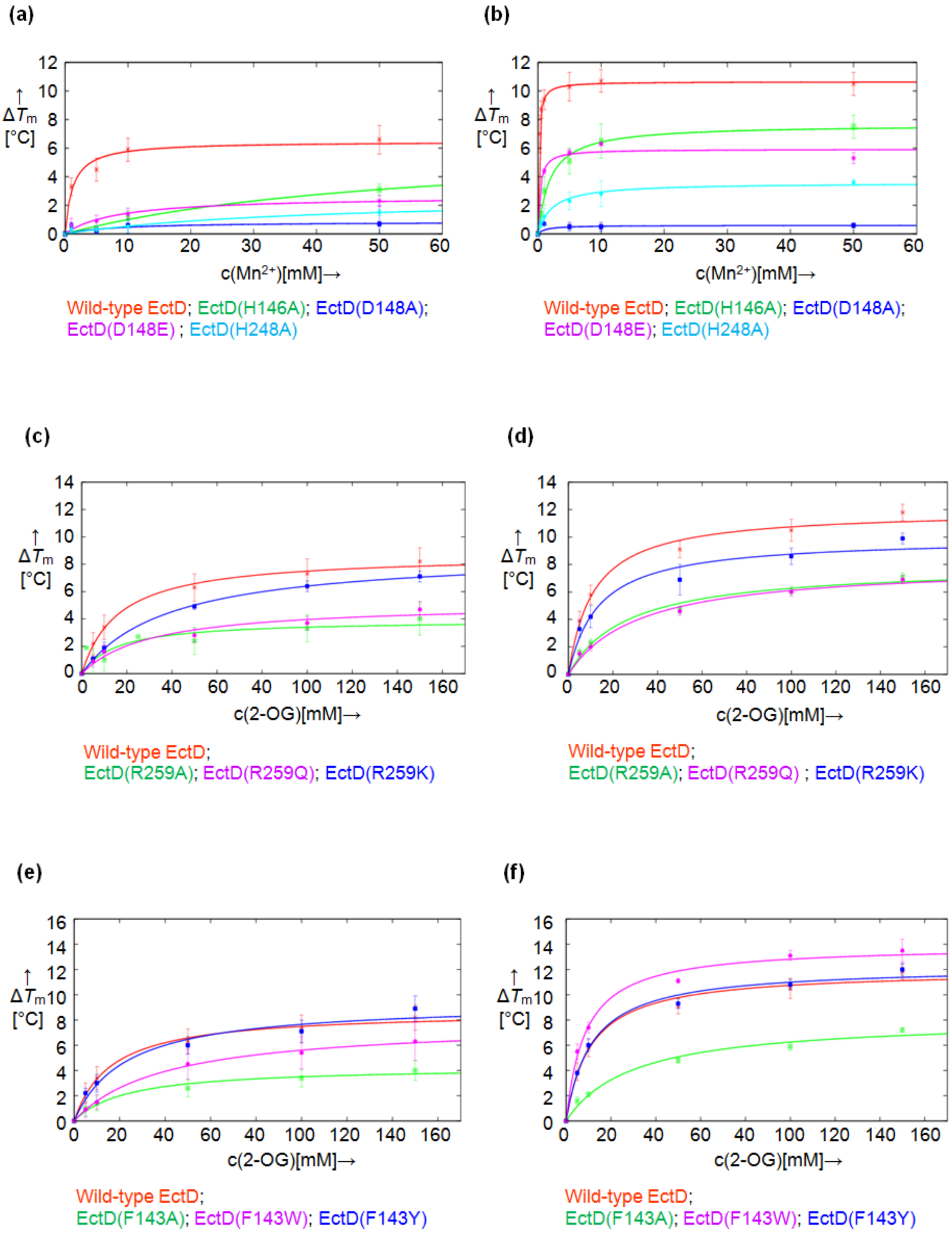


Figure 9.3: Thermostabilization of EctD and mutated derivatives dependent on ligand concentration. Each ΔT_m value represents the mean value of multiple (at least three) individual measurements. (a) Concentration dependent thermostabilizing effect of Mn^{2+} on wild-type EctD and variants with mutated Fe^{2+} chelating residues. (b) as (a), but in the presence of 100 mmol·L^{-1} 2-oxoglutarate. (c) Concentration dependent thermostabilizing effect of 2-oxoglutarate on wild-type EctD and variants in which Arg259 was changed to alanine, glutamine or lysine. (d) as (c), but in the presence of 50 mmol·L^{-1} Mn^{2+}. (e)

Concentration dependent thermostabilizing effect of 2-oxoglutarate on wild-type EctD and variants in which Phe143 was changed to alanine, tyrosine or tryptophan. (f) as (e), but in the presence of 50 mmol·L^{-1} Mn^{2+}, 2-OG, 2-oxoglutarate.

9.6 Residues Likely to be Involved in 2-Oxoglutarate Binding by EctD

In all Fe^{2+} and 2-oxoglutarate dependent dioxygenases the co-substrate is bound at the bottom of the binding cavity and participates in the coordination of the Fe^{2+} ion *via* its 1-carboxylate and 2-oxo moiety in a bidentate manner. The 5-carboxylate is typically stabilized by a salt bridge formed with the basic group of an arginine or lysine side chain and by at least one hydrogen bond formed to a hydroxyl group of the protein.[144] With few exceptions, the basic residue salt bridging the 5-carboxylate of 2-oxoglutarate protrudes from the amino-terminus of the DSBH β-strand VIII. In 57 EctD-type proteins compiled by us this residue is invariantly conserved and corresponds to Arg259 in the *V. salexigens* EctD enzyme (Figure 9.2). In order to substantiate the idea that this residue is involved in 2-oxoglutarate binding by EctD, we changed Arg259 to alanine, glutamine or lysine residues, respectively. All these EctD variants were enzymatically inactive under standard assay conditions, consistent with the predicted role of Arg259 in 2-oxoglutarate binding. We also analyzed binding of 2-oxoglutarate by the wild type EctD enzyme with the fluorescence-based thermal shift assay and compared these data with a corresponding analysis of the EctD(Arg259Ala), EctD(Arg259Gln) and EctD(Arg259Lys) variants. For the wild-type EctD protein, a substantial 2-oxoglutarate dependent thermostabilization was observed with T_m being increased by ~ 8 °C at a co-substrate concentration of 150 mmol·L^{-1}. In the presence of saturating amounts of Mn^{2+} the cofactor-dependent thermostabilization of EctD was more pronounced since a 2-oxoglutarate concentration of 150 mmol·L^{-1} led to an increase in the T_m by more than 11 °C (Figure 9.3 c - f). In accordance with the results obtained by Pavel et al.[156] for clavaminate synthase, our data suggest that an increased affinity of EctD for the co-factor 2-oxoglutarate is caused by the bivalent cation. Saturating concentrations of Mn^{2+} or the co-factor 2-oxoglutarate alone caused a thermostabilization by ~ 6 °C or ~ 8 °C, respectively. However, the combination of both Mn^{2+} and 2-oxoglutarate resulted in an increase in T_m by more than 17 °C in comparison to the apo-form of EctD, suggesting a cooperative effect on thermostabilization of EctD by the binding of the bivalent cation and the co-substrate 2-oxoglutarate.

The EctD(Arg259Ala) and EctD(Arg259Gln) variants are catalytically inactive under standard hydroxylation assay conditions. As expected, a considerably lower 2-oxoglutarate dependent thermostabilization was observed for EctD(Arg259Ala) and EctD(Arg259Gln) in comparison to the wild-type enzyme. In the absence of Mn^{2+}, an increase in T_m by ~ 4 °C was observed for both variants at 150 $mmol \cdot L^{-1}$ 2-oxoglutarate. In the presence of saturating amounts of Mn^{2+}, 150 $mmol \cdot L^{-1}$ 2-oxoglutarate led to an increase of T_m by ~ 7 °C (Figure 9.3 c and d). This is a clear indication that each of the introduced mutations causes the loss of a protein-stabilizing interaction formed between EctD and the co-substrate. A less pronounced decrease in 2-oxoglutarate dependent thermostabilization was observed for EctD (Arg259Lys) which is also catalytically inactive under standard assay conditions. A concentration of 150 $mmol \cdot L^{-1}$ 2-oxoglutarate resulted in an increase of the T_m by ~ 7 °C. In the presence of 50 $mmol \cdot L^{-1}$ Mn^{2+} this co-substrate concentration led to a raise of the T_m by ~ 10 °C (Figure 9.3 c and d). This suggests that the amino group of lysine is partially able to fullfil the role of the arginine guanidino function in salt bridging the 5-carboxylate of 2-oxoglutarate. However, the observation that a lysine residue is with respect to co-substrate binding only partially able to substitute for Arg259, emphasises the importance of the guanidino function in efficiently binding the 2-oxoglutarate. This is fully consistent with the finding that Arg259 is strictly conserved in 57 EctD-type ectoine hydroxylases compiled by us through data base searches of either completely or partially finished microbial genome sequences.

9.7 Materials and Methods: Fluorescence-Based Thermal Shift Assay

The thermal shift assay was carried out in 96-well thin-wall PCR plates (Bio-Rad) essentially as described in literature[139] using an iQ^{TM}5 Real Time PCR System (Bio-Rad). SYPRO orange served as a fluorescent dye to monitor protein unfolding within a temperature range between 20 °C and 80 °C in increments of 0.5 °C. The wavelengths for excitation and emission were 490 and 575 nM, respectively. The final mass concentration of EctD and its mutated variants was 200 µg·mL^{-1} (~ 6 µmol·L^{-1}). Each sample was buffered with 10 mmol·L^{-1} HEPES pH 7.5 in a volume of 100 µL. The thermofluor data were analyzed by means of an in-house Python script, the figures were generated using Gnuplot.[132]

10 RNA Editing Modulates the Binding of Drugs and Highly Unsaturated Fatty Acids to the Open Pore of Kv Potassium Channels.

10.1 Introductory Remarks

The present study was accomplished in cooperation with Nils Decher, Anne K. Streit and Jürgen Daut (Univ. of Marburg, Dept. of Physiological Chemistry). The experimental observation indicated that unsaturated fatty acids could block Kv potassium channels. To explain this phenomenon, we performed modelling and docking experiments. The results have been published in 'The EMBO Journal'[159] The contribution to this publication is the docking of anandamide (AEA; N-arachidonylethanolamide) into the pore channel and the modeling of a potassium ion with its square-antiprismatic solvation shell at the entrance of the blocked channel. In the following chapter, the focus will be on the modelling part of this study.

10.2 Abstract

The time course of inactivation of voltage-activated potassium (Kv) channels is an important determinant of the firing rate of neurons. In many Kv channels highly unsaturated lipids as arachidonic acid, docosahexaenoic acid and anandamide can induce fast inactivation. We found that these lipids interact with hydrophobic residues lining the inner cavity of the pore. We analyzed the effects of these lipids on Kv1.1 current kinetics and their competition with intracellular tetraethylammonium and Kvβ subunits. Our data suggest that inactivation most likely represents occlusion of the permeation pathway, similar to drugs that produce 'open-channel block'. This open-channel block by drugs and lipids was strongly reduced in Kv1.1 channels whose amino acid sequence was altered by RNA editing in the pore cavity, and in Kv1.x heteromeric channels containing edited Kv1.1 subunits. We show that differential editing of Kv1.1 channels in different regions of the brain can profoundly alter the pharmacology of Kv1.x channels. Our findings provide a mechanistic understanding of lipid-induced

inactivation and establish RNA editing as a mechanism to induce drug and lipid resistance in Kv channels.

10.3 Docking and Modelling Results

10.3.1 Arachidonic acid (AA) can physically occlude the pore

A pore homology model that was generated on the basis of the crystal structure of the open channel[160] shows that the residues identified by our Ala-scan face into the central cavity (Figure 10.1 D).

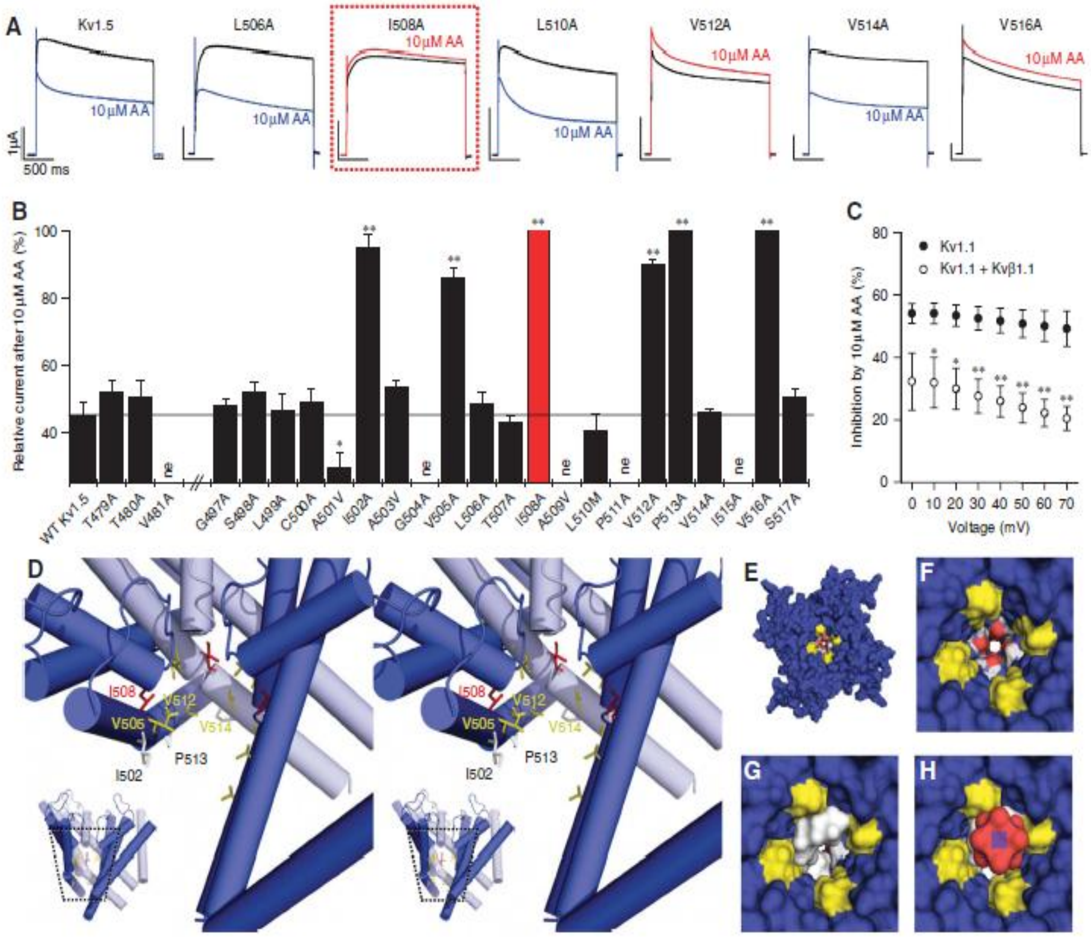


Figure 10.1: The binding site for HUFAs in Kv1 channels. (A) Currents measured during voltage steps to +40 mV in Xenopus oocytes expressing WT or mutant Kv1.5 channels. Blue traces: WT and non-pore-facing mutants in the presence of arachidonic acid (AA). Red traces: pore-facing mutants in the presence of AA. The mutation at site I508 (homologous to I400 in Kv1.1) is shown in a red box. (B) Ala-scanning mutagenesis of the pore domain of Kv1.5. Relative current (%) remaining after addition of AA, determined at the end of 1.5 s pulses to +40mV (ne, non-expressing); * indicates $P<0.05$, ** indicates $P<0.01$. (C) Relative inhibition of Kv1.1 by AA, measured with (J) or without (K) co-expression of Kvβ1.1. No significant voltage dependence of block was observed in the voltage range of 0 to +70 mV for both, Kv1.1 alone or Kv1.1+Kvβ1.1; * indicates $P<0.05$, ** indicates $P<0.01$. (D) Stereo-view of the channel shown with the lipid-binding site identified by the Ala-scan. For a better view into the central cavity only three subunits are depicted. The inset illustrates from which orientation of the homology model the close-up was made. Residue I508 is shown in red, valines in yellow, proline and isoleucines in white. (E–H) The open channel viewed from the cytosolic side. The solvent-accessible surface area of the protein is shown in blue. The PVP motif of the channel is highlighted in yellow. Residue T480 near the selectivity filter is indicated in red and white. (F) Closer view of the empty open-channel cavity shown in (E). (G) Spacefilling model of AEA (gray) in the cavity of the channel. (H) AEA (gray) and a potassium ion (purple) with its square-antiprismatic salvation shell in the entrance of the blocked channel.

To test the plausibility of our hypothesis that arachidonic acid (AA), docosahexaenoic acid (DHA) and anandamide (AEA) can enter the inner cavity of Kv1 channels, we docked AEA into this pore homology model. In most cases, AEA penetrated into the cavity, placing the ethanolamide group next to the threonine residues at the entrance of the selectivity filter and forming extended contacts with non-polar residues lining the wall of the channel cavity. The docking program GOLD[161, 162] found multiple solutions that differed in the orientation of the C20 tail of the ligand, suggesting that there is not one distinct binding orientation for the probe ligand. Some of the most prominent solutions are shown in Figure 10.2.

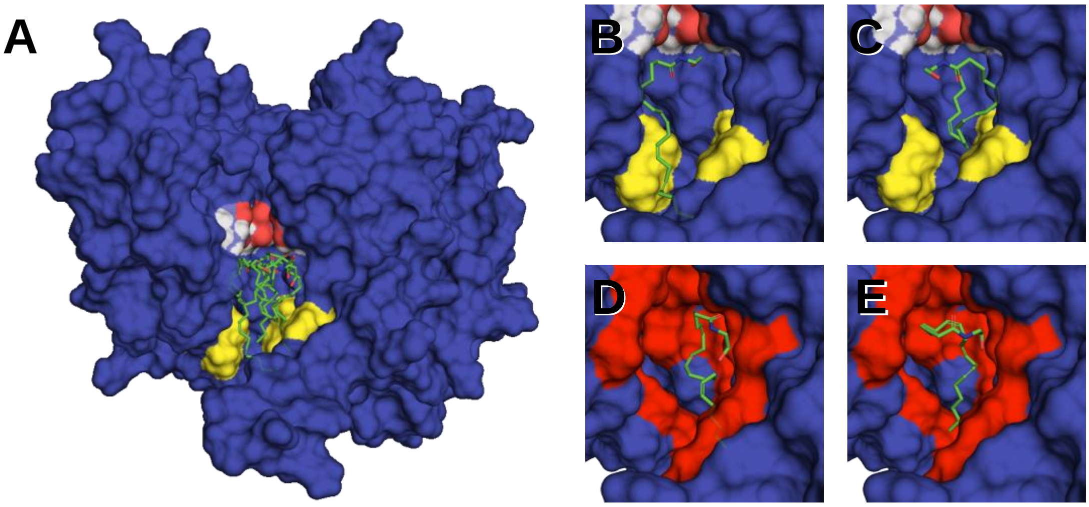


Figure 10.2 Model calculations for docking of AEA in the Kv1.5 pore. The solvent-accessible surface area of the protein is shown in blue. Yellow highlights the PVP motif of the channel. Red and white marks T480 near the selectivity filter. (A) Three subunits of the channel are shown with five different docking solutions for AEA (green). (B-E) Different docking solutions for AEA. (B and C) Docking solutions with the PVP motif highlighted in yellow. (D and E) Docking solutions with the lipid binding site identified by the Ala-scan and shown as a red surface.

These relatively crude docking considerations are consistent with the idea that the pore is wide enough to fully accommodate a molecule such as AEA (Figure 10.1 E–G) and that the binding of the ligand (Figure 10.1 G) will significantly narrow the pore width so that its remaining diameter is too small to allow a potassium ion together with its square-antiprismatic solvation shell ($K\ (H_2O)_8^+$) to pass through the channel (Figure 10.1 H).

Conclusion and Outlook

In recent years, fragment screening has become a popular approach to identify new lead structures. Fragments are usually defined by the Astex ‘rule of three’ (RO3) (molweight < 300 Da, H-bond acceptors and donors < 3, logP < 3, total polar surface area < 60 Å^2). Surface Plasmon Resonance (SPR), Nuclear Magnetic Resonance spectroscopy (NMR), biochemical assays and X-ray crystallography are efficient screening techniques to discover prospective fragments as binders. However , these methods need an assembled fragment library.

We designed an in-house fragment library, starting from approx. 380,000 commercially available fragments. During library design, we modified the RO3 and we did no strict filtering of physico-chemical properties during fragment enumeration (e.g. twice the number of H-bond acceptors was allowed). The fragments were stepwise reduced to 4,000 compounds. The last step was a visual inspection of the candidates, which lead to a final fragment library of 364 fragments. To validate the quality of the library, we screened it against endothiapepsin. The biochemical screening suggested 55 hits, which were entered into a crystallographic screen. Eleven complex crystal structures were determined, pointing out the remarkably high hit rate of the designed library.

HotspotsX is a program which predicts (based on knowledge-based potentials) the probability of a certain atom type at a certain position in the binding pocket of a target enzyme. The eleven crystal structures obtained before were used to validate the program HotspotsX. Due to chemical diversity and the different binding modes of the fragments observed for the library examples we obtained binding through aromatic- , H-bond donor- , acceptor- , doneptor- and hydrophobic interactions. The calculated HotspotsX maps coincide remarkably well with the crystallographically determined fragment positions inside the binding pocket.

The program HotspotsX has also been validated with crystal structures of molecular probes like phenol, urea and methylurea. Crystal structures of these molecular probes were determined with different targets. Overall, the experimental hotspot analysis coincided well with the computed contour maps. Thus, the calculated maps by HotspotsX have an excellent predictive power.

Based on the binding modes of the molecular probe phenol to the cAMP-dependent protein kinase A (PKA), we started a fragment growing approach. In the latter complex, three phenol molecules are bound. Two are occupying the ATP binding site and one is sitting on top of the glycine-rich loop (G-loop). A virtual screening, using the hinge binding phenol as constraint, suggested a phenol derivative for which a crystal structure could be determined. Starting from this hit, a hotspot analysis was performed. This analysis indicates that growth in the direction of the G-loop, placing an aromatic portion under the G-loop and an acceptor functionality capable to address Lys72 is desired. The first compound of this *de novo* design had an affinity of 70 µM.

In the following first design cycle, we were able to enhance the affinity to 6.5 µM. In the second design cycle an additional amino function was introduced, which did not improve affinity dramatically, but enhanced ligand efficiency to 0.38. In the last cycle, a spacer of one and two methylene groups was introduced and the affinity could be increased to about 110 nM for a diastereomeric mixture of four compounds.

The phenol-PKA complex provides a putative allosteric site of PKA. The G-loop in this structure is in a closed state which is stabilized by two H-bonds. This G-loop conformation is probably induced by the phenol molecule sitting on top of the G-loop. Therefore, several molecular dynamics (MD) studies were performed, lacking different phenol molecules, to get insights into the G-loop opening. The MD studies suggest that after removal of the phenol sitting on top of the G-loop some first side chain movements are initiated that can indicate the first steps of the G-loop opening cascade.

In a different project, a virtual screening approach was used to find new inhibitors for aldose reductase. A pre-filtered subset of the ZINC database was used as ligand dataset. For the best hit, a series of five compounds was synthesized. Among them one compound displayed an inhibition of 920 nM.

The available assays to detect fragment hits are currently not sufficient. The challenges are the low affinity of the fragments and their poor solubility. Therefore, the known thermal shift assay was applied and adapted to detect fragment hits. To validate the method, it was used to characterize variant mutations of EctD.

Lastly, a modeling study was used to get ideas about possible binding modes of arachidonic acid derivatives in a K^+ ion channel. One predominant binding pose could not be suggested. The study proposes, however, that one arachidonic acid molecule can occupy the inner pore cavity, which is consistent with experimental data.

Zusammenfassung und Ausblick

In den letzten Jahren wurden vermehrt Fragment-basierte Verfahren verwendet, um neue Leitstrukturen zu identifizieren. Fragmente werden normalerweise anhand der Astex-Dreier-Regel (RO3) (Molekülgewicht < 300 Da, Waserstoffbrücken Akzeptoren und Donoren < 3, logP < 3, gesamte zugängliche polare Oberfläche < 60 $Å^2$) definiert. Oberflächenplasmonenresonanzspektroskopie, Kernspinresonanzspektroskopie sowie biochemische Assays und Röntgenstrukturanalyse sind effiziente und gut entwickelte Verfahren um Fragmente zu entdecken. Diese Methoden müssen jedoch auf eine bereits bestehende Fragmentbibliothek angewendet werden.

Wir haben daher unsere eigene Fragmentbibliothek entwickelt, wobei als Startpunkt rund 380.000 kommerziell erhältliche chemische Verbindungen dienten. Während des Designs der Fragmentbibliothek haben wir die Astex-Dreier-Regel modifiziert. Es wurde kein strikter physiko-chemischer Filter verwendet (z.B. wurde die Anzahl der Wasserstoffbrückenakzeptoren verdoppelt). Die Fragmente wurden so schrittweise auf 4000 Strukturen heruntergefiltert. Im letzen Schritt wurden sie visuell inspiziert, was zu einer Bibliothek mit 364 Fragmenten führte. Um die Güte der Bibliothek zu überprüfen, haben wir diese gegen Endothiapepsin getestet. Als Ergebnis erhielten wir 55 Fragmente, welche den Umsatz eines fluorogenen Substrates durch das Zielenzym reduzieren. Diese wurden anschließend kristallographisch untersucht, wobei elf Kristallstrukturen bestimmt werden konnten, was eine hervorragende Trefferquote darstellt.

Das Programm HotspotsX kann aufgrund von wissensbasierten Potentialen die Aufenthaltswahrscheinlichkeit von einem definierten Atomtyp in einer bestimmten Umgebung der Bindetasche des Zielenzyms vorhersagen. Mit Hilfe der zuvor erwähnten elf Fragmentkristallstrukturen haben wir das Programm HotspotsX validiert. Durch die chemische Diversität und die diversen Posen der Fragmente erhielten wir Hinweise auf aromatische, Donor, Akzeptor, Doneptor und hydrophobe Binder. Die berechneten HotspotsX Karten passen hervorragend zu den experimentell ermittelten Bindungsposen der Fragmente innerhalb der Bindetasche.

Das Programm HotspotsX wurde ebenfalls an Kristallstrukturen von Sonden-Molekülen wie Phenol, Harnstoff und Methylharnstoff getestet. Die Bindungsposen dieser Sonden konnten in verschiedenen Zielenzymen mit Hilfe von Röntgenstrukturanalysen ermittelt werden. Die meisten experimentell bestimmten bevorzugten Bindungsregionen passten hervorragend mit den computer-vorhergesagten Positionen überein, was die exzellente Vorhersageleistung des Programmes unterstreicht.

Beginnend mit der Sondenstruktur von Phenol in der cAMP abhängigen Protein Kinase A (PKA) haben wir Fragment-basierte Prinzipien des *de novo* Designs angewendet. Sie beruhen auf dem Wachstum von inital entdeckten Fragmenten. In der Kristallstruktur findet man drei Phenolmoleküle, wobei zwei die ATP-Bindestelle besetzen und das andere auf der Glycin-reichen Schleife (G-Schleife) sitzt. Eine computer-basierte Suche wurde durchgeführt, wobei für die Leitfragmente eine Bindungsgeometrie ähnlich dem Phenolmolekül an der Scharnier-Region vorausgesetzt wurde. Es wurden Phenol-artige Strukturen vorgeschlagen und für eine konnte eine Kristallstruktur bestimmt werden. Diese Struktur wurde verwendet, um die bevorzugten Bindungsregionen der Bindungstasche auszuleuchten. Das Programm schlug vor, den Liganden in Richtung der G-Schleife zu wachsen, in dieser Region eine aromatische Gruppe zu platzieren und Lys72 mit einer Akzeptorgruppe zu adressieren. Die erste Testverbindung hatte einen Inhibitionswert von 70 µM. Im folgenden ersten Design-Zyklus konnte die Affinität auf 6.5 µM gesteigert werden. Im zweiten Zyklus wurde eine zusätzliche Aminogruppe an die Leitstruktur synthetisiert, was die Affinität nur wenig steigerte, aber das Verhältnis Schweratome zu Affinität auf 0.38 anhob. Im letzten Zyklus wurden eine bzw. zwei Methylen-Gruppen als Brücken für die Aminogruppe synthetisiert. Die Affinität für das synthetisierte Diastereomerengemisch zusammengesetzt aus vier verschiedenen Isomeren zeigte eine Affinität von ca. 110 nM.

Die Struktur des Phenol-PKA Komplexes zeigte eine mögliche allosterische Bindetasche. Die G-Schleife in diesem Komplex liegt in einer eingeklappten Konformation vor und wird durch zwei Wasserstoffbrücken stabilisiert. Diese Konformation könnte durch eines der drei Phenolmoleküle, das auf der G-Schleife sitzt, erzwungen werden. Mehrere Moleküldynamik (MD) Berechnungen wurden durchgeführt, wobei verschiedene Kombinationen bezüglich der Besetzung der drei Phenolmoleküle ausprobiert wurden, um einen Einblick in erste Schritte bei dem Öffnen der Schleife zu bekommen. Die MD Simulation, bei welcher das Phenolmolekül auf der Schleife fehlte, zeigte erste Anzeichen für ein Öffnen der Schleife, was die ersten Schritte eines kaskadenartigen Öffnens darstellen könnte.

In einem weiteren Projekt wurde eine virtuelle Suche nach neuen Leitstrukturen der Aldose Reduktase durchgeführt. Dafür wurde die gefilterte ZINC Datenbank als Ligand-Datenbank genutzt. Von den besten Leitstrukturen wurden fünf Verbindungen synthetisiert und die Affinität gemessen. Unter diesen Leitstrukturen war eine Verbindung, die eine Affinität von 920 nM aufwies.

Die etablierten Affinitäts-Testsysteme sind, um Fragmente als Binder zu finden, noch nicht ausreichend. Die Herausforderungen liegen in der schwachen Affinität und der schlechten Löslichkeit der Fragmente. Daher wurde der bekannte Temperatur-Stabilitäts-Test auf Fragmente angewendet. Um die Methode zu etablieren, wurden verschiedene Mutanten von EctD charakterisiert.

Im letzten Projekt wurden Bindungsposen von Arachidonsäurederivaten in einem K^+ Kanal erzeugt, um eine Vorstellung zu bekommen, wie die Bindung aussehen könnte. Eine genaue Bindungspose konnte nicht bestimmt werden, es konnte allerdings gezeigt werden, dass nur ein einzelnes Arachidonsäuremolekül die innere Pore des Kanals blockieren kann. Dies ist konsistent mit den experimentellen Befunden.

Literature

1. 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 Deliv Rev* **2001,** 46, 3-26.
2. Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today* **2005,** 10, 464-9.
3. Jahnke, W. *Fragment-based approaches in drug discovery*. Wiley-VCH-Verl.: Weinheim, 2006; p XXII, 369 S.
4. Blundell, T. L.; Patel, S. High-throughput X-ray crystallography for drug discovery. *Curr Opin Pharmacol* **2004,** 4, 490-6.
5. Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **2005,** 48, 403-13.
6. Tickle, I.; Sharff, A.; Vinkovic, M.; Yon, J.; Jhoti, H. High-throughput protein crystallography and drug discovery. *Chem Soc Rev* **2004,** 33, 558-65.
7. Dalvit, C.; Fasolini, M.; Flocco, M.; Knapp, S.; Pevarello, P.; Veronesi, M. NMR-Based screening with competition water-ligand observed via gradient spectroscopy experiments: detection of high-affinity ligands. *J Med Chem* **2002,** 45, 2610-4.
8. Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, G. W.; Ajay; Murcko, M. A.; Moore, J. M. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem Biol* **1999,** 6, 755-69.
9. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** 274, 1531-4.
10. Blundell, T. L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* **2002,** 1, 45-54.
11. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **2003,** 8, 876-7.
12. Siegal, G.; Ab, E.; Schultz, J. Integration of fragment screening and library design. *Drug Discov Today* **2007,** 12, 1032-9.
13. Congreve, M.; Aharony, D.; Albert, J.; Callaghan, O.; Campbell, J.; Carr, R. A.; Chessari, G.; Cowan, S.; Edwards, P. D.; Frederickson, M.; McMenamin, R.; Murray, C. W.; Patel, S.; Wallis, N. Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of beta-secretase. *J Med Chem* **2007,** 50, 1124-32.
14. de Kloe, G. E.; Bailey, D.; Leurs, R.; de Esch, I. J. Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discov Today* **2009,** 14, 630-46.
15. Erlanson, D. A. Fragment-based lead discovery: a chemical update. *Curr Opin Biotechnol* **2006,** 17, 643-52.
16. Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-based drug discovery. *J Med Chem* **2004,** 47, 3463-82.
17. Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* **2001,** 41, 856-64.

18. Jhoti, H. A new school for screening. *Nat Biotechnol* **2005,** 23, 184-6.

19. Keseru, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discov Today* **2006,** 11, 741-8.

20. Leach, A. R.; Hann, M. M.; Burrows, J. N.; Griffen, E. J. Fragment screening: an introduction. *Mol Biosyst* **2006,** 2, 430-46.

21. Makara, G. M. On sampling of fragment space. *J Med Chem* **2007,** 50, 3214-21.

22. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat Rev Drug Discov* **2004,** 3, 660-72.

23. Zartler, E. R.; Shapiro, M. J. Fragonomics: fragment-based drug discovery. *Curr Opin Chem Biol* **2005,** 9, 366-70.

24. Hajduk, P. J. Puzzling through fragment-based drug design. *Nat Chem Biol* **2006,** 2, 658-9.

25. Barker, J. J.; Barker, O.; Courtney, S. M.; Gardiner, M.; Hesterkamp, T.; Ichihara, O.; Mather, O.; Montalbetti, C. A.; Muller, A.; Varasi, M.; Whittaker, M.; Yarnold, C. J. Discovery of a novel Hsp90 inhibitor by fragment linking. *ChemMedChem* 5, 1697-700.

26. Brough, P. A.; Aherne, W.; Barril, X.; Borgognoni, J.; Boxall, K.; Cansfield, J. E.; Cheung, K. M.; Collins, I.; Davies, N. G.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Finch, H.; Fink, A.; Hayes, A.; Howes, R.; Hubbard, R. E.; James, K.; Jordan, A. M.; Lockie, A.; Martins, V.; Massey, A.; Matthews, T. P.; McDonald, E.; Northfield, C. J.; Pearl, L. H.; Prodromou, C.; Ray, S.; Raynaud, F. I.; Roughley, S. D.; Sharp, S. Y.; Surgenor, A.; Walmsley, D. L.; Webb, P.; Wood, M.; Workman, P.; Wright, L. 4,5-diarylisoxazole Hsp90 chaperone inhibitors: potential therapeutic agents for the treatment of cancer. *J Med Chem* **2008,** 51, 196-218.

27. Cheung, K. M.; Matthews, T. P.; James, K.; Rowlands, M. G.; Boxall, K. J.; Sharp, S. Y.; Maloney, A.; Roe, S. M.; Prodromou, C.; Pearl, L. H.; Aherne, G. W.; McDonald, E.; Workman, P. The identification, synthesis, protein crystal structure and in vitro biochemical evaluation of a new 3,4-diarylpyrazole class of Hsp90 inhibitors. *Bioorg Med Chem Lett* **2005,** 15, 3338-43.

28. Dymock, B. W.; Barril, X.; Brough, P. A.; Cansfield, J. E.; Massey, A.; McDonald, E.; Hubbard, R. E.; Surgenor, A.; Roughley, S. D.; Webb, P.; Workman, P.; Wright, L.; Drysdale, M. J. Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. *J Med Chem* **2005,** 48, 4212-5.

29. Huth, J. R.; Park, C.; Petros, A. M.; Kunzer, A. R.; Wendt, M. D.; Wang, X.; Lynch, C. L.; Mack, J. C.; Swift, K. M.; Judge, R. A.; Chen, J.; Richardson, P. L.; Jin, S.; Tahir, S. K.; Matayoshi, E. D.; Dorwin, S. A.; Ladror, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Fesik, S. W.; Elmore, S. W.; Hajduk, P. J. Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. *Chem Biol Drug Des* **2007,** 70, 1-12.

30. Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; Fazal, L.; Feltell, R.; Figueroa, E.; Frederickson, M.; Lewis, J.; McMenamin, R.; Murray, C. W.; O'Brien, M. A.; Parra, L.; Patel, S.; Phillips, T.; Rees, D. C.; Rich, S.; Smith, D. M.; Trewartha, G.; Vinkovic, M.; Williams, B.; Woolford, A. J. Discovery of (2,4-dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-di hydroisoindol-2-yl]methanone (AT13387), a novel

inhibitor of the molecular chaperone Hsp90 by fragment based drug design. *J Med Chem* 53, 5956-69.

31. Caldwell, J. J.; Davies, T. G.; Donald, A.; McHardy, T.; Rowlands, M. G.; Aherne, G. W.; Hunter, L. K.; Taylor, K.; Ruddle, R.; Raynaud, F. I.; Verdonk, M.; Workman, P.; Garrett, M. D.; Collins, I. Identification of 4-(4-aminopiperidin-1-yl)-7H-pyrrolo[2,3-d]pyrimidines as selective inhibitors of protein kinase B through fragment elaboration. *J Med Chem* **2008,** 51, 2147-57.

32. Donald, A.; McHardy, T.; Rowlands, M. G.; Hunter, L. J.; Davies, T. G.; Berdini, V.; Boyle, R. G.; Aherne, G. W.; Garrett, M. D.; Collins, I. Rapid evolution of 6-phenylpurine inhibitors of protein kinase B through structure-based design. *J Med Chem* **2007,** 50, 2289-92.

33. Frederickson, M.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S. R.; Matthews, J. E.; McMenamin, R.; Smith, D. M.; Vinkovic, M.; Wallis, N. G. Fragment-based discovery of mexiletine derivatives as orally bioavailable inhibitors of urokinase-type plasminogen activator. *J Med Chem* **2008,** 51, 183-6.

34. Medina, J. s. R.; Blackledge, C. W.; Heerding, D. A.; Campobasso, N.; Ward, P.; Briand, J.; Wright, L.; Axten, J. M. Aminoindazole PDK1 Inhibitors: A Case Study in Fragment-Based Drug Discovery. *ACS Medicinal Chemistry Letters* 1, 439-442.

35. Nordstrom, H.; Gossas, T.; Hamalainen, M.; Kallblad, P.; Nystrom, S.; Wallberg, H.; Danielson, U. H. Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment library. *J Med Chem* **2008,** 51, 3449-59.

36. Saxty, G.; Woodhead, S. J.; Berdini, V.; Davies, T. G.; Verdonk, M. L.; Wyatt, P. G.; Boyle, R. G.; Barford, D.; Downham, R.; Garrett, M. D.; Carr, R. A. Identification of inhibitors of protein kinase B using fragment-based lead discovery. *J Med Chem* **2007,** 50, 2293-6.

37. Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S.; Fong, D.; Zhu, Y. L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A* **2008,** 105, 3041-6.

38. Wyatt, P. G.; Woodhead, A. J.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Davis, D. J.; Devine, L. A.; Early, T. R.; Feltell, R. E.; Lewis, E. J.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Reule, M.; Saxty, G.; Seavers, L. C.; Smith, D. M.; Squires, M. S.; Trewartha, G.; Walker, M. T.; Woolford, A. J. Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J Med Chem* **2008,** 51, 4986-99.

39. Liu, G.; Szczepankiewicz, B. G.; Pei, Z.; Janowick, D. A.; Xin, Z.; Hajduk, P. J.; Abad-Zapatero, C.; Liang, H.; Hutchins, C. W.; Fesik, S. W.; Ballaron, S. J.; Stashko, M. A.; Lubben, T.; Mika, A. K.; Zinker, B. A.; Trevillyan, J. M.; Jirousek, M. R. Discovery and structure-activity relationship of oxalylarylaminobenzoic acids as inhibitors of protein tyrosine phosphatase 1B. *J Med Chem* **2003,** 46, 2093-103.

40. Liu, G.; Xin, Z.; Liang, H.; Abad-Zapatero, C.; Hajduk, P. J.; Janowick, D. A.; Szczepankiewicz, B. G.; Pei, Z.; Hutchins, C. W.; Ballaron, S. J.; Stashko, M. A.; Lubben, T. H.; Berg, C. E.; Rondinone, C. M.; Trevillyan, J. M.; Jirousek, M. R.

Selective protein tyrosine phosphatase 1B inhibitors: targeting the second phosphotyrosine binding site with non-carboxylic acid-containing ligands. *J Med Chem* **2003,** 46, 3437-40.

41. Liu, G.; Xin, Z.; Pei, Z.; Hajduk, P. J.; Abad-Zapatero, C.; Hutchins, C. W.; Zhao, H.; Lubben, T. H.; Ballaron, S. J.; Haasch, D. L.; Kaszubska, W.; Rondinone, C. M.; Trevillyan, J. M.; Jirousek, M. R. Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. *J Med Chem* **2003,** 46, 4232-5.

42. Lubbers, T.; Angehrn, P.; Gmunder, H.; Herzig, S. Design, synthesis, and structure-activity relationship studies of new phenolic DNA gyrase inhibitors. *Bioorg Med Chem Lett* **2007,** 17, 4708-14.

43. Oblak, M.; Grdadolnik, S. G.; Kotnik, M.; Jerala, R.; Filipic, M.; Solmajer, T. In silico fragment-based discovery of indolin-2-one analogues as potent DNA gyrase inhibitors. *Bioorg Med Chem Lett* **2005,** 15, 5207-10.

44. Sutherland, A. G.; Alvarez, J.; Ding, W.; Foreman, K. W.; Kenny, C. H.; Labthavikul, P.; Mosyak, L.; Petersen, P. J.; Rush, T. S., 3rd; Ruzin, A.; Tsao, D. H.; Wheless, K. L. Structure-based design of carboxybiphenylindole inhibitors of the ZipA-FtsZ interaction. *Org Biomol Chem* **2003,** 1, 4138-40.

45. Swayze, E. E.; Jefferson, E. A.; Sannes-Lowery, K. A.; Blyn, L. B.; Risen, L. M.; Arakawa, S.; Osgood, S. A.; Hofstadler, S. A.; Griffey, R. H. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J Med Chem* **2002,** 45, 3816-9.

46. Tsao, D. H.; Sutherland, A. G.; Jennings, L. D.; Li, Y.; Rush, T. S., 3rd; Alvarez, J. C.; Ding, W.; Dushin, E. G.; Dushin, R. G.; Haney, S. A.; Kenny, C. H.; Malakian, A. K.; Nilakantan, R.; Mosyak, L. Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design. *Bioorg Med Chem* **2006,** 14, 7953-61.

47. Bruncko, M.; Oost, T. K.; Belli, B. A.; Ding, H.; Joseph, M. K.; Kunzer, A.; Martineau, D.; McClellan, W. J.; Mitten, M.; Ng, S. C.; Nimmer, P. M.; Oltersdorf, T.; Park, C. M.; Petros, A. M.; Shoemaker, A. R.; Song, X.; Wang, X.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H.; Elmore, S. W. Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J Med Chem* **2007,** 50, 641-62.

48. Edwards, P. D.; Albert, J. S.; Sylvester, M.; Aharony, D.; Andisik, D.; Callaghan, O.; Campbell, J. B.; Carr, R. A.; Chessari, G.; Congreve, M.; Frederickson, M.; Folmer, R. H.; Geschwindner, S.; Koether, G.; Kolmodin, K.; Krumrine, J.; Mauger, R. C.; Murray, C. W.; Olsson, L. L.; Patel, S.; Spear, N.; Tian, G. Application of fragment-based lead generation to the discovery of novel, cyclic amidine beta-secretase inhibitors with nanomolar potency, cellular activity, and high ligand efficiency. *J Med Chem* **2007,** 50, 5912-25.

49. Murray, C. W.; Callaghan, O.; Chessari, G.; Cleasby, A.; Congreve, M.; Frederickson, M.; Hartshorn, M. J.; McMenamin, R.; Patel, S.; Wallis, N. Application of fragment screening by X-ray crystallography to beta-secretase. *J Med Chem* **2007,** 50, 1116-23.

50. Wang, Y. S.; Strickland, C.; Voigt, J. H.; Kennedy, M. E.; Beyer, B. M.; Senior, M. M.; Smith, E. M.; Nechuta, T. L.; Madison, V. S.; Czarniecki, M.; McKittrick, B. A.; Stamford, A. W.; Parker, E. M.; Hunter, J. C.; Greenlee, W. J.; Wyss, D. F. Application of fragment-based NMR screening, X-ray crystallography, structure-based

design, and focused chemical library design to identify novel microM leads for the development of nM BACE-1 (beta-site APP cleaving enzyme 1) inhibitors. *J Med Chem* 53, 942-50.

51. Fragment-Blog. http://practicalfragments.blogspot.com/

52. Howard, S.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Curry, J.; Devine, L. A.; Early, T. R.; Fazal, L.; Gill, A. L.; Heathcote, M.; Maman, S.; Matthews, J. E.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Rees, D. C.; Reule, M.; Tisi, D.; Williams, G.; Vinkovic, M.; Wyatt, P. G. Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J Med Chem* **2009,** 52, 379-88.

53. Coates, L.; Tuan, H. F.; Tomanicek, S.; Kovalevsky, A.; Mustyakimov, M.; Erskine, P.; Cooper, J. The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *J Am Chem Soc* **2008,** 130, 7235-7.

54. Cooper, J.; Quail, W.; Frazao, C.; Foundling, S. I.; Blundell, T. L.; Humblet, C.; Lunney, E. A.; Lowther, W. T.; Dunn, B. M. X-ray crystallographic analysis of inhibition of endothiapepsin by cyclohexyl renin inhibitors. *Biochemistry* **1992,** 31, 8142-50.

55. Geschwindner, S.; Olsson, L. L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; de Beer, T.; Folmer, R. H. Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J Med Chem* **2007,** 50, 5903-11.

56. Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. *J Chem Inf Comput Sci* **2004,** 44, 2157-66.

57. Molecular Operating Environment (MOE); version 2008.11; Chemical Computing Group Inc.; Montreal, Canada
http://www.chemcomp.com

58. Kramer, C.; Beck, B.; Clark, T. Insolubility classification with accurate prediction probabilities using a MetaClassifier. *J Chem Inf Model* 50, 404-14.

59. Kramer, C.; Heinisch, T.; Fligge, T.; Beck, B.; Clark, T. A consistent dataset of kinetic solubilities for early-phase drug discovery. *ChemMedChem* **2009,** 4, 1529-36.

60. Coates, L.; Erskine, P. T.; Crump, M. P.; Wood, S. P.; Cooper, J. B. Five atomic resolution structures of endothiapepsin inhibitor complexes: implications for the aspartic proteinase mechanism. *J Mol Biol* **2002,** 318, 1405-15.

61. Andreeva, N. S.; Rumsh, L. D. Analysis of crystal structures of aspartic proteinases: on the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Sci* **2001,** 10, 2439-50.

62. Coates, L.; Erskine, P. T.; Mall, S.; Gill, R.; Wood, S. P.; Myles, D. A.; Cooper, J. B. X-ray, neutron and NMR studies of the catalytic mechanism of aspartic proteinases. *Eur Biophys J* **2006,** 35, 559-66.

63. Matter, H.; Nazare, M.; Gussregen, S.; Will, D. W.; Schreuder, H.; Bauer, A.; Urmann, M.; Ritter, K.; Wagner, M.; Wehner, V. Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity. *Angew Chem Int Ed Engl* **2009,** 48, 2911-6.

64. Larson, M. K.; Whitaker, J. R. Endothia parasitica Protease. Parameters Affecting Stability of the Rennin-like Enzyme. *Journal of Dairy Science* **1970,** 53, 262-269.

65. Otwinowski, Z., & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **1997,** 276, 307-326.

66. McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **2005,** 61, 458-64.

67. Sheldrick, G. M.; Schneider, T. R. SHELXL: high-resolution refinement. *Methods Enzymol* **1997,** 277, 319-43.

68. Adams, P. D.; Afonine, P. V.; Bunkoczi, 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 D Biol Crystallogr* 66, 213-21.

69. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** 60, 2126-32.

70. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **1985,** 28, 849-57.

71. Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J Mol Biol* **1999,** 289, 1093-108.

72. Neudert, G.; Klebe, G. fconv: format conversion, manipulation, and feature computation of molecular data. *Bioinformatics*.

73. Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B* **2002,** 58, 380-8.

74. Laederach, A.; Reilly, P. J. *J. Comput. Chem.* **2003,** 24, 1748.

75. deLano, W. L. http://www.pymol.org/

76. Behnen, J. Kristallstrukturen mit kleinen Sondenmolekülen : Ausleuchten von Bindetaschen, Startpunkt für ein Fragment-basiertes Wirkstoffdesign und Erhalten von Phaseninformationen zur Strukturlösung, Doktorarbeit 2010, Universität Marburg. http://archiv.ub.uni-marburg.de/diss/z2011/0076/pdf/djb.pdf

77. Hohwy, M.; Spadola, L.; Lundquist, B.; Hawtin, P.; Dahmén, J.; Groth-Clausen, I.; Nilsson, E.; Persdotter, S.; von Wachenfeldt, K.; Folmer, R. H. A.; Edman, K. Novel Prostaglandin D Synthase Inhibitors Generated by Fragment-Based Drug Design *Journal of Medicinal Chemistry* **2008,** 51, 2178-2186.

78. Brown, D.; Superti-Furga, G. Rediscovering the sweet spot in drug discovery. *Drug Discovery Today* **2003,** 8, 1067-1077.

79. Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; de Beer, T.; Folmer, R. H. A. Discovery of a Novel Warhead against β-Secretase through Fragment-Based Lead Generations. *Journal of Medicinal Chemistry* **2007,** 50, 5903-5911.

80. Owens, J. Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discovery Today* **2003,** 8, 12-16.

81. Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is There a Difference between Leads and Drugs? A Historical Perspective. *Journal of Chemical Information and Computer Sciences* **2001,** 41, 1308-1315.

82. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nature Reviews in Drug Discovery* **2004,** 3, 660-672.

83. Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004,** 9, 430-431.

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

85. Erlanson, D. A.; McDowell, R. S.; O´Brien, T. Fragment-Based Drug Discovery. *Journal of Medicinal Chemistry.* **2004,** 47, 3463-3482.

86. Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discovery Today* **2002,** 7, 522-527.

87. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996,** 274, 1531-1534.

88. Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **2007,** 6, 211.

89. Tse, C.; Shoemaker, A. R.; Adickes, J.; Anderson, M. G.; Chen, J.; Jin, S.; Johnson, E. F.; Marsh, K. C.; Mitten, M. J.; Nimmer, P.; Roberts, L.; Tahir, S. K.; Xiao, Y.; Yang, X.; Zhang, H.; Fesik, S.; Rosenberg, S. H.; Elmore, S. W. ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res* **2008,** 68, 3421-8.

90. Mattos, C.; Ringe, D. Locating and characterizing binding sites on proteins. *Nat Biotechnol.* **1996,** 14, 595-599.

91. Ringe, D. What makes a binding site a binding site? *Curr Opin Struct Biol.* **1995,** 5, 825-829.

92. Mattos, C.; Bellamacina, C. R.; Peisach, E.; Pereira, A.; Vitkup, D.; Petsko, G. A.; Ringe, D. Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J Mol Biol.* **2006,** 357, 1471-1482.

93. Allen, K. N.; Bellamacina, C. R.; Ding, X.; Jeffrey, C. J.; Mattos, C. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. *The Journal of Physical Chemistry* **1996,** 100, 2605 - 2611.

94. Goodford, P. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem.* **1985,** 28, 849-857.

95. Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **1991,** 11, 29-34.

96. Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999,** 37, 88-105.

97. Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* **2001,** 42, 256-68.

98. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry* **1985,** 28, 849-857.

99. Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: A Knowledge-based Approach for Identifying Interaction Sites in Proteins. *Journal of Molecular Biology* **1999,** 289, 1093-1108.

100. Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology* **2000,** 295, 337-356.

101. Neudert, G.; Klebe, G. fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics* 27, 1021-2.

102. Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chemical Reviews* **2004,** 104, 3641-3676.

103. Metz, G.; Ottleben, H.; Vetter, D. *Methods and Principles in Medicinal Chemistry*. 2003; Vol. 19.

104. English, A. C.; Done, S. H.; Caves, L. S.; Groom, C. R.; Hubbard, R. E. Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins* **1999,** 37, 628-640.

105. English, A. C.; Groom, C. R.; Hubbard, R. E. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Engineering.* **2001,** 14, 47-59.

106. Casale, E.; Collyer, C.; Ascenzi, P.; Balliano, G.; Milla, P.; Viola, F.; Fasano, M.; Menegatti, E.; Bolognesi, M. Inhibition of bovine beta-trypsin, human alpha-thrombin and porcine pancreatic beta-kallikrein-B by 4',6-diamidino-2-phenylindole, 6-amidinoindole and benzamidine: a comparative thermodynamic and X-ray structural study. *Biophys Chem* **1995,** 54, 75-81.

107. Baum, B.; Muley, L.; Heine, A.; Smolinski, M.; Hangauer, D.; Klebe, G. Think twice: understanding the high potency of bis(phenyl)methane inhibitors of thrombin. *J Mol Biol* **2009,** 391, 552-64.

108. Holland, D. R.; Tronrud, D. E.; Pley, H. W.; Flaherty, K. M.; Stark, W.; Jansonius, J. N.; McKay, D. B.; Matthews, B. W. Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* **1992,** 31, 11310-11316.

109. Englert, L.; Biela, A.; Zayed, M.; Heine, A.; Hangauer, D.; Klebe, G. Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S'-pocket of thermolysin. *Biochim Biophys Acta* 1800, 1192-202.

110. Zheng, J.; Trafny, E. A.; Knighton, D. R.; Xuong, N.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M. 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallographica Section D* **1993,** 49, 362-365.

111. Carrell, H. L.; Hoier, H.; Glusker, J. P. Modes of binding substrates and their analogues to the enzyme d-xylose isomerase. *Acta Crystallographica Section D* **1994,** 50, 113-123.

112. Richard, S. B.; Bowman, M. E.; Kwiatkowski, W.; Kang, I.; Chow, C.; Lillo, A. M.; Cane, D. E.; Noel, J. P. Structure of 4-diphosphocytidyl-2-C-methylerythritol synthetase involved in mevalonate-independent isoprenoid biosynthesis. *Nature Structural Molecular Biolology* **2001,** 8, 641-648.

113. Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScoreCSDKnowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *Journal of Medicinal Chemistry* **2005,** 48, 6296-6303.

114. Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins* **1997,** 29, 32-58.

115. Velec, H. F.; Gohlke, H.; Klebe, G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem* **2005,** 48, 6296-303.

116. Akamine, P.; Madhusudan; Wu, J.; Xuong, N. H.; Ten Eyck, L. F.; Taylor, S. S. Dynamic features of cAMP-dependent protein kinase revealed by apoenzyme crystal structure. *J Mol Biol* **2003,** 327, 159-71.
117. Madhusudan; Trafny, E. A.; Xuong, N. H.; Adams, J. A.; Ten Eyck, L. F.; Taylor, S. S.; Sowadski, J. M. cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. *Protein Sci* **1994,** 3, 176-87.
118. Madhusudan; Akamine, P.; Xuong, N. H.; Taylor, S. S. Crystal structure of a transition state mimic of the catalytic subunit of cAMP-dependent protein kinase. *Nat Struct Biol* **2002,** 9, 273-7.
119. Zheng, J.; Trafny, E. A.; Knighton, D. R.; Xuong, N. H.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M. 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr D Biol Crystallogr* **1993,** 49, 362-5.
120. Taylor, S. S.; Yang, J.; Wu, J.; Haste, N. M.; Radzio-Andzelm, E.; Anand, G. PKA: a portrait of protein kinase dynamics. *Biochim Biophys Acta* **2004,** 1697, 259-69.
121. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996,** 261, 470-89.
122. Gerlach, C.; Munzel, M.; Baum, B.; Gerber, H. D.; Craan, T.; Diederich, W. E.; Klebe, G. KNOBLE: a knowledge-based approach for the design and synthesis of readily accessible small-molecule chemical probes to test protein binding. *Angew Chem Int Ed Engl* **2007,** 46, 9105-9.
123. http://publications.nigms.nih.gov/medbydesign/images/ch4_kinases.jpg
124. Huggenvik, J. I.; Collard, M. W.; Stofko, R. E.; Seasholtz, A. F.; Uhler, M. D. Regulation of the Human Enkephalin Promoter by Two Isoforms of the Catalytic Subunit of Cyclic Adenosine 3',5'-Monophosphate-Dependent Protein Kinase. *Mol Endocrinol* **1991,** 5, 921-930.
125. Gjertsen, B. T.; Døskeland, S. O. Protein phosphorylation in apoptosis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1995,** 1269, 187-199.
126. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J Comput Chem* **2005,** 26, 1668-88.
127. Vogtherr, M.; Saxena, K.; Hoelder, S.; Grimme, S.; Betz, M.; Schieborr, U.; Pescatore, B.; Robin, M.; Delarbre, L.; Langer, T.; Wendt, K. U.; Schwalbe, H. NMR characterization of kinase p38 dynamics in free and ligand-bound forms. *Angew Chem Int Ed Engl* **2006,** 45, 993-7.
128. AmberTools. http://ambermd.org/#AmberTools
129. Manchester. Contributed parameters http://www.pharmacy.manchester.ac.uk/bryce/amber
130. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber, An Accessory Software Package
For Molecular Mechanical Calculations. http://ambermd.org/antechamber/antechamber.pdf
131. GAFF. http://ambermd.org/antechamber/gaff.html#abstract
132. Gnuplot. http://www.gnuplot.info/
133. Gunera, J. New Scaffolds for Aldo-Keto Reductase: A Virtual Screening Study. Bachelor, Philipps Universität Marburg, Marburg, 2010
134. PDB. http://www.rcsb.org/

135. Klebe, G. *Wirkstoffdesign : Entwurf und Wirkung von Arzneistoffen*. 2. Aufl.. ed.; Spektrum Akad. Verl.: Heidelberg, 2009; p XX, 634 S.
136. Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005,** 45, 177-82.
137. qiagen. http://www.qiagen.com/literature/qiagennews/weeklyarticle/10_07/e07/default.aspx
138. Cummings, M. D.; Farnum, M. A.; Nelen, M. I. Universal Screening Methods and Applications of ThermoFluor®. *Journal of Biomolecular Screening* **2006,** 11, 854-863.
139. 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, 2212-21.
140. Nettleship, J. E.; Brown, J.; Groves, M. R.; Geerlof, A. Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Methods Mol Biol* **2008,** 426, 299-318.
141. Bursy, J.; Pierik, A. J.; Pica, N.; Bremer, E. Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. *J Biol Chem* **2007,** 282, 31147-55.
142. Bursy, J.; Kuhlmann, A. U.; Pittelkow, M.; Hartmann, H.; Jebbar, M.; Pierik, A. J.; Bremer, E. Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by Streptomyces coelicolor A3(2) in response to salt and heat stresses. *Appl Environ Microbiol* **2008,** 74, 7286-96.
143. Purpero, V.; Moran, G. R. The diverse and pervasive chemistries of the alpha-keto acid dependent enzymes. *J Biol Inorg Chem* **2007,** 12, 587-601.
144. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J. Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **2006,** 100, 644-69.
145. Hausinger, R. P. FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* **2004,** 39, 21-68.
146. Flashman, E.; Schofield, C. J. The most versatile of all reactive intermediates? *Nat Chem Biol* **2007,** 3, 86-7.
147. Prabhu, J.; Schauwecker, F.; Grammel, N.; Keller, U.; Bernhard, M. Functional expression of the ectoine hydroxylase gene (thpD) from Streptomyces chrysomallus in Halomonas elongata. *Appl Environ Microbiol* **2004,** 70, 3130-2.
148. Welford, R. W.; Kirkpatrick, J. M.; McNeill, L. A.; Puri, M.; Oldham, N. J.; Schofield, C. J. Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett* **2005,** 579, 5170-4.
149. Branden, C.; Tooze, J. *Introduction to protein structure*. 2. ed.; Garland: New York , NY, 1999; p XIV, 410 S.
150. 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, 289-98.
151. Roach, P. L.; Clifton, I. J.; Fulop, V.; Harlos, K.; Barton, G. J.; Hajdu, J.; Andersson, I.; Schofield, C. J.; Baldwin, J. E. Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* **1995,** 375, 700-4.

152. Koski, M. K.; Hieta, R.; Bollner, C.; Kivirikko, K. I.; Myllyharju, J.; Wierenga, R. K. The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *J Biol Chem* **2007,** 282, 37112-23.

153. Elkins, J. M.; Hewitson, K. S.; McNeill, L. A.; Seibel, J. F.; Schlemminger, I.; Pugh, C. W.; Ratcliffe, P. J.; Schofield, C. J. Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **2003,** 278, 1802-6.

154. Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **2006,** 439, 879-84.

155. Ng, S. S.; Kavanagh, K. L.; McDonough, M. A.; Butler, D.; Pilka, E. S.; Lienard, B. M.; Bray, J. E.; Savitsky, P.; Gileadi, O.; von Delft, F.; Rose, N. R.; Offer, J.; Scheinost, J. C.; Borowski, T.; Sundstrom, M.; Schofield, C. J.; Oppermann, U. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **2007,** 448, 87-91.

156. Pavel, E. G.; Zhou, J.; Busby, R. W.; Gunsior, M.; Townsend, C. A.; Solomon, E. I. Circular Dichroism and Magnetic Circular Dichroism Spectroscopic Studies of the Non-Heme Ferrous Active Site in Clavaminate Synthase and Its Interaction with Î±-Ketoglutarate Cosubstrate. *Journal of the American Chemical Society* **1998,** 120, 743-753.

157. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **2006,** 440, 368-71.

158. Hewitson, K. S.; Holmes, S. L.; Ehrismann, D.; Hardy, A. P.; Chowdhury, R.; Schofield, C. J.; McDonough, M. A. Evidence that two enzyme-derived histidine ligands are sufficient for iron binding and catalysis by factor inhibiting HIF (FIH). *J Biol Chem* **2008,** 283, 25971-8.

159. Decher, N.; Streit, A. K.; Rapedius, M.; Netter, M. F.; Marzian, S.; Ehling, P.; Schlichthorl, G.; Craan, T.; Renigunta, V.; Kohler, A.; Dodel, R. C.; Navarro-Polanco, R. A.; Preisig-Muller, R.; Klebe, G.; Budde, T.; Baukrowitz, T.; Daut, J. RNA editing modulates the binding of drugs and highly unsaturated fatty acids to the open pore of Kv potassium channels. *EMBO J* 29, 2101-13.

160. Decher, N.; Gonzalez, T.; Streit, A. K.; Sachse, F. B.; Renigunta, V.; Soom, M.; Heinemann, S. H.; Daut, J.; Sanguinetti, M. C. Structural determinants of Kvbeta1.3-induced channel inactivation: a hairpin modulated by PIP2. *EMBO J* **2008,** 27, 3164-74.

161. Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* **1995,** 245, 43-53.

162. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997,** 267, 727-48.

Danksagung

Prof. Dr. Gerhard Klebe danke ich herzlich für die interessante Themenstellung dieser Doktorarbeit und die sehr gute Betreuung während der gesamten Zeit. Ich danke ihm für die vielen Möglichkeiten an Konferenzen teilzunehmen und die vielen fachlichen und persönlichen Gespräche für die er stets Zeit gefunden hat. Auch danke ich ihm für die zahlreichen ausführlichen Korrekturen von Manuskripten sowie das Vertrauen in mich die Projekte eigenständig lenken und weiterentwickeln zu können.

Ich danke *Helene Köster, Dr. Sascha Brass, Barbara Wienen* und *Hans-Dieter Gerber* von der Universität Marburg, sowie *Dr. Per Hillertz* und *Dr. Jörg Bomke* von Merck Serono für die hervorragende Zusammenarbeit auf dem PKA Projekt.

Ich danke *Helene Köster, Dr. Sascha Brass* und *Barbara Wienen* sowie Boehringer/ Ingelheim, Merck Serono und Proteros Biostructures für die gute Zusammenarbeit beim Aufbau der Fragmentbibliothek, sowie *Helene* für die Validierung und die elf Fragment-Kristallstrukturen von Endothiapepsin.

Dr. Andreas Heine danke ich für die Beantwortung von kristallographischen Fragen, sowie das Korrigieren von diversen Manuskripten und die Bereitschaft als Zweitgutachter zu fungieren.

Ich danke *Prof. Dr. Klaus Reuter* für die hervorragende Zusammenarbeit bei der Etablierung des Temperatur-Verschiebungs-Assay.

Ich danke *Prof. Dr. Nils Decher* für die interessanten Einblicke in das Gebiet der Ionenkanäle und die Zusammenarbeit auf dem K_v-Kanal Projekt.

Dr. Sascha Brass, Thomas Kronenberger und *Hans-Dieter Gerber* möchte ich für die Synthese der Substanzen danken. *Dres. Nan-Si* und *Sascha Brass* danke ich zudem für ihre außerordentliche Hilfsbereitschaft und Unterstützung.

Frau Lydia Hartleben möchte ich für Ihre große Hilfsbereitschaft und Unterstützung danken.

Allen Administratoren unserer Arbeitsgruppe vor allem *Gerd, Sven, Micha, Andreas* und *Felix* danke ich für die Wartung der Rechner und des Netzwerkes.

Allen Vertiefungsstudenten und den *Bachelorstudenten Jakub Gunera* und *Heiner Saßmannshausen* sowie dem Mastervertiefungsstudent *Rajathees Rajaratnam* danke ich für die geleistete Arbeit.

Ich danke den *Mitgliedern der AG Klebe* für die gute Zusammenarbeit. Vor allem möchte ich *Jürgen, Johannes, Florian, Kan, Christian, Andreas, Michael, Tina, Björn, Nan-Si, Sascha, Felix, Serghei, Gerd, Conny, Sven, Kerstin, Daniela und Frank (AG Steinmetzer)* für die schöne Zeit am Institut bedanken.

Danken möchte ich *Dr. Peter Kolb und Frank Sielaff* für das Korrekturlesen des Manuskripts meiner Dissertation.

Ich danke *Sven Siebler und Gerd Neudert* für die ausgezeichnete Büroatmosphäre, die vielen interessanten und lehrreichen Gespräche, sowie *Florian Immekus* für die ebenfalls ausgezeichnete Büroatmosphäre und seine Weisheiten.

Ein Dank geht an *Marion, Joe, Frank, Ilka, Nan-Si, Sascha, Mei-Lin, Dani, Anabel, Hase* und *Diana* für die wöchentlichen Treffen.

Danken möchte ich auch *Theresa Wilhelm* und *Till Reher* für die schöne Zeit in Marburg und die vielen Kochsessions.

Meiner ganzen Familie und *meinen Freunden* danke ich für die große Unterstützung im Studium und während der Promotion. Von ganzem Herzen danke ich *meinen Eltern* und *Steffi* für die allgegenwärtige, unermüdliche Unterstützung, ihre Förderung und ihr Vertrauen.

Erklärung

Ich versichere, dass ich meine Dissertation

Fragment based Drug Discovery; Design and Validation of a Fragment Library; Computer-based Fragment Screening and Fragment-to-Lead Expansion

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

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

………………………………..

(Tobias Friedrich Craan)

Curriculum Vitae

PERSÖNLICHE DATEN

Geburtsdatum: 19. Februar 1982

Geburtsort: Korbach

Staatsangehörigkeit: deutsch

Familienstand: ledig

AUSBILDUNG

Seit 03/2008	Promotion im Fachbereich Pharmazeutischen Chemie der Philipps Universität Marburg, Arbeitskreis Prof. Dr. Klebe
01/2008	Approbation als Apotheker
06/2007 – 12/2007	wissenschaftlicher Aufenthalt am Cambridge Crystallographic Data Center CCDC, Cambridge, UK
12/2006 – 05/2007	Praktikum in der Merlin Apotheke in Köln
10/2002 – 12/2006	Studium der Pharmazie an der Philipps Universität Marburg
08/1998 – 06/2001	Geschwister-Scholl-Gymnasium in Winterberg; Abitur
08/1992 – 06/1998	Gymnasium in Medebach
08/1988 – 06/1992	Grundschule in Medebach

ZIVILDIENST

09/2001 – 06/2002	St. Elisabeth Altenheim, Meschede