

**Lead optimization for new antimalarials**  
**and**  
**Successful lead identification for metalloproteinases:**  
**A Fragment-based approach Using Virtual Screening**

DISSERTATION

ZUR

ERLANGUNG DES DOKTORGRADES

DER NATURWISSENSCHAFTEN

(DR. RER. NAT.)

dem

Fachbereich Pharmazie der

PHILIPPS-UNIVERSITÄT MARBURG

vorgelegt von

**Katrin Silber**

aus Frankfurt

Marburg an der Lahn, 2007



Vom Fachbereich Pharmazie der PHILIPPS-UNIVERSITÄT MARBURG als Dissertation

angenommen am: 1. Oktober 2007

Erstgutachter: PROF. DR. G. KLEBE, Philipps-Universität Marburg

Zweitgutachter: PROF DR. M. SCHLITZER, Philipps-Universität Marburg

Tag der mündlichen Prüfung: 2. Oktober 2007

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof.G. Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit vom Mai 2002 bis Februar 2006 durchgeführt.

## Table of Contents

<b><i>Introduction</i></b>	<b>1</b>
Current Drug Discovery.....	1
Computational drug design.....	5
Scope of the thesis.....	7
<b><i>AFMoC enhances predictivity of 3D QSAR: a case study with DOXP-reductoisomerase</i></b>	<b>11</b>
Introduction .....	11
The AFMoC Approach.....	14
Target Protein.....	18
Methods.....	23
Data Set and Alignment.....	23
CoMFA and CoMSIA Analyses.....	29
Receptor-based predictions.....	30
AFMoC analysis.....	31
Results and Discussion.....	34
Molecular alignment .....	34
CoMFA/CoMSIA analyses.....	38
Generating the AFMoC model.....	42
Application of the derived AFMoC model to predict binding affinities.....	43
Prediction using other scoring schemes .....	54
Conclusions.....	56
Supporting Information .....	58
<b><i>Methods in Receptor-based drug design</i></b>	<b>61</b>
Protein-based docking .....	61
Protein-based scoring.....	62
Applied docking programs and Scoring functions.....	65
FlexX.....	65
AutoDock.....	66

GOLD.....	67
DrugScore.....	68
SFC-Score.....	69
Addressing challenges for docking .....	69
<b><i>Optimizing binding affinities for Benzophenone-based Protein Farnesyltransferase inhibitors</i></b>	<b>73</b>
Introduction .....	73
Target Protein.....	73
Methods.....	78
Docking.....	79
Results and Discussion.....	80
QSAR in the far aryl binding sites.....	81
QSAR in the specificity pocket.....	84
Conclusions.....	91
<b><i>Successful lead identification for Metalloproteinases: A Fragment-based approach Using Virtual Screening</i></b>	<b>92</b>
Introduction.....	92
Drug-like versus Lead-like.....	93
Small is beautiful – the concept of Fragment-based ligand design.....	96
Library design.....	99
Screening.....	101
Linkage and growth.....	103
Computational support in fragment-based drug discovery.....	105
Target protein.....	109
Methods.....	115
Library composition.....	115
Virtual Screening.....	116
Docking.....	117
Crystallization.....	120
Data collection.....	121
Structure determination.....	121
Binding Assay.....	122

Results and Discussion.....	124
Library design and Virtual Screening.....	126
Docking.....	132
Crystallization.....	148
Docking and Scoring 3-Methylaspirin .....	152
Determining the inhibition constant .....	159
Conclusions.....	163
Supporting information for fragment-based virtual screening.....	165
XRAY Fragment data set.....	165
Dixon plots for aspirin derivatives.....	175
SFC Scoring function docking scores .....	176
Databases and Programs used throughout this work .....	178
PDB entries.....	179
Conformation search for aspirin.....	183
Substrate free protein structure.....	184
Refinement data.....	185
<b><i>Summary &amp; Outlook</i></b> .....	<b>187</b>
<b><i>References</i></b> .....	<b>190</b>
<b><i>Appendix</i></b> .....	<b>228</b>
Acknowledgment of work from collaboration partners.....	228
Abbreviations.....	229
Overview over described scoring function:.....	231
Virtual Screening examples .....	233
Publications arising from this work:.....	237
Articles.....	237
Oral communications.....	238
Poster presentations .....	239
Acknowledgments .....	240
Curriculum vitae.....	242
Erklärung.....	243



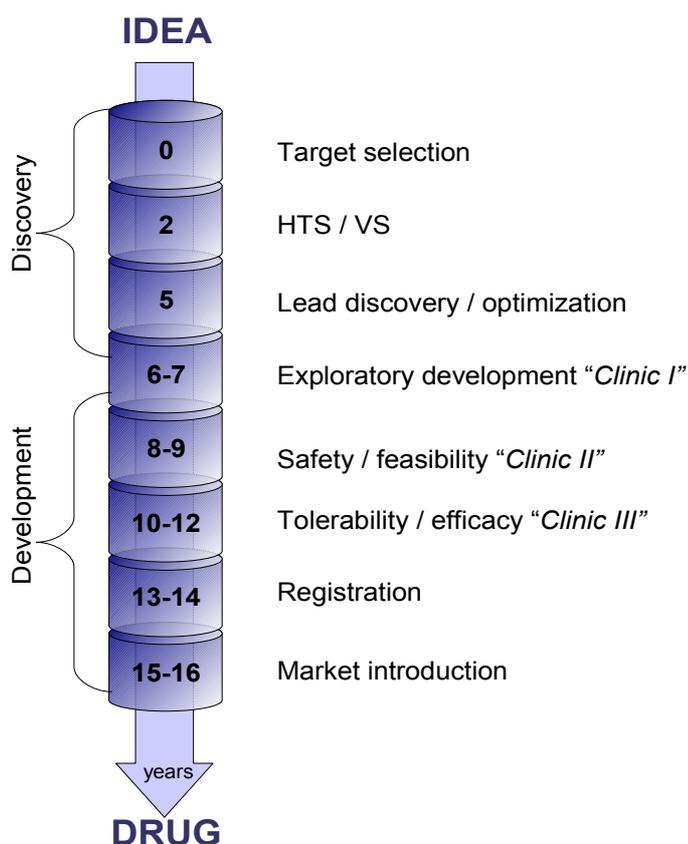
## INTRODUCTION

### ***Current Drug Discovery***

Reducing time, resources and serendipity has been the major aim for the development of modern, rational drug discovery strategies.

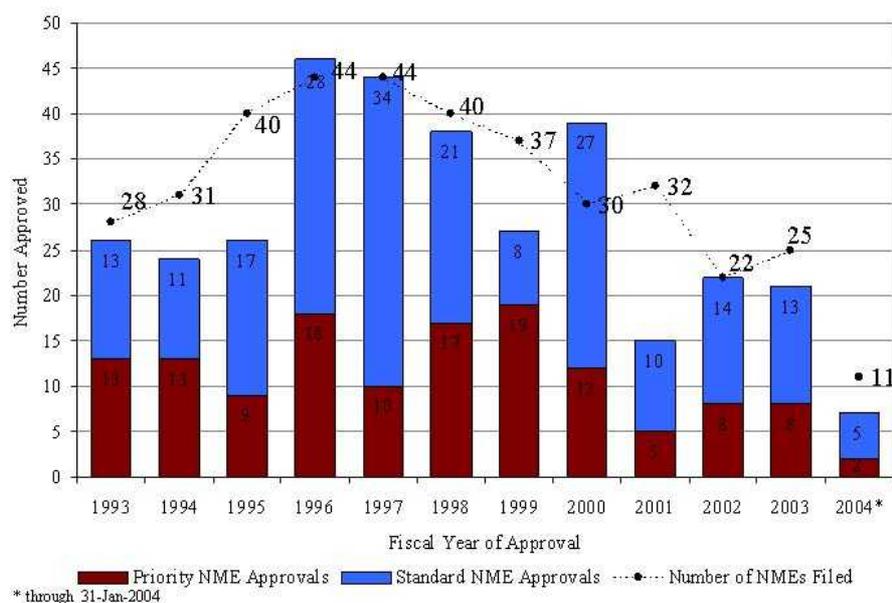
Historically, most drugs have been discovered mainly by two of the following methods. The first is to modify a known starting molecule like a natural ligand, co-factor or simply an already existing drug. The first drug to market is rarely the best. COX-2 inhibitors, HIV protease inhibitors or the “patent-busting” around PDE5 inhibitors serve as quite productive examples for this approach [Wermuth, 2004, 2006a, 2006b]. For diseases where neither a drug nor natural template structure exists, the second route is applied, random screening. Being venerable this approach requires a certain serendipity to succeed. Cyclosporine and paclitaxel are prominent examples for drugs identified by screening corporate compound selections. The invention of combinatorial chemistry increased the size of the chemical collections and assay automation reduced time and resources necessary to screen large libraries. Still a certain degree of serendipity is required. Therefore, more rational routes for drug discovery have been sought [Lundqvist, 2005], [Klebe, 2000], [Oprea, 2004].

Today, the whole drug discovery process from development to registration for the market is an impressively long, expensive, and risky challenge. On average, only one out of 10,000 originally synthesized or isolated compounds will clear all the hurdles on the way to becoming a commercially available drug [Gad, 2005]. The process from first discovery to full development takes ~15 years to complete and costs approximately 800 million dollars.



**Figure 1.1** Scheme of the pharmaceutical company's drug design and development strategy.

The process starts with finding the right molecular target to address a certain disease, i.e., a single gene or protein [Kubinyi, 2003]. Interestingly, current drug therapy is based on less than 500 targets. Molecular biology efforts help understanding disease processes at the genetic level and to determine optimal targets [Drews, 2000]. The completion of the human genome together with new technologies like bioinformatics, genomics and proteomics allow to characterize more genetic and molecular processes in humans and other species, thus providing thousands of possible enzymes, receptors and ion channels as potential new drug targets [Austin, 2004]. Unfortunately, having more targets does not necessarily lead to more drugs on the market. Contrarily, the number of new molecular entities (NMEs) approved by the Food and Drug Administration (FDA) decreased in recent years (Figure 1.2).



**Figure 1.2** Survey of recent submissions and approvals of new molecular entities to the food and drug administration by fiscal year. (<http://www.fda.gov>)

Once a target is validated, different screening strategies are applied to identify the so called lead, a prototypical structure that demonstrates an adequate activity. Considering a representative target portfolio, high-throughput screening (HTS) is presently the most widely applicable technology delivering chemistry entry points for drug discovery programmes [Baxter, 2000]. Identified lead compounds are subsequently “optimized,” i.e., altered in ways that both increase their potential efficacy as well as minimize any potential side effects [Keseru, 2006]. The tools of modern chemistry (parallel and automated synthesis) and modern biology merge in these stages to give the best candidates for the drugs of the future.

To ensure the safety of administering a new drug candidate to healthy volunteers and patients, extensive toxicological and safety pharmacological profiles are done in both, in vitro tests and animals studies. Once the initial safety profile is established, the drug candidate is given to a small group of patients or healthy volunteers to verify tolerance and possible effects to the human disease (first-man-first-visit). Subsequently, Phase I trials are undertaken to gain

information about tolerable dose ranges and the metabolism of the drug candidate. Phase II clinical trials involving up to a few hundred patients establish the range of efficacious doses. Phase III trials (up to 10,000 patients) provide final information about the drug's effectiveness and provide comparison to standard treatments. For the registration of the new drug, quality, efficacy and safety of the drug need to be demonstrated to the FDA (USA) or EMEA (EU). At times, additional Phase IV studies are undertaken after registration to add new indications or improve existing formulations of the drug.

Future challenges are to reduce development time and control costs effectively without compromising safety and quality. Therefore, the pharmaceutical industry tries to constantly integrate new research and development techniques. At drug discovery state (Figure 1.1), this means to consider "compound quality" as early as possible. Compound quality comprises drug-likeness of a molecule with respect to administration, distribution, metabolism and excretion (ADME).

Another requirement is high affinity of the drug to its target to keep the necessary dosage as low as possible. This facilitates formulation and decreases the risk to experience possible side-effects. It is well recognized that even when compounds are identified from HTS they are not always suitable to embark on further chemistry exploration [Keseru, 2006]. One promising and now frequently applied approach is to integrate computational models to predict the desired property at each step of the discovery process. In the last 20 years, those models and computational approaches to assist drug discovery have gained in importance and proven to be a resource-saving techniques to identify and optimize novel chemotypes in biologically active molecules.

## ***Computational drug design***

Computer-aided drug design is an essential part of the modern medicinal chemistry, and has led to the acceleration of many projects, and even to drugs on the market [Coupez, 2006]. Its application ranges from early target validation to toxicity models at late discovery/early development transition state.

Assessing the druggability of a target is supported by the analysis of theoretical signaling pathway networks [Coupez, 2006], comparing the putative binding pocket to existing ones [Nettles, 2006], [Weber, 2004], or generating homology models [Evers, 2004], [Hillisch, 2004], [Kairys, 2006], [Rong, 2002] if neither crystal nor NMR structures are available. Computational strategies for the target validation process have been reviewed by Wiemann, Blundell, Hajduk, and Keller [Blundell, 2006], [Hajduk, 2005a], [Hajduk, 2005b], [Wieman, 2004].

Most extensively, computer-aided strategies are applied in the lead discovery and lead optimization stage. In addition to high throughput screening (HTS), the main lead discovery technology employed by most pharmaceutical companies today is virtual screening [Walters, 1998], [Klebe, 2006]. It involves the rapid assessment of large libraries of chemical structures and can be categorized as being either ligand-based or receptor-based. The compounds are compared to previously as active identified molecules (similarity searches) or matched to a pharmacophore derived from the target binding site (pharmacophore search). When the structure of the target protein is known, receptor-based docking can be employed. This approach aims to predict correctly the structure of the intermolecular complex formed between the target receptor and the ligand. To correctly dock a molecule, two technical challenges imply, (1) the pose generation (docking) of the ligand in the active site and (2) the evaluation of the different poses (scoring). Scoring requires estimation of the binding energy between protein and ligand and produces a relative rank-ordering between different ligand,

docked to the same target. Examples of virtual screening have been published as referenced in the appendix. Different screening protocols and their successful application are reviewed by Jalaie, Pirard, Hou, Green, Langer and Schneider [Green, 2003], [Hou, 2004], [Jalaie, 2006], [Langer, 2003], [Pirard, 2005], [Schneider, 2002]. An alternative to screen for compounds is designing molecules with desired properties from scratch. Integration of so called de-novo design to drug discovery is reviewed by Schneider and Honma [Honma, 2003], [Schneider, 2005]. At lead optimization state, docking is applied to assess the modifications of the lead compound. As a series of molecules with known activity arises, three dimensional quantitative structure-activity relationships (3D QSAR) methods gain in importance. Using 3D QSAR statistical correlations are established between the potencies of a series of structurally related compounds and one or more quantitative structural parameters, such as lipophilicity, polarity, and molecular size, by using multi linear regression analysis [Gohlke, 2003], [Kuo, 2004].

For lead compound series with convincing activities computational methods are also applied to estimate pharmacokinetic properties. Again docking, homology modeling and statistical correlations methodologies are used in order to evaluate potential Cyp450 inhibition [Cruciani, 2005], [Zamora, 2003], plasma protein binding [Gleeson, 2007], HERG channel blockage [Farid, 2006], [Pearlstein, 2003], bioavailability and toxicity [Egan, 2002], [Lipinski, 2001].

Computer-aided drug design techniques are nowadays established and have emerged as key strategy to help assessing compounds [Good, 2000]. Predicting chemical and biological properties with computational models identifies compounds that are likely to fail in primary, secondary and further downstream screen at significantly lower costs. Integration of computational approaches into the drug discovery process is nicely reviewed by Chin and Oprea [Chin, 2004], [Oprea, 2004].

## ***Scope of the thesis***

The present thesis describes two aspects of modern drug discovery in detail. Firstly, the challenges of **lead optimization** are addressed considering two antimalarial drug targets. As mentioned above, one important aspect of lead optimization is increasing the target-drug affinity as much as required. This is especially important for antimicrobial drugs. To prevent secondary resistance development during the drug therapy they are usually administered at high dosages.

Using computer-aided drug design techniques, target-drug binding affinities are analyzed for the two enzymes, DOXP-reductoisomerase (DXR) and farnesyltransferase (ftase). Both targets have recently been validated as potential antimalarial drug targets [Jomaa, 1999], [Chakrabarti, 1998]. As a potential drug has to inhibit the enzymatic activity the terms 'inhibitor' or 'ligand' are used for the drug candidate as well as the term 'protein' to represent the target.

*DOXP-Reductoisomerase:* Structure-activity relationships for 43 inhibitors are established, derived from protein-based docking, ligand-based 3D QSAR and a combination of both approaches as realized by AFMoC (Adaptation of fields of molecular comparison). DXR is a key enzyme of the non-mevalonate pathway for isoprenoid building blocks. As part of an effort to optimize the properties of the established inhibitor Fosmidomycin, analogues have been synthesized and tested to gain further insights into the primary determinants of structural affinity. Herein, these molecules have been used to create a predictive model for DXR inhibition applying data taken from several DXR X-ray structures. Unfortunately, these structures still leave the active Fosmidomycin conformation and detailed reaction mechanism undetermined. This fact, together with the small inhibitor data set provides a major challenge

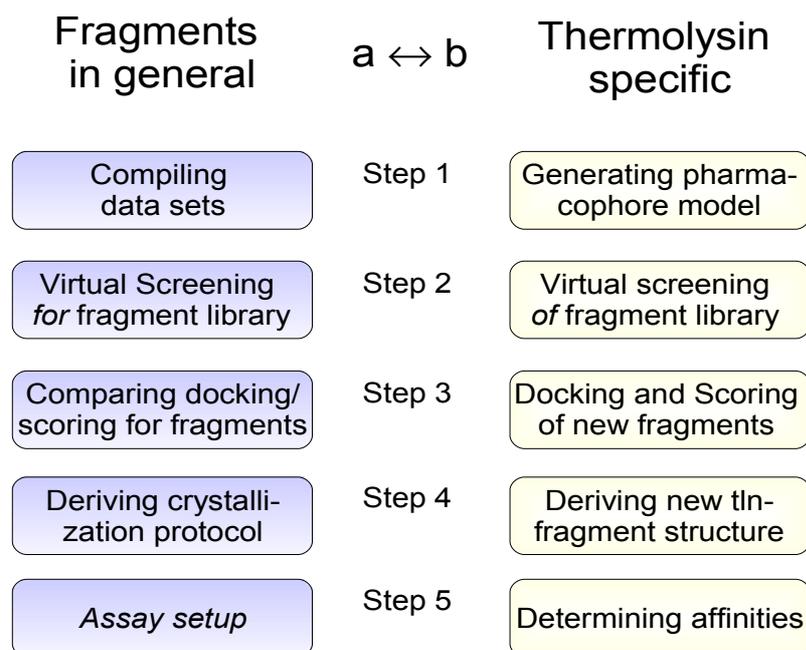
for presently available docking programs and 3D QSAR tools. Using the recently developed protein tailored scoring protocol AFMoC precise prediction of binding affinities for related ligands as well as the capability to estimate the affinities of structurally distinct inhibitors has been achieved.

*Farnesyltransferase:* Farnesyltransferase is a zinc-metallo enzyme that catalyzes the posttranslational modification of numerous proteins involved in intracellular signal transduction. The development of farnesyltransferase inhibitors is directed towards the so-called non-thiol inhibitors because of adverse drug effects connected to free thiols. A first step on the way to non-thiol farnesyltransferase inhibitors was the development of an CAAX-benzophenone peptidomimetic based on a pharmacophore model. On its basis bisubstrate analogues were developed as one class of non-thiol farnesyltransferase inhibitors. In further studies two aryl binding and two distinct specificity sites were postulated. Flexible docking of model compounds was applied to investigate the sub-pockets and design highly active non-thiol farnesyltransferase inhibitor. In addition to affinity, special attention was paid towards in vivo activity and species specificity.

The second part of this thesis describes a possible strategy for computer-aided **lead discovery**. Assembling a complex ligand from simple fragments has recently been introduced as an alternative to traditional HTS [Carr, 2002]. While frequently applied experimentally, only a few examples are known for computational fragment-based approaches. Mostly, computational tools are applied to compile the libraries and to finally assess the assembled ligands. Using the metalloproteinase thermolysin (TLN) as a model target, a computational fragment-based screening protocol has been established (Fig 1.3). As a prerequisite, it is assumed that (1) the three-dimensional structure of the target is available and (2) experimental validation assays for both, affinity and structure are established.

Starting with a data set of commercially available chemical compounds, a fragment library has been compiled considering (1) fragment likeness and (2) similarity to known drugs (Step 1a). The library is screened for target specificity (Step 1b). After analyzing the performance of multiple docking programs and scoring functions for fragments in general (Step 2a), the fragments derived from the screening are docked and the most promising candidates are selected for further analysis (2b). After purchasing both, already validated reference fragments and the selected candidates, co-crystallization and soaking experiments were performed on the reference fragment to derive a general applicable crystallization protocol (Step 3a) for TLN. Subsequently, crystallization experiments are done to derive new protein-fragment complex structures. Additionally, the affinity is validated using the formerly established inhibition assay [Silber, 2001] (Step 4a,b).

Some additional studies based on the screening results include (1) a retrospective performance analysis of the applied scoring functions and (2) some modification on the screening hit. Having discovered unexpected binding properties of a new fragment, some slightly deviating molecules were synthesized, computationally and experimentally analyzed and compared to the screening result molecule.



**Figure 1.3** Schematic overview over the development and application of a fragment-based lead discovery approach. Left panel: General approaches applied to literature-derived test sets. Right panel: Design performed to find new fragments binding to thermolysin.

# AFMOC ENHANCES PREDICTIVITY OF 3D QSAR: A CASE STUDY WITH DOXP-REDUCTOISOMERASE

## **Introduction**

In recent years, virtual screening of large compound libraries has been established as a fast and rather inexpensive alternative to experimental high-throughput screening in the lead discovery process. It has been successfully applied to suggest new inhibitors of, e.g. carbonic anhydrase, tRNA-guanine transglycosylase [Brenk, 2003] and DNA gyrase in the submicromolar range [Doman, 2002], [Gane, 2000], [Shoichet, 2002].

Of utmost importance in a virtual screening assay is the reliable filtering of putative hits in terms of their predicted binding affinity (scoring problem) which is based on the *in silico*-generated near native protein-ligand configurations (docking problem). Over the last years, a broad spectrum of first principle-, regression- or knowledge-based scoring functions for protein–ligand complexes has emerged. Although these scoring functions achieve increasing reliability in affinity prediction for some biological targets, for many others, predictions remain rather unsatisfactory [Bissantz, 2000], [Ferrara, 2004], [Wang, 2003]. In any case, scoring based on one single function remains a crucial and often limiting step. Alternatively, very pragmatic approaches have been suggested such as intersection-based consensus scoring [Ferrara, 2004], [Clark, 2002], [Charifson, 1999], [Wang, 2001a] or multivariate analysis of scoring results based on different functions [Jacobsson, 2003] to improve the reliability in ranking putative screening hits and predicting their binding affinities. A promising approach, in particular in the incipient phase of a drug-development program, where multiple aspects of

the protein-ligand binding determinants are still to be unraveled, is the weighting and tailoring of a general purpose scoring function, using information collected for some first leads known to bind to the target under consideration. These leads could correspond to known endogenous ligands or result from early hits discovered by high throughput screening (HTS). Trimming of methods by additional information has already been introduced into docking tools. The option to specify positions in the binding site to be favorably accommodated by certain atom types was first implemented in the program Targeted-DOCK [DesJarlais, 1994], [Good, 2003], [Shoichet, 1993] as an extension to DOCK [Ewing, 1997], [Kuntz, 1982], [Shoichet, 1992]. Later, a “similarity-driven approach to flexible ligand docking” described the possibility to include a similarity term to a given reference ligand or pharmacophore during docking [Fradera, 2000]. The program FlexX [Claussen, 2001], [Rarey, 1996], [Rarey, 1997], [Rarey, 1999] allows one to define starting fragments or pharmacophore restraints [Hindle, 2002] to guide the ligand placement during its construction phase. However, the subsequent affinity ranking of the generated docking solution(s) is purely based on a general scoring function. No additional information about the binding mode of already characterized ligands is considered at this stage.

To incorporate knowledge about characterized ligands, Grüneberg *et al.* [Grüneberg, 2001], [Grüneberg, 2002], analyzed the library to be docked by computing mutual superimpositions of the candidate molecules with known high affinity ligands. Only candidates showing some degree of similarity in terms of spatial physicochemical properties with the reference structures were further considered for docking. This strategy retrieved subnanomolar inhibitors for human carbonic anhydrase II. Furthermore, tailor-made scoring functions have been derived by either correlating force-field contributions with experimental affinity data of a known training set [Holloway, 1885] or by factorizing the total binding energy into individual components of van der Waals, Coulomb, and desolvation contributions per residue.

In the COMBINE (COMparative BINDing Energy) [Ortiz, 1995], [Wang, 2001b], [Wang, 2002] approach these contributions are correlating with experimentally determined binding free energies to derive a 3D QSAR model. COMBINE factorizes the protein-ligand interactions on a per residue basis into individual contributions. This can be protein residues as well as parts of a ligand. Both ligand and protein design is facilitated by the spatially resolved selection of important protein-ligand interactions. However, with respect to ligand design this more generic information of contributions to protein-ligand interactions is somewhat difficult to translate into actual design concepts of how to optimize a ligand by chemical means.

In case of missing structural information about the receptor, binding affinity predictions are usually attempted by “three-dimensional quantitative-structure-activity-relationship” (3D QSAR) methods such as “Comparative Molecular Field Analysis” (CoMFA) [Cramer, 1988], [Cramer, 1989], [Cramer, 1993] or Comparative Molecular Similarity Indices Analysis (CoMSIA) [Klebe, 1994]. Surprisingly high predictive power can be achieved. However, their success strongly depends on the composition and mutual alignment of the training set ligands. If knowledge about the protein structure is incorporated into the latter approaches, it is only used to generate a meaningful alignment of the ligands [Cruciani, 1994], [Medina-Franco, 2004], [Pastor, 1997], [Sippl, 2002].

Surprisingly enough, in the affinity prediction step, information about the surrounding protein environment is virtually neglected. This rather imprudent fact prompted Gohlke and Klebe [Gohlke, 2002a] to develop a tailor-made scoring function (which can also be termed a protein-based CoMFA). In this method, the general purpose knowledge-based pair-potentials of DrugScore [Gohlke, 2000a] are specifically adapted with respect to a given protein. The AFMoC (Adaptation of Fields of Molecular Comparison) approach [Gohlke, 2000a] allows to gradually move from generally applicable to specifically adapted affinity scoring considering

an increasing amount of information about the training set inhibitors. Important enough, already a small set of several known inhibitors can enhance the predictive power and makes AFMoC a suitable tool to support the drug development process in its early phase.

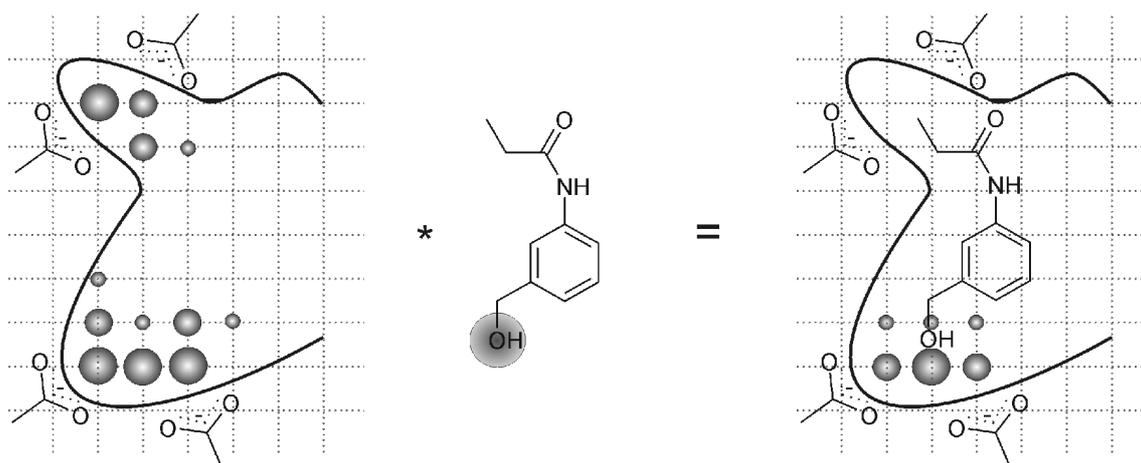
In the present contribution, AFMoC's potential to improve the predictive power in affinity prediction will be demonstrated using the first leads from a drug development program for 1-deoxyxylulose-5-phosphate (DOXP)-reductoisomerase inhibitors as potent antimalarials. The sole application of standard 3D QSAR methods did not achieve any prospectively predictive models in the present case study, although significant correlations among the training set compounds could be derived. This observation can be regarded as a caveat that comparative molecular field analyses easily achieve descriptive potential, however rarely accomplish the desired predictive power required for ligand design.

### ***The AFMoC Approach***

Distance-dependent pair-potentials of atomic protein-ligand contacts retrieved from experimentally determined protein-ligand complexes have been widely used to score protein-ligand interactions. As such, they have been successfully applied to rank docking solutions [DeWitte, 1996], [DeWitte, 1997], [Gohlke, 2000a], [Gohlke, 2001], [Muegge, 1999a], [Muegge, 1999] in de novo design or served as objective function in ligand placement [Sotriffer, 2002a]. However, as these potentials only provide an average description of a general type of interactions, special situations in biological systems that deviate significantly from this "average" cannot be described adequately. For example, an ionic hydrogen bond formed at the rim of a binding pocket, partly exposed to the solvent, contributes significantly less to affinity than a comparable contact deeply buried in a hydrophobic environment. Per se, both are not distinguished using "averaged" potentials. If additional structural and energetic knowledge about a system under investigation is considered, such "averaged" potentials can be tailored to the specific situation. In this respect, AFMoC allows to specifically adapt

DrugScore pair-potentials to one particular protein. AFMoC requires a protein structure with a predefined binding pocket and several bound ligands. If such data are available, the mutual ligand superpositioning can be achieved based on the crystallographically characterized protein-ligand complexes; if no such complex structures are available, solutions from docking runs might serve equally well, though with the clear caveat that the considered geometries are computed and not experimentally confirmed.

The AFMoC approach proceeds in three steps. First, a regularly spaced grid is embedded into the protein binding pocket. Similar to the “hot spots” analysis [Gohlke, 2000b], the interaction between a probe atom of a given type (as parameterized for the original DrugScore) and the surrounding protein atoms is calculated at every grid point. To account for missing experimental contact data at close distances, an artificial repulsion term is added to the original pair-potentials in this region. In a second step, structural information about the bound ligands is considered. By placing a Gaussian function onto a particular atom position, a distance-dependent contribution of ligand atoms of a given type with respect to each grid point is established. For each atom type and at each grid point, this contribution is multiplied with the original DrugScore value (potential field), yielding “interaction fields” (Figure 2.1). In a third step, the interaction fields are correlated with experimental binding affinities using Partial least-square (PLS) analysis.



**Figure 2.1** Concept of the AFMoC approach. As an example DrugScore “potential fields” for the ligand atom type “hydroxyl oxygen” are calculated (left). The ligand is described by assigning Gaussian functions to the positions of the individual atoms and loading them by the atom types under consideration. Subsequently, the potential field values at the different grid positions are multiplied by the values of the adjacent Gaussian functions attributed to the atoms of the accommodated ligand. Finally, the obtained product values are mapped back onto the binding site grid points. This figure is adapted from Gohlke and Klebe [Gohlke, 2002a].

In contrast to the “generic” fields in CoMFA or CoMSIA (i.e. electrostatic, steric, ...), AFMoC- “interactions fields” are based on individual ligand-atom types. Depending on the size and composition of the training set, some atom types might occur only rarely. As a consequence, the corresponding fields will hardly achieve the required statistical significance. Using the original pair potentials in DrugScore, experimentally determined  $pIC_{50}$  can be factorized into contributions originating from all considered atom types. Only those atom types which explain affinity differences are subsequently considered in PLS ( $pIC_{50}^{PLS}$ ). In consequence, the contribution of atom types, not explicitly considered in the PLS ( $pIC_{50}^{tot} - pIC_{50}^{PLS}$ ), must be regarded in terms of their original DrugScore potential contributions. Therefore, statistical results are always described by two correlation coefficients. First the coefficient considering only the *contribution to affinity* corresponding to atom types used for the PLS and second, considering the *total affinity* and using PLS results for the corresponding

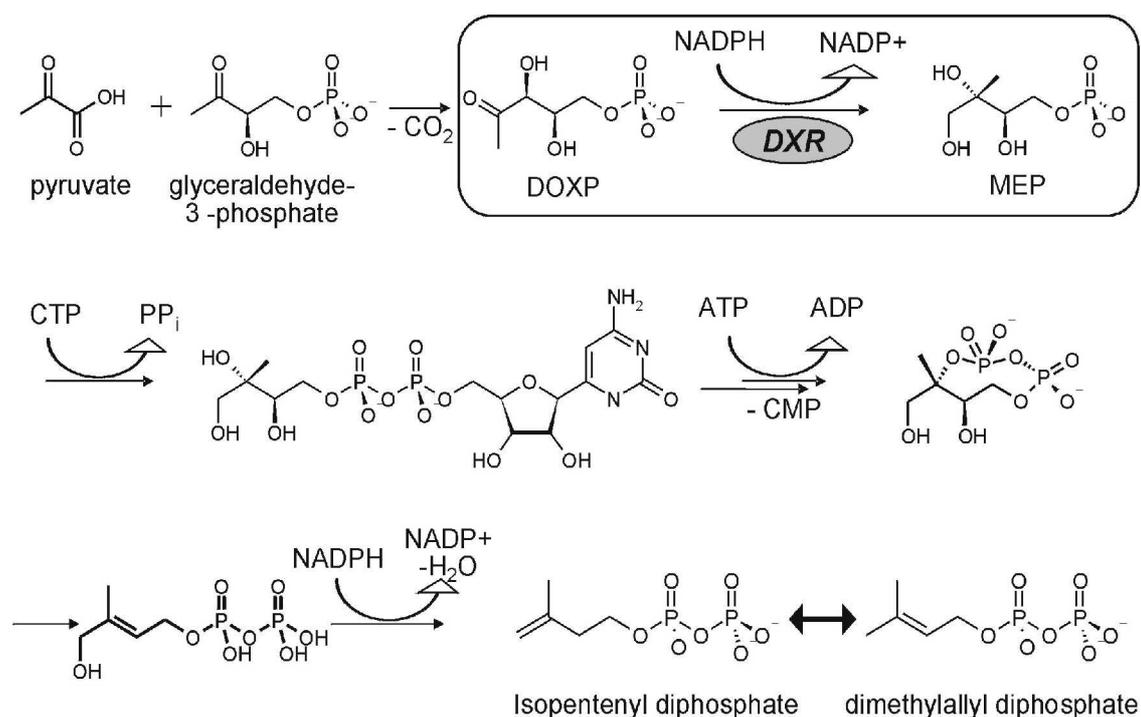
atom types and general DrugScore for the remaining ones. By use of a linear combination of the “original” DrugScore potential fields and the AFMoC-adapted ones, the approach can gradually shift between generally applicable and specifically adapted fields for prediction. Accordingly, an optional mixing coefficient  $\theta$  is used to switch from pure DrugScore ( $\theta=0$ ) to pure AFMoC fields ( $\theta=1$ ). As in CoMFA, the results of an AFMoC analysis can be interpreted in graphical terms consulting contribution maps, and binding affinities of novel ligands are predicted by evaluating the established 3D QSAR equation. Therefore, compounds to be predicted must be placed into the binding pocket by the same docking protocol as applied to the training set ligands. As major advantage of AFMoC, ligands to be predicted can structurally exceed beyond the scope of the training set compounds or comprise deviating functionalities, simply because properties not included in PLS will be scored purely based on the original DrugScore potentials.

### **Target Protein**

Malaria is world wide one of the most severe infectious diseases killing more than three million people each year [WHO, 2005]. The most fatal infections are caused by the protozoa *plasmodium falciparum*, the pathogen for malaria tropica. Due to large scale administration a widespread resistance has developed and quinine-type antimalarials have meanwhile become less effective in therapy. Moreover, some strains of *plasmodium falciparum* in Africa, Thailand, and South America have mutated in a way to produce multi-drug resistance. Thus, antimalarial research desperately seeks new targets, involving novel therapeutic principles. In this context valuable information is expected to result from the complete sequencing of the plasmodium genome [Gardner, 2002a], [Gardner, 2002b]. As a consequence, several new

pathways, unique to pathogenic microbes, have already been identified that could culminate in the development of innovative antimalarial chemotypes.

In recent years, the biosynthesis of isoprenoids being involved in a plethora of natural products, has attracted considerable attention since an alternative biosynthetic route has been discovered to the classical mevalonate pathway [Eisenreich, 1998], [Eisenreich, 2004], [Lichtenthaler, 1999], [Rohdich, 2001a], [Rohdich, 2001b], [Rohdich, 2002], [Schwender, 1996], [Seto, 1997]. Most isoprenoids are assembled from two universal precursors, dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP). These are either synthesized via the classical mevalonate pathway or along the newly discovered non-mevalonate or DOXP-pathway. As *plasmodium falciparum* relies exclusively on this second DOXP route, all enzymes involved in this pathway (Figure 2.2) could serve as potential targets for antimalarial therapy. The first intermediate, 1-deoxy-D-xylulose-5-phosphate, is a precursor not only of isoprenoids but also of the cofactors thiamine pyrophosphate and pyridoxal phosphate. Consequently, the subsequent enzyme DOXP-reductoisomerase (IspC/DXR) [Kuzuyama, 2000], [Kuzuyama, 2002], [Mueller, 2000], [Takahashi, 1998], operating on this substrate through an isomerization and reduction step, has been the first target selected for a drug development program. Recently, successful inhibitor design has also been reported for the downstream enzyme 2C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase [Crane, 2006]. Other enzymes of the pathways are also investigated regarding target suitability.

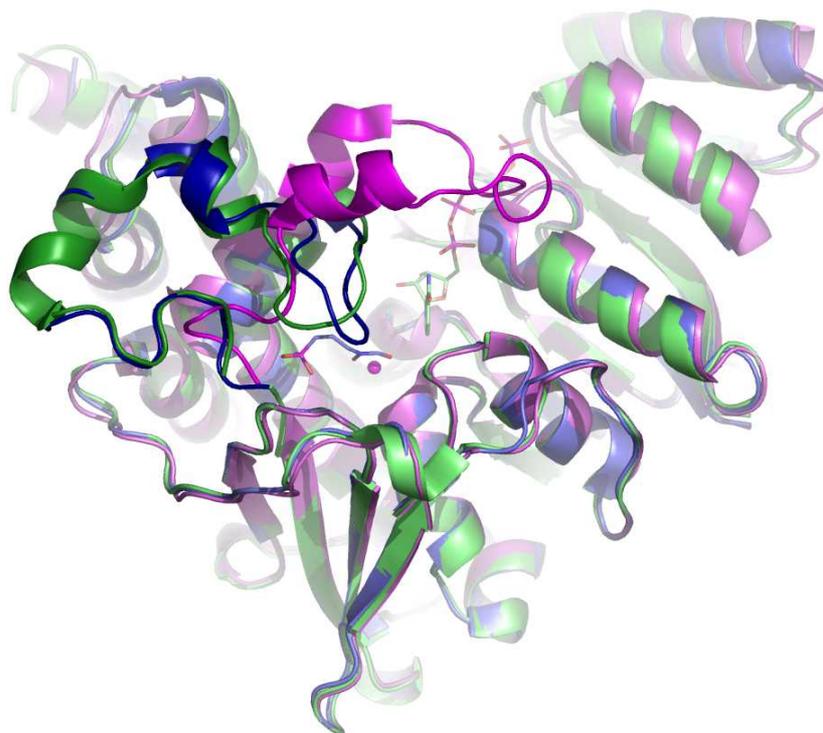


**Figure 2.2** Biosynthesis of isoprenoids precursors IPP and DMAPP via the non-mevalonate pathway. The framed step shows the reaction catalyzed by DOXP-reductoisomerase (IspC/DXR).

DOXP-reductoisomerase is inhibited by fosmidomycin (Table 2.1-1) [Jomaa, 1999], a drug which is already in clinical trials and shows promising antimalarial activity [Borrmann, 2004], [Lell, 2003], [Missinou, 2002], [Wiesner, 2003a]. Despite of its good inhibitory potency, fosmidomycin shows unfavorable properties such as complexation of essential bivalent metal cations. Therefore, a consortium of several groups began a synthesis program described in other contributions to discover improved lead structures [Reichenberg, 2001]. To support this development by rational ligand design concepts, computer methods were applied to provide some insight into the structural determinants for DXR-inhibitor binding. Although previous studies provide new insights into the binding mechanism [Proteau, 1999] and DXR has reached enormous attention [Proteau, 2004], [White, 2003] an inhibitor with higher affinity to DXR has not yet been discovered. Chemical modification of fosmidomycin include cyclization of the alkylchain to rigidify molecule [Devreux, 2006]. New substrate analogs and fosmidomycin analogs have been published [Phaosiri, 2004], [Woo, 2006] as well as prodrugs

to enhance fosmidomycin efficiency [Ortmann, 2003], [Ortmann 2005]. The lack of success to develop a non-fosmidomycin like inhibitor has motivated other experimental and computational studies to gain insights into target specificity [Fernandes, 2005], [Merkle, 2005], [Singh, 2006].

Crystal structure analyses have shown that the protein can adopt various different states. The apo structure (pdb ID: 1K5H) [Reuter, 2002] lacks the bound catalytic metal ion, the cofactor and a bound substrate or ligand molecule. This fact handicaps its direct use for structure-based drug design. Although an alternative structurally deviating entry (pdb ID: 1JVS) [Yajima, 2002] accommodates NADPH and a sulfate ion in the active site, only little information can be extracted with respect to putative protein-ligand interactions. This is mainly due to the fact that an extended loop folds down on the catalytic site, forming interactions to the bound ligand. In the apo structure three different loop conformers are observed, among them, two exhibiting rather distinct loop conformations. Steinbacher *et. al.* [Steinbacher, 2003] published the first DXR-ligand complex structure (pdb ID: 1ONP) [Berman, 2000] which shows fosmidomycin together with one  $Mn^{2+}$  ion bound to the active site. Unfortunately, this structure lacks the cofactor. The covering loop region is only poorly defined in the electron density.



**Figure 2.3** Superimposition of the apo (pdb ID: 1K5H, violet), holo (1JVS, green) comprising the cofactor NADPH and the fosmidomycin- and  $Mn^{2+}$  (magenta) bound protein (1ONP, blue). For all three structures, only the chain A is shown and the part of the flexible loop is highlighted.

Nevertheless, it clearly shows differences compared to the previously determined protein structures. A superimposition of all three structures (only chain A for 1K5H, 1JVS, 1ONP) is shown in Figure 2.3. Recently, nine new DXR structures were published. Ricagno *et al.* solved the apo (pdb ID: 1R0K) and DXR-NADPH complex (pdb ID: 1R0L) structure from *Zymomonas mobilis* [Ricagno, 2004]. Besides small differences in the recognition of the cofactor, the enzymes of both organisms share high similarity in structure and active site architecture. Very close to the *Plasmodium falciparum* protein-inhibitor-complex structure (pdb ID: 1ONP) is the apo structure of *Mycobacterium tuberculosis* [Henriksson, 2006] (pdb ID: 2C82). Although the overall RMS is 12.6 Å, the binding sites are quite similar. It has neither metal ion nor NADPH bound, yet the flexible loop is very close. Two biphosphonate inhibitors (pdb Ids: 1T1R and 1T1S) bound to DXR were published by Yajima *et al.* [Yajima,

2004]. Neither NADPH nor a metal ion are present but a buffer sulfate ion is observed which fits to the substrate phosphate pocket similarly to the pdb ID: 1JVS. Compared to the fosmidomycin-metal-DXR complex, these inhibitors adopt totally different binding modes. Their biphosphonate moieties replace the metal ion and their hydrophobic substituents orient towards the solvent which enforces an overall 90° rotation of both inhibitors compared to fosmidomycin. Additionally, parts of the flexible loop occupy a region formerly accommodated by the propylene chain of fosmidomycin. Further crystal structures show the ternary complexes of DXR-NADPH-fosmidomycin, selenomethionine-labelled DXR-NADPH-fosmidomycin and the DXR-NADPH-substrate binding [Mac Sweeney, 2005] (pdb IDs: 1Q0L, 1Q0H, and 1Q0Q). Considering the assumption of a two step binding mode of fosmidomycin [Koppisch, 2002], the ternary structure 1Q0L indicates a 12.5° domain rotation, likely to occur upon cofactor binding, thus forming the “tight-binding conformation of the DXR-fosmidomycin complex”. Additionally, the loop interacts with the ligand, not observed in the former complex structures. Unfortunately, the protein was crystallized at pH=5. As reported, DXR is not catalytically active under these acidic conditions [Mac Sweeney, 2005] and binding of the metal ion is very weak. Additionally, the carboxylic acids which normally complex the metal ion are most likely protonated and therefore have different interactions. Superimposing the metal-fosmidomycin (1ONP) and the NADPH-fosmidomycin (1Q0L) structures ( $\alpha$ -fit) the N-hydroxyl oxygen in the latter complex replaces the metal ion in the former one. In total fosmidomycin is moved by 1.5 Å towards the cofactor binding site (rmsd of both ligands: 1.84Å). Although these new complex structures give very important information about the domain and loop movements, we believe that due to the low and unlikely relevant pH conditions, the protein-ligand interactions exhibited are supposedly quite different compared to the in vivo situation. The complex structural properties along with the presence of a strong substrate- or inhibitor-metal interaction and the observed adaptability of a

large flexible loop make this enzyme a real challenge for any currently available computer tool in structure-based drug design. The protein data bank had currently seven new DXR crystal structures on hold which will be released spring 2007. They all are from *Mycobacterium tuberculosis* DXR, including mutant apo structures as well as inhibitor-metal-complex structures. They will reveal very interesting details in the interaction mechanism and the dependencies between loop conformation and ligand and/or metal ion presence (pdb ID: 2JCV, 2JCX, 2JCY, 2JCZ, 2JD0, 2JD1, 2JD2).

## **Methods**

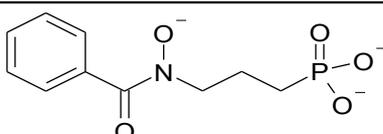
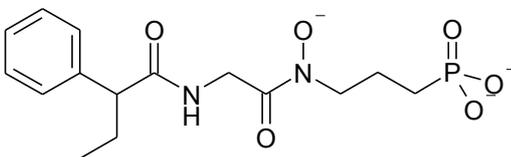
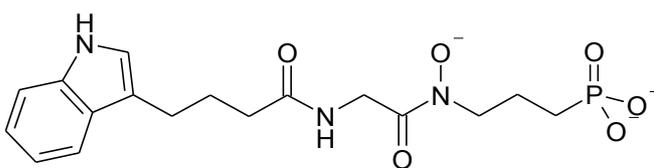
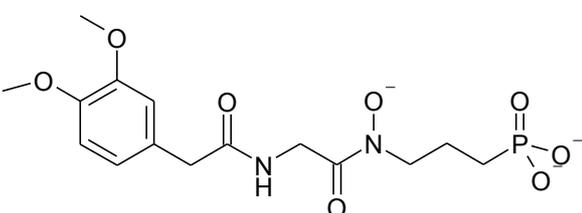
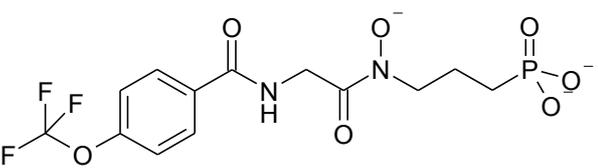
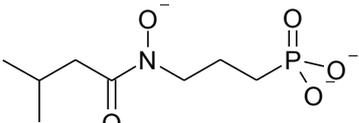
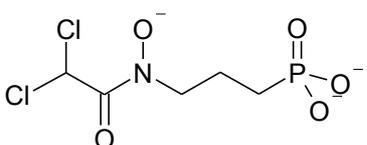
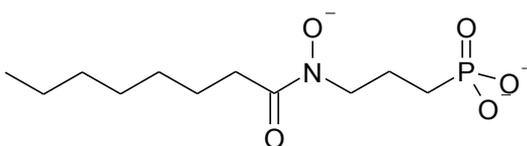
All molecular modeling and comparative molecular field analyses were performed using SYBYL 6.9 [TRIPOS, 2002] and DrugScore 1.2 [Gohlke, 2002a].

## **Data Set and Alignment**

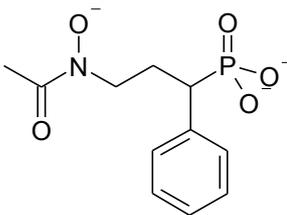
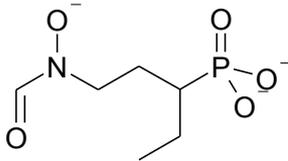
All ligands considered in this analysis were tested under the same conditions in the same laboratory. For affinity measurements as well as for crystallization and docking the *Escherichia coli* enzyme was used, which shows 74 % sequence identity to the putative active site of the parasitic enzyme. To obtain a diverse training set, we selected representative ligands from all compound classes complemented by additional examples to achieve an equal distribution of affinity data. This training set of in total 27 inhibitors was used for all CoMFA, CoMSIA, and AFMoC analyses (Table 2.1). The binding affinities covered a range of 3.6 logarithmic units.

**Table 2.1** DOXP-reductoisomerase inhibitors used as training set ligands.

No.	Name <sup>a)</sup>	Structure	<i>pIC</i> <sub>50</sub>
1	Fosmidomycin		7.70
2	BC183		4.30
3	FR900098		7.55
4	KL562		7.19
5	LD_1_1		6.59
6	LD_1_4		4.79
7	LD_1_5		4.50
8	LD_2_6		6.19
9	LD_2_7		5.19

No.	Name <sup>a)</sup>	Structure	<i>pIC</i> <sub>50</sub>
10	LD_2_8		7.00
11	LIHP_83		4.70
12	LIHP_84		4.77
13	LIHP_87		4.85
14	LIHP_88		5.29
15	LTB59		4.62
16	LTB65		6.75
17	LTB66		5.77

No.	Name <sup>a)</sup>	Structure	<i>pIC</i> <sub>50</sub>
18	LTB95		4.59
19	LTB97		4.39
20	TK54		5.63
21	TK85		5.01
22	UK153		4.30
23	UK213		5.74
24	UK243		4.00
25	UK513		7.27

No.	Name <sup>a)</sup>	Structure	<i>pIC</i> <sub>50</sub>
26	UK613		6.55
27	UK721		5.75
mean (standard deviation)			5.55 (1.0)

a) Name column names refer to compound codes as worked with during the project

Whereas the binding-site geometry of fosmidomycin(**1**) was taken as observed in the crystal structure with bound  $Mn^{2+}$  (pdb ID: 1ONP), the training set ligands were docked into the binding pocket using AutoDock 3.0 [Goodsell, 1996], [Morris, 1996], [Morris, 1998]. All ligands exhibit a hydroxamate function to mimic the  $\alpha$ -ketohydroxyl moiety of the substrate. Both oxygen atoms in the hydroxamate function show nearly the same distance to the metal (N-hydroxyl: 2.1 Å and carbonyl: 2.4 Å). Based on these structural findings along with an expected  $pK_a$  shift due to the neighborhood of a metal ion, the hydroxamate function is assumed as planar and deprotonated. The atom types of both oxygen atoms were assigned to be O.3. Additionally, most ligands exhibit a phosphonate moiety that is also considered deprotonated. Assuming these protonation states, most ligands were formally assigned a total charge of minus three. The partitioning of the total charge into individual atomic contributions was performed using the AM1 Hamiltonian [Dewar, 1985] within the semiempirical package MOPAC 6.0 [Limited, 1993], [Stewart, 1990].

From the reference crystal structure (pdb ID: 1ONP), ligand and solvent molecules were removed and  $Mn^{2+}$  was converted to  $Mg^{2+}$ . The latter exchange was performed as  $Mg^{2+}$  is better parameterized for all techniques applied in this study. Chemical experience has been

collected that both ions resemble closely in their coordination properties (as  $\text{Mn}^{2+}$  or  $\text{Mg}^{2+}$ ) and can mutually replace each other in active DXR.

To suggest a reasonable cofactor binding mode NADPH coordinates were transferred from pdb ID: 1JVS after both protein structures were superimposed, based on  $C_\alpha$  coordinates. A subsequent minimization of the transferred cofactor with the MAB force field, as implemented in MOLOC [Gerber, 1995], revealed no significant movements.

AutoDock 3.0 requires polar hydrogens for docking. They were added with the PROTONATE utility in AMBER [Case, 2005] and the generated hydrogen bonding network was visually inspected for internal consistency. AMBER united-atom charges were assigned as defined in the AMBER force field [Weiner, 1984], and solvation parameters were added using the ADDSOL utility from AutoDock3.0. All bonds except the one in the  $\text{C}(=\text{O})\text{-NO}$ -fragment, were kept rotatable. Docking runs were performed using the Lamarckian genetic algorithm as implemented in AutoDock 3.0, using an initial population of 50 randomly placed individuals, a maximum number of  $1.5 \times 10^6$  energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80 and an elitism value of 1. Generated ligand docking solutions, mutually differing by  $\text{rmsd} \leq 1 \text{ \AA}$  were clustered together and the lowest docking energy found for one entry of a cluster was used as representative. For each ligand, in total 10 solutions were generated. Out of these, one configuration was selected for the ligand alignment (Figure 2.4) by visual inspection, examining the obtained metal coordination and placement of the phosphonate group. This protocol was initially validated by docking fosmidomycin back into its original protein crystal structure (1ONP). Furthermore, docking into the model including the bound cofactor was attempted. Results are summarized in Table 2.2.

**Table 2.2** Validation of the docking protocol.

<b>Results of Docking Fosmidomycin with AutoDock</b>							
<b>DXR incl. NADPH</b>				<b>DXR excl. NADPH</b>			
<b>Rank</b>	<b>No. of molecules per cluster<sup>a)</sup></b>	<b>RMSD<sup>b)</sup></b>	<b>Energy</b>	<b>Rank</b>	<b>No. of molecules per cluster<sup>a)</sup></b>	<b>RMSD<sup>b)</sup></b>	<b>Energy</b>
<b>1</b>	4	0.96	-8.47	<b>1</b>	2	0.86	-8.63
<b>2</b>	4	1.31	-8.25	<b>2</b>	3	0.81	-8.52
<b>3</b>	1	1.23	-8.21	<b>3</b>	1	1.62	-8.46
<b>4</b>	1	1.12	-7.86	<b>4</b>	1	1.20	-8.40
				<b>5</b>	1	1.34	-8.40
				<b>6</b>	1	5.04	-8.25
				<b>7</b>	1	5.02	-8.20

a) rmsd-threshold = 1.0 Å. b) In Å compared to fosmidomycin as observed in pdb-entry 1ONP.

Additional docking with 100 runs was performed to approve significance of the previously applied parameters. Results are given in Table 2.9 (Supporting Information). As alternative, the docking program FlexX was tested to generate the required protein-ligand complexes. Default parameters were applied as well as including restraints as implemented in FlexX-Pharm. Therefore, the phosphorous-atom position was constrained as observed in 1ONP with a 1 Å threshold. Results are discussed below.

### **CoMFA and CoMSIA Analyses**

An identically oriented lattice of 1 Å grid spacing was chosen for the comparative molecular field analyses as already defined for docking with AutoDock, possessing a size of 22x25x22 Å thus guaranteeing sufficient embedding of all ligands with a margin of at least 4 Å. Steric and electrostatic CoMFA fields were calculated as implemented in SYBYL using Lennard-Jones and Coulomb potentials, respectively. All CoMFA calculations were

performed with SYBYL standard parameters (TRIPOS standard field, dielectric constant 1/r, cutoff 30 kcal/mol) using a  $sp^3$  carbon probe atom with a charge of +1.0. For CoMSIA, five physicochemical properties (steric, electrostatic, hydrophobic and hydrogen-bond donor and acceptor) were evaluated using a common probe atom of 1Å radius, charge +1, and hydrophobicity and hydrogen bond property values of +1. The attenuation factor  $\alpha$  used in the Gaussian functions was set to 0.3. PLS analyses and “leave-one-out” (LOO) cross-validations were performed using the SAMPLS [Bush, 1993], [Wold, 1984], [Wold, 1993] as implemented in SYBYL. The optimal number of components was determined by taken the following rules into account. As suggested by Thibaut *et al.* [Thibaut, 1993], the number of components has been checked to be lower than the amount of training set ligands divided by 5. The  $S_{PRESS}$  values achieved a minimum and any further component was only added if the  $q^2$  increased by at least 5%. The latter rule follows the “parsimony-principle” by selecting the smallest number of significant components. In our case, both rules suggested the same results. Accordingly, the model with the smallest number of components was subsequently used to establish the final QSAR. For CoMFA-PLS, the “minimum  $\sigma$ ” standard deviation (“column filtering” in SYBYL) was set to a threshold of 2 kcal/mol, resulting in approximately 16% of the columns considered in the PLS analysis. For CoMSIA, we compared the PLS results by selecting the same percentage of considered columns and the same  $\sigma$ -value. As both criteria produce similar results, we finally selected a  $\sigma$ -value of 2 also for CoMSIA. The  $q^2$ ,  $S_{PRESS}$ ,  $r^2$ ,  $S$  and *contribution* values were computed as described by Cramer *et al.* [Cramer, 1988], [Cramer, 1989], [Cramer, 1993]. They are summarized in Tables 2.3 and 2.4.

### ***Receptor-based predictions***

To compare the predictive power of the applied 3D QSAR techniques with those of protein-based scoring functions, the ligand orientations considered for the alignment were also

evaluated by three different scoring functions. As force-field based function, scores from AutoDock 3.0 were used. Additionally, the regression-based Boehm-scoring function [Bohm, 1994a] and DrugScore were applied. For DrugScore, affinities considering and neglecting the solvent-accessible-surface area dependent contributions were calculated (Table 2.7).

### ***AFMoC analysis***

Interaction fields for the superimposed DOXP-reductoisomerase inhibitors were calculated using the protein atom coordinates given in 1ONP as reference. The box size applied was identical to the prior studies. A value of 10 for the height of the added Gaussian repulsion function at short atom-atom distances was taken as suggested by Gohlke and Klebe. Subsequent to the calculation of DrugScore pair-potentials for the training set ligands, atom-type based Gaussian functions located at the spatial positions of each ligand atom were mapped onto the neighboring grid points and multiplied with the potential values to convert them into interactions fields. This step is only performed for those atom types that are frequently populated in the training set ligands and/or that show a variance in space to avoid false correlations when deriving the model ( $\text{pIC}_{50}^{\text{PLS}}$ ). For example, all phosphorous atoms in the training data are closely aligned in space. Therefore, phosphorous fields are not expected to provide significant information to the model and, hence, are not considered. PLS analysis extracts the contribution of each considered atom-type specific interaction field as given in Table 2.4. The correlation of interaction fields with the binding affinities of the training set ligands was calculated using the (SAM)PLS algorithm as implemented in AFMoC. The statistical significance of the obtained model was assessed by performing “leave-one-out” cross-validation runs. The optimal number of components was determined by selecting the smallest  $S_{\text{PRESS}}$  value and the same number of components was subsequently used to derive the

final AFMoC models. Statistical results for all 3D QSAR and AFMoC are summarized in Table 2.3 and 2.4.

**Table 2.3** Statistical results for PLS obtained for a leave-one-out analyses.

	CoMFA	CoMSIA	AFMoC
spacing <sup>a)</sup>	1.0	1.0	1.0
column filtering	2.0	2.0	
$\sigma$			0.85
No. components	5	4	4
$q^2$ <sup>b,c)</sup>	0.58	0.39	0.46 (0.36)
$s_{press}$ <sup>d,e)</sup>	0.80	0.97	0.97 (0.97)
$r^2$ <sup>b,f)</sup>	0.98	0.92	0.88 (0.86)
$S$ <sup>d,g)</sup>	0.20	0.35	0.46 (0.46)
$F$ <sup>b,h)</sup>	164.90	63.5	40.0 (32.8)

a) In Å. b) Values are given considering only the part of binding affinity ( $pK_i^{PLS}$ ) used in PLS analysis or considering the total binding affinity (values in parentheses). c)  $q^2 = 1 - PRESS/SSD$  as obtained by “leave-one-out” cross-validation. *PRESS* equals to the sum of squared differences between predicted and experimentally determined binding affinities, *SSD* is the sum of the squared differences between experimentally determined binding affinities and the mean of the training set binding affinities. d) In logarithmic units. e)  $s_{PRESS} = \sqrt{PRESS/(n - h - 1)}$  as obtained by “leave-one-out” cross-validation. *n* equals to the number of data points, *h* is the number of components. f) squared correlation coefficient. g)  $S = \sqrt{RSS/(n - h - 1)}$ . *RSS* corresponds to the sum of squared differences between fitted and experimentally determined binding affinities. h) Fisher’s F-value.

**Table 2.4** Field contributions to the models.

	CoMFA	CoMSIA	AFMoC <sup>b)</sup>	
<b>no. comp.</b>	5	4	4	
<b>fraction</b>				
<b>steric</b>	0.62	0.16	C.3	0.45
<b>electrostatic</b>	0.38	0.12	C.ar	0.25
<b>hydrophobic</b>		0.31	C.2	0.15
<b>donor</b>		0.28	O.3	0.06
<b>acceptor</b>		0.14	O.co2	0.05
			N.ar <sup>a)</sup>	0.04

a) For AFMoC, N.2 is converted to N.ar. b) The denotation of the combination of interaction fields follows the atom type convention of SYBYL, further details are given in the original paper describing DrugScore.

Plots of predicted versus experimentally determined  $pIC_{50}$  values are shown in Figure 2.5. Ten runs of a “leave-five-out” procedure were also performed. For this rather rigorous statistical test with respect to the amount of training set ligands, five arbitrarily selected compounds were discarded from the training set and subsequently used for prediction. Statistical results are shown in Table 2.10 (Supporting Information).

Additionally, the biological data were randomly scrambled and the model derivation was repeated, which allows the detection of possible chance correlations. Only negative  $q^2$  values in the PLS analyses were observed after randomizing the data by several protocols. Results are given in Table 2.11 (Supporting Information). Besides statistical analyses, AFMoC results can also be interpreted graphically. Therefore, the least squares coefficients and standard deviations (STDEV\*COEFF) assigned to each grid intersection were contoured within the binding pocket (Figure 2.10-12).

In order to validate the predictive power of the derived models, a test set of 16 additional DOXP-reductoisomerase inhibitors was selected (Table 2.5). To obtain a reasonable ligand

conformation, the same ligand placement protocol as described for the training set was applied. In contrast to any of the training set ligands, the test set includes a sulfonamide moiety to coordinate the  $\text{Mg}^{2+}$  ion as well as ligands that protrude beyond the commonly occupied volume accommodated by the training set ligands. The scaling factor  $c_s^{Pair}$  which adjusts DrugScore values to the  $pIC_{50}$  range was originally derived based on four targets covering a broad range of inhibitors. For the present study, it converged to a slightly different value ( $-1.97 \times 10^{-5}$  compared to  $-1.25 \times 10^{-5}$ ) possibly because the affinity data within the DXR data set differs in range and order of magnitude. AFMoC predictions were performed adjusting the mixing factor  $\theta$  in 10 steps of 0.1 from 0 to 1. To compare the performance of AFMoC with respect to the other approaches,  $pIC_{50}$  values for the test set ligands were predicted solely using protein-based scoring and comparative molecular field analyses (CoMFA, CoMSIA; see Table 2.8).

## **Results and Discussion**

### **Molecular alignment**

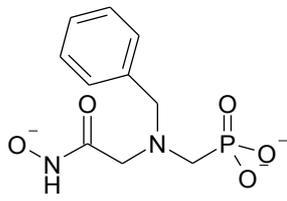
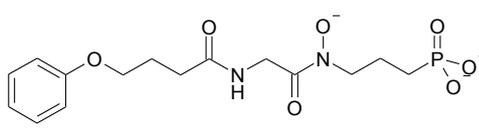
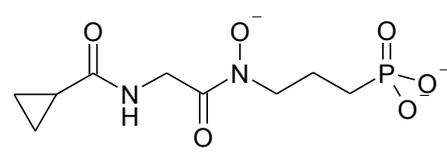
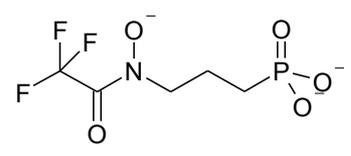
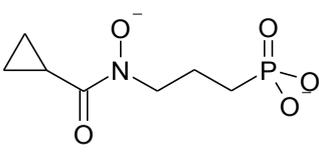
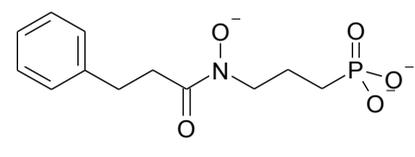
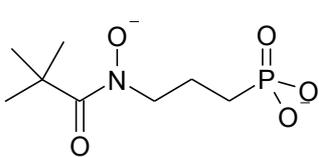
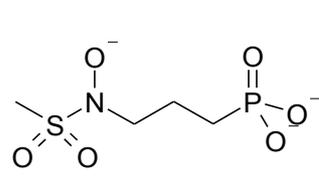
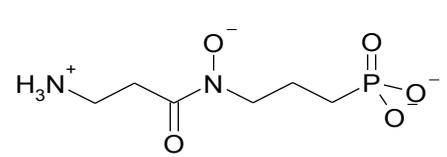
In the present study, we considered the crystal structure of DXR with fosmidomycin and merged the cofactor NADPH into the complex with a geometry as found in a related structure. For a training (27 compounds) and test set (16 compounds) of inhibitors with known binding affinity reasonable binding modes were generated by docking the ligands into the active site of DXR. The reliability of the applied protocol using AutoDock was validated by docking fosmidomycin successfully (rmsd = 0.86 Å) back into its crystal structure. Considering the modeled complex with NADPH, a very similar geometry is generated (rmsd with respect to the original orientation in the crystal structure = 0.96 Å). Taking the frequency how often a

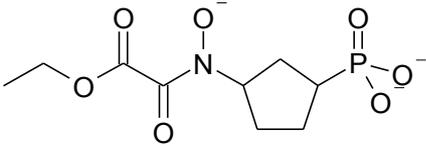
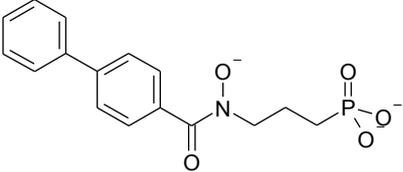
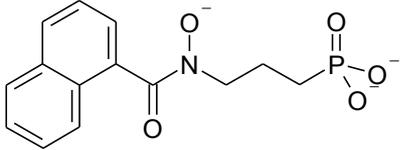
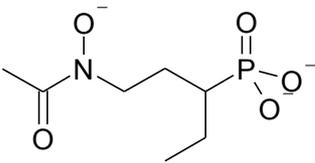
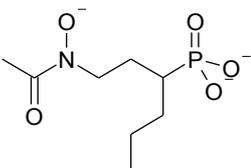
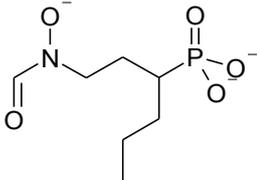
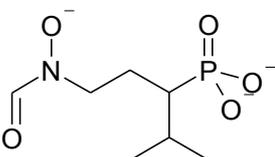
particular solution is found *within one cluster* by AutoDock as an additional figure-of-merit, the relevant binding modes are repeatedly detected, obviously even more frequently if the cofactor is bound. Therefore, we included NADPH in our docking simulations. Apparently, the cofactor restrains the binding pocket in space and, as a consequence, more relevant binding modes are produced. A repetition with 100 docking runs to increase statistical significance revealed similar results (see Table 2.9, Supporting Information).

As a matter of fact, FlexX did not generate ligand poses of comparable quality, neither based on its original scoring function nor on the implemented DrugScore. To our opinion, this deficiency can be traced back to an inadequate reproduction of the geometry between the hydroxamate moiety of fosmidomycin and the  $Mg^{2+}$  ion. Even by restraining the placement of the phosphorous atom using FlexX-Pharm, the metal hydroxamate interaction was not correctly generated. The better performance of AutoDock 3.0 in this respect might be due to its explicit consideration of electrostatic interactions which are apparently of overwhelming importance in the present example.

The protocol developed for fosmidomycin was similarly applied to all ligands. A suitable placement of the phosphonate and a correct coordination of hydroxamate moiety to the  $Mg^{2+}$  ion served as criteria to select the most appropriate docking solutions. For 16 (55%) out of 31 training set ligands, the top-ranked docking solutions were selected to generate the alignment. In four cases, no acceptable docking solutions could be obtained. Accordingly, those structures were excluded from further analyses, resulting in a training set of 27 ligands. A similar procedure was followed to construct the binding modes of the 16 test set ligands.

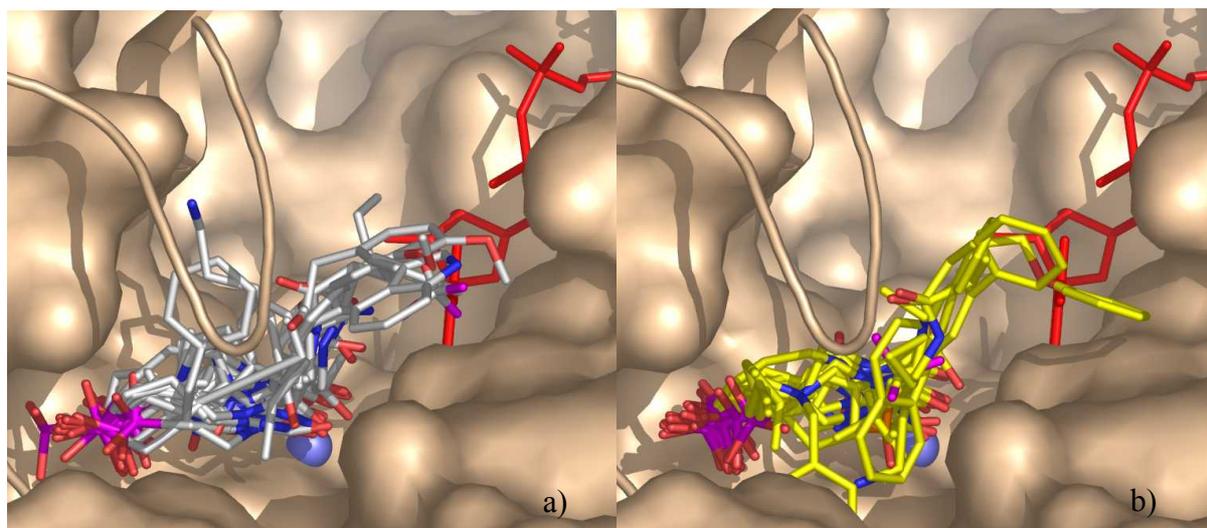
**Table 2.5** DOXP-reductoisomerase inhibitors used as test set ligands.

No.	Name	Structure	<i>pIC</i> <sub>50</sub>
28	LD_1_6		5.74
29	LIHP_86		6.00
30	LIHP_89		5.15
31	LTB104		4.85
32	LTB60		4.44
33	LTB71		4.30
34	LTB75		4.96
35	LTB77		4.42
36	LTB96		4.80

No.	Name	Structure	<i>pIC</i> <sub>50</sub>
37	UK163		4.80
38	UK2203		5.62
39	UK2213		6.64
40	UK711		6.15
41	UK811		5.18
42	UK821		5.09
43	UK921		5.18
mean (standard deviation)			5.2 (0.65)

a) Name column names refer to compound codes as worked with during the project

The finally achieved molecular alignment is shown in Figure 2.4. This alignment has been used in all comparative molecular field analyses and in protein-based affinity scoring.



**Figure 2.4** Docking solutions of DOXP-inhibitors, superimposed in the binding pocket of pdb ID: 1ONP. Amino acids Gly185-Ile218 which compose the flexible loop are shown as cartoon to allow a better view into the active site; a) 27 inhibitors of the training data set; b) 16 inhibitors of the test data set.

### **CoMFA/CoMSIA analyses**

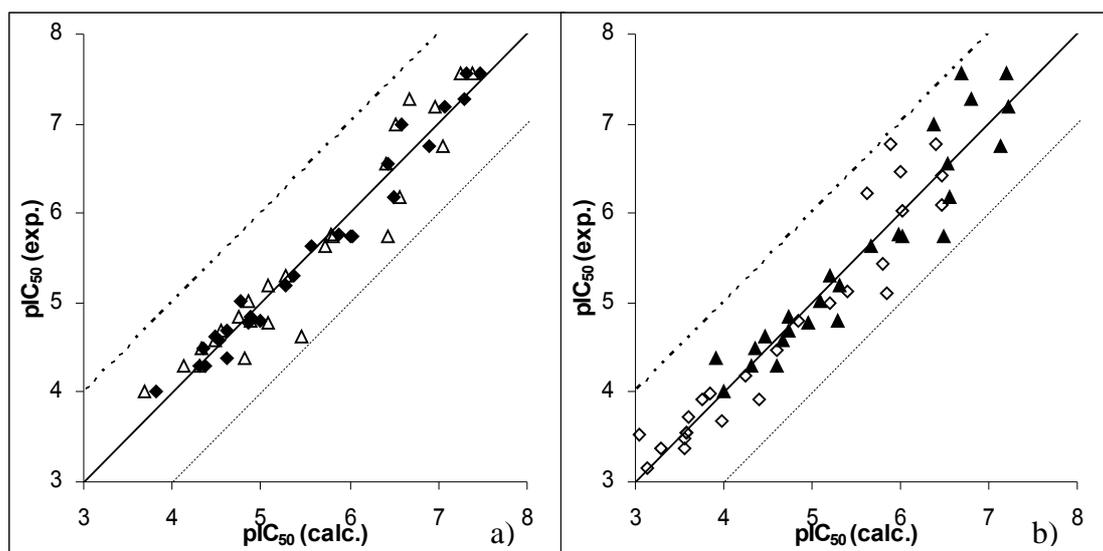
For CoMFA, a satisfactory training set model could be established whereas for CoMSIA a model at the borderline to significance has been obtained (Tables 2.4, 2.5, 2.9 and Figure 2.5a). The correlation coefficients for the leave-one-out analyses are  $q^2=0.58$  and  $q^2=0.39$  respectively, indicating “good” models. For CoMFA, steric interactions dominate, as their fraction contributes with 62% (Table 2.4).

Both 3D QSAR models lack any predictive power to correctly estimate the binding data of the 16 test set ligands (Table 2.7, 2.9). CoMFA as well as CoMSIA are known to be only valid for interpolation. Accordingly no satisfactory prediction is achieved for test set ligands comprising functional groups not present in any of the training set ligands. Additionally,

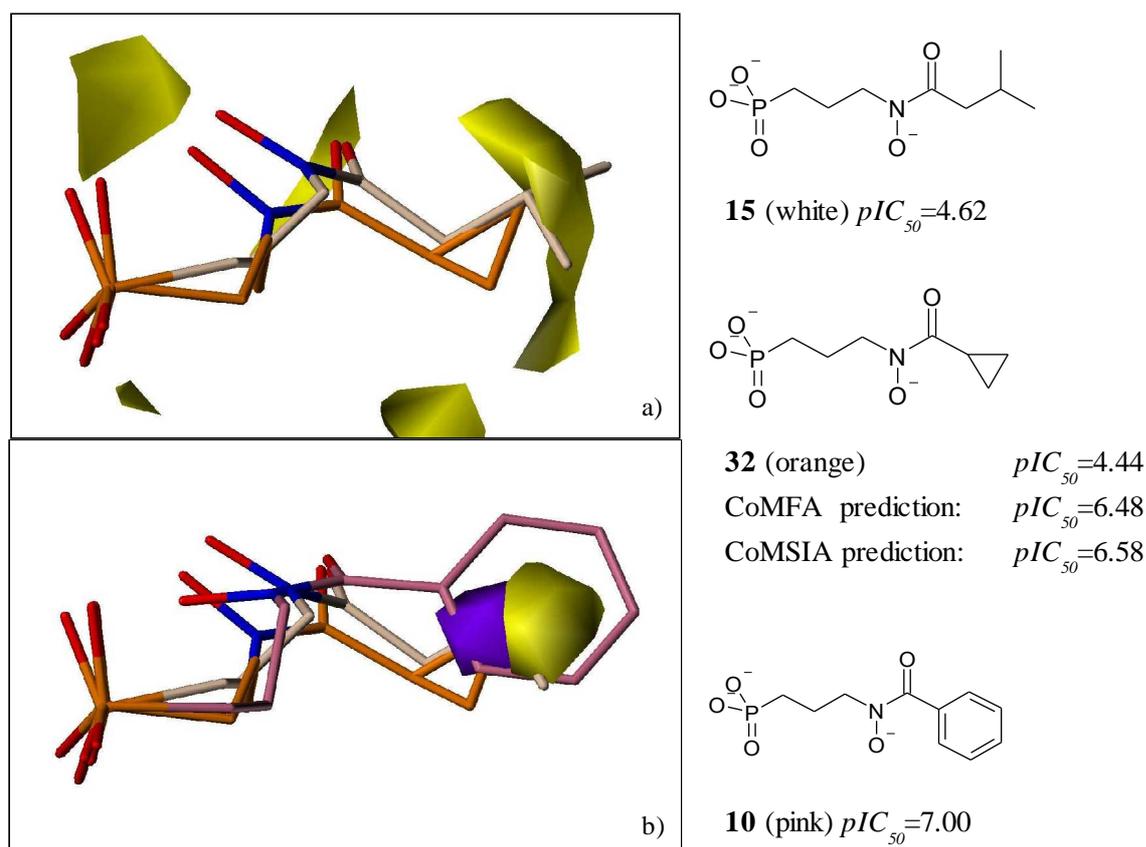
prediction fails for test set ligands that extend in space into areas never experienced by any of the training set ligands.

However, surprisingly enough, predictions also fail for ligands exhibiting only minor structural differences to the training set references. For example, training set inhibitor **15** ( $pIC_{50}=4.62$ ) and test set entry **32** ( $pIC_{50}=4.44$ ) differ only slightly in affinity. This parallels with minor structural differences; **15** has an isopropyl-group, whereas **32** exhibits a cyclopropyl moiety. Nevertheless, CoMFA predicts **32** by two orders of magnitude too strong in binding ( $pIC_{50}=6.48$ ). Remarkably, the smallest inhibitors of the training set exhibit the best affinity, whereas any increase in size parallels with a loss in affinity. Any spatial extension beyond the hydroxamate is detrimental to affinity. Nevertheless, if a test set ligand's placement of the phosphonate and hydroxamate functions matches exactly with those of fosmidomycin, this ligand is predicted as high in affinity. Test set inhibitor **32** ( $pIC_{50}=4.44$ ) serves again as an illustrative example. In Figure 2.6b it is shown together with the strong binding training set ligand **10** ( $pIC_{50}=7.00$ ). The phenyl moiety of the latter contributes to a favorable hydrophobic field which suggests the incorrect affinity prediction of **32** in CoMSIA.

Compared to **15** or **10**, the hydrophobic portion of inhibitor **32** does not accommodate an area indicated as unfavorable for steric occupancy but points to the field contributions favorable for hydrophobic properties. In consequence, CoMSIA ( $pIC_{50}=6.58$ ) overestimates **32** even more than CoMFA. Possibly a limited training set of 27 ligands captures insufficiently the molecular diversity supposedly allowed for ligands accommodating the binding site of DXR. We believe that this limited view of CoMFA and CoMSIA mainly arises from the fact that such field-based methods solely exploit ligand information and completely ignore the properties of the surrounding binding pocket. This detrimental disregard of any influences caused by the local binding site becomes increasingly evident once CoMFA or CoMSIA are trained by rather small or uniformly distributed data sets.



**Figure 2.5** a) Experimentally determined binding affinities versus fitted predictions using the derived CoMFA model for the training set. Results are shown applying the optimal number components which is 5 for CoMFA (◆) and 4 for CoMSIA (△), respectively. b) Experimentally determined binding affinities versus fitted AFMoC predictions for the 27 training set DOXP-reductoisomerase inhibitors. Both, experimental and calculated values are shown considering only the part of binding affinity ( $pIC_{50}^{PLS}$ ) used in PLS analysis (◇) or considering the total binding affinity (▲). In addition to the line of ideal correlation, dashed lines are depicted to indicate deviations of one logarithmic unit from ideal prediction.



**Figure 2.6** a) Stdev\*coeff contour plots elucidating the steric features as obtained by CoMFA analysis. Yellow contours (contour level 0.006 kcal/mol) highlight areas that should remain in their vicinity unoccupied, otherwise affinity will decrease. Despite its structural similarity to training set inhibitor **15** (white), test set ligand **32** (orange) is overestimated by CoMFA, because it does not penetrate into the unfavourable steric field region (yellow contour). b) Stdev\*coeff contour plots elucidating hydrophobic and steric features as obtained by CoMSIA analysis. Magenta contours (contour level -0.021) encompass regions favourable to be occupied by hydrophobic groups. Yellow contours (contour level -0.0083) highlight areas unfavourable for any steric occupancy of a ligand. Additional to a), also the strong binding training set ligand **10** (pink) is shown. It contributes in the training model to the favourable hydrophobic field which on the opposite is responsible for the overestimation of **32**.

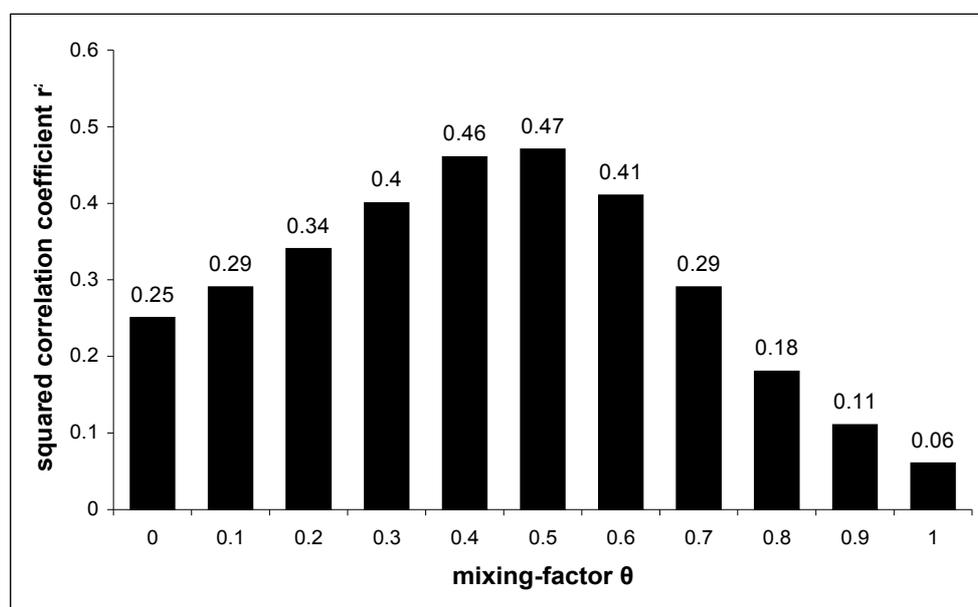
This observation motivated us to use AFMoC, since it considers simultaneously ligand and protein structural information. This enhancement should equip AFMoC with improved predictive power to correctly handle also test set ligands strongly deviating from their training set references.

## **Generating the AFMoC model**

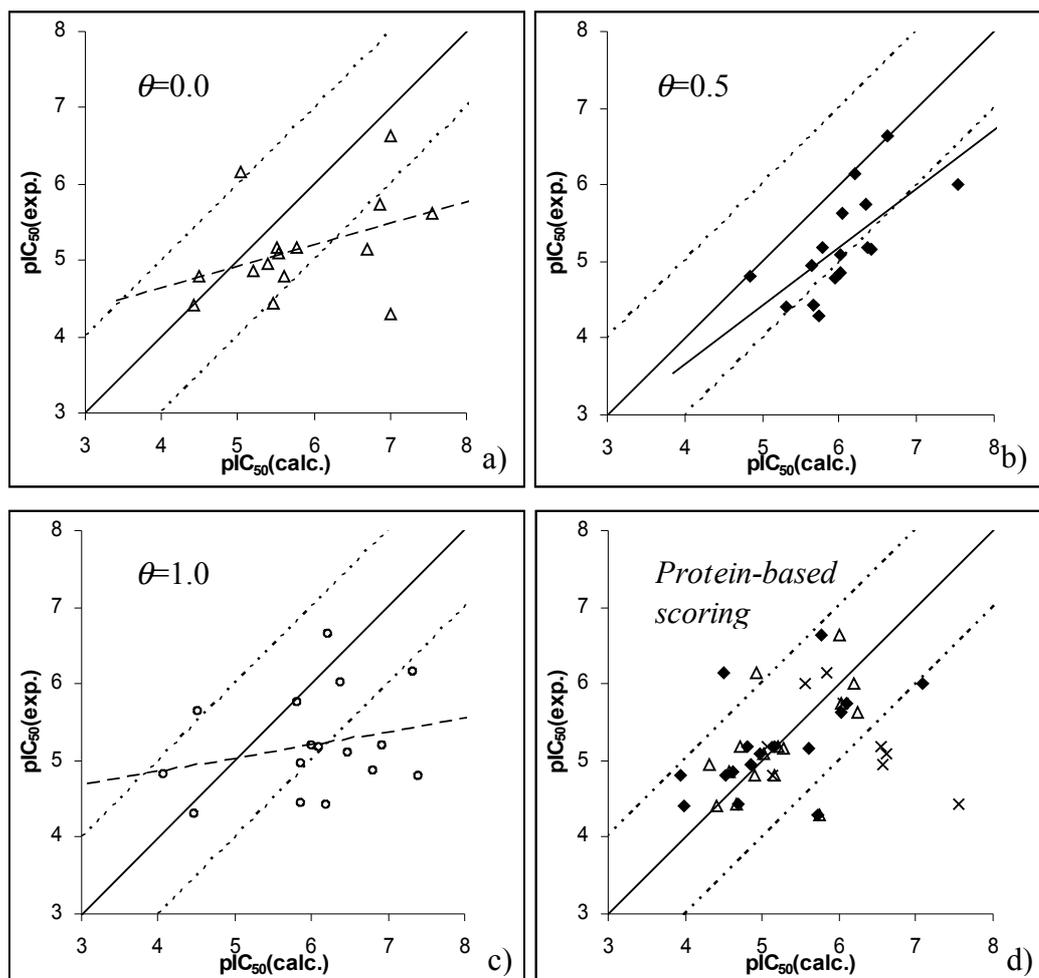
Results for the PLS analyses are given in Table 2.3. The LOO-analysis suggests optimal statistical results when considering four components in the model. With respect to the limited number of 27 training set ligands, a model based on a small number of components appears most reasonable. The correlation of predicted and experimentally determined affinity data of the training set is shown in Figure 2.5b), resulting in an  $r^2$  of 0.86. The final AFMoC model reveals a  $q^2=0.46$  considering only the atom-type contributions to affinity ( $pK_i^{PLS}$ ) used in PLS analysis, or  $q^2=0.36$  considering all contributions to binding affinity (Table 2.3, Figure 2.5). Ten runs of “leave-five-out” (LFO)-analyses resulted in  $q^2$  values between 0.41-0.61 for  $pK_i^{PLS}$  and 0.31-0.54 for the total binding affinity with optimal numbers of components varying between 3 and 5. This variance indicates the inappropriate size of the training set. Leaving out five rarely represented functionalities reveals models of only minor quality. Random assignment of the affinity data to the ligands of the training set resulted in only negative  $q^2$  (see Table 2.11, Supporting Information). These results indicate that the models derived with the properly assigned affinity data are statistically significant. As apparent from PLS, contributions of C.2, C.3, and C.ar fields dominate the correlation. In all cases the  $Mg^{2+}$  ion is coordinated by a hydroxamate group through the O.3 and N.2 atom. For convenience, the N.2 atom type has been converted into N.ar in our analysis. As the field contributions for both atom types correspond to 6% or 4%, resp., the correct placement of the group coordinating  $Mg^{2+}$  is important for affinity correlation in our model. Similarly the 5% contribution attributed to the O.co2 field experiences differences in the orientation of the phosphate groups across the data set. To some degree also the placement of these groups correlates with the affinity data.

### ***Application of the derived AFMoC model to predict binding affinities***

To assess the predictive power of the AFMoC model in particular in comparison to CoMFA and CoMSIA, binding affinities of 16 inhibitors not included in the training set were predicted. As mentioned, the linear combination of DrugScore and PLS contributions fields allows AFMoC to gradually move between generally applicable and specifically adapted interaction fields (Figure 2.7 and Figure 2.8). Accordingly, the mixing factor  $\theta$  was augmented in steps of 0.1 from 0 to 1. The best result was obtained for  $\theta=0.5$ , with a squared correlation coefficient  $r^2=0.47$ .



**Figure 2.7** Dependence of squared correlation coefficient ( $r^2$ ) for AFMoC models on the mixing-coefficient  $\theta$  between original DrugScore potentials ( $\theta=0$ ) and specifically adapted AFMoC fields (PLS-model,  $\theta=1$ ).



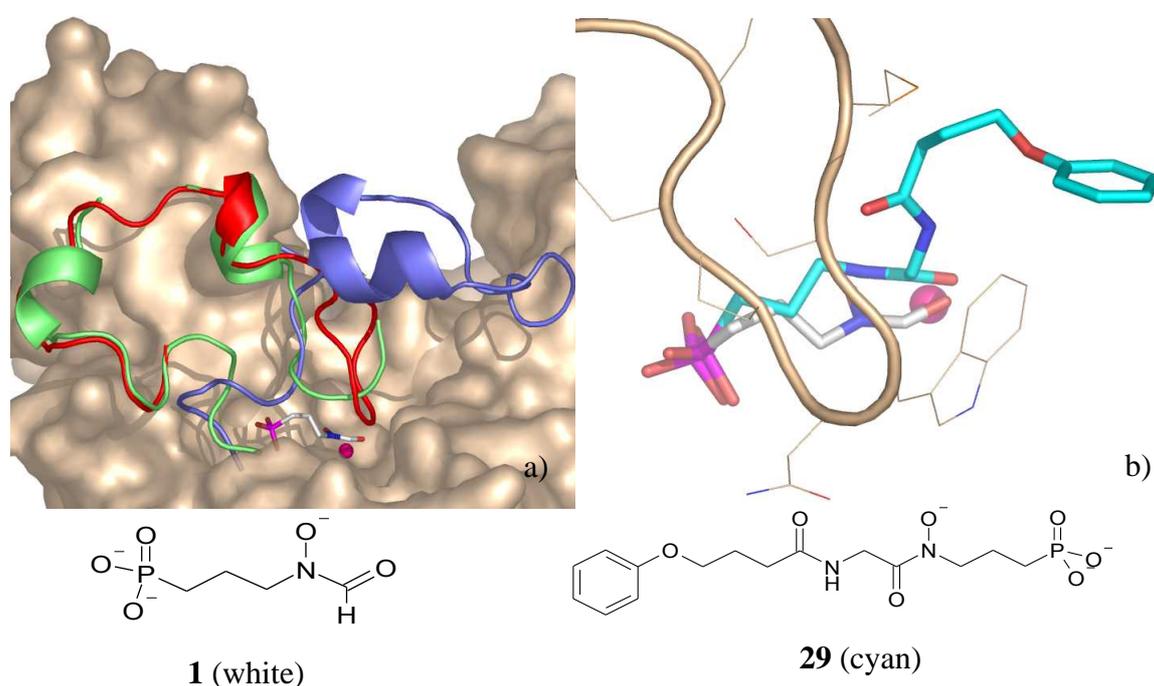
**Figure 2.8** a) Experimentally determined binding affinities versus fitted predictions for the test set molecules not considered in the training set. Correlations are shown for varying mixing-factor  $\theta$ .  $\theta=0$  (a,  $\Delta$ ),  $\theta=0.5$  (b,  $\blacklozenge$ ),  $\theta=1.0$  (c,  $\circ$ ). Best regression is obtained for  $\theta=0.5$ , indicated by the slope of the regression line. d) Experimentally determined binding affinities versus predictions for the test set using the following additional methods, based on the same binding modes as used for the comparative molecular field analyses. The diagram shows scoring using the Boehm function ( $\times$ ), AutoDock ( $\Delta$ ), and DrugScore considering only pair-potentials ( $\blacklozenge$ ) (including the solvent accessible surface term results only in minor deviations, data not shown).

**Table 2.6** Affinity predictions for the test set ligands, using AFMoC and the mixing factor  $\theta$ .

ligand	Mixing factor $\theta$												
	$pIC_{50}$	pure DrugScore						pure AFMoC					
		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	
28	5.74	6.86	6.76	6.66	6.55	6.45	6.35	6.38	6.15	6.04	5.94	5.84	
29	6.00	8.65	8.42	8.20	7.97	7.75	7.52	7.28	7.07	6.85	6.62	6.40	
30	5.15	6.71	6.65	6.59	6.53	6.47	6.41	6.34	6.29	6.23	6.18	6.12	
31	4.85	5.20	5.36	5.52	5.68	5.84	6.00	6.18	6.33	6.49	6.65	6.81	
32	4.44	5.47	5.51	5.55	5.59	5.63	5.67	5.70	5.75	5.79	5.83	5.88	
33	4.30	7.00	6.75	6.50	6.25	5.99	5.74	5.51	5.24	4.99	4.74	4.49	
34	4.96	5.39	5.44	5.49	5.54	5.59	5.64	5.77	5.73	5.78	5.83	5.88	
35	4.42	4.41	4.59	4.77	4.95	5.13	5.31	5.53	5.67	5.85	6.03	6.21	
36	4.80	4.49	4.78	5.07	5.36	5.65	5.95	6.23	6.53	6.82	7.11	7.40	
37	4.80	5.60	5.44	5.29	5.14	4.99	4.83	4.71	4.53	4.38	4.23	4.07	
38	5.62	7.55	7.25	6.94	6.64	6.34	6.04	5.74	5.44	5.14	4.84	4.54	
39	6.64	7.01	6.93	6.85	6.78	6.70	6.62	6.51	6.47	6.39	6.31	6.24	
40	6.15	5.05	5.28	5.51	5.74	5.97	6.19	6.43	6.65	6.88	7.11	7.34	
41	5.18	5.77	5.89	6.00	6.12	6.24	6.35	6.46	6.59	6.70	6.82	6.93	
42	5.09	5.54	5.63	5.73	5.83	5.92	6.02	6.15	6.21	6.31	6.41	6.50	
43	5.18	5.52	5.57	5.62	5.67	5.72	5.77	5.82	5.87	5.92	5.98	6.03	
$r^2$		0.25	0.29	0.34	0.40	0.46	0.47	0.41	0.29	0.18	0.11	0.06	

The minor performance of pure DrugScore ranking (Figure 2.8d) clearly underlines the need to include specific target information to create more predictive models. Since detrimental entropic contributions due to conformational immobilization of both binding partners are only rudimentarily taken into account by DrugScore, larger molecules usually receive a better scoring than smaller ones. However, in the present DXR example some of the smaller ligands exhibit high affinities. Possibly they mimic the substrate more closely allowing the flexible loop region to properly fold upon the substrate binding site

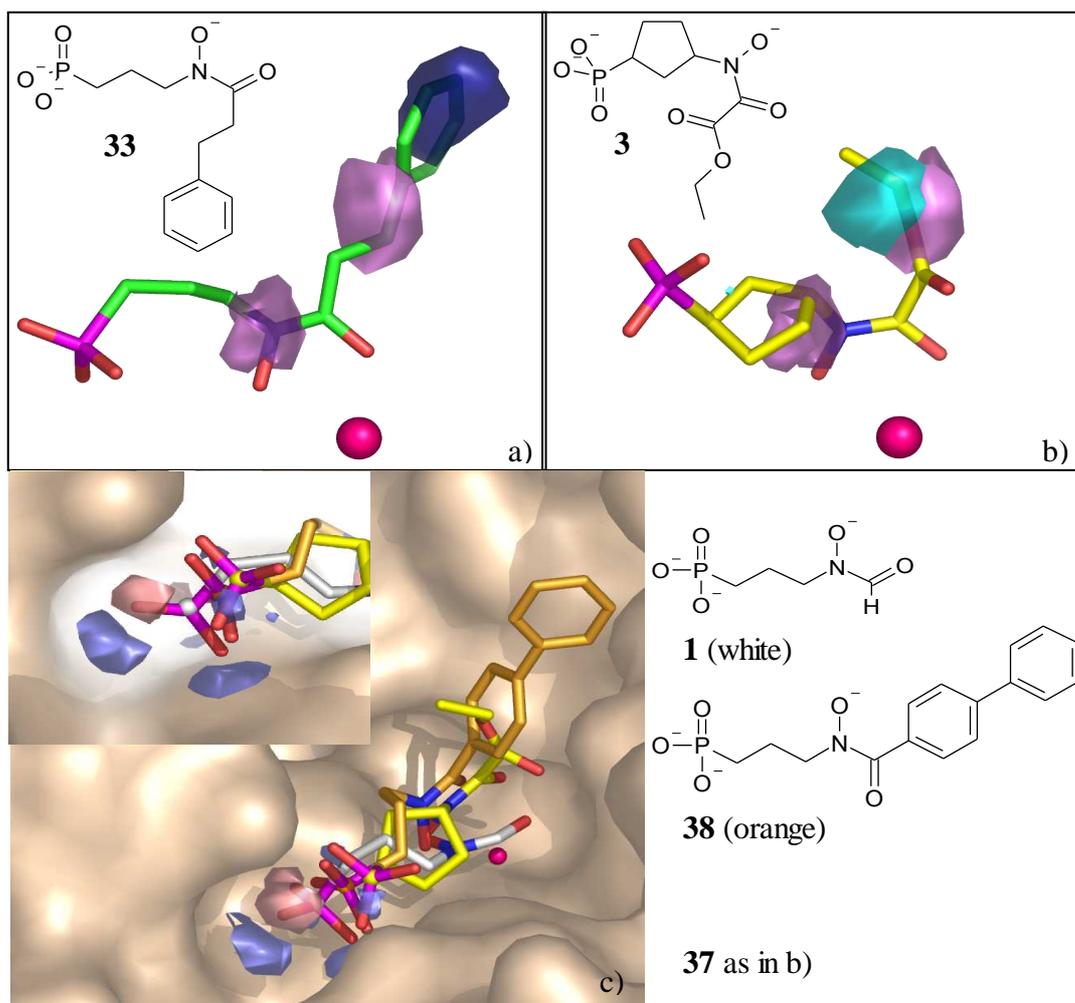
(Figure 2.9a). It might well be that in particular the substrate-like inhibitors gain in affinity due to the fact that the protein can adopt a fully relaxed conformation corresponding to its natural stage along the catalytic path. For larger molecules adequate loop closure is supposedly prevented. Possibly this explains why some of the larger DXR inhibitors frequently exhibit reduced binding affinities, even though they should be capable to form more favorable interactions compared to the smaller substrate analogs (Figure 2.9b).



**Figure 2.9** a) Different conformations of the flexible loop closing-up the active site of DOXP-reductoisomerase. In red the loop conformation is shown as determined for the fosmidomycin-Mn<sup>2+</sup> complex (pdb ID: 1ONP). Superimposed are the loop conformations adopted either in the apo structure (pdb ID: 1K5H, blue) and in the cofactor-bound complex (pdb ID: 1JVS, green). b) Superimposition of the proposed binding mode of the rather bulky inhibitor **29** onto fosmidomycin. Obviously, in addition to the shape of the binding pocket (not shown), the estimated conformation from docking strongly depends on the loop conformation (shown in wheat) selected for the docking studies.

In such a situation, the tailored AFMoC model is capable of dealing with this difference as the contribution of the various moieties in the ligands are weighted dissimilarly. For example, enhancing contributions next to the putative loop position adopted along the catalytic pathway

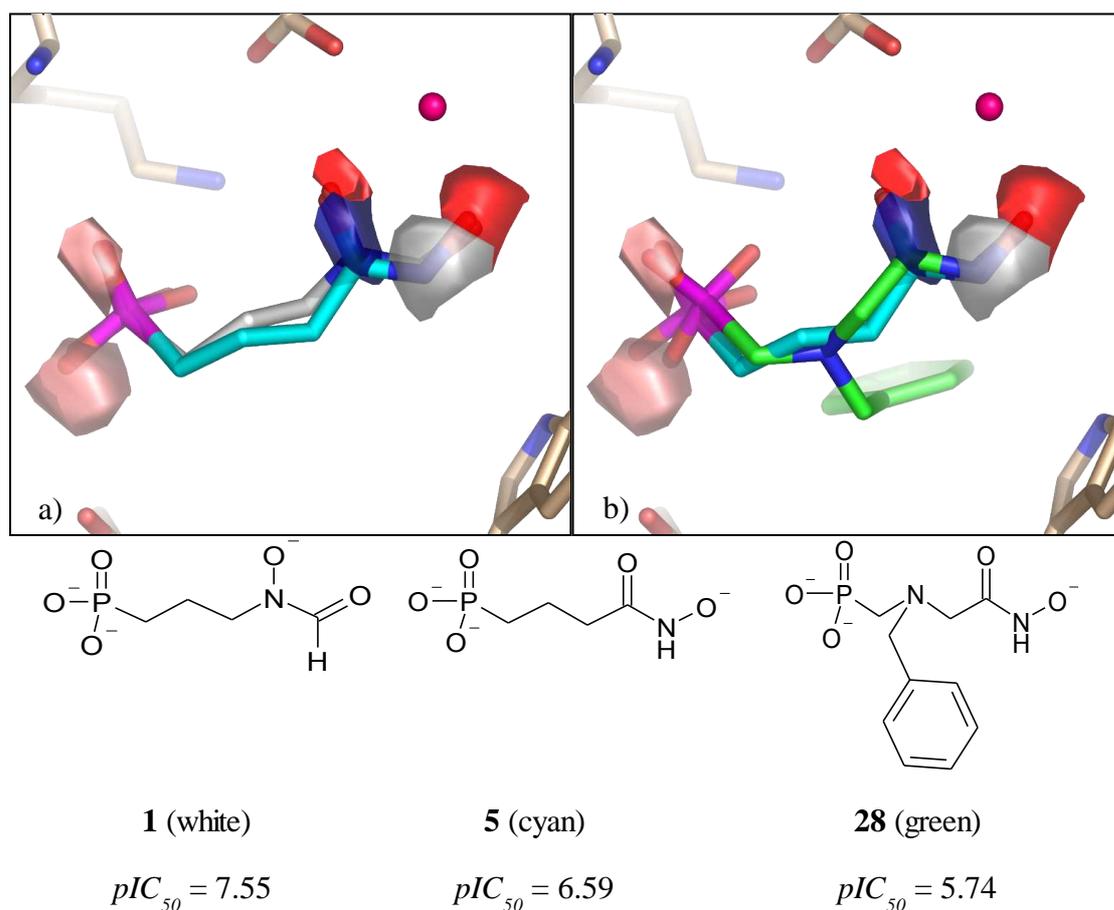
are reduced. Affinity predictions of the test set inhibitors **33**, **37**, and **38** is exaggerated by DrugScore (Table 2.6). Most likely, their large hydrophobic portions interfere with the putative loop conformation adopted in the fully closed state. In the AFMoC model, this repulsive impact is penalized with increasing  $\theta$  (Figure 2.10a+b). One might speculate that the pure AFMoC ( $\theta=1.0$ ) would reveal the best predictions. However, this is not the case. Possibly, the slightly shifted phosphonate groups of these inhibitors compared to the phosphonate positions in the training set are penalized (Figure 2.10c).



**Figure 2.10** Test set inhibitors **33**, **37**, and **38** as predicted based on AFMoC using DrugScore potentials only. The binding modes of **33** (a) and **37** (b) are shown together with AFMoC STDEV\*COEFF contour plots where the presence of ligand atom type C.2 (violet), C.ar (blue), and C.3 (cyan) will *reduce* binding affinity. Contour levels are 0.005, 0.015 and 0.04, respectively. c) In comparison to fosmidomycin (white), **38** and **37** obviously do not adopt an ideal interaction geometry with respect to the phosphonate group in our model. In addition AFMoC STDEV\*COEFF contour plots are displayed that elucidate regions in the binding pocket where the presence of O.co2 will enhance (pink) or reduce (blue) affinity. Contour levels are  $-0.007$  and  $0.002$ , respectively. The upper left image is a blow-up of the phosphonate binding modes. The corresponding phosphorous atoms are indicated as sphere shown in the same color as the corresponding ligand carbon atoms. The flexible loop capping the active site (is clipped off in the figure for reasons of clarity).

In a pure AFMoC model such a deviating placement is handled very strictly. In contrast, mixing with information from the original DrugScore potentials allows one to score appropriately this deviating placement of the phosphonate groups.

The key binding interaction of DXR inhibitors is the interaction with the  $Mg^{2+}$  ion. Upon visual inspection of original DrugScore versus adapted AFMoC contribution maps, it becomes obvious that the O.3, N.ar, and C.2 fields get increasingly restricted in space. Inhibitors **5** ( $pIC_{50}=6.59$ ) and **1** ( $pIC_{50}=7.70$ ), both included in the training set, differ only in their stereochemistry at the hydroxamate moiety because the sequence of hydroxylamino and carbonyl functions is reversed. This reversal parallels a loss in affinity by one order of magnitude compared to fosmidomycin (Figure 2.11a) possibly due to a different desolvation energy.

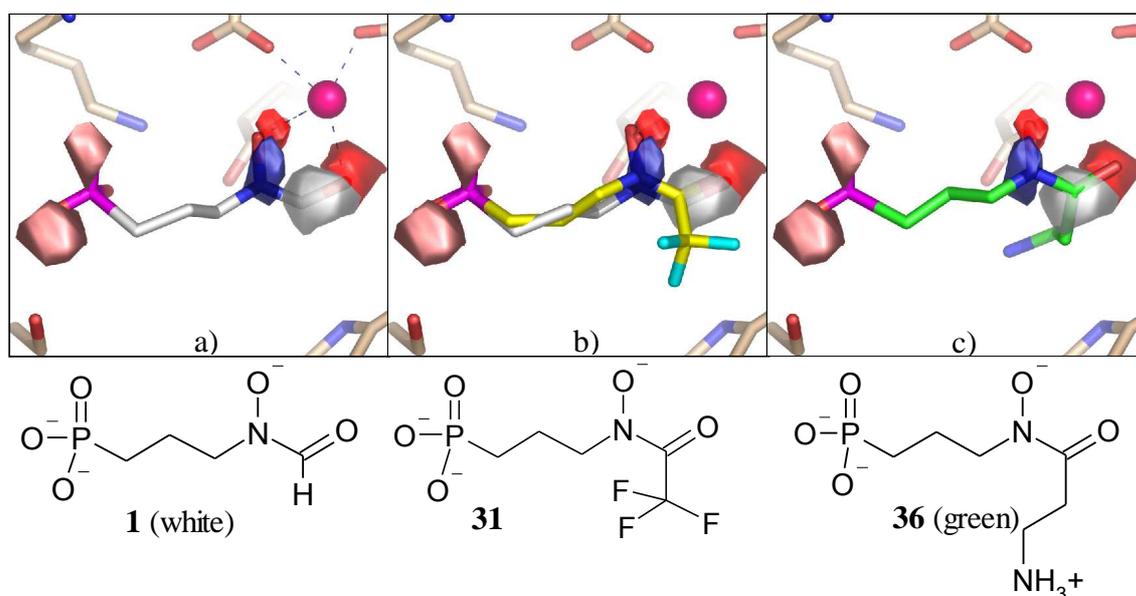


**Figure 2.11** AFMoC STDEV\*COEFF contour plots for ligand atom types N.ar (blue), C.2 (white), O.co2 (pink) and O.3 (red) representing regions where the presence of the corresponding atom type will enhance binding. Contour levels are  $-0.02$ ,  $-0.005$ ,  $-0.008$  and  $0.008$ , respectively. The position of the  $Mg^{2+}$  ion is indicated by the magenta sphere. a) Fosmidomycin (white) and training data set inhibitor **5** (cyan) which only differ by the reversal of their hydroxamate functions are distinguished by one order of magnitude in affinity. b) Test data set inhibitor **28** which exhibits a similar reversal hydroxamate superimposes well with the training set inhibitor **5**. AFMoC based on pure DrugScore potentials ( $\theta=0.0$ ), overestimates binding affinity for **28**. However, with increasing mixing of adapted AFMoC fields ( $\theta>0\dots1$ ), the presence of the reversed hydroxamate moiety is sufficiently penalized and a more accurate prediction is achieved.

Obviously this reversal does not affect the position of the terminal O.3 atoms that interacts with the  $Mg^{2+}$  ion. Adapted AFMoC fields for N.ar and C.2 next to the  $Mg^{2+}$  site are sharply contoured in a narrow spatial area (Figure 2.11). The N.2 and C.2 atoms of **5** coincide exactly with the encompassed regions. This fact of very localized field contributions using AFMoC is

convincingly demonstrated by the test set ligand **28** ( $pIC_{50}=5.74$ ; Figure 2.11b). This inhibitor shows the same reversed hydroxamate geometry as **5** and the accuracy of its affinity prediction improves with increasing contribution of the adapted AFMoC fields to the total score, to achieve an optimum (predicted  $pIC_{50}=5.84$ ) at  $\theta=1.0$  (Table 2.6).

Based on a sufficiently diverse training set covering a broad affinity range, one would expect that the adapted fields ( $\theta=1.0$ ) reveal the highest predictive power, because they include the largest amount of knowledge about the actually contributing target-ligand interactions. However, such an ideal spread of training set data is never given. Accordingly, a mix between AFMoC adapted fields and original DrugScore fields helps to overcome model restriction or over-fitting in particular in case of a small and restricted training set. For example, binding affinities of **31** and **36** are best predicted applying pure DrugScore potentials only (Table 2.6). As shown in Figure 2.12(a-c), they perfectly align with fosmidomycin and their N.2 and C.2 atoms coincide exactly with the corresponding AFMoC contribution maps. The same holds for the O.co2 atoms of the phosphonate group, in total these features results in a high affinity prediction of both compounds.



**Figure 2.12** Comparison of the binding mode of fosmidomycin (a, ligand shown with atom type coding), with two inhibitors **31** (b, ligand carbon atoms in yellow) and **36** (c, ligand carbon atoms in green). The latter two examples from the test set are overpredicted using the AFMoC-adapted fields ( $\theta=1$ ) because they firstly satisfy very similar interactions as the potent inhibitor fosmidomycin. Secondly, due to their additional substituent at the hydroxamate group, they can experience additional interactions with the protein. For analysis purposes the AFMoC STDEV\*COEFF contour plots are shown that highlight favorable regions for occupancy by a C.2 (white), N.ar (blue), O.3 (red) and O.co2 (pink) type ligand atom. Contour levels are  $-0.02$ ,  $-0.005$ ,  $-0.008$  and  $-0.008$ , respectively. The additional decorations of **31** and **36** are detrimental to affinity, however since none of the training set ligands reflect such information, the adapted AFMoC model cannot predict the properties of these additional substituents correctly. In the present case, the original DrugScore potentials perform somewhat better.

Nevertheless, both ligands comprise a side chain, e.g. a trifluoromethylene (**31**) and an aminopropyl group (**36**). These decorations are detrimental to affinity, however since none of the examples in the training set indicates such behavior, the AFMoC model cannot reflect this information. As a consequence, AFMoC fields fail to predict **31** and **36** correctly, both are highly overestimated in affinity. Considering purely the fields based on original DrugScore reveals much better predictions. For **36**, a slight mixing with adapted AFMoC fields ( $\theta=0.1$ ) achieves the best predictive power.

### Prediction using other scoring schemes

The predictive power of AFMoC (Table 2.8) has to be faced to prediction rates achieved by applying either the CoMFA or CoMSIA model or other scoring functions. A summary is given in Table 2.7, detailed information on scoring the test set ligand is given in Table 2.8.

**Table 2.7.** Statistical parameters for the prediction of binding affinities of the test set ligands applying docking, AFMoC and 3D QSAR.

	DrugScore <sup>a)</sup>	AutoDock	Boehm-fct	AFMoC <sup>d)</sup> ( $\theta=0.5$ )	CoMFA <sup>d)</sup>	CoMSIA <sup>d)</sup>
$r^2$ <sup>b)</sup>	0.28 (0.30)	0.33	<i>0.0002<sup>e)</sup></i>	0.47	<i>“0.33”<sup>e)</sup></i>	<i>“0.03”<sup>e)</sup></i>
SD <sup>c)</sup>	0.75 (0.74)	0.60		0.99		

a) Values are given considering only the pair potentials for the prediction or considering also the solvent-accessible surface term (values in parentheses). b) Squared correlation coefficient. c) In logarithmic units. d) Values are given for the optimal number of components, which are 4 for AFMoC, 5 for CoMFA, and 4 for CoMSIA. e) Values in italics denote lacking correlation.

**Table 2.8:** Affinity predictions for the test set ligands, using CoMFA, CoMSIA, DrugScore. AutoDock and the Boehm function.

ligand	$pIC_{50}$			DrugScore	Auto	Boehm-	
		CoMFA	CoMSIA	DrugScore	Dock	function	
28	5.74	5.08	5.70	6.11	6.20	6.05	6.06
29	6.00	4.95	4.37	7.08	7.06	6.18	5.57
30	5.15	5.38	4.99	5.60	5.55	5.27	10.18
31	4.85	6.38	7.39	4.63	4.57	4.57	8.06
32	4.44	6.48	6.58	4.70	4.76	4.68	7.56
33	4.30	5.58	6.23	5.72	5.65	5.75	9.26
34	4.96	6.59	6.75	4.85	4.84	4.32	6.57
35	4.42	7.01	6.04	3.98	3.91	4.40	8.12

ligand	$pIC_{50}$	CoMFA	CoMSIA	DrugScore		Auto	Boehm-
				DrugScore	incl. SAS	Dock	function
<b>36</b>	4.80	6.01	5.26	3.95	3.99	5.16	5.1
<b>37</b>	4.80	4.94	4.14	4.52	4.52	4.89	2.22
<b>38</b>	5.62	5.15	6.00	6.02	6.09	6.25	9.12
<b>39</b>	6.64	5.36	6.42	5.78	5.83	6.01	9.04
<b>40</b>	6.15	6.62	6.90	4.50	4.50	4.93	5.85
<b>41</b>	5.18	6.96	7.04	5.16	5.09	5.20	6.55
<b>42</b>	5.09	6.47	7.03	4.97	4.97	5.01	6.61
<b>43</b>	5.18	6.34	6.60	4.82	4.78	4.71	5.06
<b>r<sup>2</sup></b>		<i>“0.33”<sup>a)</sup></i>	<i>“0.03”<sup>a)</sup></i>	0.28	0.30	0.34	0.00

a) Values in italics denote lacking correlation.

As mentioned above, both 3D QSAR methods do not reveal any predictive power for the test set ( $r^2_{\text{CoMFA}}=“0.33”$ ,  $r^2_{\text{CoMSIA}}=“0.03”$ ). The AutoDock scoring function correlates the experimental data to docking results with an  $r^2=0.33$  (Table 2.7, Figure 2.8b). Major deviations can be attributed to a score-to-size dependency. E.g. compared to the actually more potent methyl derivative **40** ( $pIC_{50}=6.15$ ), the ethyl analog **41** ( $pIC_{50}=5.18$ ) obtains a slightly higher score, although **40** matches better with the fosmidomycin binding mode. Unfortunately, the Boehm scoring function is of no predictive power across the test data set. Considering the examples in detail shows that the Boehm function performed poor in particular for inhibitors with large hydrophobic moieties (**32**, **39**). Obviously, a general purpose scoring function as the Boehm function does not perform sufficiently well in the present case, since special features in the correlation cannot be regarded properly. DrugScore shows a weak correlation at the borderline to significance ( $r^2=0.25$ ). In summary it can be concluded that AFMoC with a mixing factor  $\theta=0.5$  achieves the highest predictivity across the different methods applied in this study.

## **Conclusions**

Comparative molecular field analyses were applied to correlate binding data of DOXP-reductoisomerase (IspC/DXR) inhibitors. The target enzyme provides some challenges for structure-based drug design: it (1) contains a metal-ligand coordination, (2) requires cofactor binding and (3) possesses a large flexible loop, that folds upon the active site. Although X-ray structures of the apo enzyme (pdb ID: 1K5H), the holo enzyme with bound cofactor (pdb ID: JVS), and a protein-inhibitor complex (pdb ID: 1ONP) lacking the bound cofactor have been solved, they do not provide enough information about the protein conformation adopted during the critical hydride transfer reaction step. This is mainly due to the fact that none of the crystal structures shows the flexible loop in a conformation relevant for the situation when metal ion, cofactor and substrate are simultaneously bound. It is assumed that in this situation ligands experience tightest binding. This assumption appears justified for the following reasons. First of all, the hydride transfer step has to avoid any access of water molecules from the surrounding solvent environment. Second, the close-up of the loop regions on top of the catalytic center will enhance the strength of the electrostatic interactions to the  $Mg^{2+}$  ion as a result of the shift in the dielectric conditions.

To establish a relevant structure activity relationship, in particular to predict affinity data for new inhibitors, we applied comparative molecular field analyses. Whereas CoMFA and CoMSIA succeed to produce reasonable models based on a training set of 27 aligned inhibitors, they do not exhibit any predictive power with respect to 16 ligands not included in the training set. To enhance comparative molecular field analysis, AFMoC can be considered as a knowledge-based scoring function tailored to a specific protein using additional inhibitor data. Similar to the other two 3D QSAR methods, the training set of 27 ligands produces a statistically significant model. In addition, AFMoC performs well on the 16 test set inhibitors, which comprise functional groups, not contained in the training data or address areas of the

binding pocket not experienced by any of the training set ligands. In contrast to CoMFA and CoMSIA, AFMoC provides reasonable affinity estimates. Best results are obtained if the original DrugScore fields are equally mixed with the PLS-derived AFMoC fields. Compared to pure DrugScore scoring, the tailored AFMoC fields also allow to better handle the affinity prediction of some stronger binding ligands of reduced size.

Especially at the beginning of a structure-based drug design project, when only one or two protein-ligand complexes are available and binding data of a small set of inhibitors has been characterized, AFMoC offers the opportunity to include a maximum of knowledge about the studied target protein into a computer model. With an increasing amount of additional data, the training set can be extended and in consequence the affinity prediction will improve the adopted model (i.e.  $\theta$  can be chosen larger). Furthermore, the AFMoC model can be applied as a tailored scoring function in virtual screening. Finally, by contouring the atom-type based contribution maps, AFMoC provides in a very similar fashion as usual CoMFA or CoMSIA the opportunity to create new ideas for drug design. However, as major improvement to the classical comparative molecular field analyses, AFMoC incorporates information about the interaction properties of the surrounding protein.

## Supporting Information

**Table 2.9** Approval of statistical significance of the docking protocol

Results of 100 docking runs for Fosmidomycin with AutoDock							
DXR incl. NADPH				DXR excl. NADPH			
Rank	No. of molecules per cluster <sup>a)</sup>	RMSD <sup>b)</sup>	Energy <sup>c)</sup>	Rank	No. of molecules per cluster <sup>a)</sup>	RMSD <sup>b)</sup>	Energy <sup>c)</sup>
1	80	0.83	-8.77	1	66	0.86	-8.99/
2	3	1.38	-8.59	2	4	1.34	-8.87
3	4	5.63	-8.43	3	11	5.54	-8.68
4	1	1.98	-8.41	4	2	5.44	-8.67
5	1	5.12	-8.37	5	3	1.47	-8.74
6	3	5.40	-8.30	6	3	5.07	-8.58
7	2	5.10	-8.21	7	3	5.49	-8.60
8	2	5.31	-8.11	8	2	5.52	-8.47
9	4	5.23	-8.06	9	1	4.94	-8.48
				10	2	5.26	-8.41
				11	1	5.37	-8.41
				12	1	5.17	-8.33
				13	1	1.80	-8.29

a) rmsd-threshold = 1.0 Å. b) In Å compared to fosmidomycin as observed in pdb ID: 1ONP. c) In kcal/mol.

**Table 2.10** Statistical results for 10 runs of „leave-five-out“ cross-validation, averaged over (no. of training set compounds/5)+1 runs.

no. of run	Leave-five-out		
	$q^2$ <sup>a)</sup>	$S_{press}$ <sup>b)</sup>	components
1	0.47 (0.36)	0.97	5
2	0.60 (0.53)	0.84	5
3	0.54 (0.45)	0.89	4
4	0.47 (0.37)	0.98	5
5	0.61 (0.54)	0.82	4
6	0.43 (0.32)	0.99	4
7	0.41 (0.31)	1.00	4
8	0.51 (0.42)	0.93	5
9	0.44 (0.35)	0.90	4
10	0.46 (0.36)	0.94	3
LOO <sup>c)</sup>	0.46 (0.36)	0.96	4

a) Values are given considering only the part of binding affinity ( $pK_i^{PLS}$ ) used in PLS analysis or considering the total binding affinity (values in parentheses). b) In logarithmic units. c) For comparison, results of the “leave-one-out” cross-validation are shown.

**Table 2.11**  $q^2$  values for AFMoC models obtained with randomly scrambled affinity data.

<b>no. of components</b>	<b><math>q^{2\text{ a)}</math></b>
<b>1</b>	-0.44 (-0.70)
<b>2</b>	-0.19 (-0.42)
<b>3</b>	-0.30 (-0.54)
<b>4</b>	-0.19 (-0.41)
<b>5</b>	-0.18 (-0.40)
<b>6</b>	-0.29 (-0.53)
<b>7</b>	-0.38 (-0.64)
<b>8</b>	-0.45 (-0.72)
<b>9</b>	-0.60 (-0.90)
<b>10</b>	-0.66 (-0.97)

a) Values are given considering only the part of binding affinity ( $pK_i^{PLS}$ ) used in PLS analysis or considering the total binding affinity (values in parentheses).

## METHODS IN RECEPTOR-BASED DRUG DESIGN

As outlined in the introduction, the present work comprises examples for computer-aided lead optimization. These challenges are mostly addressed via protein-based docking [Kitchen, 2004], [Sotriffer, 2002b]. This chapter gives a brief summary of currently available approaches and applications. Programs applied for the present work are discussed in detail.

### ***Protein-based docking***

Protein-based docking is defined as to computationally predict structures of protein-ligand complexes. The orientation that maximizes the interaction reveals the most accurate structure of the complex. Docking a data set of ligands comprises two problems; (1) generation of the ligand-binding mode that corresponds best to the experimentally given situation, and (2) reasonable estimate of the expected binding affinity and thus, ranking a series of different compounds for one target.

Early docking approaches were done orienting a ligand inside a three-dimensional structure of the protein and simply reducing steric clashes. Both, protein and ligand were treated as rigid bodies and only six degrees of translational and rotational freedom were sampled. With increasing computational power it is now possible to consider conformational degrees of freedom. Solving the configuration generation problem, three categories of search algorithms are used.

1. The ligand is divided into multiple fragments. The core fragment is docked into the binding site and additional fragments added in an incremental way. Examples are the

programs DOCK4.0 [Ewing, 2001], FlexX [Rarey, 1996], LUDI [Bohm, 1992a], [Bohm, 1994b], [Bohm, 2002] and HammerHead [Welch, 1996]. Being very fast, the pitfall of the approach is the high number of docking geometries possible for multiple core fragments. Unfortunately favorable protein-core fragment interactions might prevent the correct placement of the whole ligand.

2. Random changes are applied to a population of ligand geometries, mostly via Monte Carlo search or genetic algorithms. Examples are the programs GOLD [Jones, 1997], AutoDock [Goodsell, 1996], [Morris, 1996], [Morris, 1998] and ICM [Totrov, 1997]. Key issue here is the number of iterations necessary to correctly sample the binding site.

3. Full simulation of both, ligand and receptor using force fields. Examples are the programs AMBER [Case, 2005], CHARMM [Brooks B. R. Bruccoleri, 1983] and GROMOS [Osterberg, 2002]. Caveats here are the long computation times and the possibility of getting trapped in local minima.

One docking algorithm which combines the protocols is GLIDE [Friesner, 2004]. In a first step, active site properties are mapped on a grid, a set of ligand conformation is placed using a Monte Carlo approach and minimized. In the second step only a selected number (3-6) of conformations is considered for the subsequent Monte Carlo simulation. Comprehensive reviews on docking algorithms and their implementations have been given by Taylor, Halperin, Leach and others [Brooijmans, 2003], [Chen, 2006], [Coupez, 2006], [Halperin, 2002], [Leach, 2006a], [Sousa, 2006], [Taylor, 2002].

### ***Protein-based scoring***

Using structural information from both the protein target and the docked ligand geometry protein-based scoring functions are applied to achieve two goals. For a calculated set of

different docking solutions for one ligand to one target, the scoring function has to recognize the binding mode that corresponds best to the experimentally derived complex structure (structure prediction problem), thus, discriminating between correct and incorrect binding modes with reasonable accuracy and speed [Abagyan, 2001], [Kramer, 1999]. Based on the selected complex geometry, an ideal scoring function should be able to estimate as accurately as possible binding free energies. This way, different ligands can be ranked with respect to their affinity to one target (affinity prediction problem) [Stahl, 2001], [Wang, 2003]. Unfortunately, current scoring functions are a trade-off between fast evaluation of protein-ligand interaction needed for database screening and the accuracy. The limited accuracy is due to an incomplete understanding of complex interactions, to an incomplete training set for parameterization, and to a highly simplified approximation (or omission) of important variables in binding such as solvation and entropy [Chin, 2004]. There are three main classes of scoring functions.

1. Molecular mechanics functions are the closest to first principles that are used in docking. Examples are those implemented into the programs AMBER [Cornell, 1995], CHARMM [MacKarell, 1998], AutoDock [Morris, 1998], DOCK4.0 [Ewing, 2001] and GOLD [Jones, 1997]. These functions are based on additive atomic parameters (van der Waals, electrostatics, bonds, angles and torsions) usually derived from quantum mechanical calculations, which are assumed to be transferable between molecules containing similar atom types. Molecular mechanics scoring functions have a number of limitations. They model gas-phase contributions and neglect solvation and significant contributions to entropy. To account for speed, cut-off distances are applied which prevent the accurate treatment of long range effects.
2. Empirical or regression-based scoring functions are calibrated to associate numbers of atomic or molecular features (e.g. rotatable bonds, hydrogen-bonding parameters,

lipophilic interactions, solvation and entropic effects) with a training set of ligands with known affinity to their targets. Coefficients are determined via multiple linear regression, partial-least square or neural-network analysis. Examples are the Boehm scoring function applied in FlexX [Bohm, 1994a], [Bohm, 1998], [Bohm, 1999], X-Score/X-CCore [Wang, 2003] and ChemScore [Eldridge, 1997]. Often successfully applied it remains questionable, whether it is possible to predict binding affinities for ligands that are structurally different from those used in the training set.

3. Knowledge-based scoring functions apply a Boltzmann weighting to distributions of atom-atom distances observed in crystal structures. They are originally developed to reproduce binding geometries rather than binding energies. Binding effects, difficult to describe explicitly are somewhat captured implicitly in the derivation of interaction pair potentials. Examples for knowledge-based scoring function are DrugScore [Gohlke, 2000a], [Gohlke, 2001] and PMF [Muegge, 1999b], [Muegge, 1999], [Muegge, 2006] SmoG [Ishchenko, 2002] and BLEEP [Mitchell, 1999]. Their advantage is speed, better independence from a training set than empirical functions and a certain amount of robustness towards small changes in the protein target structure compared to molecular mechanics scoring functions. However, their capabilities are restricted to the information implicitly contained in the training set which can lead to a limited differentiation of atom types and rare distribution effects. Using the CSD instead of the PDB, this restriction has been overcome by deriving DrugScore<sup>CSD</sup> [Veleg, 2005]. The better resolved small molecule structures provide more contact data and produce potentials of superior statistical significance and more detailed shape.

A common pitfall for most scoring functions is the size dependency: the larger the ligand, the better the score [Kirtay, 2005]. Correction methods to account for stickiness of a molecule have been proposed [Vigers, 2004]. Protein-based scoring function have been reviewed by

Gohlke, Tame, and others [Gohlke, 2002b], [Kirtay, 2005], [Stahl, 2001], [Tame, 1999], [Tame, 2005].

### ***Applied docking programs and Scoring functions***

The three most popular [Sousa, 2006] docking programs have been applied in the present work: FlexX, GOLD and AutoDock. With respect to scoring, we additionally considered Chemscore as implemented in GOLD, DrugScore and SFC-Score.

#### ***FlexX***

In FlexX [Kramer, 1999], [Rarey, 1996], [Rarey, 1997] protein-ligand interactions are describes by certain interaction types and their corresponding geometry [Bohm, 1992b], [Klebe, 1994]. Interaction groups and their compatibilities are assigned to the ligands to be docked. For each interaction pair, a special geometry and a surface at a certain threshold is defined. An actual protein-ligand interaction is considered if the interaction center of one group contacts the surface of the counter group. The ligand placement is performed in three steps. The first step is the selection of the base fragment. The second step is the base fragment placement into the binding site. Conformers are created using the MIMUMBA torsion angle database [Klebe, 1999], [Sadowski, 2006]. The energy cutoff allowing divers conformers can be manually controlled as well as the number of possible base fragment poses. Subsequently, the remaining fragments are added in following a tree-search technique and a number of best partial solution based on assessment of the scoring function are kept until the ligand is completely assembled. The weakness of FlexX is docking very flexible and large molecules. A potentially reasonable base fragment placement might prohibit any further of the remaining

ligand to the binding site. This can be overcome in two different ways. For a series of similar compounds it might be useful to apply the *mapref* option. It allows to manually place the base fragment and therefore increases the probability of producing a reasonable binding mode. A second option is to apply user-defined constraints via the FlexX-Pharm option [Hindle, 2002]. This way, either a certain interaction type or a certain atom type at a specific position can be defined as essential or partial constraint. Checking the possibility to obey these constraints upon ligand construction, falsely placed fragments are discarded early and more reasonable geometries are further considered. The FlexX scoring function belongs to the class of empirical scoring functions. It comprises terms for hydrogen bonding, ionic interactions, lipophilic interactions as well as penalties for unfavorable angles. Coefficients were derived adapting the function to a data set of originally 45 protein-ligand complexes *SCORE1* [Bohm, 1994a] and later updated to 84 complexes *SCORE2* [Bohm, 1998], [Rarey, 1995].

### **AutoDock**

An example of the implementation of a genetic algorithm into protein-ligand docking is the program AutoDock [Goodsell, 1996], [Morris, 1996], [Morris, 1998]. To model putative interactions to be formed by the protein an user defined affinity grid is calculated for the binding site. For each atom type present in the ligand data set one grid is calculated with expected protein-ligand affinities attributed to each grid point. Additionally, one grid is calculated for the electrostatic potential. A population of ligand conformers is placed randomly onto the grid. Using either a Lamarckian genetic algorithm, a Monte Carlo simulated annealing or a traditional genetic algorithm ligand conformations are generated, evaluated using the affinity grids and kept or discarded. Stop criterion is either a defined number of energy evaluations or reaching a energy minimum with a further improvement

within a given energy threshold. Evaluating the three different algorithms, a combination of the Lamarckian genetic algorithm with a traditional genetic algorithm to perform a local energy minimization had been reported to perform best [Morris, 1996]. Scoring implemented in AutoDock is a force field-based function comprising five terms based on the AMBER force field [Case, 2005]. It comprises a Lennard-Jones 12-6 dispersion term, a directional 12-10 hydrogen bonding term and an intermolecular pairwise desolvation term. The scaling factor for each of these five terms is empirically calibrated from a set of 30 structurally known protein-ligand complexes.

## **GOLD**

Also GOLD uses a genetic search algorithm and allows for full ligand flexibility. Furthermore, it considers multiple orientations of protein polar hydrogens [Kairys, 2006]. Each possible protein-ligand interaction is defined as a chromosome. Starting from an initial randomly generated population of chromosomes the fitness is evaluated by the scoring function and the least fittest members is replaced. Ligands are placed into the binding site based on fitting points. Then, the program adds fitting points to hydrogen-bonding groups on protein and ligand, and maps acceptor points in the ligand on donor points in the protein and vice versa. Additionally GOLD generates hydrophobic fitting points in the protein cavity onto which ligand CH groups are mapped. The generic algorithm uses a least-square routine to attempt to form as many hydrogen bonds as possible. Therefore, rotatable ligand dihedral angles, ligand ring geometries and protein OH and  $\text{NH}_3^+$  groups are optimized. Many new options in GOLD now allow to keep specified scaffolds rigid, allowing to introduce user-specific interaction (e.g. metal interactions) [Verdonk, 2003]. The GOLD scoring function is force-field based and includes four terms, a hydrogen bonding term, a 4-8 intermolecular

dispersion potential, and a 6-12 intramolecular potential for the internal energy of the ligand and an intramolecular hydrogen bonding term which is mostly switched off in applications. Integrated into GOLD is the scoring function Chemscore [Verdonk, 2003]. It is a regression-based function considering scores for hydrogen-bonding, acceptor-metal, and lipophilic interactions, and the loss of conformational entropy. The coefficients were derived from a 82 protein-ligand complexes [Eldridge, 1997].

### **DrugScore**

DrugScore [Gohlke, 2000a], [Gohlke, 2001] is one of the most widely used knowledge-based scoring function. The basis is the comparison of the observed number of contacts between certain atom types to the number of contacts one would expect if there were no interaction between the atoms. For the development of DrugScore ~1400 protein-ligand complexes have been used. Due to missing information about non-binding events, the derived potentials are adapted using a reference state, representing an “average interaction”. For 17 atom types, the number of contacts between protein atom of type  $i$  and ligand atom of type  $j$  in the database (ReliBase+ <http://relibase.ccdc.cam.ac.uk/>) are tabulated as a function of their distance. To consider only direct protein-ligand contacts, the upper sample radius has been set to 6 Å. In addition to the statistical atom-atom pair potentials, DrugScore also contains a solvent-accessible surface (SAS)-dependent term. Non-polar surfaces on both protein and ligand in both complexed and uncomplexed state are calculated. They express the probability of protein and ligand atoms to get buried upon formation of the complex. The term rewards burial of certain atom types and penalizes the burial of others. Additionally to scoring final docking geometries, DrugScore potentials were also applied directly as energy function upon docking [Sotriffer, 2002a], [Radestock, 2005].

An extension to DrugScore was reported which uses the CSD instead of the PDB as database. The new pair-potentials (DrugScore<sup>CSD</sup>) are based on a larger number of crystal structures. Especially for ligand atom types rarely found in the PDB (like S, P, Cl) this led to a better recognition of “relevant” binding geometries and affinity prediction [Velec, 2005]. Formulas for all described scoring functions are given in the supporting information.

### ***SFC-Score***

In order to derive the empirical scoring function SFC score data was collected from both, public and industrial contributions. Over 60 descriptors were applied and evaluated for affinity prediction. Via correlation and regression analysis the most suitable descriptors were selected to calibrate the empirical function. Using selected training datasets, general functions as well as target class biased (serine proteases) or ligand feature biased (fragment-like sized) functions were derived [Sotriffer, 2007]. Applying the functions to several test data set superior performance of the SFCscore functions was observed in many cases, but the results also illustrate the need for further improvements.

### ***Addressing challenges for docking***

This work addresses a protein target showing flexibility in the active site and the docking of fragment-sized ligands. Therefore, two additional aspects of docking should be discussed. The first is how to consider protein flexibility during the docking process and the second is how to incorporate water molecules in the case the ligand occupies only parts of the pocket or protein-ligand interactions are transmitted via water contacts.

Most commonly used docking methods treat the ligand as flexible and the protein as rigid. However, in most cases it is not justified to neglect protein flexibility. McGovern and

Shoichet docked a data set against the holo (ligand bound), apo (empty) and protein binding sites derived via homology modelling [McGovern, 2003a] for several targets. Achieving best docking performances upon using the holo structures clearly demonstrates the importance of the particular representation of the target used in the screen. However, using only one particular protein structure will clearly bias virtual screening. Same conclusions can be drawn from so-called “cross-docking” experiments. For a single target and a set of experimentally determined protein-ligand complexes each ligand is docked into its own holo structure as well as into the other available protein structures. The lowest energy configuration for a ligand should be found when docking it back to the protein structure it was originally complexed with [Sotriffer, 2005], [Zentgraf, 2006].

To meet the need for a docking algorithm which considers certain degrees of protein flexibility while keeping computational complexity at a reasonable level, several approaches have been introduced recently. There are reviewed by Carlson [Carlson, 2000a], [Carlson, 2002]. Examples for programs which consider target flexibility are Glide, FlexE, Slide. The methodological foundation is (1) sampling binding site conformational space via molecular dynamic simulations [Sivanesan, 2005], (2) using rotamer libraries for binding site side chains [Alberts, 2005], (3) docking into an ensemble representation of the binding site [Carlson, 2000b] or (4) permitting some steric clashes with the protein [Ferrari, 2004]. However, it has also been reported that considering multiple protein structures for a virtual screening can result in reduced hit rates [Barril, 2005]. Successful virtual screening for a flexible binding site has also been reported by pragmatically only targeting one specific representation of the binding site [Kraemer, 2004].

Another aspect currently discussed with respect to docking of especially small ligands is the contribution made by water molecules. Water can form hydrogen bonds between the ligand and the protein and thus contributing to the affinity or it can be displaced by the ligand

[Ladbury, 1996]. Recently implemented into GOLD, water molecules can now be switched on and off [Verdonk, 2005]. The authors claim to be able to correctly predict water mediation/ or displacement in 93% of the tests.

Improving scoring function performance [Verdonk, 2004] has led to several new concepts. A pragmatic strategy to enhance the reliability of predicted binding affinities results from simultaneously considering multiple scoring functions [Clark, 2002]. Theoretically, this should be more robust and accurate than only applying a single method [Charifson, 1999], [Clark, 2002], [Wang, 2001b]. Proven successful in several applications [Bissantz, 2000], [Oda, 2006], [Wang, 2003], *consensus scoring* has been implemented in the FlexX-docking suite. A subsequent procedure, introduced by Terp [Terp, 2001], is to correlate several predicted affinities to experimental data via PLS to determine the best scoring function combination [Jacobsson, 2003]. Alternatively to consensus scoring, less relevant docking poses can be filtered out to avoid artificial scores [Springer, 2005]. More sophisticated, but much more time consuming is the application of MM/PBSA to protein-ligand docking. First studies [Kuhn, 2005], [Michel, 2006] have been reported as successful. However, an extensive analysis of comparing detailed kinetic data retrieved from isothermal titration calorimetry to MM/PBSA energy terms revealed certain shortcomings of the method [Zentgraf, 2006].

Many groups have addressed the task of comparing the performance of docking programs and scoring functions. Providing useful information there are certain caveats to consider [Cole, 2005]. Test sets must be large and appropriate. Numerical quality measurements need careful statistical interpretation. For affinity prediction, compatibility of the experimental data must be given. Systematic studies comparing docking engines and scoring functions are given by Ferrara, Perola, Kellenberger, Warren and others [Cummings, 2005], [Ferrara, 2004], [Kellenberger, 2004], [Perola, 2004], [Warren, 2006].

# OPTIMIZING BINDING AFFINITIES FOR BENZOPHENONE-BASED PROTEIN FARNESYLTRANSFERASE INHIBITORS

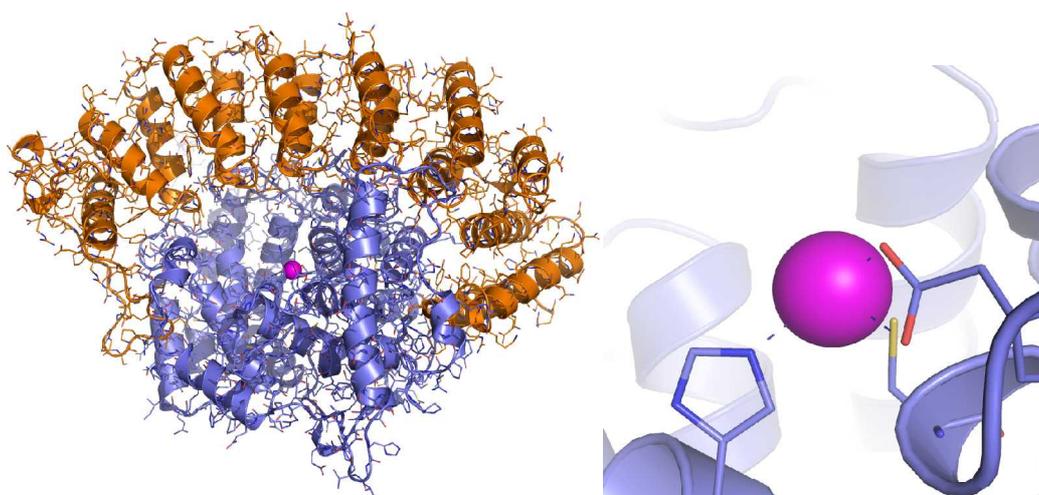
## ***Introduction***

During lead optimization, properties of a compound such as potency, selectivity, bioavailability, absorption and metabolism must be assessed and optimized. Although some strategies towards parallel optimization of these properties have been developed in the last ten years, enhancement of the potency (resp. activity) is most essential. Given a data set of ligands with experimentally determined activities computational models for quantitative structure-activities relationships are derived. These models are subsequently applied to prioritize compounds suggested for synthesis. Two key technologies aimed at achieving this goal are protein-based docking and ligand-based 3D QSAR.

## ***Target Protein***

Over the last several years protein prenylation has been the subject of intense study and has been found to be critical for the function of key proteins involved in signal transduction [Marshall, 1993]. Prenylation is a form of lipid modification in which either a C<sub>15</sub> farnesyl or C<sub>20</sub> geranylgeranyl group is covalently attached via a thioether linkage to the cysteine residue of proteins near the carboxy terminus. Proteins in mammalian cells that are prenylated fall into four classes, depending on their size [Reese, 1991]. The best characterized isoprenyltransferases is protein farnesyltransferase (ftase). Ftase transfers a C<sub>15</sub> farnesyl

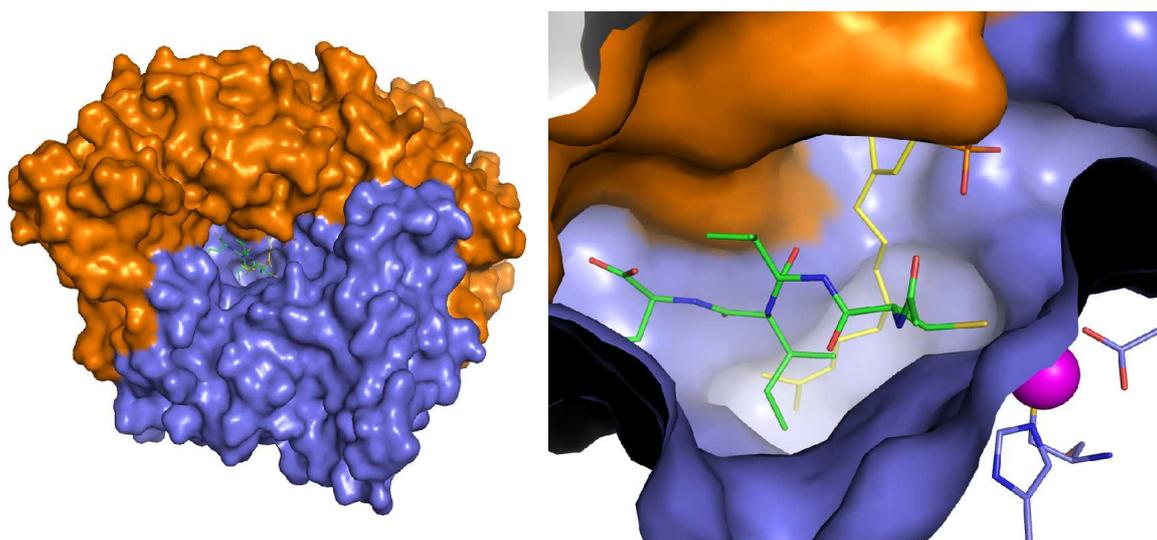
residue to a cysteine carbon atom, which is the fourth amino acid from the carboxy terminus of the CA<sub>1</sub>A<sub>2</sub>X tetrapeptide sequence. The residues A<sub>1</sub> and A<sub>2</sub> are generally aliphatic amino acids and X may be methionine or serine. In similar fashion protein geranylgeranyl transferase transfers C<sub>20</sub> and the selectivity is given by the identity of the X amino acid. Ftase was discovered following the incorporation of mevalonic acid to polypeptides. It is a heterodimer with a molecular weight of 48 kDa ( $\alpha$ -subunit) and 46 kDa ( $\beta$ -subunit) respectively [Reiss, 1990]. Ftase is a metalloenzyme requiring a single zinc ion for activity [Moomaw, 1992]. The zinc ion is located in the  $\beta$ -subunit, near the subunit interface [Park, 1997] and is coordinated by residues Asp297, Cys299, and His362. The precise reaction mechanism needs yet to be fully explored.



**Figure 4.1** Representation of protein ftase as observed in pdb ID: 1FT1. Left gives the overall structure, showing the  $\alpha$ -subunit in orange, the  $\beta$ -subunit in blue. Right gives the zinc coordination in the active site, showing Asp287, His362, and Cys299.

Combining biochemical assays, X-ray crystallography and computational approaches the following mechanisms have been proposed [Long, 2002]: Comparable to common zinc metalloproteinase, a water molecule is bound to the zinc ion. The water molecule gets replaced by the peptide substrate during the reaction. In due course of the chemical

transformation this water molecule returns to its coordination site to be available for a proceeding reaction turnover. The other hypothesis suggests that residue Asp297 is in fact an almost symmetrical bidentate ligand. Upon peptide binding, one oxygen atom is displaced by the peptide cysteine thiol. These structural alteration could facilitate the nucleophilic addition during the prenylation reaction [Tobin, 2003].



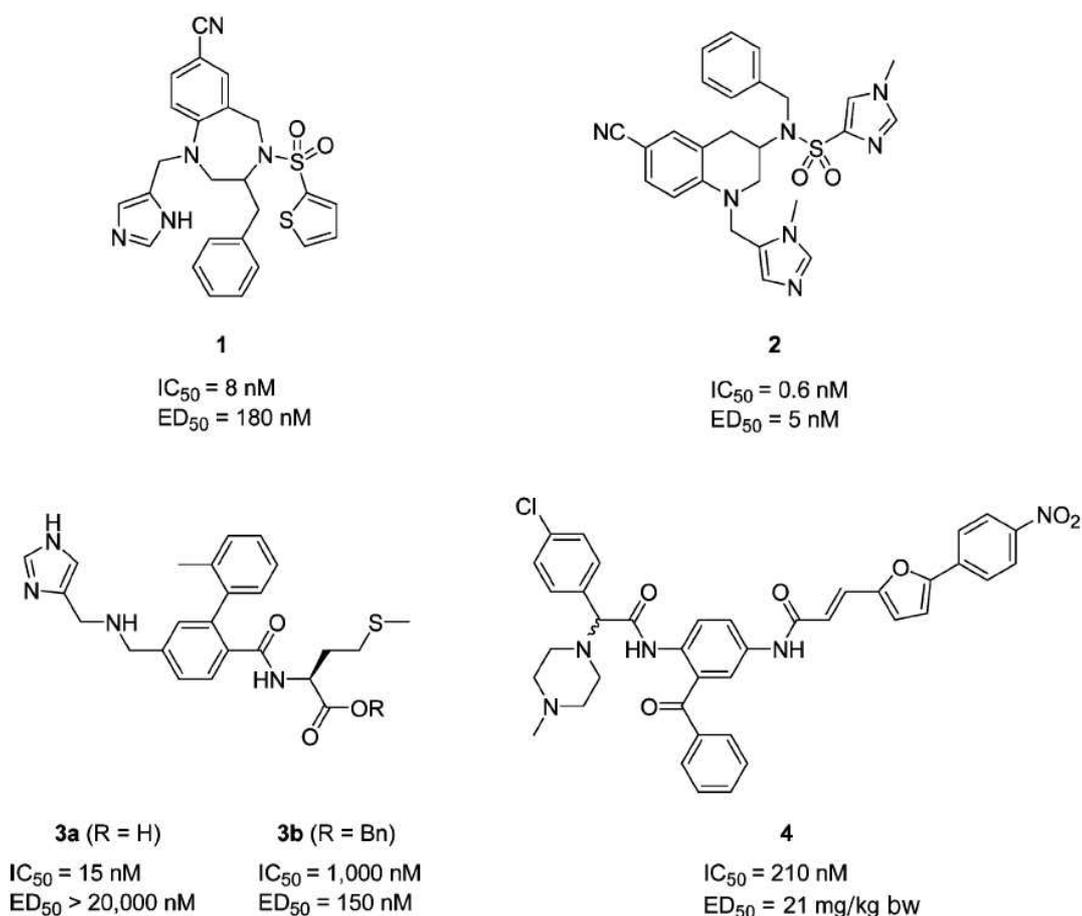
**Figure 4.2** Coordination of active site components of protein ftase as observed in pdb ID: 1QBQ. Left, given a surface representation, identifying the active site at the interface of the  $\alpha$ -subunit (orange) and  $\beta$ -subunit (blue). Right, a close-up of the binding site is shown. Shown with yellow carbon atoms is the farnesylpyrophosphate and with green carbon atom, the tetrapeptide, Selenomethionine- Isoleucine-Valine-Cysteine. The zinc atom is represented by the pink sphere.

Ftase was investigated as a potential target for oncology therapy. Special ftase activity was recognized in tumor cells and more than 70% of cancer cell lines have shown sensitivity to treatment with ftase inhibitors [Haluska, 2002]. Medicinal chemistry efforts to find ftase inhibitors have already led to several potential anticancer agents currently in clinical trials [Dinsmore, 2003]. Recently it has been shown, that clinically relevant red blood cell (RBC) stages of *Plasmodium falciparum* contain high ftase activity [Ohkanda, 2001]. The application of ftase inhibitors to cells infected by *Plasmodium falciparum* result in a decrease of

farnesylated proteins and to associated lysis of the parasites. Additionally, the completion of the *Plasmodium falciparum* genome (www.plasmodb.org 2003) revealed no existence of geranylgeranyltransferase-1. One might speculate that possessing no alternative enzyme makes inhibiting ftase in malaria is lethal compared to mammals. Ftases have now been identified in other pathogens like Trypanosoma, Leishmania and Toxoplasma [Esteva, 2005]. Therefore inhibition of ftase has also been suggested as a treatment of parasitic infections [Cherkasov, 2006].

Different classes of protein ftase inhibitors are known so far. Some having their origins from the oncology efforts. Reviews to the current statement of ftase inhibitors in cancer treatment are given by Mesa, Haluska, and Swanson [Haluska, 2002], [Mesa, 2006], [Park, 1997]. Compounds derived from other drug design programs comprise favorable or even already optimized drug-like features. However, they target the human protein and differences with respect to the parasite target structures might provide the opportunity to develop specific inhibitors to the latter diseases. Ftase inhibitors with respect to antimalarial treatment have been primarily investigated by two research groups, the teams around A. Hamilton and M. Schlitzer [Bohm, 2001], [Esteva, 2005], [Kettler, 2005], [Kettler, 2006], [Schlitzer, 1999], [Schlitzer, 2000], [Wiesner, 2003b], [Adnane, 2000], [Clark, 1997], [Glenn, 2005], [Lerner, 1995], [Qian, 1994], [Qian, 1996], [Sebti, 1998], [Vasudevan, 1999].

Examples for ftase inhibitors:



**Figure 4.3** Examples of ftase inhibitors. Figure is taken from Glenn *et al*, [Glenn, 2006].  $IC_{50}$  values represent the doses that inhibit 50% of the ftase enzyme activity.  $ED_{50}$  values represent the doses that inhibit 50% of *P. falciparum* growth.

As shown in Figure 4.3, most ligand classes comprise functional groups, known to be able to bind to zinc ions, e.g. sulfonamides or thiols. Current ftase inhibitors are lacking the free thiol incorporated into the early inhibitors because of adverse drug effects (skin, nausea) associated with free thiols. Most of these so-called non-thiol ftase inhibitors have nitrogen-containing heterocycles. Here, the ring nitrogen coordinates the enzyme-bound zinc similarly to the cysteine thiol group [Hunt, 1996]. However, nitrogen heterocycles can be replaced by aryl residues lacking the ability to coordinate metal atoms, without losing too much of their ftase inhibitory activity [O'Connor, 1999]. Inhibitor series with a benzophenone (Figures 4.3

and 4.6) scaffold show an interesting example of high affinity ligands without targeting the zinc ion directly.

Previous work on benzophenone-based ftase inhibitors showed the possibility to apply structure-based drug design approaches. Using ligand-based 3D QSAR methods, a model to explain and predict protein-ligand binding affinities could be established [Sakowski, 2001]. Applying flexible docking and GRID binding site analyses two regions in the ftase binding site could be identified [Bohm, 2001]. This way, benzophenone-based ftase inhibitors were optimized from  $IC_{50}$  of 2.7  $\mu$ M to 200 nM. The present work shows further optimization of this compound series, specifically focusing on the (1) far aryl binding site, (2) the specificity pocket and (3) the tetrapeptide region in the ftase binding site (see Figure 4.4). Now, benzophenone-based ftase inhibitors with  $IC_{50}$  up to 4 nM are available.

Other docking studies have been performed on ftase: Cui *et al.*, Henriksen *et al.* and Pedretti *et al.* investigated possible ftase natural ligands and calculated possible ternary complex structures [Cui, 2005], [Henriksen, 2005], [Pedretti, 2002]. Guida *et al.* compared docking generated geometries to crystal complex structures for two ligands, showing little side chain movements in the protein [Guida, 2005].

## **Methods**

All molecular modeling and comparative molecular field analyses were performed using SYBYL 6.9 and DrugScore 1.2. The protein crystal structure used for all investigations in the present thesis was determined for the rat enzyme. The *Plasmodium falciparum* ftase differs significantly in size of both the  $\alpha$ - (472 vs 379 residues) and  $\beta$ - subunit (621 vs 437). However these differences address mainly insertions in the protein sequence. The active site residues show a high degree of similarity allowing the rat protein structure to serve as a

template. The slight differences in eight amino acids might be targeted for specificity purposes (Figure 4.4).

## **Docking**

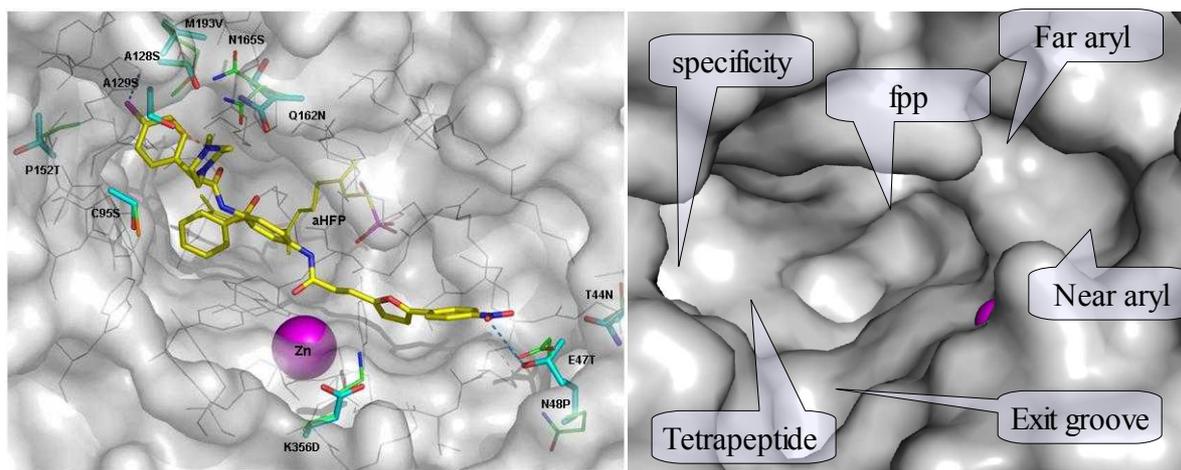
Compatible to previously performed docking analyses, ligands were initially docked using FlexX. Based on the coordinates of the published crystal structure [Strickland, 1998] of a ternary complex of rat ftase with a farnesylpyrophosphate analogue and with N-acetyl-Cys-Val-Ile-selenoMetOH (pdb ID: 1QBQ). For initial evaluation, the position of the benzophenone peptidomimetic substructure, calculated in a previous study, was used as a starting fragment for the docking of this inhibitors via the MAPREF command. Default parameters were employed except the MAX\_ENERGY value, that was set to 10 kJ mol<sup>-1</sup>. Subsequently, the remaining fragments of the inhibitors were placed into the active site using the incremental construction algorithm of FlexX. The docking runs provided sets of solutions which were inspected according to their calculated energy score. The resulting geometries were evaluated using the knowledge-based scoring function DrugScore.

Due to misplacements of the initial base fragments, ligand construction of large molecules in the binding site could not be finalized using FlexX. Therefore, subsequent docking of inhibitors was performed with AutoDock3.0. Again, the protein structure was taken from pdb ID: 1QBQ. Ligands and solvent molecules were removed, but the zinc ion and farnesylidiphosphate were included as part of the protein. For the use within AutoDock3.0 [Goodsell, 1996], [Morris, 1996], [Morris, 1998] polar hydrogens were added with the PROTONATE utility from AMBER. AMBER united atom force field charges were assigned [Weiner, 1984], and solvation parameters were added using the ADDSOL utility from AutoDock3.0. Ligand structures were built in mol2-format, Gasteiger partial atomic charges

were assigned [Gasteiger, 1980] and all bonds except for amides were kept rotatable. For some ligands, in contrast to the biological testing, only one enantiomer instead of the racemates were considered for molecular docking. Docking runs were performed with the Lamarckian genetic algorithm included in AutoDock3.0, performing 50 independent runs per ligand, using an initial population of 50 randomly placed individuals, a maximum number of  $1.5 \times 10^6$  energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80 and an elitism value of 1. Resulting ligand conformations that differ by less than 1 Å rmsd from each other were clustered together and represented by the result with the best docking energy. The results of the highest ranked clusters were chosen to compare structure-activity relationships. The relative binding affinity was predicted, using the knowledge-based scoring function DrugScore.

## **Results and Discussion**

Rat ftase can be used as a reasonable homology model for the development of antimalarial ftase inhibitors because differences in the structure compared to *Plasmodium falciparum* do not majorly address the binding site. Only a Tyr166Phe and His201Asn mutation affect the peptide and farnesylpyrophosphate binding [Schlitzer, 2005]. The binding site can be subdivided in six different regions, near and far aryl binding site, farnesylpyrophosphate (fpp) and tetrapeptide binding site, specificity pocket and exit groove.

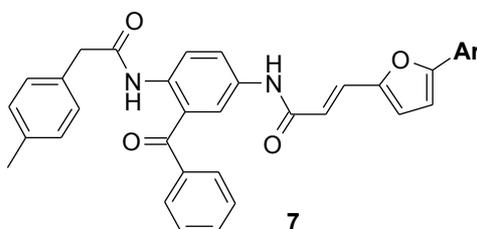


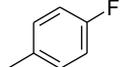
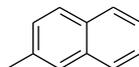
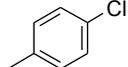
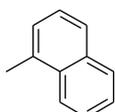
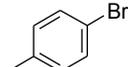
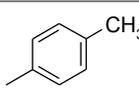
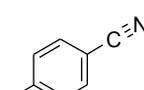
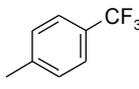
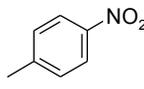
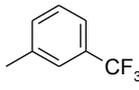
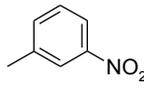
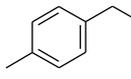
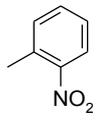
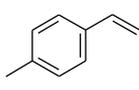
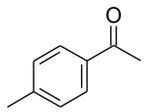
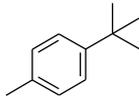
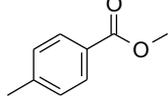
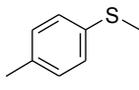
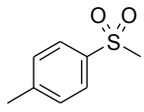
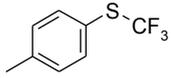
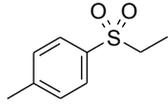
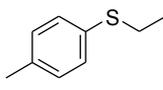
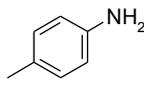
**Figure 4.4** Left: Homology model of *P. falciparum* ftase (turquoise carbon atoms) in comparison to rat/human ftase (green carbon atoms). Differences in amino acids which might affect binding of inhibitor 7t (yellow carbons) are shown in stick mode. Right: Orientation of the six different binding site regions.

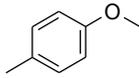
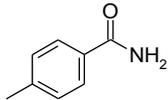
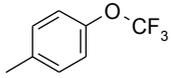
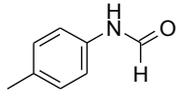
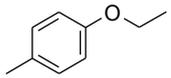
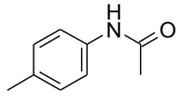
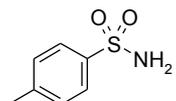
### QSAR in the far aryl binding sites

Several arylfurylacryl-substituted benzophenones targeting the far aryl binding site have been designed. In the study [Mitsch, 2004] extensive variation of the terminal aryl residue of the arylfuryl moiety are performed. Using flexible docking, two possible binding modes could be determined which enabled to establish structure-activity relationships for the series.

**Table 4.1** Ftase inhibitory activity of 5-arylfuryl derivatives 7a-af.

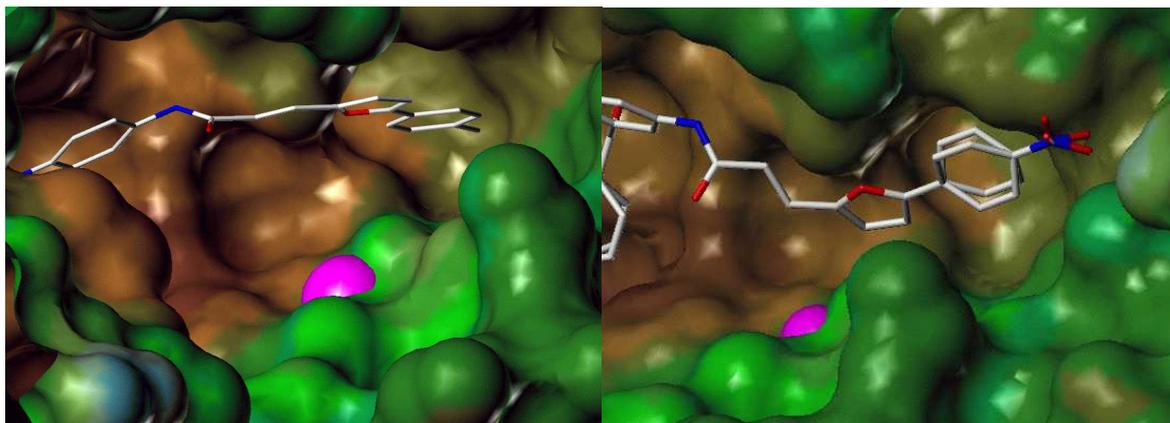


Compound	R	IC <sub>50</sub> (nM)	Compound	R	IC <sub>50</sub> (nM)
7a		100 ± 32	7p		30 ± 5
7b		165 ± 22	7q		150 ± 28
7c		190 ± 16	7r		925 ± 45
7d		4 ± 1	7s		105 ± 65
7e		4 ± 2	7t		35 ± 5
7f		220 ± 10	7u		710 ± 40
7g		90 ± 27	7v		925 ± 60
7h		130 ± 34	7w		225 ± 65
7i		115 ± 45	7x		95 ± 10
7j		170 ± 65	7y		30 ± 18
7k		140 ± 22	7z		125 ± 6
7l		180 ± 34	7aa		30 ± 8

Compound	R	IC <sub>50</sub> (nM)	Compound	R	IC <sub>50</sub> (nM)
<b>7m</b>		510 ±76	<b>7ab</b>		330 ±62
<b>7n</b>		225 ±48	<b>7ac</b>		285 ± 38
<b>7o</b>		840 ±16	<b>7ad</b>		545 ±107
			<b>7ae</b>		40 ± 7

Experimental data is given in [Mitsch, 2004].

While some substitutions show affinity differences that are beyond what can be distinguished via docking (comparing halogen substitutions **7p**, **7q**, **7r**) some of the structure-activity relationships described above may be explained. Figure 4.5 shows a binding mode of the para methyl substituted inhibitor **7d** obtained by flexible docking. The methyl group is deeply buried in a hydrophobic area in the far aryl binding site. Any substituent larger than methyl or trifluoromethyl would prevent this favourable binding mode explaining the low activity of any extended alkyl (other than methyl or trifluoromethyl), alkylsulfanyl or alkyloxy substituted derivative. However, considerable activity has also been obtained with substituents larger than methyl as for instance with nitro or methylsulfonyl. For these inhibitors docking indicates a different binding mode with the para substituent not that deeply buried in the aryl binding site (what is prevented by its size) but directed to a more hydrophilic area closer to the rim of the aryl binding site as shown for the nitro derivative **7t** in Figure 4.5.



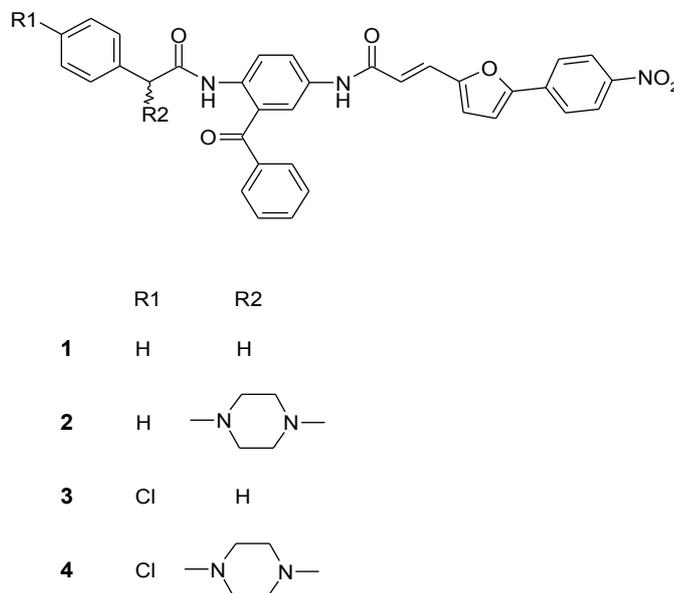
**Figure 4.5** Views of a part of ftase's active site and the far aryl binding site from slightly different angles. Lipophilic (brown) and hydrophilic (green to blue) properties are displayed on the Connolly surface. The structural zinc is shown as magenta sphere. Left: Methyl-substituted inhibitor **7d** deeply buries its methyl group into a lipophilic area in the far aryl binding site. Right: Nitro-substituted inhibitor **7t** directs its nitro group to a more hydrophilic area more towards the rim of the far aryl binding site.

Common feature of the nitro or sulfonyl substituents is a hydrogen bond acceptor property. At the rim of the aryl binding site are two lysine side chains (Lys294 $\beta$  and Lys353 $\beta$ ) which might act as appropriate hydrogen-bond donors. Inhibitors having residues too large to fit deeply into the binding site but lacking hydrogen bonding properties get penalized in terms of affinity. The presence of two amino acid side chains acting as hydrogen bond donors might also explain the relatively low activity of inhibitors **7w** and **7x**, since in contrast to the nitro or sulfonamide moieties their substituents are capable of forming only one but not two hydrogen bonds. As demonstrated by the amino substituted derivative **7aa** also a hydrogen bond donor property seems to be favorable. Possible bonding partners for this moiety are provided by the side chains of Glu47 $\beta$  and Asp233 $\beta$ .

### ***QSAR in the specificity pocket***

Two compounds with very promising activity in vitro turned out to be inactive in the malaria model, quite likely due to limited solubility. Introducing a methyl piperazinyll residue into the

$\alpha$ -position of the phenyl acetyl moiety of the 2-amino group of the benzophenone scaffold. This modification resulted in a significantly improved water solubility (Table 4.2).



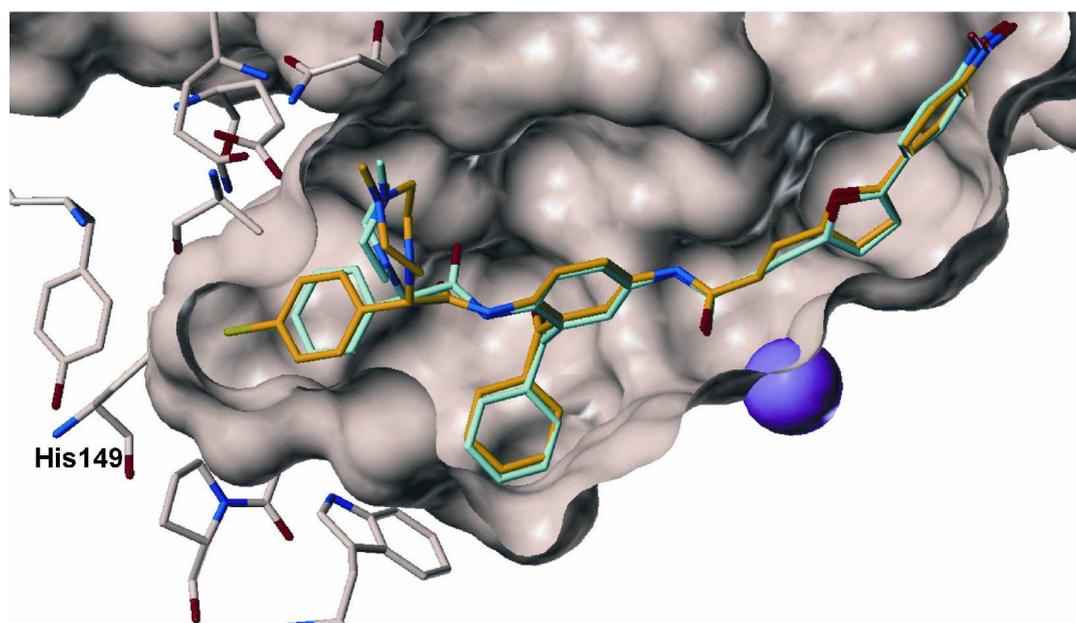
**Figure 4.6** Substitution on benzophenone-based ftase inhibitors addressing the specificity pocket (R1 substitution) and to the tetrapeptide region (R2 substitution).

**Table 4.2** *In vitro* and *in vivo* activity and solubility of compounds 1-4.

Comp.	IC <sub>50</sub> (nM)	ED <sub>50</sub> (mg/kgBW)	ED <sub>90</sub> (mg/kgBW)	Solubility <sup>[b]</sup> (mM)	Solubility <sup>[c]</sup> (mM)
<b>1</b>	270 ± 30	--	--	< 0.04	< 0.06
<b>2</b>	270 ± 35	30	40	0.35	> 3.33
<b>3</b>	64 ± 11	--	--	< 0.04	< 0.06
<b>4</b>	210 ± 21	21 <sup>[a]</sup>	25 <sup>[a]</sup>	0.25	1.25

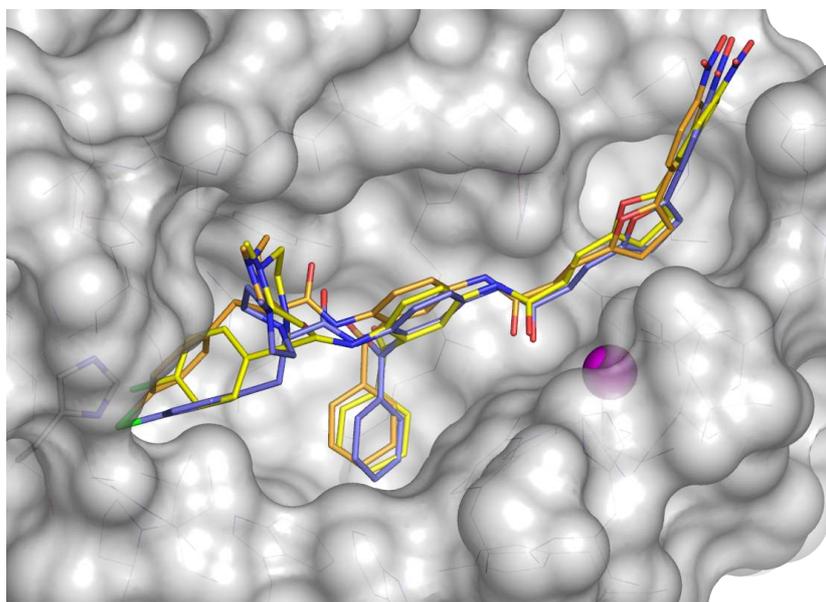
[a] The experiment was repeated under comparable conditions with compound **4** resulting in an ED<sub>50</sub> value of 22 mg/kgBW and ED<sub>90</sub> value of 28 mg/kgBW; [b] in phosphate buffer pH 7.2; [c] in water.

The introduction of the methyl piperazinyl residue had no influence on the *in vitro* activity in case of compound **2** with unsubstituted phenyl ring. In the series of  $\alpha$ -unsubstituted compounds significantly increased activity was achieved with a *p*-chloro-substituent. Although less pronounced, this effect was also observed with the  $\alpha$ -piperazinyl derivatives (comparing compound **3** and **4**). The docking results clearly show that both inhibitors adopt a similar binding mode (Figure 4.7). Yet, the chlorine-substituted inhibitor **4** leads to better interactions with the protein, in particular it forms a N-H $\cdots$ Cl interaction with His149. As an additional piperazinyl-moiety would sterically hinder this interaction, less increase in activity is observed for piperazinyl derivatives compared to the unsubstituted inhibitors.



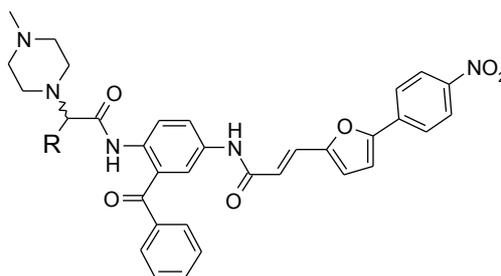
**Figure 4.7** Superposition of selected docking results of the inhibitors **2** (turquoise) and **4** (orange) in the binding pocket of the farnesyltransferase. Only the amino acids forming the pocket responsible for the difference in affinity are shown. The zinc ion is indicated by the blue ball. Both inhibitors possess a comparable binding mode. The slightly increased activity of compound **4** can be explained by an additional N-H $\cdots$ Cl interaction with His149 (in part hidden by the surface).

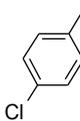
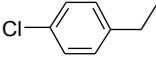
To overcome this effect and to obtain more active inhibitors, two approaches were envisioned. First, a methylene group was inserted between the  $\alpha$ -position and the phenyl residue thereby obtaining  $\alpha$ -piperazinyl- $\beta$ -phenylpropionic acid derivative **5**. In docking studies, this compound behaved as planned with its piperazinyl moiety found in the same region as for the lead **4** and directing the terminal phenyl residue towards His149 $\beta$  but reaching closer to its imidazole because of the additional methylene group (Figure 4.8).



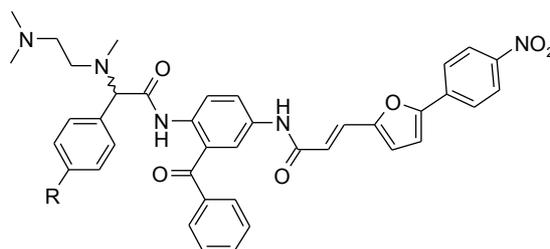
**Figure 4.8** Docking of both enantiomers of inhibitor **5** (orange and blue) into ftase in comparison to inhibitor **4** (yellow). The docking yielded a binding mode for both enantiomers in which the additional methylene group facilitates a better chlorine His149 interaction compared to **4**.

However, *in vitro* activity of inhibitor **5** did not improve since it was virtually equipotent to inhibitor **4** against blood stages of *P. falciparum* (Table 4.3). Furthermore, inhibitor **5** displayed higher cytotoxicity than **4**. Therefore, this approach was discontinued.

**Table 4.3** Activity of N-methylpiperaciny-substituted compounds **2**, **4** and **5**.

	R	IC <sub>50</sub> (nM) FTase	IC <sub>50</sub> (nM) <i>P. falcip</i>	ED <sub>50</sub> (mg/kg)	ED <sub>90</sub> (mg/kg)	CC <sub>50</sub> (μM) <i>HeLa cells</i>	CC <sub>50</sub> /IC <sub>50</sub>
<b>2</b>		10 ± 3	270	30	40	37.0	137
<b>4</b>		4 ± 2	210	21	25	38.6	184
<b>5</b>		124 ± 15	210	nd	nd	18.0	86

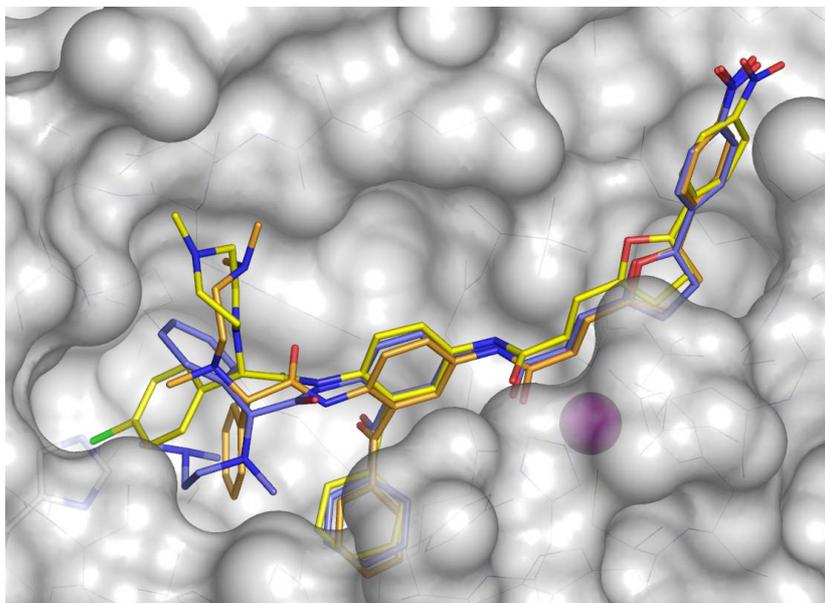
A second strategy envisioned to overcome the steric hinderance by the piperazinyl moiety was the replacement of the piperazinyl moiety by an open chain ethylene diamine hoping that the enhanced flexibility to this residue would allow for both, an optimal chlorine imidazol interaction as well as for an interaction of the terminal amine with the amino acids in the upper part of the binding site. Indeed, when assayed *in vitro* against *P. falciparum* blood cultures the ethylene diamine derivatives **20** and **22** displayed considerably improved *in vitro* activity with IC<sub>50</sub>-values of 40 and 50 nM (Table 4.4).

**Table 4.4** Activity of N,N'-trimethylethylenediamine-substituted compounds **20** and **22**.

	R	IC <sub>50</sub> (nM) FTase	ED <sub>50</sub> (mg/kg)	ED <sub>90</sub> (mg/kg)	CC <sub>50</sub> (μM) <i>HeLa cells</i>
<b>20</b>	H-	21 ± 3	16	20	14.4
<b>22</b>	Cl-	28 ± 3	26	30	6.6

An important feature can be delineated from these *in vitro* results. The cytotoxicity of the para chloro-substituted derivatives **22** is significantly higher than that of the phenyl-unsubstituted derivative **20**, while anti-plasmodial activity of the phenyl-unsubstituted derivatives is higher. Although the absolute cytotoxicity of inhibitor **20** is higher than that of the lead **13** (CC<sub>50</sub>-values 14.4 and 38.6 μM, respectively), the toxicity/activity is still considerably improved because of the significantly improved antiplasmodial activity of inhibitor **20**. More striking is the difference in the *in vivo* activity between these two compounds. The phenyl unsubstituted inhibitor **20** displays a significantly higher activity in the mouse model (ED<sub>50</sub> = 16 mg/kg b.w.) than the para chloro substituted derivative **22** (ED<sub>50</sub> = 26 mg/kg b.w.). Furthermore, inhibitor **20** is also more active than the piperazinyl derivatives **2** and **5**. The binding modes suggested by AutoDock3.0 vary considerably from the originally planned mode (Figure 4.9). AutoDock3.0 proposes several solutions. In one mode the flexible ethylene diamine moiety is directed into the same region as the piperazinyl residue in **5** but the chlorophenyl residue is not directed towards His149β. Instead it is oriented more steeply downwards towards Trp102β resembling a binding mode which we obtained earlier for simple α-amino acid

derivatives [Kettler, 2005]. For the corresponding epimer, the aryl residue is now directed upwards into the region formerly occupied by the piperazinyl moiety and the ethylene diamine pointing towards His149 $\beta$ .



**Figure 4.9** Superimposition of docking results of the methylpiperazinyl-substituted inhibitor **5** (yellow) and two different docking solutions of inhibitor **20**. In one docking solution (orange) ethylene diamine residue points roughly into the same direction as the methylpiperazinyl moiety of inhibitor **5**, while the phenyl residue is markedly shifted downwards forming a edge to face interaction with Trp102. In the second docking solution (blue) the phenyl and ethylene diamine moieties exchanged their positions.

Because energy values obtained for both solutions are virtually identical, no orientation can be favored at this time. Regardless which binding mode is realized in the enzyme, both binding modes are considerably different from the mode obtained for inhibitors **2** and **5**, explaining the different response as a result of the introduction of a para-chlorine atom at the terminal phenyl residue.

## Conclusions

At first, benzophenone-based ftase inhibitors with high *in vitro* antiplasmodial activity but no *in vivo* activity were discovered. Via the introduction of a methylpiperazinyl moiety into the  $\alpha$ -position of the phenylacetic acid substructure, the first ftase inhibitors with *in vivo* anti-malarial activity have been obtained. Molecular docking studies supported some presumption about possible binding modes but the results also illustrate the need for further structural analyses. Based on the docking results, in a second design cycle the piperazinyl moiety has been replaced by a *N,N'*-trimethylethylene diamine moiety. This resulted in an inhibitor with significantly improved *in vitro* and *in vivo* anti-malarial activity. Furthermore, this inhibitor displays notable selectivity towards malaria parasites in comparison to human cells. This is a particular important result for the development of specific anti-malarial ftase inhibitors. Whether this difference is due to binding site composition is not yet clear. Homology modelling revealed slight differences between the active sites of rat/human and plasmodial ftases but without further experimental structural data an analysis of deviating binding modes would be too speculative. Still, the support via docking provided helpful guidance for further inhibitor design. In this example the lowest energy structure determined by DrugScore was chosen to select an appropriate binding mode for each ligand. The predicted binding mode was given attention while predicted affinities played only a minor role. In most cases, ranking inhibitors with respect to their protein-ligand affinity did not match the experimental data. However, compared to their overall large size the compound's structural substitutions are humble. Therefore trying to estimate affinity differences is beyond the level of accuracy. And the main objective of a knowledge-based scoring function is recognizing a possible binding mode from a set of multiple solutions for one ligand. In that case, DrugScore outperformed other applied scoring functions and showed to be a supportive tool for ligand optimization.

# SUCCESSFUL LEAD IDENTIFICATION FOR METALLOPROTEINASES: A FRAGMENT-BASED APPROACH USING VIRTUAL SCREENING

## ***Introduction***

Over the past decade, an enormous effort has been spent to enhance the success rate of the drug discovery process. One aspect has been the use of high-throughput technologies, applied to synthesis and screening. The conventional approach involves screening with respect to a disease, where a particular target protein is tested for inhibition against a large collection of drug-sized compounds. The collection typically consists of 200,000-2 millions compounds. The readout is typically a biochemical assay and compounds inhibiting the target with  $IC_{50}$ s at or below 20  $\mu$ M are usually regarded as *hits* which get developed into *leads* and further optimized for potency and other drug-like properties. However, the success rate has not been as high as expected and one important aspect are inappropriate lead structures [Lahana, 1999], [Oprea, 2002]. Due to the diversity and huge size of chemical space one cannot expect to immediately detect a drug candidate in the primary screen. Therefore, a process of evolution by selection is inevitable, starting with a compound eligible for multiple substitutions [Leach, 2006b]. Recently, several new technologies have emerged which screen small drug fragments (also referred to as needles, shapes, binding elements or seeds) and thereby open the perspective towards a combinatorial approach to lead discovery. The goal is to build drug leads from pieces, by identifying small molecular fragments and then either linking or expanding them. The definition of a 'fragment' varies, but usually refers to molecules

weighing less than 200-300 Da, with fewer than 15-20 heavy atoms [Erlanson, 2004a], [Erlanson, 2004b], [Rees, 2004], [Zartler, 2005].

### ***Drug-like versus Lead-like***

The definition of drug-likeness and lead-likeness serve their own ends: Drug-like molecules are suitable for drug development and clinical candidacy. Lead-like molecules instead bind in a non-covalent, reversible manner to their targets in a biochemically assay and show chemical stability towards proteins. However, they should not be a “promiscuous inhibitor” [McGovern, 2002], a “frequent hitter” [Roche, 2002] or a “warhead compound” [Rishton, 2003]. The latter is extremely challenging for the herein applied target class of metalloproteinases.

Not every marketed drug is derived from a lead structure. Alternative sources for development are already existing drugs or more important, natural products (neuropeptide Y, taxol). Contrarily, there are also lead-like drugs on the market, e.g. propranolol or paracetamol. The key contribution to what defines a drug was an examination of clinically tested drug molecules in order to determine their shared properties [Lipinski, 1997], [Lipinski, 2001]. The resulting “rule of five” tries to predict absorption problems due to poor solubility and/or poor permeability. Compounds that do NOT obey two of the four following parameters are likely to have poor absorption and/or distribution.

1. Molecular weight  $\geq 500$  Da
2.  $\log P > 5$
3. More than five hydrogen bond donors (defined as the sum of OH and NH groups)
4. More than ten hydrogen bond acceptors (defined as the number of N and O atoms)

Several other groups reported analyses of collections of drugs. The number of rotatable bonds [Veber, 2002] and polar surface area [Clark, 2000] have also contributed to predict drug-likeness of a molecule. Analyzing sets of corresponding lead and drug molecules Oprea *et al.* and Hann *et al.* reported statistics over the increase of several physicochemical properties in the lead optimization step [Hann, 2001], [Oprea, 2001], [Oprea, 2002].

**Table 5.1** Changes in molecular properties from leads to drugs

<b>Property</b>	<b>Average values for leads</b>	<b>Average values for drugs</b>	<b>Increment</b>
<b>Molecular weight (Da)</b>	272.0	314.0	42
<b>Hydrogen bond acceptors</b>	2.2	2.5	0.3
<b>Hydrogen bond donors</b>	0.8	0.8	0
<b>Number of heavy atoms</b>	19.0	22.0	3.0
<b>Clog P</b>	1.9	2.4	0.5

Table is derived from [Hann, 2001]

This shows that physicochemical properties required for drug-like compounds are different from those required for lead-like compounds which are followed-up in early stages of drug discovery [Rishton, 2003]. Unlike the drug-like scores, the lead-like concept [Teague, 1999] is based on smaller data sets [Oprea, 2001], [Proudfoot, 2002]. Emerging from studying leads, the “rule of three” has been established [Congreve, 2003a]. Fragments that were positively tested against several target proteins seem to obey:

1. Molecular weight  $\leq 300$  Da
2. Number of hydrogen bond donors  $\leq 3$
3. Number of hydrogen acceptors  $\leq 3$
4. ClogP is  $\leq 3$

5. Number of rotatable bonds  $\leq 3$
6. Polar surface area  $\leq 60 \text{ \AA}^2$

These rules could be useful when constructing a fragment library. Another parameter often applied with respect to lead discovery is “ligand efficiency” [Hopkins, 2004]. Andrews *et al.* first introduced the concept of estimating potential activities as function of the presence of certain protein-ligand interacting groups [Andrews, 1984]. Considering the maximal affinity for ligands and calculating the binding affinity per atom [Kuntz, 1999] has led to the following definition for ligand efficiency: Converting  $K_d$  into free energy of binding and dividing by the number of heavy atoms.

$$\text{Free energy of binding: } \Delta G = -RT \cdot \ln K_d$$

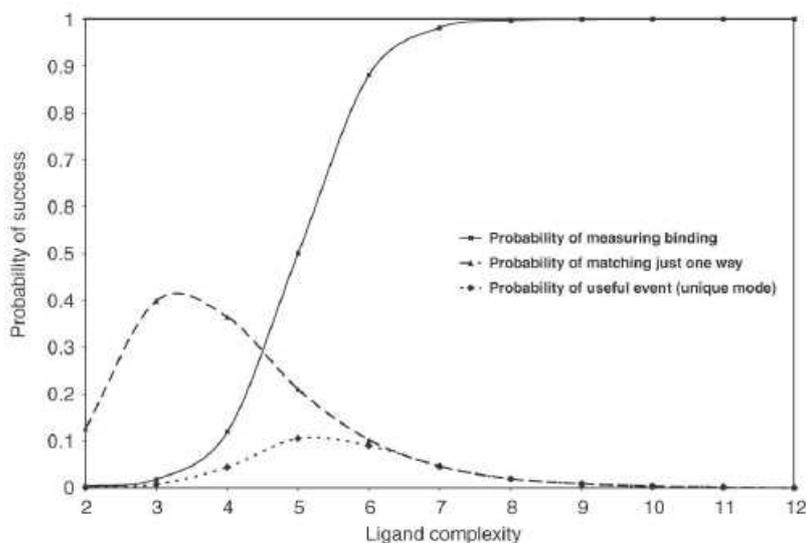
$$\text{ligand efficiency: } \Delta g = \Delta G / N_{\text{non-hydrogen atoms}}$$

Considering an average mass of 13.29 Da for a non-hydrogen atom, a 500 Da compound with 10 nM binding affinity possesses a ligand efficiency of 1.21 kJ/mol. Considering these figures it should be mentioned that fosmidomycin from the previous AFMoC study has a ligand efficiency of 3.68 J/mol per non-hydrogen atom. This might explain the difficulties in optimizing its structure. Taking ligand efficiency one step further is reported by Abad-Zapatero and Metz. They describe affinities as 'percentage of inhibition' measured in primary HTS and  $K_i$ ,  $K_d$  or  $IC_{50}$  measured in secondary assays. Crucial physicochemical properties in the lead selection process are molecular weight and PSA. To provide a mathematical link they introduce a percentage efficiency index (%percent inhibition at a given [compound]/MW), a binding efficiency index ( $pK_i$ ,  $pK_d$  or  $pIC_{50}/MW$ ) and a third surface-binding efficiency index ( $pK_i$ ,  $pK_d$  or  $pIC_{50}/PSA$ ). Comparing these indexes to marketed drugs and molecules along the project pipeline shows how numerical rules can work as guidance in the drug candidate decision process [Abad-Zapatero, 2005].

### ***Small is beautiful – the concept of Fragment-based ligand design***

The concept that underlies the chemical fragments approach was first proposed by Jencks in 1981 [Jencks, 1981]. Difficulties in identifying and linking fragments delayed its success but further development in screening for low affinity compounds led to the first practical demonstration in 1996 [Shuker, 1996]. Since then, the technology has become very popular and numerous fragment discovery, growing and linkage strategies have been published [Erlanson, 2006].

The basic approach of fragment-based ligand design is identifying small compounds that bind to a target protein with low affinity. They are subsequently 'grown' or decorated to generate larger, high-affinity leads, often using structure-guided medicinal chemistry. Fragment-based strategies aim to achieve three goals. (1) Reducing the size of screening libraries, (2) chemical optimization can exploit the structural understanding of the protein-ligand binding interactions and (3) starting with small fragments allows to further increase molecular weight during optimization without exceeding the desired limit for drug-likeness. A simple model presented by Hann *et al.* shows the relationship between the probability of finding a hit and the complexity of the molecule [Hann, 2001].



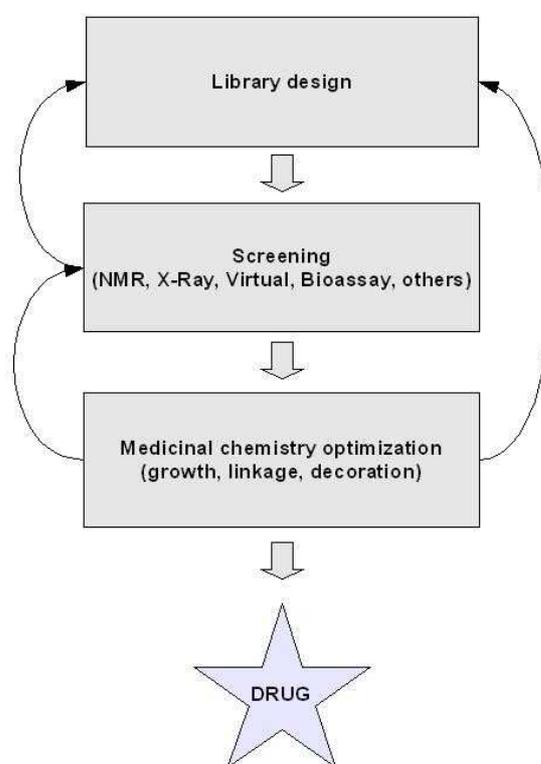
**Figure 5.1** Multiplying the probability of a single match with the probability of detecting the binding event. The result is the “useful event” which passes through a maximum at rather low complexity. (data from [Hann, 2001])

The useful event is defined as determining unique molecular recognition in a biochemical assay. This depends not only on the chance of measuring binding but also on how the complexity of the molecule affects its binding mode. The chance of determining the binding event increases with the complexity of the molecule. Contrary, the chance of finding a match between ligand and binding site properties is much higher for simpler molecules. Finding the useful event in HTS with a randomly assembled library of molecules with greater complexity is quite unlikely. For less complex molecules, multiple binding modes will dominate and thus confound the affinity measurement. The aim of this study is not to predict the optimal complexity for a library but to show that starting simple has still a reasonable chance for detection while offering chemical space for optimization.

Applying biophysical methods to fragments enables better correlation of structure with affinity and establishes details about the interactions. To actually exhibit measurable activity a fragment has less possibilities to compensate entropy compared to a larger molecule. Binding to the protein requires a penalty in rigid-body translational and rotational entropy. This must be compensated by protein-ligand interaction energy. The entropic penalty is independent of

molecular weight [Finkelstein, 1989] and is estimated to be 15 to 20 kJ/mol [Murray, 2002]. The fragment requires more binding energy per atom to display comparable affinity. Considering the affinity of the final drug-like molecule, less additional interaction will be required for growing from a small fragment instead of substituting a drug-sized molecule [Verdonk, 2004].

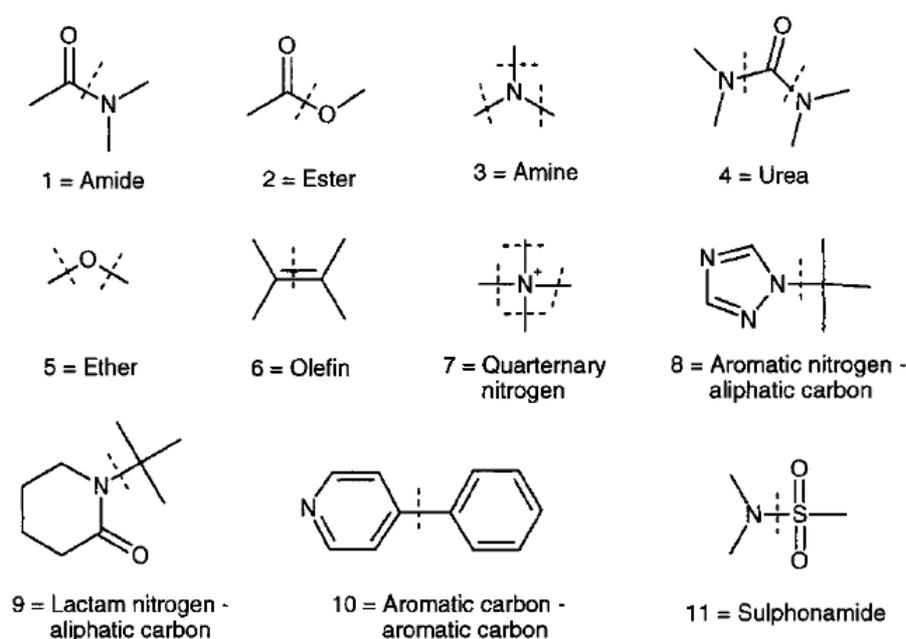
The process of a fragment-based drug discovery project consists of three stages:



**Figure 5.2** The fragment-based drug discovery process involves integration and parallel application of multiple technologies.

## Library design

The low affinity of the fragment screening hit makes an optimization necessary. Independent of which optimization strategy is applied, it is difficult to modify molecules lacking suitable linker groups [Jacoby, 2003]. Avoiding functional groups which might react covalently with surface-exposed amino acids of the protein, fragments should be decorated with the groups that allow for subsequent chemical reaction. Typically these are carboxylic acids, amines and alcohols. However, in some cases, the potential linker group participates in non-covalent protein binding. Selecting fragments for specific linker reactions (e.g. self-assembly) or having multiple possible linker groups might avoid these linkage complication [Schuffenhauer, 2005]. Targeting a library with respect to a specific set of chemical reactions is frequently applied. RECAP (Retrosynthetic Combinatorial Analysis Procedure) is an algorithm which enables partitioning of molecules in fragments corresponding to the following eleven chemical reactions.



**Figure 5.3** Eleven default bond cleavage types used in the RECAP algorithm [Lewell, 1998].

While breaking the bonds, the size of the remaining fragments is assessed to create suitable starting materials [Lewell, 1998]. The retrosynthetic analyses of given data sets do not only enable library design for combinatorial synthesis. They also allow statistics about common functionalities especially when applied to a data set of *drug* molecules. One might speculate that fragments derived from approved drugs show desirable properties with respect to toxicity, common decoration groups to improve ADME properties and synthetic feasibility. Multiple studies about properties of known drugs have been reported. The work of Bemis and Murcko focuses on organizing substructures of drug molecules into the categories ring, linker, framework and side chains and assigning shape descriptors. Subsequent cluster analyses show low diversity in frameworks based on two-dimensional molecular structure. Contrarily, clustering based on atom types, bond types and hybridization reveals a large diversity [Bemis, 1996]. In a sequel, a detailed study on the features and attributes of side chains to the drug molecule revealed an average number of four side chains per core and common classes of side chains [Bemis, 1999]. Considering molecular structure but also administration and physicochemical properties, Vieth *et al.* describe common scaffolds and features which distinguish orally administered drug from others routes [Vieth, 2004]. Studying properties of drugs provides guidance for future library design projects. Baurin *et al.* used a reference drug molecule dataset, crystallized drug-sized molecules, water solubility and several synthesis guidelines to develop a fragment library for primary NMR screening [Baurin, 2004]. Kolb *et al.* used the decomposition program DAIM (Decomposition and Identification of Molecules) to compare commercially available fragments to fragments derived from drug molecules and thus, generated a screening library and successfully applied it to  $\beta$ -secretase [Kolb, 2006]. In another study, drug molecules of the National Cancer Institute dataset were analyzed for co-occurrence of fragments [Lameijer, 2006].

## **Screening**

There are various methods that can be used to detect the binding of a compound to a protein. However, one drawback of fragments is their low expected target affinity, making their binding not always compatible with classical screening strategies. Higher compound concentrations are required to detect low affinity binding than the usually applied micromolar cut-off. This can cause problems due to the incompatibility of most proteins to DMSO which is generally applied to dissolve the screening fragments [McGovern, 2003b].

**X-ray crystallography:** Knowledge of the binding modes assists the selection which fragments to follow up (e.g. whether the potential linker is involved in interactions). Advances in parallel cloning, optimizing expression, nanolitre crystallization, automated phasing and model building and the increasing number of available protein structures opens the perspective to use protein crystallography as screening technology [Blundell, 1996], [Jhoti, 2001]. Determining the structure of a fragment-protein complex is either done by soaking or by co-crystallization. Interpreting the difference electron density usually requires manual manipulation which can make this step time-consuming and is one bottleneck of using X-ray crystallography as a first screening approach. Recently, the algorithm AutoSolve has been shown to automatically refine a protein-ligand complex structure if the protein can be represented by molecular replacement [Mooij, 2006]. A disadvantage of the method is that it can only prove but never disprove binding, since unsuccessful soaking or co-crystallization with a ligand can have various reasons apart from lacking binding affinity. Additionally, crystallization conditions and crystal packing can influence the complex structure.

**NMR:** In target-based NMR screening, chemical shift changes in the protein spectrum, caused by the binding of a ligand, are measured. This allows not only to detect binding but given appropriate isotope labeling and full assignment of the spectra the detection of the

protein atoms involved in the interactions is possible. The pioneering approach using target-based NMR is SAR by NMR [Shuker, 1996]. A disadvantage of this technique is the need of large amounts of labeled protein as well as time-consuming resonance assignments. Ligand-based NMR screening detects changes in the ligand spectrum upon binding. It is independent of any labeling and is particularly suitable for detecting weak binding. Examples are the SHAPES strategy and STD method [Fejzo, 1999], [Meyer, 2003], successfully applied to fragment-based screening approaches.

**Mass spectroscopy:** Employing suitable ionization methods such as electrospray ionization, weakly binding of ligands can be detected by mass spectroscopy. Complexes will stay intact thus the observed mass indicates ligand binding and conclusions about stoichiometry can be drawn [Breuker, 2004], [Moy, 2001]. The big advantage is the low protein concentration needed and a well developed automation. However, compared to NMR or crystallography, no structural information is available and it is in question whether interactions detected in the gas phase are relevant for interactions in solution.

**Tethering:** A special fragment screening approach developed by Sunesis Pharmaceuticals Inc combines protein engineering, mass spectroscopy and protein crystallography [Erlanson, 2000], [Erlanson, 2004c]. A residue close to the active site is mutated to cysteine and fragments are covalently bound via disulfide bonds under reducing conditions. Facilitating subsequent mass spectroscopy or crystallography, the bound ligand can also be used to further explore the binding site. The approach has almost no false positives but it misses possible ligands due to steric coordination of the linker to the active site.

**SPR:** Another label free binding assay is surface plasmon resonance. Immobilizing the target on an optical sensor surface, compound binding can be detected by measuring the change in mass concentration. Incoming light is partly reflected and partly refracted. Changes in mass, for example caused by ligand binding, cause a change in the refractive index and are thus

detected. The advantage is the low protein consumption and access to kinetic data such as association and dissociation rates [Neumann, 2005].

### ***Linkage and growth***

The strategies to transform a fragment to a lead range considerably, from small change to the fragment itself to dramatic substitutions and linkage with other fragments. Most reports classify fragment progression in four classes (Figure 5.4), (I) fragment evolution (II) fragment linking (III) fragment self-assembly and (IV) fragment optimization [Rees, 2004]. However, these distinctions are more semantic than determined by rational criteria. The same final molecule could be created by linkage of two fragments, optimizing a single fragment or through some combination [Erlanson, 2004a].

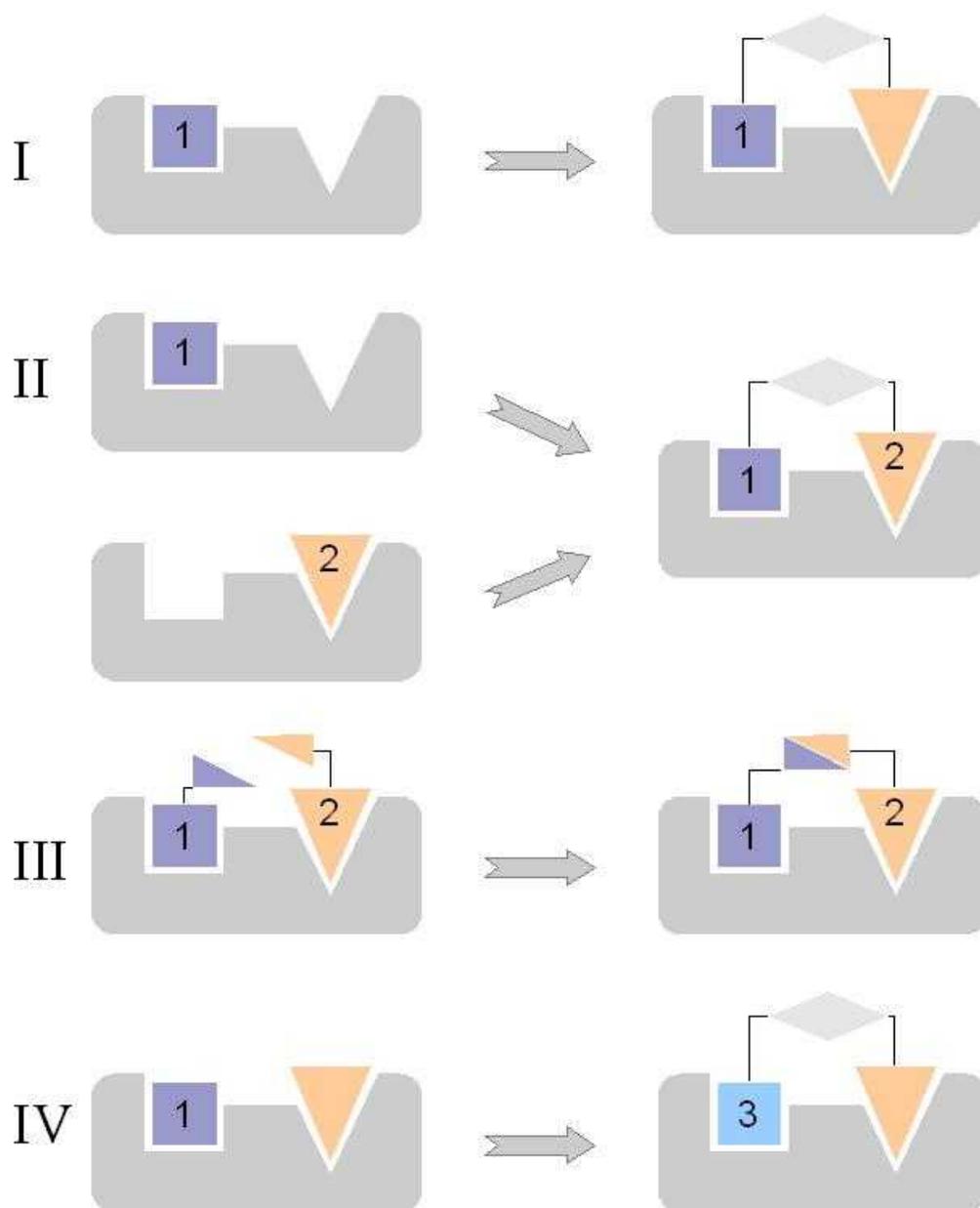
(I) Fragment evolution: An initial fragment is optimized by adding functionality to bind to adjacent region of the binding site. In principle, this approach is similar to the “common” drug discovery approach but starting with a smaller fragment. It is useful to get as much information as possible about the first fragment, thus direct binding techniques are favorable. The “needle screen” example for DNA gyrase inhibitors followed this approach [Boehm, 2000]. Assaying the ATP binding site, a first fragment was identified, characterized with NMR and X-ray crystallography and further fragments were attached to get a > 10,000 fold more active inhibitor. More examples for fragment evolution are given below.

(II) Fragment linking: Two fragments, binding to proximal parts of the binding site, are linked together. This is the classical approach to fragment-based strategies. The key issue for this approach is to find the appropriate linker. The availability of structural

information is almost a prerequisite, otherwise an exhaustive combinatorial search following random conditions to find an effective linking scheme is required. The mass independent loss of rigid-body entropy suggests a favorable additivity effect for increasing affinity if a optimal linker is applied. An impressive example is the linkage of two stromelysin binding fragments via NMR exploiting structure-based design. Starting with single affinities of 17 mM and 280  $\mu$ M respectively for the fragments, the finally linked inhibitor showed 15 nM. Further examples are given in Table 5.2.

(III) Fragment self- assembly: Fragments with matching functional groups are forced to react with each other in the presence of the protein target. The two most potent fragments are most likely to react enabling a selection of potent fragments out of a mixture. This approach can be seen as a specially designed way to form a fragment linking. The linkage reaction is predefined which determines the type of fragments suitable for this approach. Also referred to as dynamic combinatorial chemistry it has been successfully applied to proteins, exhibiting deep buried binding sites. Applying “click chemistry” fragment self-assembly led to tight binding inhibitors of acetylcholine esterase and others (Table 5.2).

(IV) Fragment optimization: Fragment approaches are used to optimize or modify only parts of the molecule, enhancing features others than just binding affinity. Again strategies are followed that do not differ significantly from those in a classical structure-based optimization process. An example is the development of an urokinase inhibitor with 100 fold improved affinity and additionally achieved oral bioavailability of 38%.



**Figure 5.4** Summary of fragment-based approaches. I Fragment evolution, II Fragment linking, III Fragment self-assembly and IV Fragment optimization

### ***Computational support in fragment-based drug discovery***

Computational tools have been developed to support both, fragment identification and fragment optimization. Calculating probability maps for the putative binding of particular

ligand atoms next or in the active site are usually summarized as so-called 'hot spots'. Programs to perform such analyses are GRID (energy map derived from force field) [Goodford, 1985], DrugScore (mapping knowledge-based potentials) [Gohlke, 2000b] and SuperStar (mapping interaction groups based on crystallographic data) [Verdonk, 1999]. Their application helps building a targeted library and guiding the starting placement for the first fragment. Docking in fragment-based drug design is rarely applied for the initial fragment screening. The available scoring functions are calibrated towards drug-sized molecules and perform considerably worse on small compounds [Mooij, 2005]. Mostly, virtual fragment libraries are docked to search for attachments to an previously determined ligand, either considering the fragment-grown ligand as a new completely flexible molecule or keeping the known binding part as a rigid constraint. In the case of a missing protein crystal structure, computational strategies for fragment-based approaches are ligand similarity searches and classification tools. They are reviewed by Oprea [Oprea, 2006]. Recently, the computational strategy *BREED* for fragment linking was reported. So-called "hybrid" compounds are generated by virtually combining previously known inhibitors [Pierce, 2004]. Well suited for later stage optimization, it is not faced with the problem of finding the correct linker fragment. Another attempt to address de novo design computationally via fragments is the program FlexNovo [Degen, 2006], a advancement to FlexX. Compared to the original program it consists of fragments derived from the World Drug Index and allows to define filters for property, pose geometry and diversity upon docking.

**Table 5.2** Examples of applications of fragment-based drug design

Target	Strategy	Screening technique	Achievements	Reference
DNA gyrase	I	VS, SBD	Maximal non-effective concentration optimized to 30 ng/ml.	[Böhm, 2000]
thymidylate synthase	I	Tethering, SBD	Affinity to $IC_{50} = 330$ nM	[Erlanson, 2000]
P38 kinase	I	NMR	Affinity to $K_i = 200$ nM	[Fejzo, 1999]
Anthrax lethal factor protease	I	NMR	Affinity to $K_i = 320$ nM	[Forino, 2005]
Urokinase	I, IV	X-ray, bioassay, NMR	Affinity to $K_i = 6.3$ nM; affinity to $K_i = 370$ nM, bioavailability	[Wendt, 2004]
Erm methyl transferase	I	NMR	Affinity to $7.5$ $\mu$ M	[Hajduk, 1999]
Interleukin 2	II	SPR, x-ray	Affinity	[Hyde, 2003]
FK506-binding protein	II	NMR	Affinity to $K_d = 49$ nM	[Shuker, 1996]
Avidin	II	Bioassay	Affinity to $K_i = 4$ pM	[Green, 1975]
Hepatitis C protease	II	Bioassay, NMR	Affinity to $K_i = 800$ nM	[Wyss, 2004]
Vancomycin	II	Bioassay	Affinity to $K_i = 1.1$ nM	[Rao, 1997], [Rao, 2000]
Acetylcholinesterase	II	Bioassay, SBD	Affinity to $K_i = 0.4$ nM	[Pang, 1996]
C-SRC kinase	II	Bioassay	Affinity to $IC_{50} = 64$ nM	[Maly, 2000]
Stromelysin	II	NMR, SBD	Affinity to $K_d = 15$ nM	[Hajduk, 1997]

Target	Strategy	Screening technique	Achievements	Reference
Collagenase	II	HTS, NMR	Selectivity towards MMP-1, MMP-9, TACE	[Chen, 2000]
Caspase	II	Tethering	Affinity to $K_i = 200$ nM	[Erlanson, 2003]
Protein tyrosine phosphatase	II	NMR, SBD	Affinity to $K_d = 22$ nM, $K_d = 7$ $\mu$ M	[Liu, 2003], [Szczepankiewicz, 2003]
Bacterial 23S rRNA	II	MS, SAR by MS	Affinity to $K_d = 6.5$ $\mu$ M	[Swayze, 2002]
Carbonic anhydrase	III	Virtual combinatorial chemistry	Affinity	[Huc, 1997]
Neuraminidase	III	Dynamic combinatorial libraries	Affinity to $K_i = 85$ nM	[Hochgurtel, 2003]
CDK2	III	X-ray	Affinity to $IC_{50} = 30$ nM	[Congreve, 2003b]
Acetylcholinesterase	III	Click chemistry	Affinity to $K_d = 77$ fM	[Lewis, 2002]
Factor Xa	IV	Bioassay, SBD	Bioavailability	[Liebeschuetz, 2002]
SH2 domain of pp60Src	IV	X-ray, SPR, SBD	Affinity to $IC_{50} = 3$ nM, Plasma stability	[Lange, 2003], [Lesuisse, 2002]
Adenosine kinase	IV	NMR	Affinity to $IC_{50} = 10$ nM, enhanced in vivo data	[Hajduk, 2000]
Leukocyte Function-Associated Antigen 1	IV	NMR, Bioassay	Affinity, Solubility, Oral Bioavailability	[Liu, 2001]

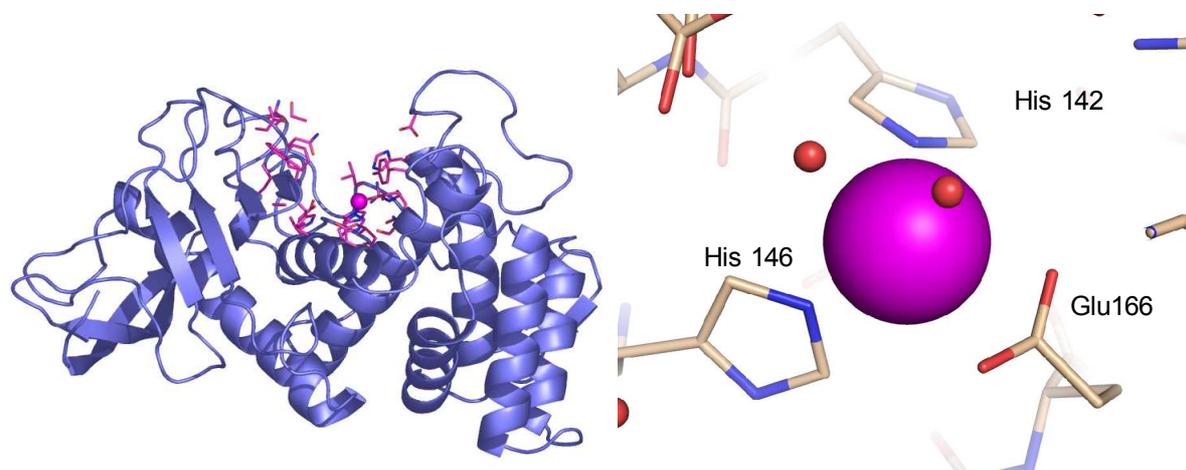
The numbers I, II, III, IV represent one of the four routes to fragment-based lead discovery. X-ray stands for the detection of the initial fragment by protein-ligand complex crystal structure, MS for mass spectroscopy, VS for virtual screening, SBD for structure-based design.

More detailed reviews are given by Rees, Carr and Erlanson [Carr, 2002], [Erlanson, 2004a], [Erlanson, 2006], [Rees, 2004]. Recently, an example showing the limitations of fragment-based drug design has been reported. Deconstructing an AmpC  $\beta$ -lactamase inhibitor revealed three possible fragments. The full ligand as well as the three fragments bind to the protein and all four corresponding protein-ligand complex structures were determined. They showed that fragments of a larger molecule did not recapitulate its binding [Babaoglu, 2006]. Obviously, the recognition and binding properties of the fragments are strongly apart from those of the finally composed ligand. This example points to some deficiencies following a strategy starting with a set of individual fragments that are subsequently linked to a larger molecule.

### **Target protein**

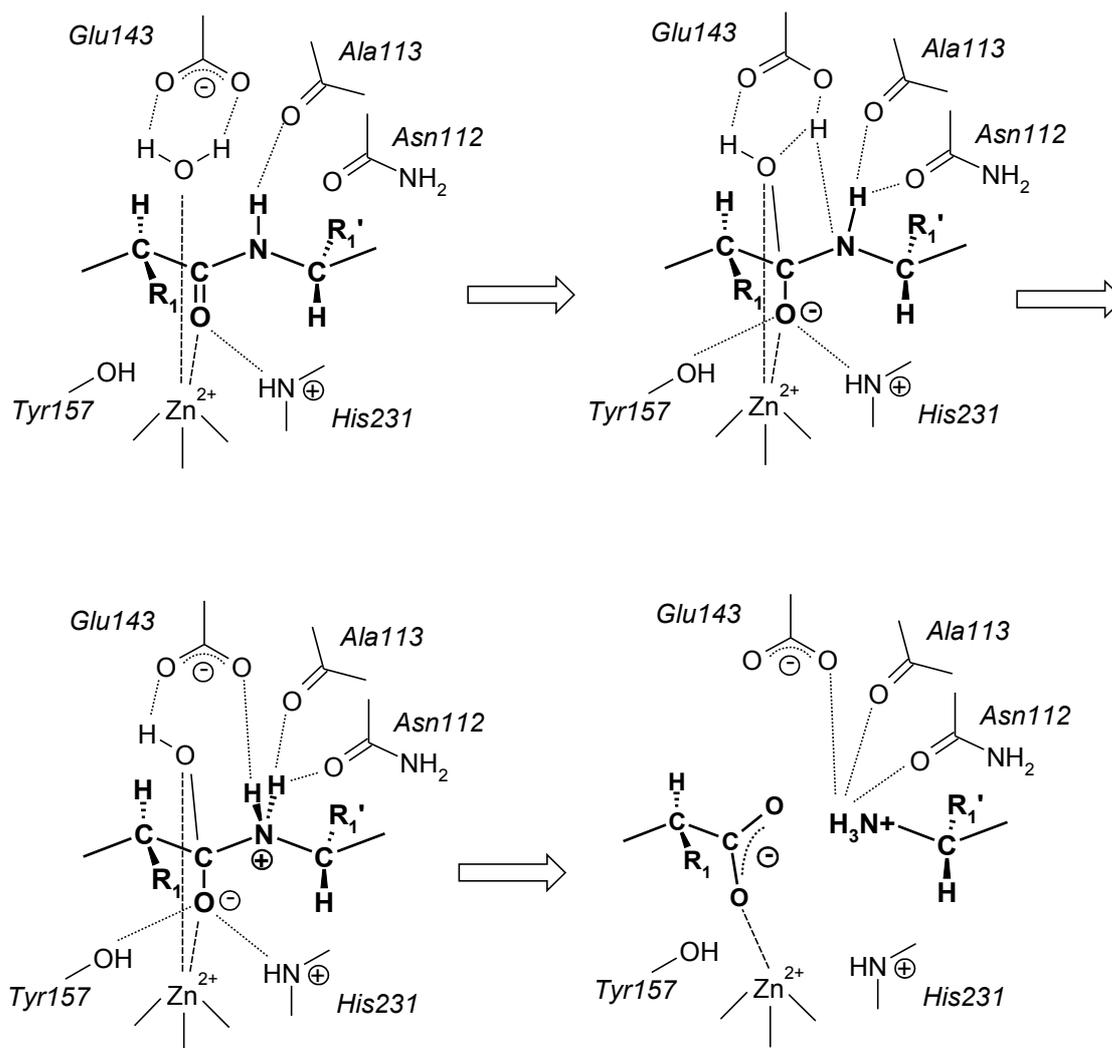
Zn<sup>2+</sup>-containing metalloproteinases constitute a long list of structurally related proteases which are widely distributed in nature. They are involved in biochemical events of high importance such as digestion (carboxypeptidase A, astacin), tissue remodeling and extracellular matrix degradation (matrix metalloproteinases), blood-pressure regulation (neprilysin, angiotensin converting enzyme), formylation and deformylation in bacterial protein synthesis (peptide deformylase), etc. Thermolysin (TLN) is a heat-stable proteolytic enzyme isolated from *Bacillus thermoproteolyticus* and carries a catalytically essential zinc ion at the active site [Matthews, 1988]. Four additional calcium ions are required for stability. The enzyme is an extracellular endoproteinase, hydrolyzing the peptide bond on the imino side of the amino acid residue having a large hydrophobic side chain, such as Leu, Ile, and Phe. Thermolysin is bilobal and the active site is situated in a cleft between the two lobes. The zinc ion is tetrahedrally ligated in the apo structure, binding His142, His146, Glu166 and one

water molecule. Additionally, there is a Val-Lys dipeptide bound to the active site, presumably a product of self-proteolysis. It occupies the hydrophobic specificity pocket.



**Figure 5.5** Structure of thermolysin. Left: Overall structure, showing two domains (blue and as cartoon representation) and the binding cleft in the center. The zinc ion is shown as pink sphere and residues forming the binding site are shown in purple. Right: Binding site geometry of the catalytic zinc ion and observed in pdb ID: 2TLX.

The enzyme has been extensively studied as a prototypical zinc enzyme and used as model target system for developing design strategies of enzyme inhibitors that can be transferred to other zinc proteinases of medicinal interest such as the angiotensin converting enzyme or matrix metalloproteinases. Most small molecule inhibitors for zinc proteases carry a functional group that coordinates the active site zinc ion, such as carboxylate, thiol or hydroxamate. Of these zinc ligating groups, hydroxamates are frequently applied as they show strong chelating properties towards zinc. The thermolysin and zinc metalloproteinase reaction mechanism is well studied [Bertini, 2006] and reversing its proteolytic activity TLN is also used for peptide synthesis or sugar-protein linkage.

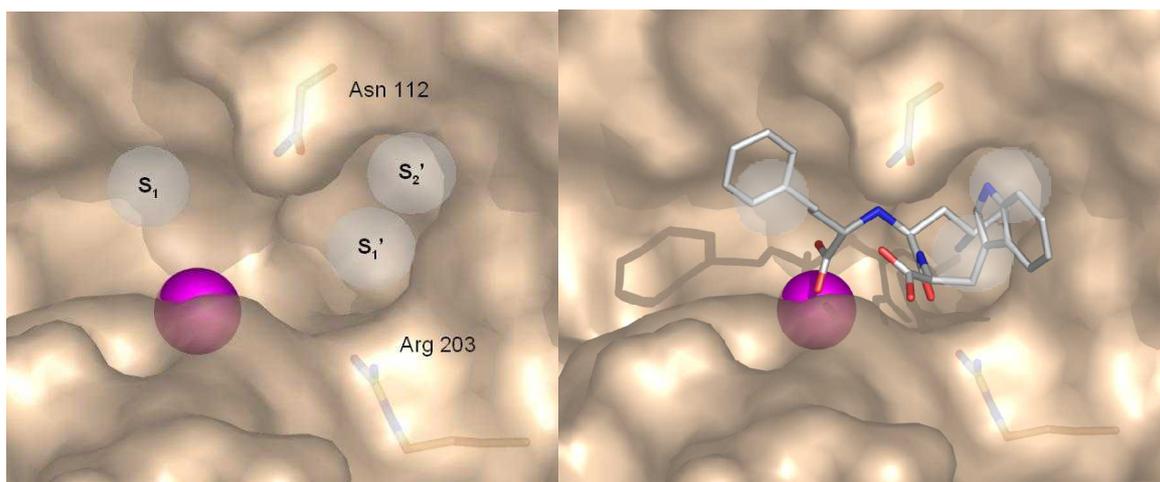


**Figure 5.6** Detailed proteolysis mechanism of thermolysin. Throughout Glu143, the O-H bond gets polarized. This enables a nucleophilic attack from the water oxygen atom to the carbon atom of the amide bond. The proton, previously temporarily bound to Glu143 is transferred to the amide nitrogen atom. The positive charge is additionally stabilized by Asn112 and Ala113. Finally, the amide bond breaks and the upon product dissociation, a new water molecule binds to the zinc ion.

According to the classification from Berger and Schlechter [Schlechter, 1967], protease binding sites can be subdivided. For thermolysin that involves most prominently:

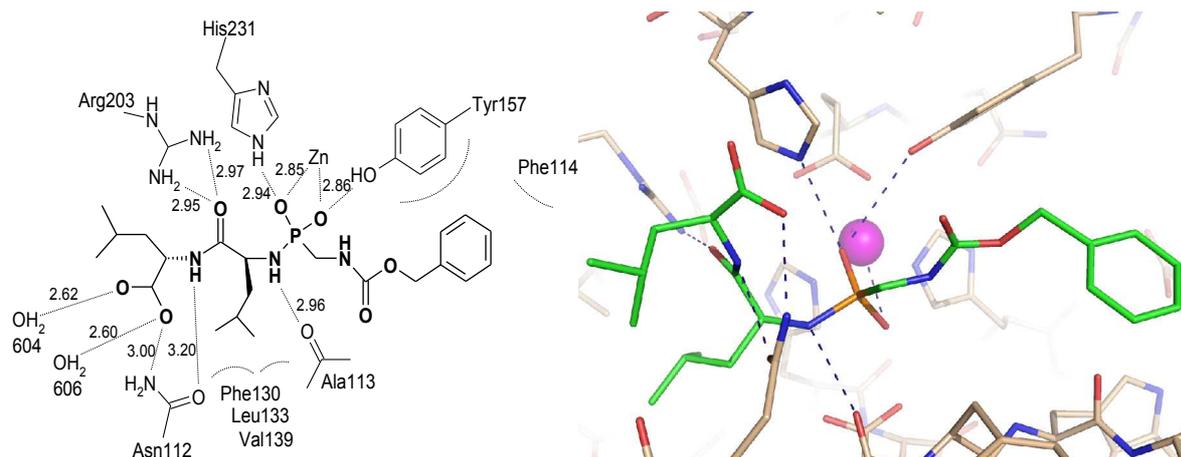
- Zinc binding region, formed by His142, His146, Glu166 and a water molecule. The water molecule gets activated by Glu143.
- $S_1'$ -pocket defining thermolysin specificity. It accommodates short lipophilic chains (up to four carbon) or a phenyl moiety.

- Huge, rather unspecific  $S_1$ -pocket. Various fragments can bind here, often forming a hydrogen bond to the backbone nitrogen of Trp115.
- Far  $S_2'$ -pocket. Again rather unspecific it offers the possibility for a lipophilic fragment and eventual hydrogen bonds to Asn111.



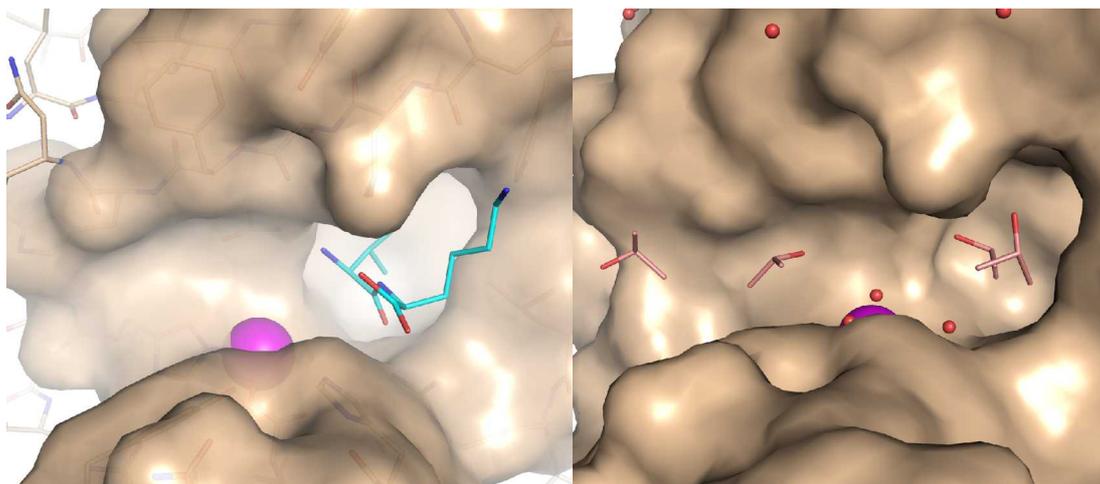
**Figure 5.7** Binding site composition. Left: The binding site can be divided into several sub-pockets. Amino acids, strongly involved in substrate or inhibitor binding are highlighted. While the  $S_1$  and  $S_2'$ -pocket can accommodate structurally diverse ligands, the  $S_1'$ -pocket is specific for short hydrophobic side chains. Right: Protein-ligand geometries as observed in pdb ID: 1TMN. Tight binding ligands address all described sub-pockets.

Multiple protein-ligand crystal structures have been solved for thermolysin. They indicate the importance of the different sub-pockets and they help to derive a pharmacophore model. Binding to the zinc ion is either mono or bidentate. Bidentate ligands bind with a trigonal bipyramidal geometry representing transition state analogs. Possible coordination groups are hydroxamate, phosphate, carboxylate functions. The monodentate ligands bind in a tetragonal geometry representing substrate analogs. Examples are thiol, urea or carbonyl functions.



**Figure 5.8** Example of ligand binding to thermolysin. Inhibitor ZG<sup>P</sup>LL binds with  $K_i = 9$  nM in a competition assay. Left: Schematic overview of protein-ligand interactions including distances in Å. Right: Protein-ligand geometry as observed in pdb ID: 5TMN [Holden, 1987]. Protein residues are shown with beige carbon atoms. Ligand carbon atoms are shown in green. The pink sphere represents the zinc ion and dashed lines represent protein-ligand interactions.

Despite tight binding ligands, there are also several dipeptides known to inhibit thermolysin. The most prominent is Val-Lys. Feder *et al.* investigated a series of Phe, Tyr, Gly, Leu combinations [Feder, 1974]. Most of these bind to the protein in the millimolar range and several protein-dipeptide complex structures could be solved by protein crystallography. Strikingly, these ligands do not bind to the zinc ion, demonstrating that targeting one part of the binding site can be sufficient to provide a starting point for ligand design.



**Figure 5.9** Small molecule inhibitors bound to thermolysin. Left: Val-Lys dipeptide as observed in pdb ID: 2TLX [English, 1999]. It occupies the  $S_1'$  and  $S_2'$  pocket whereas the zinc ion remains uncomplexed. Right: Thermolysin structure crystallized in 100 % isopropanol as observed in pdb ID: 8TLI. The structure clearly shows favorable interactions sites in the binding pocket for possible hydrogen bonds.

Thermolysin is a thermostable protein, thus comprising a core built extensively by hydrophobic amino acids compared to temperature sensible proteins. This higher stability makes the protein structure and the assembled crystals more resistant to organic solvents. Thermolysin could be crystallized from 100% isopropanol [English, 1999] and in the presence of 70% acetone [English, 2001], in buffers mixed with acetonitril, thiocyanat and phenol (supporting material). Even though these molecules possess only minor binding affinity they could be detected in the difference electron density. The possibility to grow protein-ligand complexes showing solvent molecules and other small fragments in different sub-pockets, renders thermolysin a promising target for fragment-based lead discovery.

## Methods

### Library composition

Different data sets were assembled to represent molecular properties required for either thermolysin inhibition, or in agreement with criteria such as fragment likeness or accessibility as commercially available compounds.

**X-ray reference data set (XRAY):** Ligands from thermolysin-ligand complex structures were extracted to represent a reference data set. Based on the ligand molecular size compounds were divided into 'ligands' and 'solvents'. To generate an additional reference fragment data set, the ligands were split in agreement with rules implemented into the RECAP algorithm.

**Drug fragments data set (WDI):** To compile drug-like molecules the World Drug Index 2001 (WDI) (<http://scientific.thomson.com/products/wdi/>) was chosen as a second data source. From this set, containing approximately 70,000 biologically active molecules, all diagnostics, vitamins, supplements and drugs administered differently than oral were discarded. A further selection criterion was that the given molecular weight should be  $70 \text{ g/mol} < x < 750 \text{ g/mol}$ . Compounds in agreement with these conditions were split using the RECAP [Lewell, 1998] algorithm. For both kinds of RECAP sets, the XRAY and the WDI, two different labels were defined to the scissile atoms. They were either marked with an R group dummy atom (*Rgroup*) or saturated by a hydrogen atom (*RsubH*). The  $R_x$  label refers to the type of RECAP bond cleavage. To remove duplicates, all labels were later set to  $R_1$ . The obtained chemical topologies were translated into 3D structures using CORINA.

**Commercially available compounds data set (CAC):** The 2D connection tables of the following publicly accessible databases have been converted into 3D structures using CORINA: Asinex, IBS, Specs, Leadquest, Maybridge, and Sigma (addresses in supporting information). Any counter ions have been discarded. For virtual screening purposes

protonation states were used as defined by CORINA. For docking experiments a protonation state at pH=7.3 has been assumed. Hereby, phosphate, sulfonate and carboxylate functions are considered deprotonated, amines protonated. Corresponding to the respective vendor, molecules were either claimed HTS ligands or building blocks.

### **Virtual Screening**

Three hierarchical filter were applied to the data sets corresponding to library design rules either suggested by Baurin *et al.* [Baurin, 2004] or in agreement with (1) fragment likeness, (2) target specificity (3) unwanted functional groups.

In a fast initial step, a filter based on definition of a fragment was applied using the *Selector* compound filtering utility of SYBYL. Molecular size and rotatable bonds were considered. The number of rotatable bonds criteria with respect to the published “rule of 3” for fragments was exceeded. In an intensive study, Hopkins *et al.* [Hopkins, 2003] correlated the ligand size to the corresponding target family. It turned out that proteinase inhibitors are distinctly larger compared to other enzyme family inhibitors. We did not include hydrogen bond acceptor or donor functions in the screen due to limited definitions of those in the Selector tool of SYBYL 7.0. To avoid solvents or ions in the data set, the minimum number of atoms composing a considered candidate fragment was set to seven. Additional filter criteria were molecular weight  $70 \leq x \leq 250$  g/mol and the number of rotatable bonds:  $\leq 8$ . A highly diverse fragment library increases the chance to find a hit during screening. This is important when only little is known about requirements for target inhibition. In case of prior knowledge about inhibitors a targeted fragment library may serve as a better approach [Zartler, 2005]. A protein targeted fragment library was successfully applied for factor Xa inhibitors [Fielding, 2003]. To generate a protein targeted library for thermolysin, two distinct areas in the

thermolysin binding site were addressed, the zinc binding region and the S<sub>1</sub>'-pocket. To generate appropriate pharmacophores, a database search with ReliBase+ [Bergner, 2001], [Cramer, 1989], [Gunther, 2003], [Hendlich, 2003] was performed.

1. Zn-binder: Defined as Zn- nitrogen, Zn- oxygen, or Zn- sulfur contacts with a distance restraint of  $1.5\text{\AA} < x < 2.0\text{\AA}$ . The zinc ion is defined as one ligand, the interacting molecule as a second. Both must be part of ligand entries (LIG) in the database. All functional groups capable of coordinating a zinc ion with tetra- or penta-valent geometry were selected.
2. S<sub>1</sub>'-pocket: Thermolysin and matrix metalloproteinase structures were analyzed for ligand moieties, addressing the S<sub>1</sub>'-pocket. Matrix metalloproteinases and thermolysin belong to the same superfamily of zinc proteases, the zincins. They also share a hydrophobic specificity pocket. Studying inhibitors and protein cavities, the maximal size and possible branching for the ligand was obtained.

Molecules comprising the following functional groups were discarded: Either they contain an atom different from H, C, N, O, F, Cl, S, or sugars, anhydrides, aziridines, epoxides, ortho esters, nitroso compounds, isonitriles, acetals, thioacetals, N-C-O acetals, nitro compounds.

All functional groups and pharmacophores were defined as 'Markush atom' and the program UNITY was used to perform a 2D connectivity screening. Passing compounds were sorted corresponding to the data bank/vendor they originated from and the matching pharmacophore.

## **Docking**

An appropriate strategy was chosen based on the docking of a benchmark data set. It consists of several drug-sized thermolysin inhibitors for which the protein-ligand complexes have been experimentally determined. The protein structure was retrieved from pdb ID: 1TMN

[Monzingo, 1984]. Water molecules assigned to the pdb complex structure were removed. 20 ligands were tested using FlexX, AutoDock and GOLD using the scoring functions implemented in FlexX, GOLD, AutoDock, along with Chemscore and DrugScore. Based on the achieved performance in regenerating the native the protein-ligand binding mode, the most appropriate docking/scoring combination was selected for further analyses. Additionally, the fragment data set was docked as well as several solvent molecules for which the protein-solvent structure was crystallographically determined.

**FlexX:** Standard settings were applied as suggested for FlexX 1.13. The binding site was defined by all residues coinciding within an 8Å distance around the bound ligand in pdb entry 1TMN. This comprises S<sub>1</sub>, S<sub>1</sub>' and S<sub>2</sub>' pocket. The FlexX scoring function was used during the complex construction phase. No threshold for the number of accepted solutions was defined. The finally generated geometries were scored with both, the FlexX function and DrugScore.

**GOLD:** Protein hydrogen atoms and lone pairs were added using GOLD 3.0. The binding site was chosen equivalent to that used in FlexX. Correct protonation states of histidine residues, especially at the zinc ion were manually adjusted. Allowed geometries to complex the zinc ion were tetrahedral and trigonal bipyramidal. Standard parameters for the genetic algorithm were applied and ligands were docked using the 'optimal accuracy' setting without any speed-up. 10 docking runs were performed per ligand. Scoring was done with the GOLD function initially. Rescoring with Chemscore included a local optimization. Finally the GOLD scored geometries were rescored with DrugScore.

**AutoDock:** Ligands were docked into the binding pocket using AutoDock 3.0. The binding site box was set to be similar to FlexX and GOLD. The partitioning of the total charge into individual atomic contributions was performed using the Gasteiger-Marsili charges [Gasteiger, 1980]. AutoDock 3.0 requires polar hydrogen atoms on the protein for docking. They were added with the PROTONATE utility in AMBER [Case, 2002] and the generated

hydrogen bonding network was visually inspected for internal consistency. AMBER united-atom charges were assigned as defined in the AMBER force field [Weiner, 1984], and solvation parameters were added using the ADDSOL utility from AutoDock3.0. Docking runs were performed with the Lamarckian genetic algorithm as implemented in AutoDock 3.0, using an initial population of 50 randomly placed individuals, a maximum number of  $1.5 \times 10^6$  energy evaluations (for fragments:  $0.75 \times 10^6$ ), a mutation rate of 0.02, a crossover rate of 0.80 and an elitism value of 1. Generated ligand docking solutions, mutually differing by  $\text{rmsd} \leq 1 \text{ \AA}$  were clustered together and the lowest docking energy found for one entry of a cluster was used as representative. For each reference ligand, initially 50 solutions were generated. For the reference fragment 20 solutions were generated and 10 in the subsequent screening. Docking results were rescored using DrugScore. All applications of DrugScore included the solvent accessible surface term.

**Fragment Docking:** The XRAY-fragment data set was docked and the obtained geometries were faced to the corresponding placement of this fragment in the parent ligand complex structure. For an endoproteinase most likely the strongest interaction will occur in an area next to the scissile amide bond. Therefore, fragments addressing the  $S_1$  and  $S_1'$  pocket were considered. Docking was done either with the Rgroup and the RsubH ligand setup. The WDI fragments and ligands from the CAC data set were docked using AutoDock3.0 as described with 20 docking solutions per ligand. The Rgroup was treated as dummy atom in GOLD and FlexX. For AutoDock a special atom type R was introduced to prevent Rgroup-protein interactions. Parameters were chosen similar to the non-aromatic carbon atom type and all attracting interaction points on the energy grid were set to 0.

**SFC score:** For a retrospective analysis of the generated docking solutions a recently developed scoring function was applied [Sotiffer, 2007]. Different to the establishment of the

original function, parameterized on the base of large overall protein-ligand complexes [Sotiffer, 2007], only complexes with fragment-like ligands were considered. Selection criteria were molecular weight < 300 g/mol and number of rotatable bonds  $\leq 3$ .

**Post-filtering:** Inherent to the docking smaller molecules is the increasing number of possible placements of the ligand into the pocket. This leads to multiple, structurally diverse docking modes for which approximately the same affinities are predicted. Therefore, the generated docking geometries were examined whether they agree to a predefined UNITY 3D pharmacophore. During this post-docking process, ligands were kept rigid and had to penetrate either into a sphere around the zinc ion or to address the S<sub>1</sub>'-pocket. Mismatching docking modes were discarded.

### ***Crystallization***

Using native thermolysin (purchased from Calbiochem) crystals suitable for X-ray diffraction were prepared as described by Holmes and Matthews [Holmes, 1982] with the following slight modifications. TLN was dissolved in 0.05 M tris/HCl buffer (pH 7.3), containing DMSO (50% (v/v)) and calcium acetate (0.05 M) and caesium chloride (2.5 M). The final protein concentration was 4.0-4.6 mM. The protein is predissolved in DMSO and then mixed with the buffer. Crystals were grown by the sitting-drop vapor diffusion method using water as reservoir solution. Crystals grew to their final size (0.4 mm) after 3 days. Protein-ligand complex crystals were obtained by soaking. Fully grown crystals were transferred to 0.01 M tris/acetate (pH 7.3) soaking buffer solution containing calcium acetate (0.01 M) and 5% (v/v) DMSO. Ligand concentrations were 10 mM for benzylsuccinat, 20-50 mM for the fragment soaking or saturated when low solubility was given and crystals were left at least two days in the soaking solution prior to freezing.

### **Data collection**

Data were collected at temperature = 100K using a cryo-protectant solution of 10 mM tris/acetate buffer containing 10 mM calcium acetate, 5% DMSO and 20% glycerol (pH 7.3). The data sets were collected on a RIGAKU RU-300 copper rotating anode (Molecular Structure Cooperation) at 50kV, wavelength = 1.5418 Å, 90mA using a R-AXIS IV++ image plate system. For each frame the exposure time and oscillation rate were set to 7 minutes and  $\Delta\phi = 0.5^\circ$ , respectively. The crystal to detector distance was 120 mm. Data processing and scaling was performed using the HKL2000 package [Otwinowski, 1997]. The 3-methylaspirin complex structure was additionally determined at the synchrotron BESSYII in Berlin on PSF beam line II equipped with a MAR-CCD detector. In total, 120 frames with  $\Delta\phi = 0.5^\circ$ , wavelength = 0.91841 Å at a crystal to detector distance of 175 mm were collected at  $-170^\circ\text{C}$ . Exposure time was 8 seconds. Data processing and scaling were performed using the HKL2000 package [Otwinowski, 1997].

### **Structure determination**

The coordinates of thermolysin in complex with Val-Lys (pdb ID: 2TLX) were used after removal of ligand, metal ions and water molecules for initial rigid-body refinement of the protein atoms followed by repeated cycles of conjugate gradient energy minimization, simulated annealing and B-factor refinement using the CNS program package [Brunger, 1998]. Refinement at later stages was performed with the program SHELXL [Sheldrick, 1997]. Here, at least 20 cycles of conjugate gradient minimization were performed with default restraints on bonding geometry and B-values. Five percent of all data were used for  $R_{\text{free}}$  calculation. Amino acid side-chains were fitted into sigmaA-weighted  $2F_o - F_c$  and  $F_o - F_c$  electron density maps using O and Coot [Jones, 1991], [Emsley, 2004]. After the first

refinement cycle, water molecules and subsequently the zinc atom and ligand were located in the electron density and added to the model. Restraints were applied to bond lengths and angles, chiral volume, planarity of aromatic rings and van der Waals contacts. Multiple side-chain conformations were built in case an appropriate electron density was observed and maintained during the refinement, and if the minor populated side-chain showed at least 10% occupancy. During the last refinement cycles, riding H atoms were introduced without using additional parameters. The final models were validated using PROCHECK [Laskowski, 1993]. Data collection, unit cell parameters and refinement statistics are given in Table s.11.

### **Binding Assay**

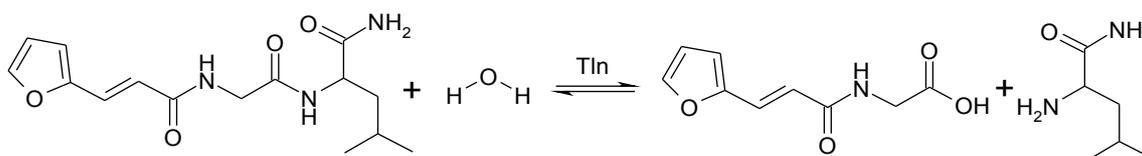
Determining enzyme and inhibitor kinetics was performed as described by Morgan *et al.* [Morgan, 1978] with slight modifications.

*Kinetic Parameters for thermolysin hydrolysis:* To a thermolysin solution (13 nM) FAGLA was added from 0.1 to 8.0 mM. For higher concentrations, 5 mM and 2 mM path length cells were used to keep the absorbance of the solution below 2.5. At each concentration, the initial velocity was determined by measuring the molar extinction coefficient at  $\epsilon=345$  nm.

*Determination of inhibition constant  $K_i$ :* Kinetics were followed with protein concentration of 13 nM and two substrate concentration of 2 mM and 4 mM. The number of applied inhibitor concentrations ([I]) varied, depending on solubility and availability of the compounds. The  $K_i$  values were determined from  $v_0/v_i$  versus [I] plots with concentrations over a range of at least  $0.5 (K_i) - 10 (K_i)$ . Depending on solubility, inhibitors were dissolved in standard buffer, 50% or 100% DMSO.

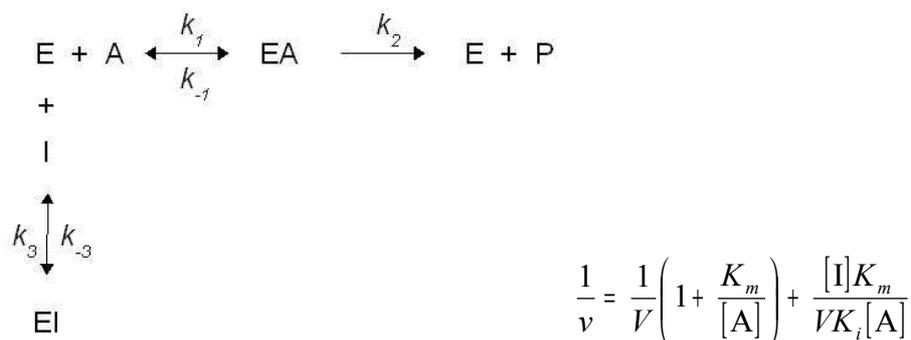
*General procedure:* Phosphoramidon and thermolysin were purchased from Calbiochem and FAGLA (3-2-furylacryloyl)-L-glycyl-L-leucine-amide from Bachem. All stock solutions

were prepared with doubled distilled water and filtered through 0.45  $\mu\text{M}$  pore size Millipore filter. The buffer for FAGLA and thermolysin was 0.1 M 3-(*N*-morpholino)propanesulfonic acid (MOPS), 2.5 M NaBr and 10 mM  $\text{CaCl}_2$ . FAGLA was pre-dissolved in DMF to yield a final DMF concentration of 2.5% (v/v) at pH 7.0. Assays were performed at 25°C, monitored by absorbance change at 345 nm and followed for five minutes. At least double measurements were applied to each substrate and inhibitor concentration.



**Figure 5.10** Hydrolysis reaction for thermolysin inhibition assay.

In an extensive study analyzing thermolysin sensitivity to organic solvents, DMSO showed the least effect on the residual activity and protein structure. However, the influence of organic solvents on the inhibitor kinetics were determined for phosphoramidon.  $K_i$  values for phosphoramidon with 5% and 10% DMSO were determined. 10% corresponds to dissolving the ligand in pure DMSO. The determined affinities of  $K_i(5\%) = 226 \mu\text{M}$  and  $K_i(10\%) = 558 \mu\text{M}$  demonstrate the strong solvent influence on protein inhibition. Assuming competitive inhibition kinetics the following mechanism and equations were applied to calculate inhibition constant  $K_i$ .



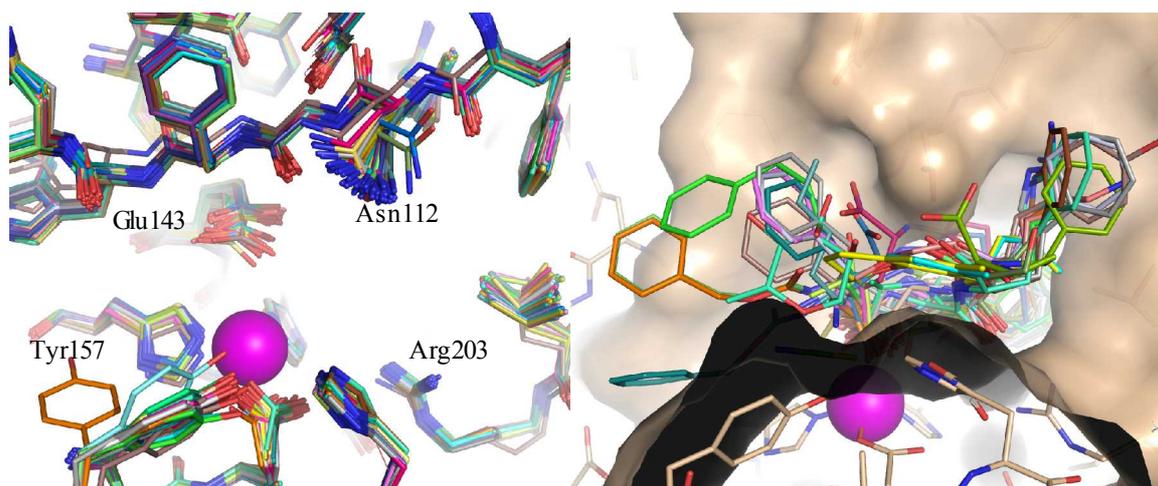
**Figure 5.11** Scheme for competitive enzyme inhibition assuming that only either substrate or inhibitor can bind to the enzyme.  $v$  is velocity measured per substrate  $[A]$  and inhibition concentration  $[I]$ .  $V$  is the maximal enzyme velocity and  $K_m$  is the Michaelis-Menten constant.

$V$  is not changed in this inhibition mechanism, as FAGLA present in large extent displaces the inhibitor (on the opposite large quantities of inhibitor displace FAGLA). This feature is indicative for this type of inhibition and in the Dixon plot [Dixon, 1953], [Dixon, 1972] the straight lines meet in a joint intercept in the second quadrant with  $-K_i$  as  $x$ -coordinates.

## Results and Discussion

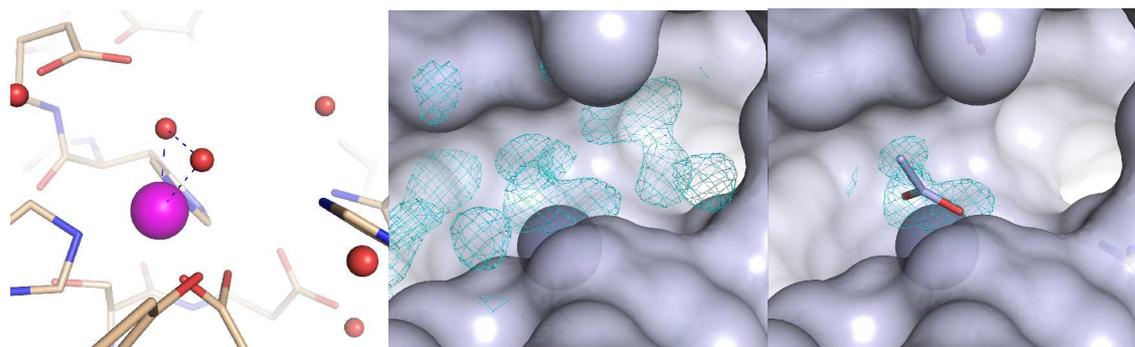
The present study describes a fragment-based virtual screening and the experimental confirmation of the detected hits. To generate the appropriate ligand and fragment libraries we divided already known ligands stored in the WDI. Obtained fragments were requested to be commercially available. The XRAY data set comprised 58 thermolysin complexes deposited in the protein data bank (<http://www.rcsb.org>). NMR structures, apo structures, non-Zn metal structures were removed, yielding 49 protein-ligand and 6 protein-solvent complex structures. Superimposition of all complex structures showed rigidity in the binding site, proving thermolysin to be a auspicious target for molecular docking. However, there are three macrocyclic inhibitors (pdb ID's 1PE5, 1PE7, 1PE8 [Hansen, 2003]) forcing Asn112 to a

slight adjustment in the  $\chi^2$  angle. Analyzing the superimposed ligands, all inhibitors address the  $S_1'$ -binding site and most are involved in zinc ion coordination. While the narrow  $S_1'$  site only allows limited access, ligands addressing the  $S_1$  site show a broad variety in functionalities and conformations.



**Figure 5.12** Superimposition of 49 protein-ligand complex structures. Left: Superimposition of protein chains only, showing some but only minor flexibility in the binding sites. Right: Superimposition of 49 different ligands into the binding site of pdb ID: 1TMN. While ligand moieties for the  $S_1'$  are limited in size and orientation, the  $S_2$  and  $S_1$  pocket get addressed by a variety of functional groups and sizes.

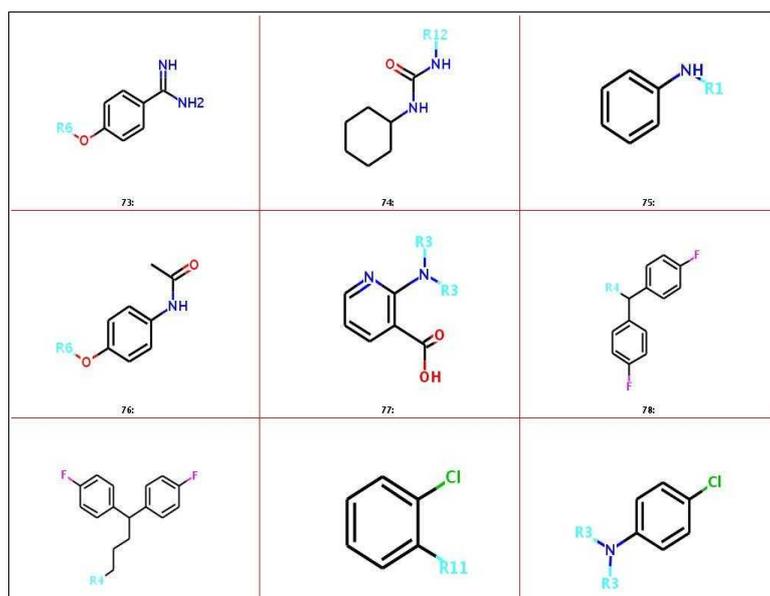
In case of a ligand not binding to the  $Zn^{2+}$  ion, two water molecules are found to occupy the remaining two positions to form a bidentate coordination (for example, pdb ID series (1-8)TLI). However, these two water molecules are at a very close mutual distance of on average 2.32 Å. Superimposing these structures with the difference density from our thermolysin Val-Lys complex structures, we observed these two oxygen positions to superimpose very well with two density peaks in our  $F_o-F_c$  density. However, we believe that instead of two water molecules in very close proximity, these oxygen atoms rather belong to the same molecule, which is likely to be a picked-up acetate from the crystallization or cryo buffer (Figure 5.13).



**Figure 5.13** Coordination next to the  $\text{Zn}^{2+}$  ion in thermolysin. Left: pdb ID: 1TLI [English, 1999] shows two water molecules, binding to the  $\text{Zn}^{2+}$  ion with a mutual distance of 2.35 Å. Center:  $F_o-F_c$  density at 2.0  $\sigma$  found in a thermolysin crystal structure measured in the context of this study. An acetate ion has been assigned to the zinc ion for one refinement cycle. Shown is  $2F_o-F_c$  density at 1.5  $\sigma$ .

### ***Library design and Virtual Screening***

Fragmenting the 21 XRAY data set ligands into small building blocks yielded 46 fragments. Unfortunately, non of the resulting fragments coordinating zinc or occupying in the complete parent ligand the  $S_1'$ -pocket were commercially available. Solvents and cosolvents determined to bind to the thermolysin binding site by X-ray crystallography are DMSO, acetonitril, acetone, isopropanol and phenol. Applying the filter criteria described above to the World Drug Index yielded 4086 unique fragments of which 3976 remained after setting the Rgroup in all scissile bonds to R1 to remove duplicates. Figure 5.14 shows examples of size and functionalities.



**Figure 5.14** Example fragments from the data set derived from the World Drug Index. Molecules were divided using the RECAP algorithm. To avoid duplicates, all cuts were subsequently set to R1.

Assembling data from six commercial suppliers revealed ca 800,000 available compounds and 100,000 building blocks. Virtual screening was performed in a stepwise fashion using Selector and Unity with hierarchical filters of increasing complexity.

**Table 5.3** Summary of screening results after Selector filter

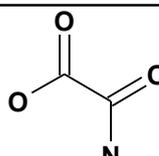
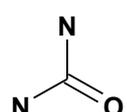
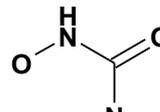
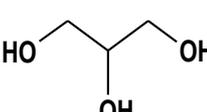
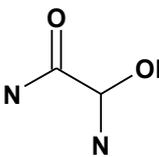
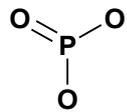
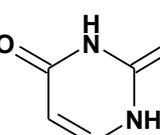
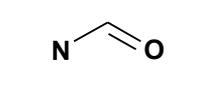
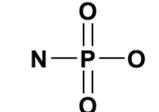
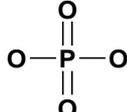
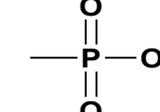
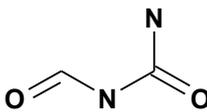
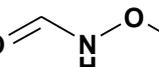
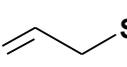
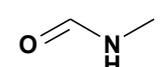
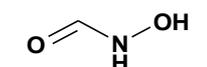
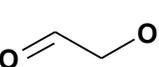
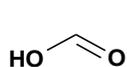
Data set	Number of passing compounds	(%)
<b>XRAY</b>	46	100
<b>WDI</b>	3,090	77
<b>Building blocks</b>	10,034	10
<b>HTS compounds</b>	53,825	6.7

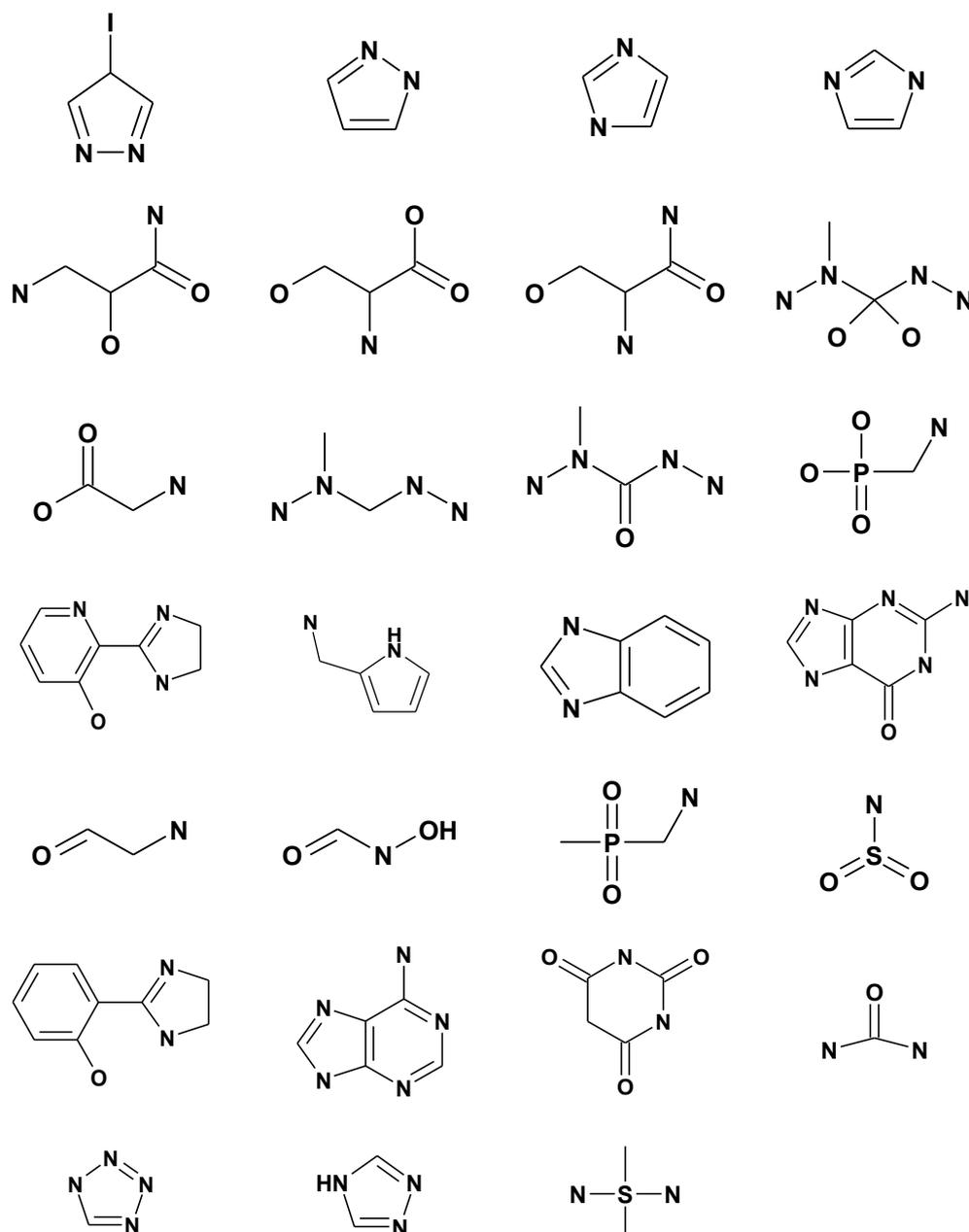
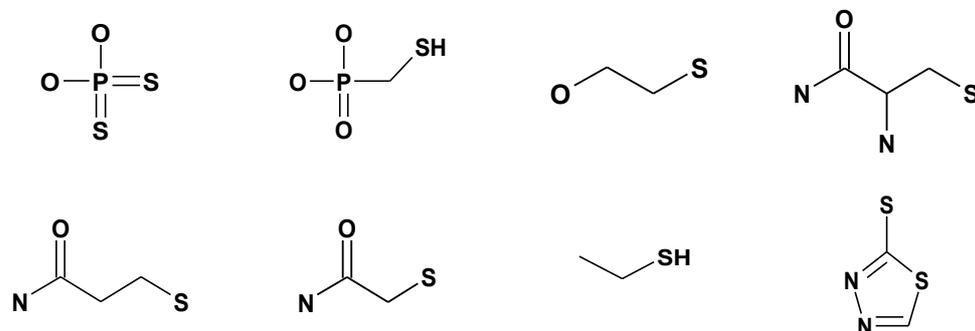
While molecules of the XRAY and WDI set were already fragment biased, maximal 10% of the ligands originating from the database of commercially available compounds passed the filter. These compounds represent a non target biased fragment set. However, as only 10%

passed the predefined fragment criteria, the low success rate indicates the tendency towards inadequate entries not only for the library of 'complete' ligands used for HTS but also for building block libraries. It also suggests that a well designed fragment library requires novel synthesis. Compared to the estimated number of  $10^{60}$  possible drug-like molecules [Villar, 2000] the here considered randomly assembled 65,000 compounds cover only a small part of chemical space.

The target-based pharmacophores have been applied as a second filter step in the screening scenario, using the 2D search tool in UNITY. The applied pharmacophore hypothesis was based on current knowledge about established and crystallographically studied compounds binding to thermolysin and matrix metalloproteinases. The resulting Zn contact groups reflect the current knowledge of experimentally determined functional groups, that complex  $Zn^{2+}$  ions in proteinases.

**Table 5.4** Functional groups found to complex  $Zn^{2+}$  ions in metalloproteinases

Zn-O Binder				
				
				
				
				

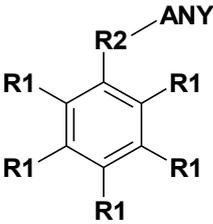
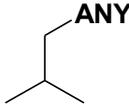
**Zn-N  
Binder****Zn-S  
Binder**


---

Zn-binding fragments are drawn as they are translated into SYBYL markush atoms.

Additionally to the zinc binding fragments, a second pharmacophore for ligand moieties addressing the S<sub>1</sub>' specificity pockets was generated. In summary, they show limited space and the bias for hydrophobicity for this area. Branching seems almost impossible.

**Table 5.5** Pharmacophores representing ligands found to address the S<sub>1</sub>'-pockets.

<b>Core structures</b>			
<b>Substitutions</b>	R1 = H, F, Cl, Br, I R2 = all but N	Incl. N,O planar	

All pharmacophores were translated to UNITY markush atoms.

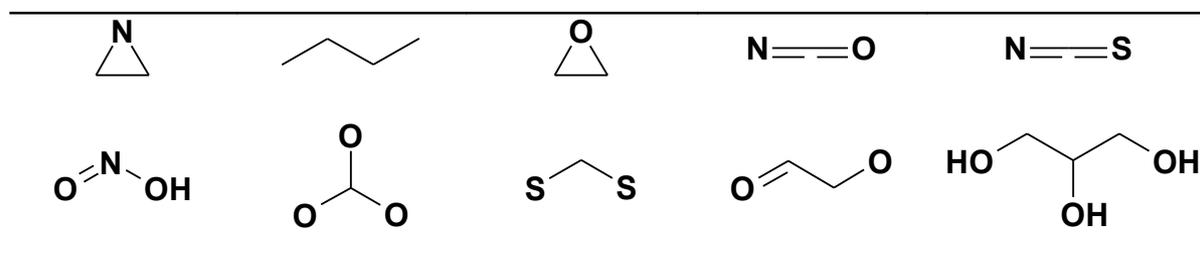
**Table 5.6** Results from applying the second filter criteria for virtual screening.

<b>Data sets</b>	<b>Zn-binder</b>	(%)	<b>S<sub>1</sub>'-binding site</b>	(%)
<b>XRAY</b>	37	80	28	63
<b>WDI</b>	958	24	577	14
<b>Building blocks</b>	2,900	2.9	1,473	1.47
<b>HTS compounds</b>	21,404	2.7	10,004	1.25

The high amount of WDI Zn-binder fragments passing this filter can be explained by the fact that many metalloproteinases have already been studied as potential drug targets and thus, such examples should pass this filter. Compared to the Zn-binder, the amount of WDI fragments passing the S<sub>1</sub>'-pocket filter is lower. However, the pharmacophores are retrieved

from molecules which are inhibitors for their targets but not necessarily drugs. Marketed drugs are optimized not only for affinity but for pharmacokinetic parameters as well. In order to achieve better ADME properties, the hydrophobic moieties are often further substituted. The amount of commercially available compounds passing the pharmacophore filter is surprisingly high indicating that a broad variety of ligands is in agreement with the pharmacophore hypothesis. The final 2D filter step was introduced to discard molecules comprising functional groups not likely to be favorable in drugs.

**Table 5.7** Functional groups applied in the unwanted functionalities filter in virtual screening.



This last filter was not applied to the XRAY data set, because it served as a reference data set only. Table 2.8 shows the final hit lists derived from virtual screening.

**Table 5.8:** Final compound data sets after virtual screening

Data base	Zn-Binder	S <sub>1</sub> '-pocket	Intersection (Zn/S <sub>1</sub> )	Intersection (Zn/S <sub>1</sub> )
<b>XRAY</b>	37	28	} 112 / 75	} 0 / 1
<b>WDI</b>	958	577		
<b>HTS/BBlocks</b>	13 990	7 952		

We were especially interested in fragments found as intersection between the CAC and WDI data set. Even though a relatively small number of 200 entries remained for virtual

screening it is still a reasonable number to perform docking experiments. Docking fragment-sized molecules has been described as difficult and correctly predicting binding affinities is hardly possible [Mooij, 2005]. We focused on the XRAY/CAC intersection data set for further docking studies. Comparing the lists of intersections for the XRAY data set shows disappointingly no intersection for the Zn-binder and only benzylic alcohol is found in common for the S<sub>1</sub>'-pocket.

### **Docking**

To select the most appropriate docking tool, we performed docking for the XRAY ligands first using three different docking programs in combination with several scoring functions. XRAY data set duplicate ligands were removed yielding 20 ligands. Splitting those into fragments resulted in 47 fragments.

**Table 5.9** Docking comparison of 20 Xray data set ligands

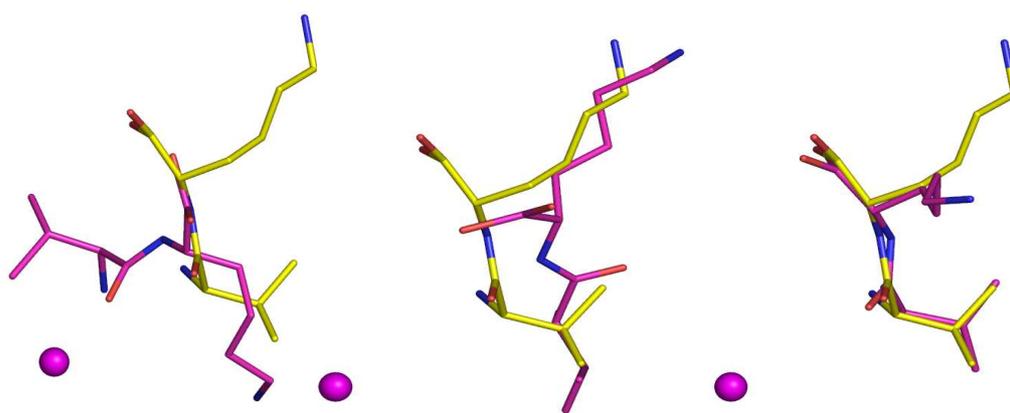
	<b>AutoDock</b>	<b>GOLD</b>	<b>GOLD-Chemscore</b>	<b>FlexX</b>	<b>FlexX DrugScore</b>
<b>Number of best solutions with RMSD &lt; 2Å</b>	20	9	14	16	16
<b>Ø RMSD of rank 1 solution</b>	2.2	4.9	2.9	5.6	5.1
<b>Number of rank 1 solution with RMSD &lt; 2Å</b>	12	7	12	4	5

Rescoring results are only given if better results were obtained.

The docking results clearly show the superior performance of AutoDock. It was not only capable of producing at least one good solution per ligands, but the AutoDock scoring

function was best capable to identify a placement of the fragments that best approximated its orientation as found for the unfragmented parent ligand (Table 5.9 row 2). Rescoring with other scoring functions did not improve the performance generally. This might be due to the final minimization step. Docking modes are minimized into the respective function. Using these modes without adjusted minimization for other scoring functions complicates their scoring. In the case of GOLD-docking/Chemscore-scoring a post-docking minimization with respect to the new potentials is performed. This minimization contributes to the better rescoring performance of Chemscore compared to DrugScore.

Quality assessment of generated docking solutions was based on calculated RMSD between docking mode and crystal structure geometry. However, relying on calculated RMSD values might not be sufficient as a criteria for correct docking. Val-Lys dipeptide docking with AutoDock serves as a good example. While the best RMSD might show the atoms close to their respective position in the crystal structure, clearly the solution given on the right of Figure 5.15 identifies best the correct binding mode.



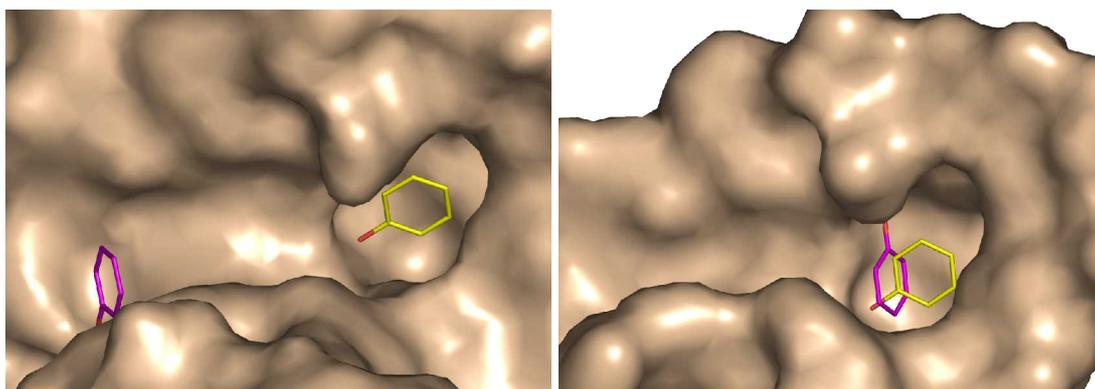
<b>RMSD</b>	2.73	1.01	3.25
<b>Rank</b>	1	47	33
<b>Cluster size</b>	2	1	12

**Figure 5.15** Docking solution of Val-Lys. Ligand conformation as experimentally derived from pdb ID: 2TLX is given with yellow carbon atoms. The docking solution are given with pink carbon atoms. Left: Docking solution selected for rank1 by AutoDock. Center: Docking solution with the best RMSD. Right: Docking solution which identifies the protein-ligand interactions the best. Far away from the binding site, the Lys side chains not not match.

As expected the ligand conformation shown on the left receives the highest score. The docking mode has a strong interaction with the zinc ion and if in fact so, this should contribute significantly to the interaction energy. One possible explanation might be that in the crystal structure, two water molecules or an acetate ion are complexed to the zinc ion and block a direct access, whereas in docking such interaction partners are not regarded and the docking program seeks to find a coordination to the metal ion. The docking solution on the right recognizes all necessary protein-ligand interactions. However, the overall RMSD results in a fairly large value. The terminal lysine side chain which mainly interacts with solvent water molecules is not correctly placed. It should be noted that this solution was found multiple times, in fact for AutoDock this geometry represents the largest cluster of solutions. While not available in the lab, the application of the IBAC (interactions-based accuracy

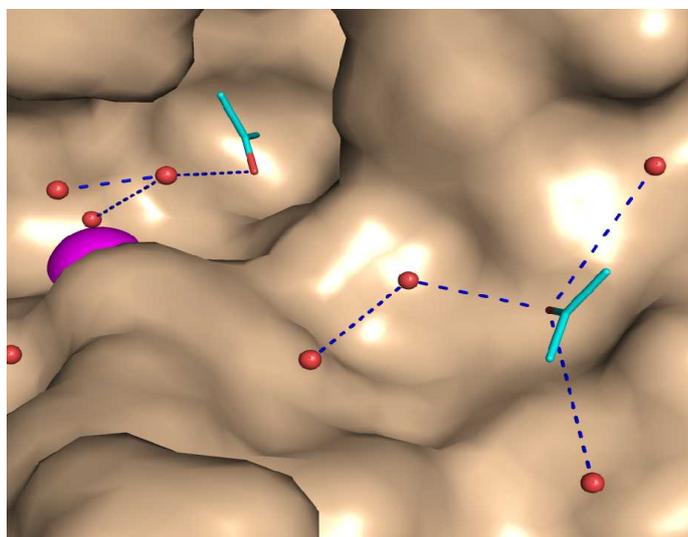
classification) scoring system might serve as a further addition or alternative [Kroemer, 2004]. For IBAC, a two step scoring is applied. First, the ligand position is compared regarding the atom-by-atom match to the crystal structure. In a second step, the docking solution is checked for the presence of key interactions with the protein. Docking poses are thus classified as correct (all interactions present), nearly correct (up to a quarter of the relevant interactions are not present) or incorrect. As an advantage, docking poses which are close by RMSD but inadequate by means of protein-ligand recognition are discarded. In the given Val-Lys example, IBAC would recognize the solution on the right as the best one. Due to these findings, all docking solutions were visually inspected for identified interactions to the zinc ion and possible hydrogen bonds to Arg203.

Pushing the fragment approach to its limit we tried docking six solvent and cosolvent molecules, for which complex crystal structures are available. One apparent problem to those very small ligands is the energetic differentiation between multiply generated binding modes. However, already the generation of docking solutions turned out to be difficult.



**Figure 5.16** Docking the cosolvent to thermolysin is difficult. Two docking solutions (pink carbon atoms) for phenole are compared to the corresponding experimental structure (yellow carbon atoms). Left: AutoDock docking solution with the highest score. Right: Docking solution with the best RMSD.

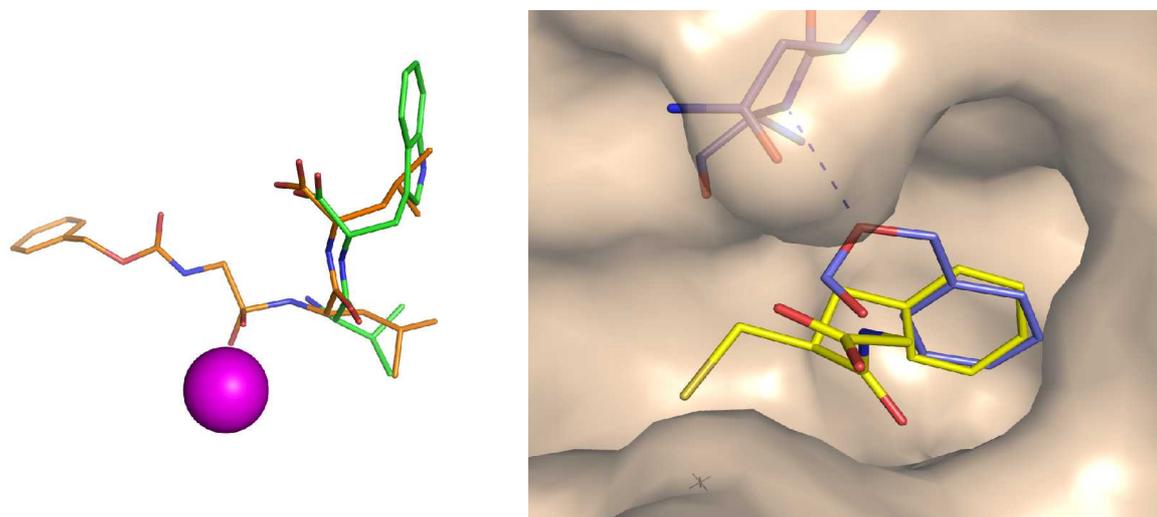
As shown in Figure 5.16, there was no better solution generated than the one shown in the right figure. However, the docking mode shown on the left achieved higher scores. A hydrogen bond is formed which yields more interaction energy than burying the hydrophobic part into the  $S_1'$ -pocket. Analyzing the thermolysin solvent complex structures shows that the solvent molecules are embedded into a network of water molecules that also solvate the binding pocket. While this might be true to some degree for all ligands in general, those interactions strongly influence the placement of the organic solvent binding to the protein.



**Figure 5.17** Water network around a ligand. Shown is pdb ID: 1FJQ, which is thermolysin soaked in 70% acetone. While there are some protein acetone interactions, more are generated between acetone and the adjacent water molecules.

Despite the described shortcomings, docking of complete drug-sized thermolysin ligands worked very well. Most docking algorithms managed to generate suitable ligand conformations. However, the selection of “the best” conformation from multiple geometries for one ligand did not work sufficiently. Further improvement in scoring and ranking is required to enhance the chance of ranking the biologically relevant conformation on rank 1. For a fragment docking reference, the XRAY data set ligands were split into fragments based on the RECAP algorithm and docked to thermolysin. This approach is based on the

assumption that the assembly of drug-size molecules from fragments can be reversed. Babaoglu and Shoichet have demonstrated in an insightful example for  $\beta$ -lactamase, that this assumption does not hold in general. However, while we were not able to obtain complex crystal structures with any of the fragments disassembled from the XRAY references we consulted protein-ligand complex structures binding fragments, which are structurally close to fragments obtained from our RECAP splitting of the known inhibitors [Senda, *to be published*]. Quite convincing results are found for these cases. If a fragment-protein complex structure can be determined the found fragment geometry provides a reasonable estimate of the conformation of this fragment when it is integrated into a larger and more complex ligand. Same functional groups, even in smaller ligands or different ligands, address the same binding site in similar fashion.



**Figure 5.18** Same ligand moieties address binding sites in similar fashion. Left: Superimposition of ligands derived from pdb IDs 3TMN (green carbon atoms) and 5TMN (orange carbon atoms). Val and Leu address  $S_1'$ -pocket similar. The same applies for the peptide backbone. Right: Superimposition of ligand moieties from pdb ID 1ZDP (yellow carbon atoms) and 1KJP (light blue carbon atoms). The strong hydrogen bond between the backbone nitrogen and the ester oxygen forces the ligand to adopt a different binding mode. However, the aromatic ring to address the  $S_1'$ -pocket aligns convincingly.

In a first attempt the docking of the XRAY fragments was performed using three docking programs in combination with several scoring functions. The RsubH (scissile atom set to hydrogen) data set was used. All XRAY fragments were docked into the whole and empty binding site and wrong placements were observed for fragments which bind to the rim of the pocket (S<sub>1</sub>- and S<sub>2</sub>'-pocket) in the parent structure. The strong binding capabilities of the zinc ion and the S<sub>1</sub>' pocket guide this misplacement into the center of the pocket. As a consequence, all but the ZN-binder and the S<sub>1</sub>'-binder were discarded. In total 32 fragments remained to be considered for docking. A detailed list is given in the supporting information.

**Table 5.10** Docking comparison of the 32 X-ray data set fragments

	Auto Dock	AutoDock DrugScore (CSD)	GOLD	GOLD-Chemscore	FlexX	FlexX DrugScore
<b>Number of best solutions with RMSD &lt; 2Å</b>	23	23 (23)	9	15	16	16
<b>Ø RMSD of rank 1 solution</b>	3.28	2.83 (3.49)	6.0	5.6	7.8	7.4
<b>Number of rank 1 solutions with RMSD &lt; 2Å</b>	12	14 (9)	6	7	3	4

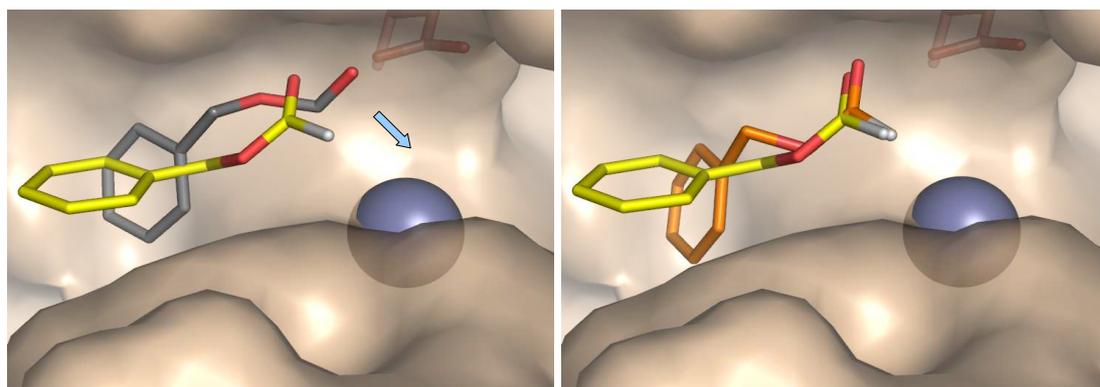
Rescoring results are only given if better results were obtained. Values in parantheses represent DrugScore when taking the CSD instead of the PDB as crystal structure database.

Again the docking performance of AutoDock is superior to the other programs. However, compared to the docking of the entire ligands, the detection of the right binding mode is less successful. But one has to keep in mind that the geometry in the present case is built on a hypothesis and not on an experimental confirmed structure. Still, as the fragments clearly can adopt those conformations, their crystal structure geometry is *one* possibility. Again, the analysis of docking results was not only based on RMSD values but also a visual inspection

has been accomplished. DrugScore performed best in recognizing the binding modes. As we have specific interest in the performance of the scoring functions DrugScore and DrugScore<sup>CSD</sup>, we analyzed their performance in depth. Two major problems were observed:

1. The fragments form interactions to the protein via the atoms at the RECAP splitting site. This way, fragment growth at this location would be impossible.
2. Fragments, supposed to bind to the  $S_1$ -pocket are falsely dragged towards the zinc ion. Especially oxygen atoms placed close to the zinc ion achieve significant scoring and result in false ranking.

In order to avoid potential interaction of RECAP splitting sites with protein residues a new atom type R was introduced. An additional AutoDock energy grid was derived with parameters similar to non aromatic carbon atoms and all attractive interactions set to 0. This way, the RECAP splitting sites of the fragments were considered in a way to use them subsequently for further exploration of the binding site.



**Figure 5.19** Fragment docking performance. The ligand (pdb ID: 4TMN) is cleaved to represent the part addressing the  $S_1$  site. The crystal structure is given in yellow and the white atom indicates the R group. Left: Docking using RsubH (gray) allows the prospective growing site to adopt a geometry which hampers the possibility to grow the fragment (arrow). Right: Docking with repulsive R atom type prevents (orange) interaction at the growing site.

**Table 5.11** Docking / Scoring results for XRAY fragments with either hydrogen or R-groups

	AutoDock DrugScore	AutoDock DrugScore	AutoDock DrugScore <sup>CSD</sup>	AutoDock DrugScore	AutoDock DrugScore	AutoDock DrugScore <sup>CSD</sup>
	RECAP sites substituted to H			RECAP site with atom type R		
<b>Zn-binder (13)</b>						
Number of best solutions with RMSD < 2Å	6	6	6	8	8	8
Ø RMSD of rank 1 solution	3.58	3.37	3.28	3.56	3.60	3.31
Number of rank 1 solution with RMSD < 2Å	2	2	2	0	0	1
<b>S<sub>1</sub>'-pocket (11)</b>						
Number of best solutions with RMSD < 2Å	10	10	10	7	7	7
Ø RMSD of rank 1 solution	3.2	2.92	4.51	3.32	4.13	4.7
Number of rank 1 solution with RMSD < 2Å	6	7	4	6	1	1
<b>both (8)</b>						
Number of best solutions with RMSD < 2Å	7	7	7	7	7	7
Ø RMSD of rank 1 solution	3.06	2.19	2.67	1.57	1.60	2.66
Number of rank 1 solution with RMSD < 2Å	4	5	3	6	7	5

A detailed list of all scores per fragment and scoring function are given in the supporting information.

Introduction of the R-group increased the docking performance for Zn-binder and fragments, addressing both, the zinc ion and the S<sub>1</sub>'-pocket. It reduced placements of ligands addressing the S<sub>1</sub>'-pocket only. Unfortunately, even if better solutions are generated, the scoring function cannot pick them and rank them correctly. There is only a slight improvement in ranking for the fragments addressing Zn<sup>2+</sup> and S<sub>1</sub>'. Subsequently docking has been performed with both systems.

In order to avoid artificially performed contacts between the zinc ion and oxygen atoms of a fragment, the charge at the zinc ions was removed. Secondly, the coordination site at the zinc ion was blocked by a water molecule. Following this strategy better placement of some fragments was obtained assessed by the achieved RMSD values. However, visually inspecting the generated binding modes and found protein-ligand interactions it did not result in an overall improved performance.

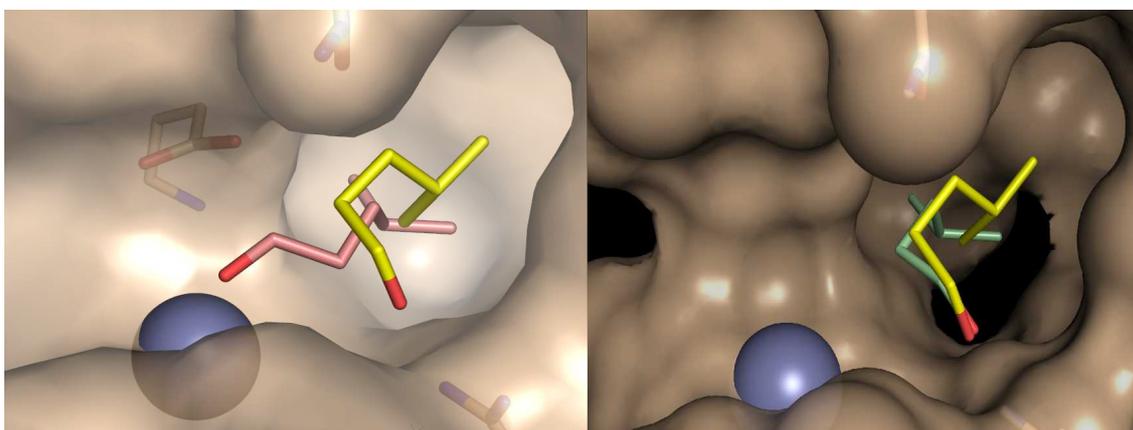
**Table 5.12** S<sub>1</sub>'-pocket fragments docked with zinc ion uncharged

	<b>AutoDock DrugScore</b>	<b>AutoDock DrugScore</b>	<b>AutoDock DrugScore<sub>CSD</sub></b>	<b>AutoDock DrugScore</b>	<b>AutoDock DrugScore</b>	<b>AutoDock DrugScore<sub>CSD</sub></b>
	<b>RECAP sites substituted to H</b>			<b>RECAP site with atom type R</b>		
<b>Number of best solutions with RMSD &lt; 2Å</b>	11	11	11	9	9	9
<b>Ø RMSD of rank 1 solution</b>	3.52	4.36	4.51	3.9	4.47	6.3
<b>Number of rank 1 sol. with RMSD &lt; 2Å</b>	4	4	3	3	4	3

The introduction of the R-atom type did not result in more reasonable docking modes of the S<sub>1</sub>'-pocket fragments either for the straight-forward docking or using the model with the uncharged zinc ion.

**Table 5.13** S<sub>1</sub>'-pocket fragments docked with permanent water at the zinc ion

	AutoDock	AutoDock DrugScore	AutoDock DrugScore <sup>CSD</sup>
	RECAP sites substituted to H		
Number of best solutions with RMSD < 2Å	10	10	10
Ø RMSD of rank 1 solution	2.8	2.57	2.57
Number of rank 1 solution with RMSD < 2Å	9	10	10



**Figure 5.20** Docking of the S<sub>1</sub>'-pocket fragment extracted from the ligand in the pdb structure 1TMN. Upon calculating the AutoDock energy grids, a water molecule occupying the fourth coordination site at the zinc ion was considered. Therefore, the docking geometries did not form undesired zinc contacts. Crystal structure geometry is given by yellow carbon atoms. Left: Rank 1 docking solution (AutoDock) with Zn<sup>2+</sup> without bias (pink carbon atoms). Right: Rank 1 docking solution when docked with permanent water molecules complexing the zinc ion (green carbon atoms).

During the resting state of the protein a water molecule is coordinated to the zinc ion. Considering this water molecule upon docking facilitates the placement of the S<sub>1</sub>'-pocket fragments. However, it restricts the binding site size dramatically and the remaining space is so small that all but one fragment had at least one docking solution with RMSD < 2Å.

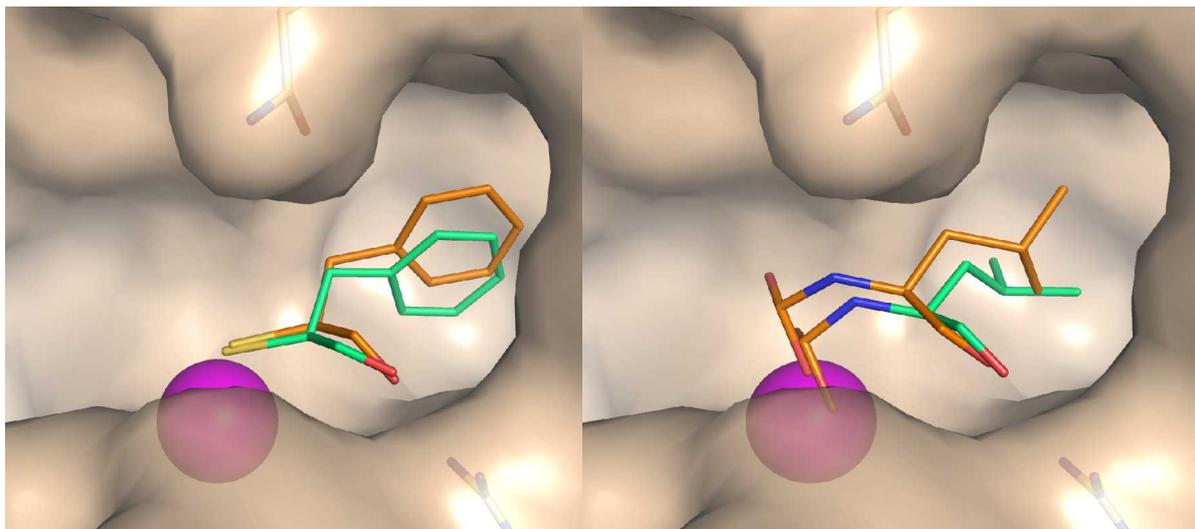
The different binding site modification provided some insightful knowledge about the docking algorithm and enhanced the docking results by means of RMSD matches slightly. However, either restricting accessibility to the only relevant sub-pocket by placing a water molecule or discarding metal interactions are huge and artificial modification. Therefore, they were both not considered for the further screening of the CAC/WDI interaction. One could think of applying these binding site representations as a post docking filter when docking larger data sets of molecules.

An alternative is to filter the generated docking modes according to a pharmacophore hypothesis. This post docking filter was performed with the rigid 3D search algorithm as implemented in Unity. The zinc binding area is represented as a sphere of 3Å around the zinc ion. The part of the fragment which satisfied the Zn-binder pharmacophore in virtual screening must address this sphere with at least one atom to pass the filter. Representing the required depth in the S<sub>1</sub>'-pocket is a  $\gamma$ -carbon atom of the Leucine residue (pdb ID: 1TMN) experimentally determined to address the pocket. The docking mode passes the filter if at least one atom of the S<sub>1</sub>'-binder region penetrates into a sphere with 2Å radius around this atom.

**Table 5.14** Results for passing docking modes through a post dock pharmacophore filter

		AutoDock	DrugScore	DrugScore <sup>CSD</sup>	AutoDock	DrugScore	DrugScore <sup>CSD</sup>
		Before pharmacophore			After pharmacophore		
<b>Zn- binder</b>	<b>Ø of best RMSD overall</b>	1.8	1.8	1.8	1.52	1.52	1.52
	<b>Ø rank of best RMSD</b>	9.6	7.8	9.3	7.6	5.5	5.0
	<b>Ø RMSD of rank1</b>	3.31	2.84	4.29	2.57	2.46	2.67
<b>S<sub>1</sub>'- binder</b>	<b>Ø of best RMSD overall</b>	1.1	1.1	1.1	1.0	1.0	1.0
	<b>Ø rank of best RMSD</b>	14.27	7.0	7.18	7.9	3.9	1.4
	<b>Ø RMSD of rank1</b>	4.13	4.36	4.51	1.1	1.0	0.9

There is no pharmacophore post filter for the fragments which bind to both, the zinc ion and the S<sub>1</sub>'-pocket, because docking them works very well without any further adjustment or filtering. For seven out of eight structures, a solution with RMSD <2 Å was generated and for five even better than RMSD = 1 Å. Probably, these fragments comprise the minimal size and interaction capabilities to enable docking analyses.



**Figure 5.21** Docking solutions for fragments, binding to both the zinc ion and  $S_1'$ -pocket. Docking runs included the R-group at the RECAP splitting site. Crystal structure geometries are shown with orange carbon atoms, docking modes with green carbon atoms. Left: Fragment 1QF2\_f1: best docking solution has  $RMSD = 0.76\text{\AA}$ . Right: Fragment 1TLP\_f2: best docking solution has  $RMSD = 0.96\text{\AA}$ .

After benchmarking this fragment reference data set, the 187 intersection ligands between the WDI and CAC data set were docked with AutoDock and scored with AutoDock, DrugScore and DrugScore<sup>CSD</sup>. The RECAP splitting site was capped by the R-group. Ideally, the selection of virtual screening hits should be solely based on the ranking of the scoring function used to examine the interactions geometry. Facing the shortcomings of current scoring functions and particularly their minor performance with respect to fragments stimulated us to consider a consensus scoring by applying several functions in parallel [Wang, 2003]. We selected the 10% highest scored docking solutions from each scoring function. They were post-filtered as described above. The obtained hits were visually inspected, especially regarding the following characteristics:

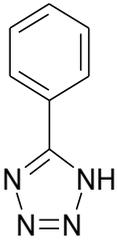
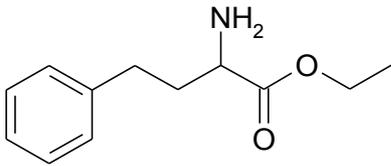
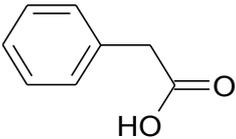
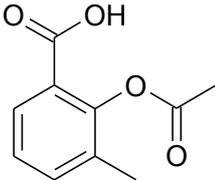
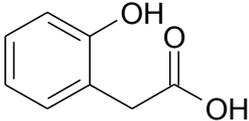
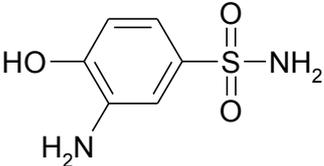
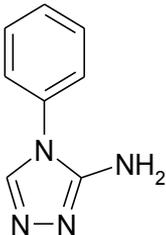
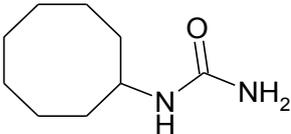
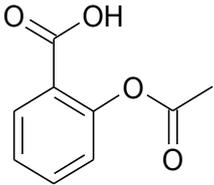
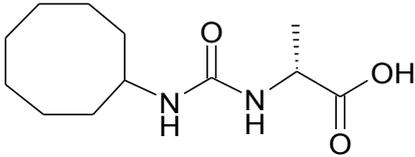
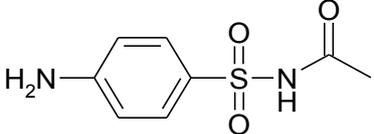
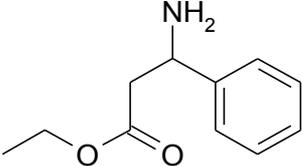
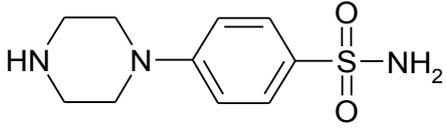
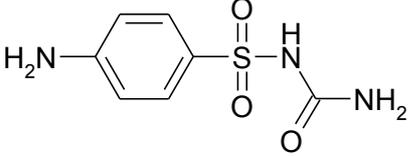
- reliability of the ligand conformation,
- formed protein-ligand interactions to Arg203 and Asn112,
- geometry of the generated metal-ligand interaction,

- penetration into the the  $S_1'$  pocket of buried fragments and
- in case of AutoDock, the amount of docking modes found in each cluster of solutions.

Applying these criteria, finally 14 compounds were selected for experimental validation. They comprise a variety of functional groups (Table 5.15). Validation of virtual screening hits usually starts with the determination of the binding affinity of the hit list compounds. In a second step soaking and co-crystallization experiments are done with candidates showing relevant binding affinities. In the case of fragment-like molecules one expects affinities in the millimolar to micromolar range which require special assay conditions to detect low affinity binders. As mentioned NMR or SPR could serve this purpose. As there was no access to neither of these methods, we decided to test the hit list compounds directly by crystallography. Protein complex crystal structures for low binding fragments can be obtained if soaking is performed at high ligand concentrations.

**Table 5.15** Compounds, selected for experimental binding studies using crystallography

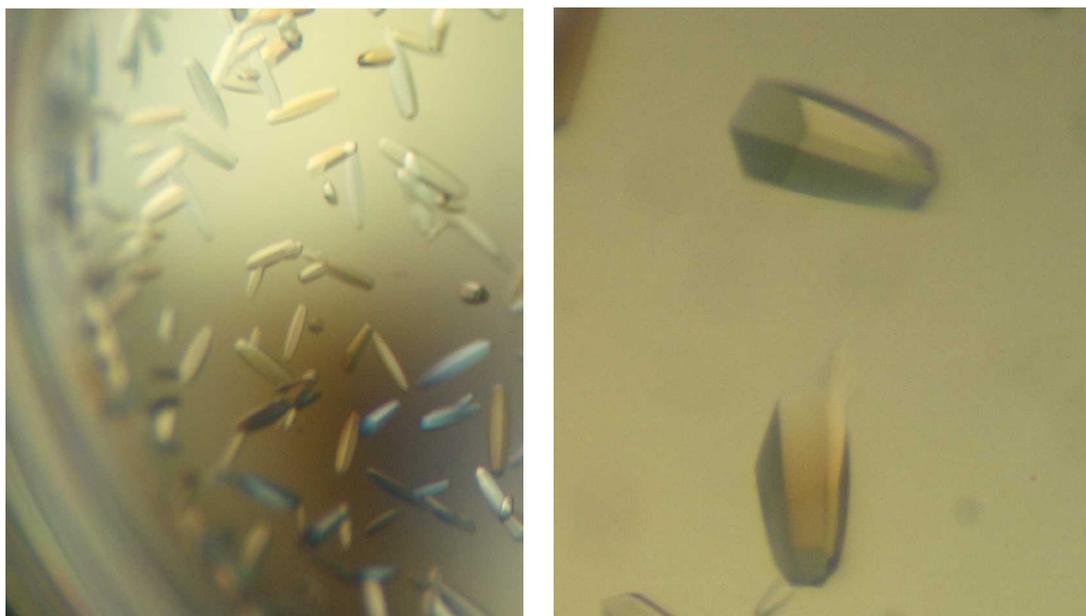
---

1		8	
2		9	
3		10	
4		11	
5		12	
6		13	
7		14	

---

## Crystallization

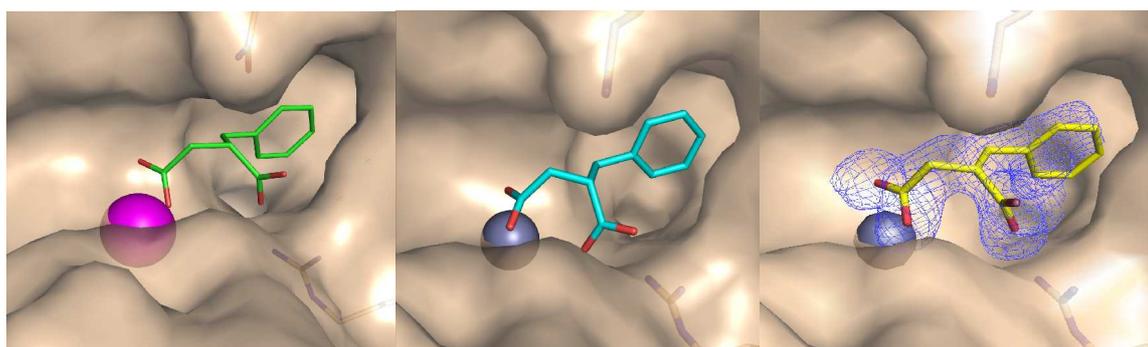
Thermolysin crystals were generated as described above and soaking was started with crystals at least five days old. Thermolysin crystallizes in different shapes but only hexagonal crystals were selected for the soaking experiments.



**Figure 5.22** Thermolysin crystals after 5 days growing. Left: Thermolysin crystallizes in different sizes and shapes. Right: Hexagonal crystals chosen for soaking experiments.

An alternative method to obtain protein-ligand complexes would be co-crystallization. For thermolysin, this method is not applicable: Thermolysin is soluble in DMSO. The crystallization drop initially contains 50% DMSO. Due to pronounced osmotic pressure present in the buffer with high salt concentration, water will diffuse from the mother liquor into the drop. Upon decreasing DMSO concentrations, thermolysin crystallizes and after five days the initial 1  $\mu\text{l}$  drop increases to 300  $\mu\text{l}$ . The resulting ligand concentration would be significantly reduced and especially for fragments with low affinity, detection in the electron density would be difficult. Adding the ligand constantly over five days would be an option but would most likely interfere with crystal growth. According to these concerns, we focused on

soaking and started with ligand concentration of 50 mM if possible or with a saturated ligand solutions. As a decrease in ligand affinity is experienced in DMSO solution, organic solvent concentrations were kept as low as possible. To validate our soaking protocol, we re-determined the complex structure of the fragment-like thermolysin inhibitor benzylsuccinic acid.



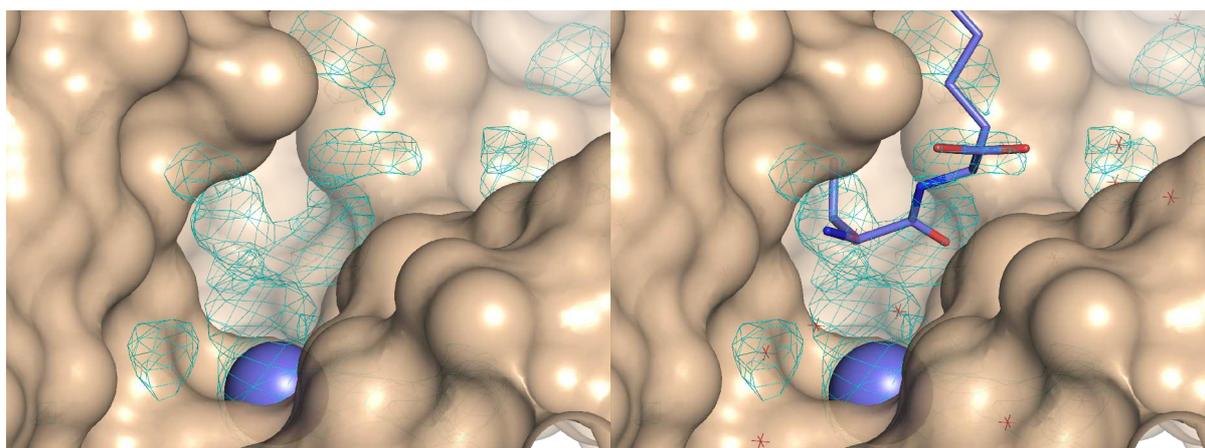
**Figure 5.23** Confirmation of soaking conditions by determining complex structure for benzylsuccinic acid. Left: Complex as determined from pdb ID: 1HYT [Hausrath, 1994]. Center: AutoDock Rank1 docking solution for benzylsuccinic acid. Right:  $2F_o-F_c$  density at  $\sigma$ -level 2.5.

**Table 5.16** Refinement parameter for thermolysin-benzylsuccinic acid complex.

<b>Resolution (Å)</b>	1.9	<b><math>\langle B \rangle_{\text{Protein}}</math> (Å<sup>2</sup>)</b>	14.4
<b>R<sub>sym</sub> (%)</b>	10.0	<b><math>\langle B \rangle_{\text{Ligand}}</math> (Å<sup>2</sup>)</b>	20.1
<b>R<sub>work</sub> (%)</b>	17.2		
<b>R<sub>free</sub> (%)</b>	24.2		

The benzylsuccinic acid example shows that applying the herein described experimental conditions, similar complex structures compared to previously reported ones can be obtained. It also shows correct prediction of the binding mode using AutoDock.

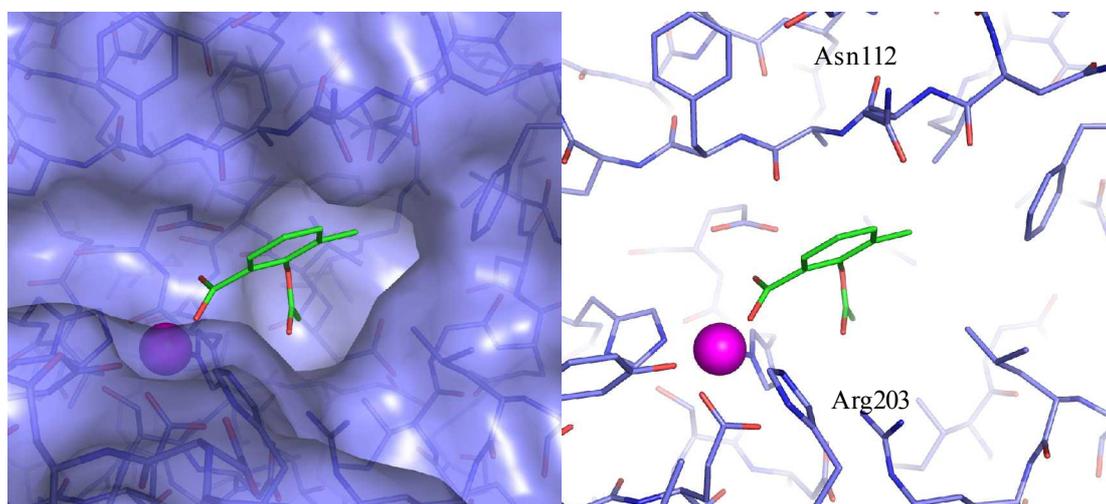
Most of the purchased compounds were not soluble at 50 mM and therefore soaking was performed in saturated ligand solutions. Compounds **3**, **11** and **13** were completely insoluble in the buffer. The at ambient temperature liquid compound **2** was not miscible with the soaking buffer. Thermolysin crystals were transferred directly into the liquid where they immediately dissolved. For compound **8**, the crystals dissolved in the soaking solution. After soaking compounds **1** and **11**, the exposed crystals did not diffract anymore. Upon soaking compound **14**, the soaking drop turned yellow. The crystals still diffracted but after the first refinement,  $F_o-F_c$  density clearly indicated that the Val-Lys dipeptide was still bound to the protein. Furthermore, the Val-Lys dipeptide could also be detected after soaking compounds **6** and **12**.



**Figure 5.24** Detecting Val-Lys dipeptide in  $F_o-F_c$  density. Left:  $F_o-F_c$  density at  $\sigma = 2.0$  as determined after initial refinement of a potential thermolysin-ligand complex. Only protein atoms are considered in the calculated model. Right: Fitting Val-Lys into the density clearly indicates that the tested ligand could not replace the dipeptide from the binding site.

For compound **4**, **5**, **7** and **10**, some electron density was observed which could **not** be explained by the presence of the Val-Lys dipeptide. Obviously, the affinity of the screening hits was strong enough to displace the dipeptide. However, the affinity was too weak to achieve sufficient occupation to determine their complex structure. Additionally to a lack of

affinity, poor solubility might be the reason for only partially successful soaking. Compound **9** (3-methylaspirin) was soaked at 100 mM. A clearly visible difference electron density of the bound fragment could be determined.



**Figure 5.25** 3-Methylaspirin binds to thermolysin. Left: 3-methylaspirin addresses both, the zinc ion and the  $S_1'$ -pocket. Right: Additionally to the zinc complex, the ligand performs a hydrogen bond to Arg203.

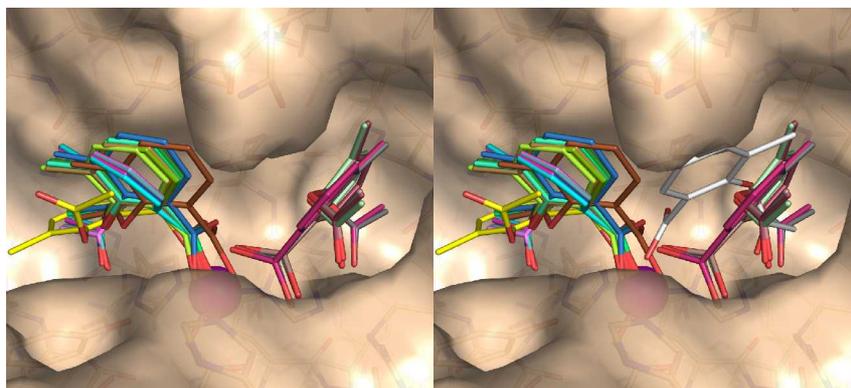
**Table 5.17:** Refinement parameter for thermolysin-3-methylaspirin complex.

<b>Resolution (Å)</b>	1.95 (1.75)	<b><math>\langle B \rangle_{\text{Protein}}</math> (Å<sup>2</sup>)</b>	16.3 (13.6)
<b><math>R_{\text{sym}}</math> (%)</b>	11.6 (8.8)	<b><math>\langle B \rangle_{\text{Ligand}}</math> (Å<sup>2</sup>)</b>	26.9 (18.4)
<b><math>R_{\text{work}}</math> (%)</b>	16.1 (15.7)		
<b><math>R_{\text{free}}</math> (%)</b>	24.7 (21.7)		

Additional crystal structure parameters are given in the supporting information. Values in parentheses represent a second structure determination at ligand concentration = 50 mM and data collection performed at the synchrotron radiation facility BESSY in Berlin.

### **Docking and Scoring 3-Methylaspirin**

Validation of the binding of 3-methylaspirin to thermolysin stimulated us to check how well docking had been capable to generate the correct binding mode. Furthermore, the ranking of docking modes approximating most closely the subsequently determined structure was performed to assess the prediction power of the applied scoring functions. Considering the formerly applied binding site of pdb ID 1TMN, 20 docking solutions for 3-methylaspirin were generated by AutoDock. All address the zinc ion with the carboxylic acid moiety; 16 docking solutions address the S<sub>1</sub>-pocket and only four the S<sub>1</sub>'-pocket.



**Figure 5.26** Docking solutions for 3-methylaspirin. Left: 20 docking solutions generated by AutoDock superimposed into the protein structure of pdb ID: 1TMN. Right: Comparison of all docking geometries and experimentally determined binding mode (white carbon atoms).

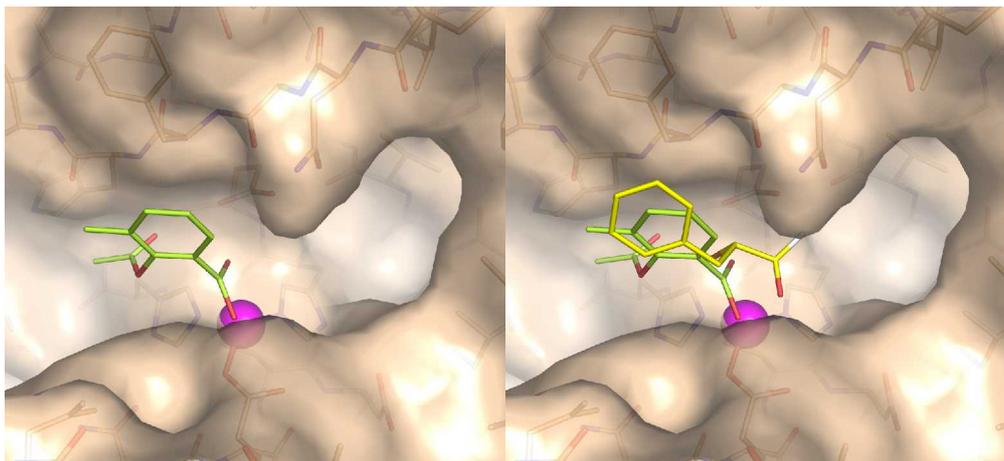
As structural reference and to assess the quality of this docking experiment, the 3-methylaspirin complex structure was superimposed onto the protein chain from pdb ID: 1TMN. The superimposition shows the ligand clashing into the side chain of Asn112 in the binding site used for docking. This demonstrates that without considering side chain flexibility, 3-methylaspirin cannot be correctly docked. The docking mode closest to the obtained crystal structure has an RMSD of 1.89Å.



**Figure 5.27** Thermolysin slightly adapts upon ligand binding. Left: Protein structure as determined with 3-methylaspirin (blue carbon atoms) superimposed to 1TMN (beige) Center: Blow-up of the superimposition of Asn112 highlighting the side chain movement. Right:  $F_o-F_c$  density for 3-methylaspirin (no model refinement for the ligand) superimposed with the 1TMN protein structure. It demonstrates that ligand and protein clash into each other at Asn112 if it would adopt the conformation present in the structure used as reference in docking.

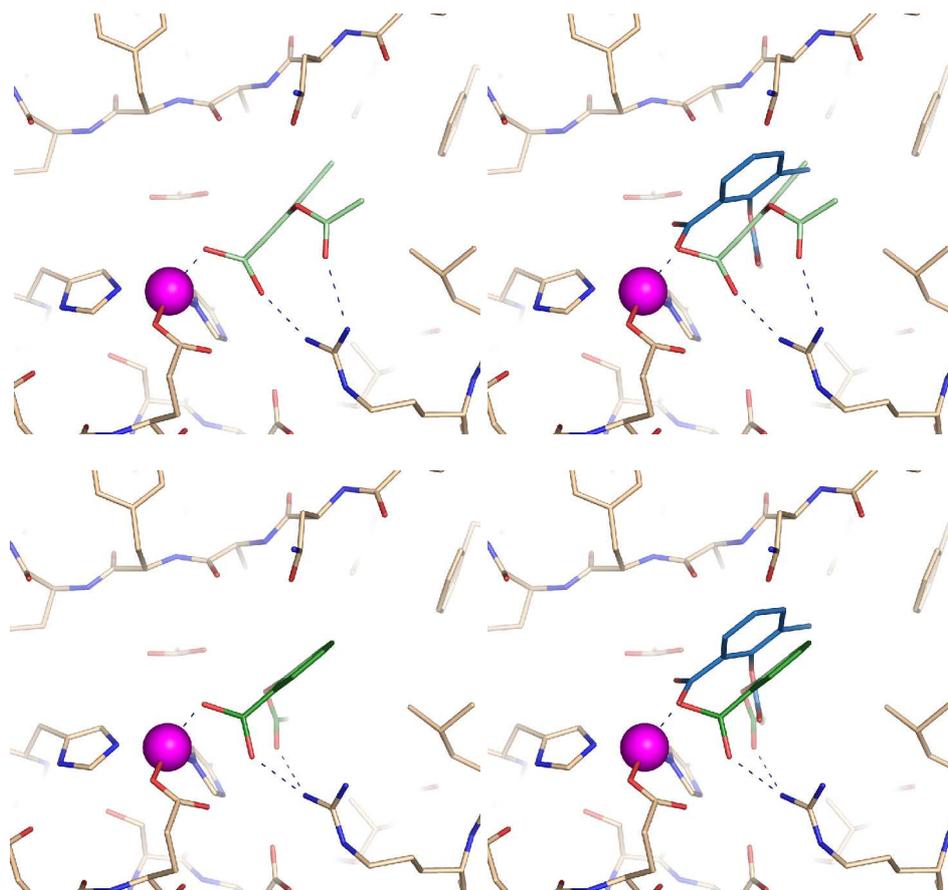
The side chain adaption (Figure 5.28 Center) is analog to the one observed upon binding of benzyloxycarbonyl D-aspartic/glutamic inhibitors in 1KR6, 1KS7. This demonstrated the need for the consideration of side chain flexibility upon docking even for binding sites which have been described as '*rigid*'. Although some initial success has been reported, consideration of target flexibility remains a challenging obstacle in protein-ligand docking. We therefore analyzed the performance of the herein applied docking and scoring methods while docking 3-methylaspirin to the rigid 1TMN binding pocket.

DrugScore and DrugScore<sup>CSD</sup> ranked one of the 16  $S_1$ -pocket geometries best (Figure 5.28). Even though not approximating the orientation found in the crystal structure, this selected binding mode does not appear unreasonable. There are examples in the XRAY ligand reference data set, addressing the  $S_1$ -pocket with an aromatic moiety. Additionally, the acetate carbonyl oxygen atom is placed to interact with the backbone nitrogen of Trp116 (distance = 3.13Å), a position usually occupied by a water oxygen atom.



**Figure 5.28** Scoring 3-methylaspirin with DrugScore. Left: Docking solution ranked highest by both, DrugScore and DrugScore<sup>CSD</sup>. Right: Superimposition of DrugScore predicted binding mode (green carbon atoms) with S<sub>1</sub>-pocket fragment of pdb ID: 1QF0 [Gaucher, 1999] (yellow carbon atoms, R-group white).

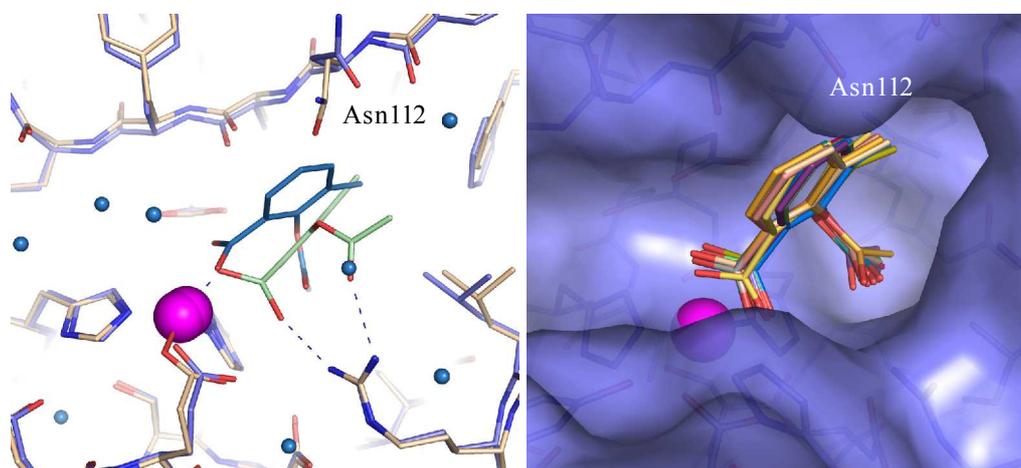
AutoDock and SFC-Score both recognize docking solution as best which address the zinc ion and the S<sub>1</sub>'-pocket. In case of AutoDock, a hydrogen bond between Arg203 and the acetate carbonyl oxygen determines the geometry. It forces the methyl substituent to orient towards the solvent which would require a desolvation penalty. However, the generated docking mode is not unreasonable as the oxygen position of the acetate to Arg203 contact is highly favorable for an hydrogen bonding acceptor. In the crystal structure a water molecule occupies exactly this position (Figure 5.29 and 5.30).



**Figure 5.29** Docking solutions for 3-methylaspirin in the protein structure derived from pdb ID 1TMN (beige protein carbon atoms). Upper row: Left: AutoDock rank 1 docking mode (light green), forming a hydrogen bond to Arg203; Right: Superimposition of the ligand crystal structure geometry (blue) and the AutoDock rank 1 solution. Lower row: Left: The docking solution selected on rank 1 with the SFC-Score scoring function (dark green) recognizes all interactions; Right: Superimposition of experimental ligand crystal structure geometry (blue) and SFC-Score rank 1. The only difference is a shift in the z-axis towards the Asn112 side chain. Superimpositions are calculated based on the protein backbone atoms of the experimentally observed complex structure and the 1TMN protein structure used for docking.

Applying the fragment specific SFC-Score to rank the AutoDock docking solutions, another conformation addressing the zinc ion and the  $S_1'$ -pocket is selected on rank 1 (Figure 5.29, lower row). In this conformation the acetate residue is placed in a way that enables a hydrogen bond to Arg203 as well as buries the methyl moiety of the acetate into the  $S_1'$ -pocket. Additionally, the 3-methyl substituent points favorable into a hydrophobic area.

This binding is more favorable and the closest to the experimental binding mode. All scoring function results with deviations from the experimentally observed binding mode are given in the table below.



**Figure 5.30** Sensitivity of docking solution to protein crystal structure. Left: Rank 1 docking solution from AutoDock (light green) into the protein structure from pdb ID 1TMN. Superimposed (based on protein backbone) is the experimentally determined 3-methylaspirin complex crystal structure (blue carbon atoms, blue spheres represent water molecules). The docked ligand conformation enables a hydrogen bond to Arg203; the acetate oxygen of the AutoDock solution and the crystal structure water directly superimpose. Right: 20 AutoDock runs for 3-methylaspirin back into its **parent** protein crystal structure. All solutions recognize correct binding mode.

In a subsequent attempt, 3-methylaspirin is redocked in the newly determined protein structure. The Asn112 side chain conformation allows to allocate more space in the binding site. 20 out of 20 docking runs adopt a good geometry (Figure 5.30). They all cluster within an RMSD of 1Å. This demonstrates that given the right conditions, the computational prediction is highly accurate. However, the 'right condition' is hard to predict. Potential solutions to accommodate the protein structure is the application of soft docking approaches [Ferrari, 2004] or including a certain degree of protein flexibility explicitly (see chapter 3) into the docking process.

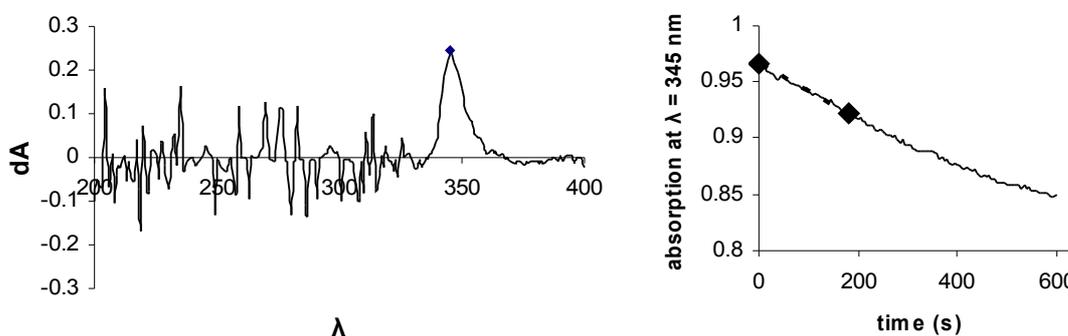
**Table 5.18:** Docking/Scoring results for 3-methylaspirin

	<b>DrugScore</b>	<b>DrugScore<sup>CSD</sup></b>	<b>AutoDock</b>	<b>SFC-Score</b>	<b>SFC-Score</b>	<b>SFC Score</b>
				<b>C-Func</b>	<b>CAS I</b>	<b>CAS II</b>
<b>RMSD</b>	6.00	6.12	3.48	1.89	1.89	1.89
<b>first rank</b>						
<b>pK<sub>i</sub> predicted</b>			5.88	6.94	6.11	5.4

The first crystal structure of the 3-methylaspirin-thermolysin complex was obtained on the in-house device. Curious about the contribution of the methyl group to thermolysin binding additional soakings with aspirin and 3-methylaspirin were performed now at 50 mM. Crystal structures were determined with synchrotron radiation this time. Again, 3-methylaspirin binding could be detected whereas no binding of aspirin could be observed. However, unlike the previous structure, ligand occupancy was only ~80%. Compared to the previous experiment only half the ligand concentration has been applied for soaking. This clearly demonstrates the need for highly soluble ligands considering X-ray crystallography as a screening method. In order to study the influence of a meta substituents on binding, 3-chloroaspirin was synthesized which turned out to be less soluble than 3-methylaspirin. Therefore, soaking was performed in a saturated ligand solution. The  $F_o - F_c$  density map showed no indication for a bound ligand. To determine binding capabilities we tested the two aspirin derivatives and aspirin itself in an photometric affinity assay.

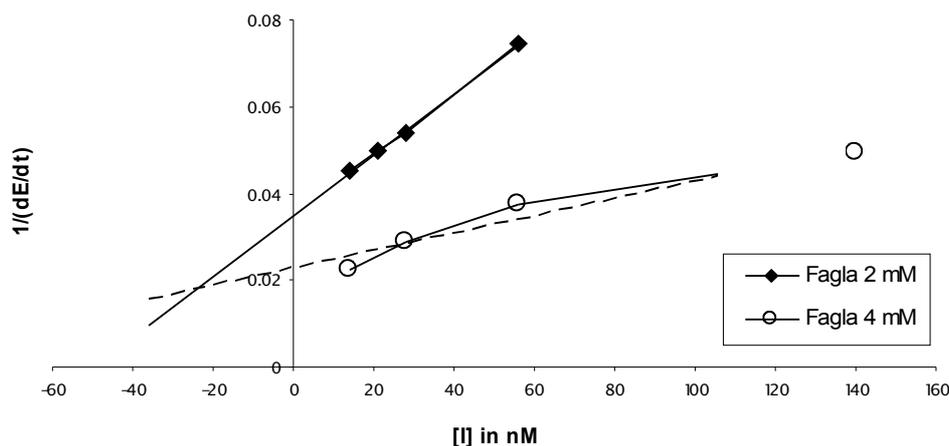
### **Determining the inhibition constant**

Kinetic measurements have been performed using a photometric affinity assay with 3-(2-furylacryloyl)-L-glycyl-L-leucine-amide (Fagla) as substrate. Due to the sensitivity of the protein activity to the presence of DMSO, ligands were dissolved in TRIS-buffer solution. Protein activity was monitored at  $\lambda=345\text{nm}$ .



**Figure 5.31** Detection wavelength for thermolysin kinetics. Left: Difference spectrum from pure Fagla 2 mM and a mixture of Fagla 2 mM and TLN 13 nM after 10 minutes. Right: A mixture of Fagla 2 mM and TLN 13 nM followed over 10 minutes. Highlighted is the time frame for determining the initial velocity.

To validate the applied protocol the inhibition constant of phosphoramidon was re-determined. Phosphoramidon acts as a competitive inhibitor and has been published with an inhibitions constant  $K_i=28\text{ nM}$  [Tronrud, 1986].



**Figure 5.32** Validation of the inhibition constant for phosphoramidon. As competitive binding kinetics apply, the inhibition constant is graphically determined via the Dixon plot [Dixon, 1953], [Dixon, 1972]. Shown data points are averages from double measurements.

Using the Dixon plot and the above described assay setup, the  $K_i$  for phosphoramidon was determined as 27 nM. Subsequently, inhibition constants were determined for aspirin, 3-methylaspirin and 3-chloroaspirin.

**Table 5.19:** Assay results and physicochemical properties for aspirin derivatives

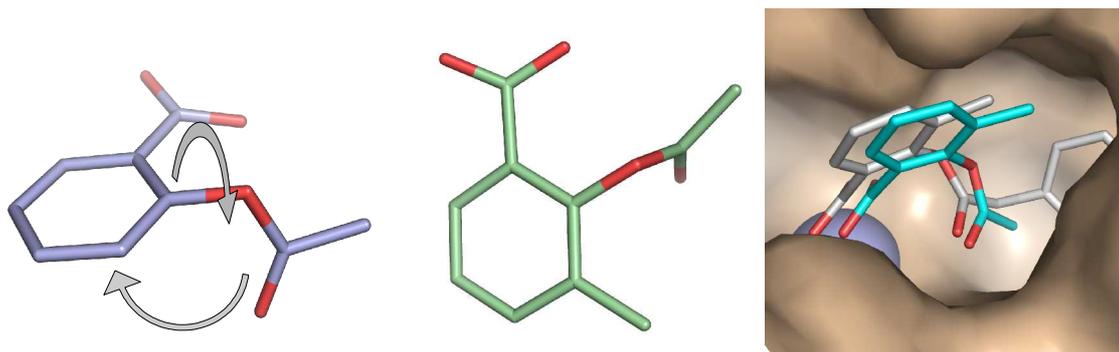
Ligand	Inhibition constant ( $K_i$ )	Molecular weight (Da)	pKa
Aspirin	2.42 mM $\pm$ 0.23	180.04	3.49 <sup>1</sup>
3-methylaspirin	1.73 mM $\pm$ 0.2	194.06	3.48 <sup>1</sup>
3-chloroaspirin	522 $\mu$ M $\pm$ 0.06	214	3.05 <sup>1</sup>

<sup>1</sup>Pka values were estimated using the prediction model as provided online via <http://www.chemaxon.com>

As expected, the binding affinity of the methyl derivative is slightly better than that of aspirin itself. One possible explanation would be that the 3-methyl group fixes a favorable conformation. The rotation about the phenyl carbon - ester oxygen bond in 3-methylaspirin is

restricted compared to unsubstituted aspirin (Figure 5.33, left). However, a conformational search showed that the preferred angle for the acetate moiety- aromatic ring angle is  $90^\circ$  even if the ring is unsubstituted (see supporting information). Thus, both compounds adopt this preferred conformation independent of protein binding. As a result, the ligand conformation prior to binding cannot serve as cause for the different affinities. Another explanation could be the modified acidity. Again, hardly any difference can be expected following the pKa estimations. Therefore, the difference likely arises from the meta substituent and its interactions with the protein. The methyl group binds in the proximity of the hydrophobic residues Leu133, Val192, Val139 and Leu202. It does not penetrate deeply into the pocket. Considering a similar binding mode for aspirin, this position would possibly be occupied by a water molecule. The only substrate free crystal structure of thermolysin (pdb ID 1L3F) shows a water molecule at the corresponding position (see supporting information). Water buried in a hydrophobic area is unfavorable and its replacement explains enhanced affinity. However, for a water replacement, the affinity difference is too subtle. Also in the substrate-free structure, the relative orientation of the two domains that define the active-site cleft differ by a five degrees rotation relative to their positions in the previously studied ligand-bound structures. This rotation makes a direct comparison of water positions difficult. Most likely, the shape of 3-methylaspirin provides the better counterpart to fit opposite the shape of the brink of the hydrophobic pocket (Figure 5.33, right).

3-chloroaspirin binds three times better than 3-methylaspirin. Possibly the decreased solubility prevents soaking at sufficiently high concentration to successfully populate the fragment in crystals. Assuming a similar binding mode, the chlorine atom would address the hydrophobic pocket. Additionally, the meta-chloro substitution increases the acidity of the molecule. Thus, the capability of the carboxylic acid to coordinate the zinc ion is improved which possibly allows for tighter binding.



**Figure 5.33** Analyzing the screening results. Upper Row: Left: Rotatable bond in salicylic acid. Preferred conformation of aspirin as obtained by analyzing the energy in steps of  $30^\circ$ . This rotation is not restricted in aspirin. Right: Conformation of 3-methylaspirin. Rotation around the phenol oxygen bond is sterically hindered. Lower Row: Left: Superimposition of 3-methylaspirin complex crystal structure (blue carbon atoms) and pdb ID 1L3F. Right: A suggested salicylic acid derivative expected to exhibit enhanced affinity towards thermolysin.

Based on the newly discovered lead structure, further inhibitors can be designed. In the 3-methylaspirin complex the methyl substituent points towards the  $S_1'$ -pocket. Larger substitutions could better accommodate the remaining space. This could be achieved by larger ester side chains that could accommodate more efficiently the  $S_1'$ -pocket. Docking the benzylic acid ester reveals (Figure 5.33, right) a promising binding mode.

## **Conclusions**

A strategy for fragment-based screening for metalloproteinase inhibitors is presented which led to the discovery of a new lead structure for optimization. We focused on virtual screening and docking to select a limited number of compounds for further validation using crystallography. Applying filter criteria with respect to chemical properties of fragments, metalloproteinase inhibition and undesired reactivity, data sets of commercially available compounds and fragments derived from the World Drug index were compared. Subsequent docking experiments and geometry assessment with several scoring functions resulted in 14 compounds selected for crystallographic screening.

Based on the determined protein-fragment structure and a subsequent affinity assay we discovered aspirin derivatives as novel thermolysin inhibitors. They inhibit in the low millimolar range. No ligand density could be found for unsubstituted aspirin. The crystal structure additionally revealed an unexpected protein flexibility induced by the discovered inhibitors. The side chain of Asn112 has to move out of space to accommodate the ligand in the binding site.

Interested in the contribution of the meta-substituent to binding capabilities of aspirin, the 3-chloroaspirin analog was synthesized. It inhibits thermolysin with a  $K_i=522 \mu\text{M}$ . Considering the small size of the investigated inhibitors, their affinity and the crystal structure are a starting point to assess whether aspirin derivatives could be promising leads for the development of novel metalloproteinase inhibitors. To improve affinity a possible structural modification would be the synthesis of esters with larger side chains.

Considering the limited number of compounds actually tested experimentally, the herein applied strategy demonstrates the power of computational lead discovery. Docking

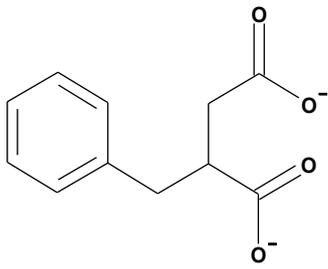
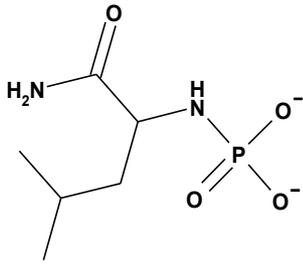
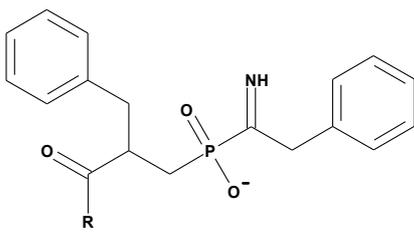
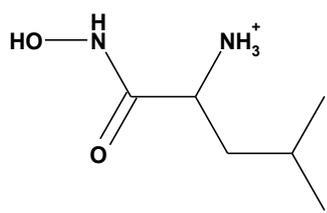
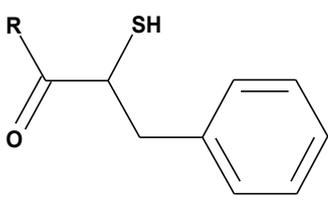
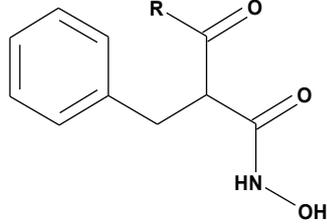
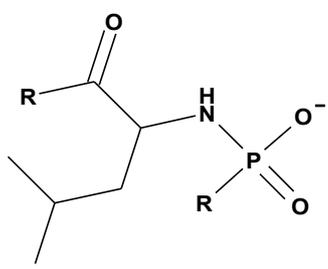
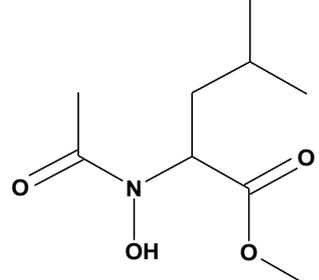
thermolysin ligands illustrates that experimentally observed binding modes can be predicted and scoring functions are capable to recognize important protein-ligand interactions. However, predicting interactions for ligands which only address a small part of a much larger binding site has been proven difficult. In these cases, water interactions and the amount of buried ligand surface is of increasing importance. The growing number of crystallographically studied protein-ligand complexes covering a larger range of functionalities and properties will certainly support our understanding of the physical nature of ligand binding.

This study focused on commercially available fragments which have also been described in drug molecules. Future applications using the developed screening protocol can be based on larger data sets of fragments, exploring better and more exhaustively chemical space of small organic molecules. This way, new surprising lead structures can be identified.

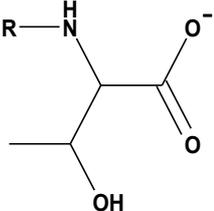
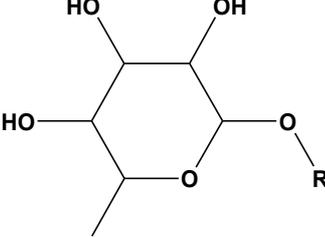
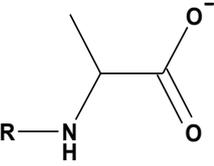
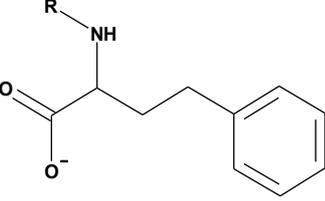
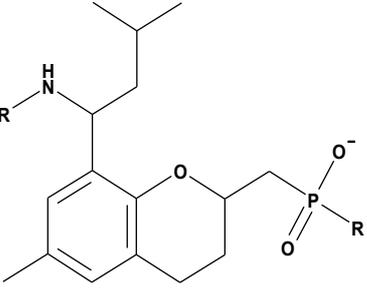
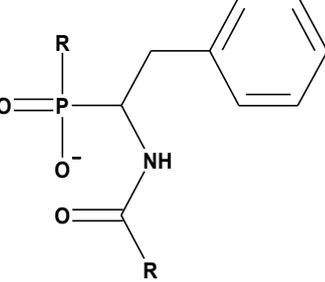
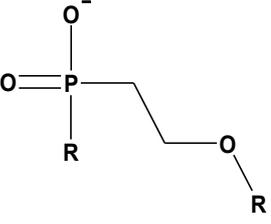
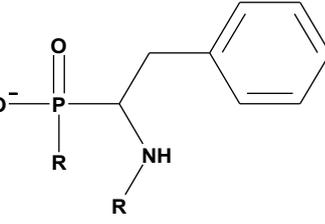
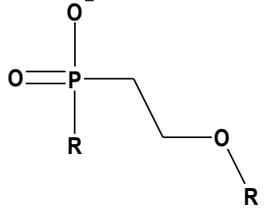
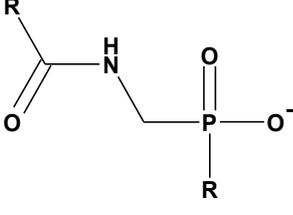
## Supporting information for fragment-based virtual screening

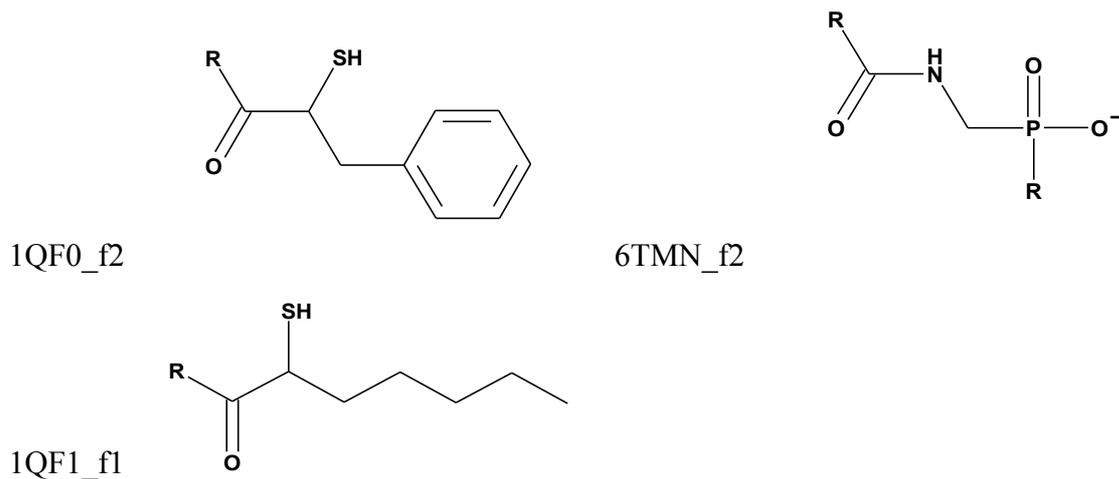
### XRAY Fragment data set

Table s.1: Fragments, binding to both the zinc ion and the S<sub>1</sub>'-pocket.

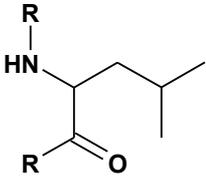
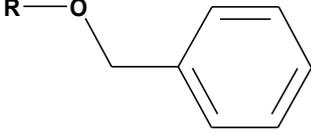
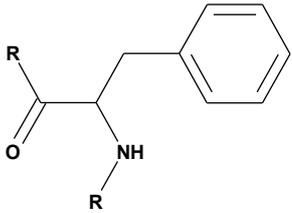
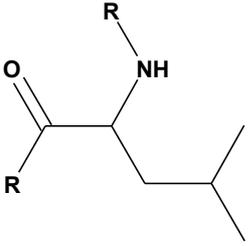
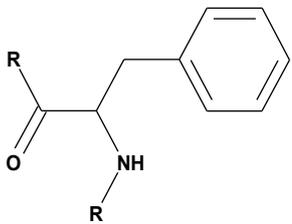
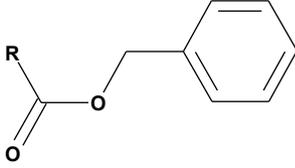
Zinc and S <sub>1</sub> '-pocket binder			
Code	Structure	Code	Structure
1HYT		2TMN	
1NO0_f2		4TLN	
1QF2_f1		5TLN_f1	
1TLP_f2		7TLN	

**Table s.2:** Fragments, binding to the zinc ion only but not to the S<sub>1</sub>'-pocket. In most cases, these fragments address the S<sub>1</sub>-pocket.

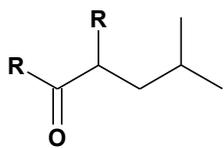
<b>Zinc-binder</b>			
Code	Structure	Code	Structure
1KRO_f1		1TLP_f1	
1KTO_f1		1TMN_f1	
1PE5_f1		4TMN_f2	
1PE7_f1		4TMN_f5	
1PE8_f2		5TMN_f3	



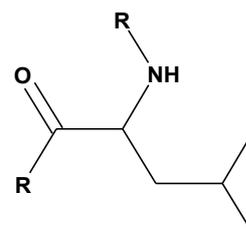
**Table s.3:** Fragments, binding to the S<sub>1</sub>'-pocket only

S <sub>1</sub> '-pocket binder			
Code	Structure	Code	Structure
1PE5_f2		4TMN_f1	
1QF0_f1		4TMN_f3	
1QF1_f2		4TMN_f4	

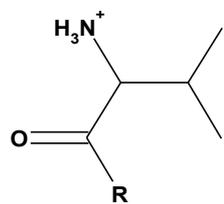
1TMN\_f2



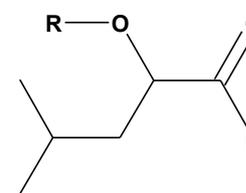
5TMN\_f2



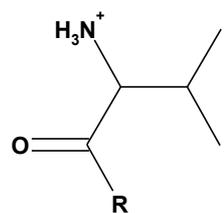
2TLX\_f2



6TMN\_f1



3TMN\_f1



**Table s.4:** Results for docking fragments with RECAP site set to hydrogen

	fragment	AutoDock		DrugScore		DrugScore <sup>CSD</sup>		Best overall
		Run	RMSD (Å)	Run	RMSD (Å)	Run	RMSD (Å)	RMSD (Å)
<b>Zn<sup>2+</sup> and S<sub>I</sub>'</b>	1HYT	14	0.73	8	0.93	12	5.97	0.69
	1NO0_f2	1	2.26	16	2.95	12	4.24	2.26
	1QF2_f1	7	9.87	1	2.15	1	2.15	1.84
	1TLP_f2	4	4.67	7	0.99	7	0.99	0.99
	2TMN	12	1.35	20	6.74	17	2.36	0.99
	4TLN	16	3.07	12	1.33	7	2.83	1.33
	5TLN_f1	8	1.32	16	1.23	6	1.64	1.19
	7TLN	14	1.21	14	1.21	14	1.21	1.21
<b>average</b>			3.06		2.19		2.67	1.31
<b>number of solutions with rank1 &lt;2 Å</b>			4		5		3	7
<b>Zn-binder</b>	1KRO_f1	16	1.87	5	0.97	10	1.01	0.94
	1KTO_f1	6	1.86	7	1.73	7	1.73	1.68
	1PE5_f1	7	2.47	18	3.86	16	2.67	2.47
	1PE7_f1	18	2.27	10	2.97	10	2.97	1.57
	1PE8_f2	4	3.1	10	2.54	1	2.64	2.54
	1QF0_f2	2	5.34	7	5.35	7	5.35	2.84
	1QF1_f1	10	2.98	18	3.57	14	3	2.95
	1TLP_f1	7	5.2	11	2.73	19	5.62	2.73
	1TMN_f1	9	6.81	3	6.21	10	3.46	2.91
	4TMN_f2	14	3.44	14	3.44	7	3.39	1.31

fragment		AutoDock		DrugScore		DrugScore <sup>CSD</sup>		Best overall
		Run	RMSD (Å)	Run	RMSD (Å)	Run	RMSD (Å)	RMSD (Å)
	4TMN_f5	4	5.89	3	5.22	3	5.22	1.4
	5TMN_f3	9	2.99	9	2.99	9	2.99	1.79
	6TMN_f2	3	2.32	15	2.16	12	2.55	2.1
<b>average</b>			3.58		3.36		3.28	2.09
<b>number of solutions with rank1 &lt;2 Å</b>			2		2		2	6
<b>S<sub>1</sub>'-binder</b>	1PE5_f2	16	4.69	15	1.43	10	1.37	1.33
	1QF0_f1	14	7.85	10	1.98	19	11.11	1.05
	1QF1_f2	10	5	17	2.63	17	2.63	0.99
	1TM_f2	17	1.89	10	1.7	12	9.56	1.1
	2TLX_f2	16	1.98	12	1.49	6	2.01	0.93
	3TMN_f1	1	0.93	8	0.92	4	0.9	0.9
	4TMN_f1	5	3.45	20	8.47	20	8.47	2.93
	4TMN_f3	10	1.14	18	0.97	18	0.97	0.97
	4TMN_f4	19	1.61	12	4.33	12	4.33	1.54
	5TMN_f2	13	4.84	9	1.62	20	1.7	0.9
	6TMN_f1	3	1.75	17	6.59	17	6.59	1.23
<b>average</b>			3.2		2.92		4.51	1.26
<b>number of solutions with rank1 &lt;2 Å</b>			6		7		4	10

**Table s.5:** Results for docking fragments with RECAP site set to R-group

	fragment	AutoDock		DrugScore		DrugScore <sup>CSD</sup>		best overall
		Run	RMSD (Å)	Run	RMSD (Å)	Run	RMSD (Å)	RMSD (Å)
<b>Zn<sup>2+</sup> and S<sub>1</sub>'</b>	1HYT	14	0.73	4	0.87	16	0.77	0.67
	1NO0_f2	11	4.5	11	4.5	11	4.5	2.18
	1QF2_f1	15	0.91	3	0.76	2	1.51	0.76
	1TLP_f2	14	1.06	10	1	10	1	0.96
	2TMN	20	0.96	9	1	11	0.91	0.86
	4TLN	9	2.11	7	1.85	7	1.85	1.45
	5TLN_f1	9	1.28	7	1.57	5	7.04	0.72
	7TLN_	9	1.04	5	1.26	3	3.72	1.04
<b>average</b>			1.57		1.60		2.66	1.08
<b>number of solutions with rank1 &lt;2 Å</b>			6		7		5	7
<b>Zn-binder</b>	1KRO_f1	15	3.09	10	2.54	9	3.01	2.38
	1KTO_f1	11	3.86	19	2.26	19	2.26	2.26
	1PE5_f1	17	2.49	18	3.16	13	2.34	1.04
	1PE7_f1	5	2.91	9	3.06	13	2.82	1.66
	1PE8_f2	2	2.97	15	3.09	3	2.87	1.39
	1QF0_f2	19	6.19	8	6.34	20	0.95	0.94
	1QF1_f1	14	4.1	10	4.26	14	4.1	2.9
	1TLP_f1	11	6.14	16	3.73	15	6.04	3.72
	1TMN_f1	12	5.03	9	5.98	9	5.98	0.64
	4TMN_f2	10	0.78	3	4.47	3	4.47	0.69
	4TMN_f5	8	3.43	19	3.56	19	3.56	0.73
	5TMN_f3	6	2.59	18	2.14	18	2.14	1.83

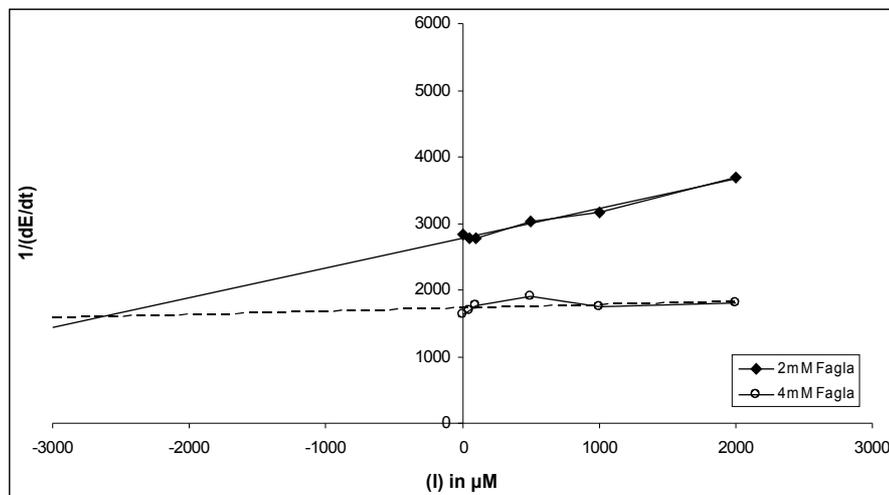
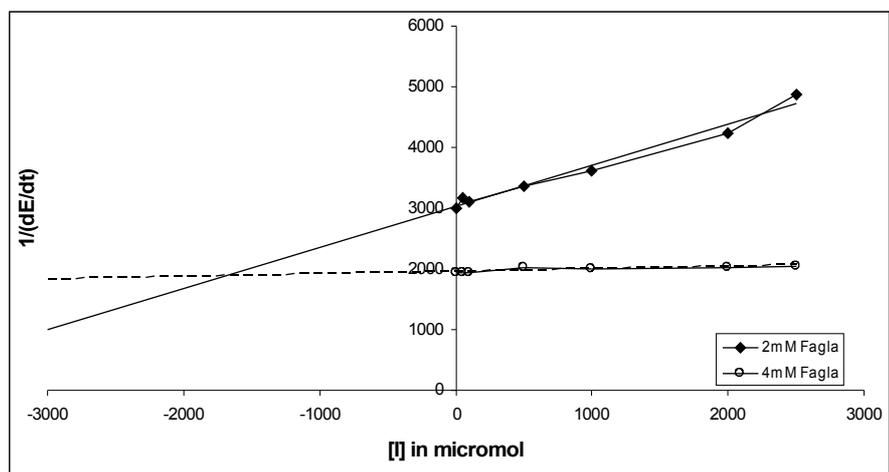
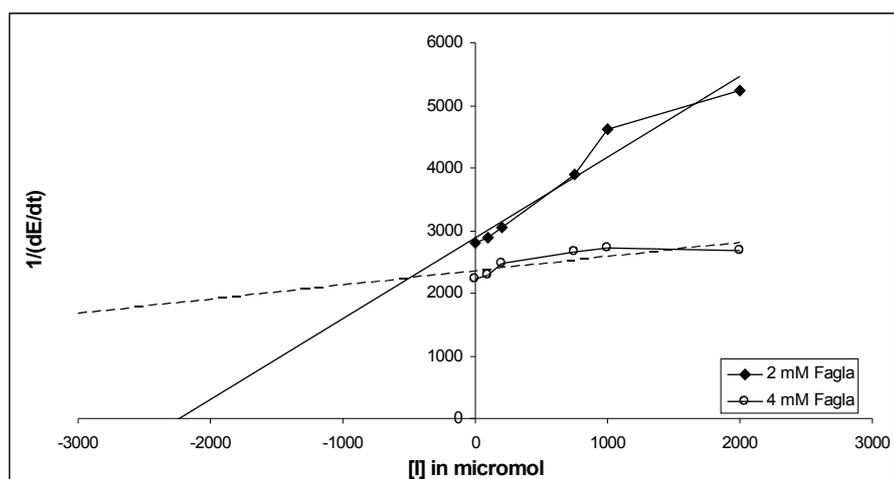
	<b>fragment</b>	<b>AutoDock</b>		<b>DrugScore</b>		<b>DrugScore<sup>CSD</sup></b>		<b>best overall</b>
		Run	RMSD (Å)	Run	RMSD (Å)	Run	RMSD (Å)	RMSD (Å)
	6TMN_f2	3	2.7	8	2.3	8	2.3	2.3
	<b>average</b>		3.56		3.60		3.30	1.73
	<b>number of solutions with rank1 &lt;2 Å</b>		0		0		1	8
<b>S<sub>1</sub>'-binder</b>	1PE5_f2	9	5.41	13	5.24	14	3.75	2.31
	1QF0_f1	3	10.17	12	2.79	18	10.49	2.25
	1QF1_f2	11	0.86	8	2.59	8	2.59	0.86
	1TMN_f2	13	1.56	8	2.45	8	2.45	1.55
	2TLX_f2	10	3.54	14	1.34	14	1.34	1.34
	3TMN_f1	3	0.93	20	3.48	20	3.48	0.89
	4TMN_f1	19	3.58	20	3.71	20	3.71	3.4
	4TMN_f3	12	5.56	8	4.97	6	3.87	2.34
	4TMN_f4	1	1.64	13	7.99	13	7.99	1.53
	5TMN_f2	12	1.79	10	2.81	1	3.92	1.72
	6TMN_f1	16	1.51	3	8.06	3	8.06	1.48
	<b>average</b>		3.32		4.13		4.70	1.79
	<b>number of solutions with rank1 &lt;2 Å</b>		6		1		1	7

**Table s.6:** Results for docking  $S_1'$ -pocket fragments with RECAP site set to R-group and zinc ion uncharged

	fragment	AutoDock		DrugScore		DrugScore <sup>CSD</sup>		Best overall
		Run	RMSD (Å)	Run	RMSD (Å)	Run	RMSD (Å)	RMSD (Å)
<b><math>S_1'</math>- binder</b>	1PE5_f2	7	2.61	18	7.86	18	7.86	2.54
	1QF0_f1	20	10.15	8	1.54	18	10.28	1.49
	1QF1_f2	9	0.85	9	0.85	1	9.99	0.85
	1TMN_f2	11	1.59	4	2.42	19	1.63	1.52
	2TLX_f2	17	4.93	15	0.71	15	0.71	0.71
	3TMN_f1	5	2.31	11	0.6	11	0.6	0.59
	4TMN_f1	12	9.77	2	5.28	4	9.58	1.52
	4TMN_f3	6	2.66	17	8.25	7	8.5	2.41
	4TMN_f4	13	3.72	15	5.57	10	3.71	1.43
	5TMN_f2	11	2.72	2	8.25	6	8.33	1.78
	6TMN_f1	10	1.61	17	7.89	20	8.05	1.46
<b>average</b>			3.908		4.50		6.30	1.48
<b>number of solutions with rank1 &lt;2 Å</b>			3		4		3	9

**Table s.7:** Results for docking  $S_1'$ -pocket fragments with RECAP site set to hydrogen and zinc ion uncharged

	fragment	AutoDock		DrugScore		DrugScore <sup>CSD</sup>		Best overall
		Run	RMSD (Å)	Run	RMSD (Å)	Run	RMSD (Å)	RMSD (Å)
<b><math>S_1'</math>- binder</b>	1PE5_f2	4	1.19	5	9.54	5	9.54	1.02
	1QF0_f1	13	5.31	6	1.16	2	10.87	1.01
	1QF1_f2	5	5.01	15	1.02	15	1.02	1.02
	1TMN_f2	14	1.47	9	3.05	9	3.05	1.11
	2TLX_f2	15	3.9	13	0.62	12	0.62	0.62
	3TMN_f1	1	2.32	18	2.42	18	2.42	0.74
	4TMN_f1	11	3.38	2	10.79	18	5.73	1.83
	4TMN_f3	6	1.14	1	0.99	17	0.97	0.97
	4TMN_f4	16	3.31	12	5.18	9	3.25	1.48
	5TMN_f2	15	1.3	20	2.72	20	2.72	1.2
	6TMN_f1	1	10.43	1	10.43	19	9.47	1.03
<b>average</b>			3.52		4.35		4.51	1.10
<b>number of solutions with rank1 &lt;2 Å</b>				4		4		3

**Dixon plots for aspirin derivatives****aspirin****3-methylaspirin****3-chloroaspirin**

**SFC Scoring function docking scores****Table s.8:** Statistics for the derived SFC-Score fragment-based functions

<b>Function CAS I</b>		<b>Function CAS II</b>	
<b>N_hb</b>	+0.426	<b>N_hb</b>	+0.484
<b>RMSScore</b>	+0.793	<b>TotBurSurf</b>	+0.013
<b>TotBurSurf</b>	+0.0114		
<b>HydBurSurf</b>	+0.00297		
<b>Statistics</b>			
<b>n</b>	130	<b>n</b>	137
<b>r</b>	0.81	<b>r</b>	0.745
<b>s</b>	0.973	<b>s</b>	1.142
<b>F<sup>a)</sup></b>	59.762	<b>F</b>	55.267
<b>q<sup>2</sup><sup>b)</sup></b>	0.627	<b>q<sup>2</sup></b>	0.525
<b>S-PRESS<sup>c,d)</sup></b>	1.015	<b>S-PRESS</b>	1.180

$n$  = number of data points,  $r$  = correlation coefficient;  $s$  = standard deviation a) Fisher's F-value. b)  $q^2 = 1 - PRESS/SSD$  as obtained by "leave-one-out" cross-validation.  $PRESS$  equals to the sum of squared differences between predicted and experimentally determined binding affinities,  $SSD$  is the sum of the squared differences between experimentally determined binding affinities and the mean of the training set binding affinities. c) In logarithmic units. d)  $s_{PRESS} = \sqrt{PRESS/(n - h - 1)}$  as obtained by "leave-one-out" cross-validation.  $h$  is the number of components. For detailed information of the descriptors see in Sotriffer *et al.*

**Table s.9:** SFC-function scores for 3-methylaspirin to thermolysin

<b>Docking</b>	<b>C-Function</b>			<b>CAS I Function</b>			<b>CAS II Function</b>		
	<b>Score</b>	<b>RMS to xray</b>	<b>RMS to run3</b>	<b>Score</b>	<b>RMS to xray</b>	<b>RMS to run3</b>	<b>Score</b>	<b>RMS to xray</b>	<b>RMS to run3</b>
<b>Run 1</b>	6.69	3.43	3.33	5.69	3.43	3.33	5.03	3.43	3.33
<b>Run 2</b>	6.7	3.42	3.33	5.68	3.42	3.33	5.00	3.42	3.33
<b>Run 3</b>	6.94	1.78	0.00	6.11	1.78	0.00	5.40	1.78	0.00
<b>Run 4</b>	6.62	6.24	7.05	4.57	6.24	7.05	3.86	6.24	7.05
<b>Run 5</b>	6.92	5.61	6.38	4.72	5.61	6.38	4.09	5.61	6.38
<b>Run 6</b>	6.7	6.15	6.96	4.62	6.15	6.96	3.96	6.15	6.96
<b>Run 7</b>	6.53	6.22	7.03	4.31	6.22	7.03	3.57	6.22	7.03
<b>Run 8</b>	6.6	6.25	7.07	4.53	6.25	7.07	3.83	6.25	7.07
<b>Run 9</b>	6.00	7.85	8.75	3.84	7.85	8.75	3.16	7.85	8.75
<b>Run 10</b>	6.35	6.18	6.97	4.18	6.18	6.97	3.54	6.18	6.97
<b>Run 11</b>	6.48	5.93	6.76	4.49	5.93	6.76	3.87	5.93	6.76
<b>Run 12</b>	6.94	1.78	0.14	5.86	1.78	0.14	5.12	1.78	0.14
<b>Run 13</b>	6.5	6.33	7.14	4.55	6.33	7.14	3.90	6.33	7.14
<b>Run 14</b>	6.55	6.22	7.04	4.3	6.22	7.04	3.62	6.22	7.04
<b>Run 15</b>	6.63	3.36	3.30	5.55	3.36	3.3	4.88	3.36	3.3
<b>Run 16</b>	6.47	6.17	6.98	4.52	6.17	6.98	3.91	6.17	6.98
<b>Run 17</b>	6.55	6.16	6.97	4.85	6.16	6.97	4.24	6.16	6.97
<b>Run 18</b>	6.35	6.15	6.95	4.22	6.15	6.95	3.57	6.15	6.95
<b>Run 19</b>	6.64	6.27	7.07	4.57	6.27	7.07	3.88	6.27	7.07
<b>Run 20</b>	6.52	6.22	7.00	4.64	6.22	7.00	3.98	6.22	7.00

***Databases and Programs used throughout this work***

Asinex:	<a href="http://www.asinex.com/libraries.html">http://www.asinex.com/libraries.html</a>
IBS:	InterBioScreen <a href="http://cds.dl.ac.uk/cds/datasets/orgchem/isis/ibscreen.html">http://cds.dl.ac.uk/cds/datasets/orgchem/isis/ibscreen.html</a>
Specs:	<a href="http://www.specs.net/">http://www.specs.net/</a>
Leadquest:	<a href="http://www.leadquest.com/">http://www.leadquest.com/</a>
Maybridge	<a href="http://www.maybridge.com/default.aspx">http://www.maybridge.com/default.aspx</a>
Sigma:	<a href="http://www.sigmaaldrich.com/">http://www.sigmaaldrich.com/</a>
World Drug Index	<a href="http://scientific.thomson.com/products/wdi/">http://scientific.thomson.com/products/wdi/</a>
Unity	<a href="http://www.tripos.com">http://www.tripos.com</a> (Version 4.0)
Sybyl	<a href="http://www.tripos.com">http://www.tripos.com</a> (Version 7.0)
ReliBase+	<a href="http://www.ccdc.cam.ac.uk/products/life_sciences/relibase/">http://www.ccdc.cam.ac.uk/products/life_sciences/relibase/</a> (Version 2.0)
Pymol	<a href="http://pymol.sourceforge.net/">http://pymol.sourceforge.net/</a> (Version 0.98)
FlexX	<a href="http://www.biosolveit.de/FlexX/">http://www.biosolveit.de/FlexX/</a> (Version 1.13)
AutoDock	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a> (Version 3.0)
GOLD	<a href="http://www.ccdc.cam.ac.uk/products/life_sciences/gold/">http://www.ccdc.cam.ac.uk/products/life_sciences/gold/</a> (Version 3.0)
SFC-Score	<a href="http://www.agklebe.de">http://www.agklebe.de</a> (Version 1.0)
CORINA	<a href="http://www.molecular-networks.com/software/corina/index.html">http://www.molecular-networks.com/software/corina/index.html</a> (Version 2.6)
DrugScore	<a href="http://www.agklebe.de">http://www.agklebe.de</a> (Version 1.2)

**PDB entries****Table s.10:** Protein data bank entries for thermolysin used in this work.

<b>pdb ID</b>	<b>Description</b>	<b>Resolution</b>	<b>Authors</b>
1FJ3	thermolysin (50% acetone soaked)	2.0	English, A.C., Groom, C.R., Hubbard, R.E.
1FJO	thermolysin (60% acetone soaked)	2.00	English, A.C., Groom, C.R., Hubbard, R.E.
1FJQ	thermolysin (70% acetone soaked)	1.70	English, A.C., Groom, C.R., Hubbard, R.E.
1FJT	thermolysin (50% acetonitrile soaked)	2.20	English, A.C., Groom, C.R., Hubbard, R.E.
1FJU	thermolysin (80% acetonitrile soaked)	2.00	English, A.C., Groom, C.R., Hubbard, R.E.
1FJV	thermolysin (60% acetonitrile soaked)	2.00	English, A.C., Groom, C.R., Hubbard, R.E.
1FJW	thermolysin (50 mM phenol soaked)	1.90	English, A.C., Groom, C.R., Hubbard, R.E.
1GXW	thermolysin in presence of potassium thiocyanate	2.18	Gaucher, J.F., Selkti, M., Prange, T., Tomas, A.
1HYT	benzylsuccinis acid with thermolysin	1.70	Hausrath, A.C., Matthews, B.W.
1KEI	thermolysin (substrate-free)	1.60	Senda, M., Senda, T., Kidokoro, S.
1KJO	thermolysin complexed with Z-L-threonine (benzyloxycarbonyl-L-threonine)	1.60	Senda, M., Senda, T., Kidokoro, S.
1KJP	thermolysin with Z-L-Glutamic acid (benzyloxycarbonyl-L-glutamic acid)	1.60	Senda, M., Senda, T., Kidokoro, S.
1KKK	thermolysin Z-L-aspartic acid (benzyloxycarbonyl-L-aspartic acid)	1.60	Senda, M., Senda, T., Kidokoro, S.

<b>pdb ID</b>	<b>Description</b>	<b>Resolution</b>	<b>Authors</b>
1KL6	thermolysin with Z-L-alanine (benzyloxycarbonyl-L-alanine)	1.80	Senda, M., Senda, T., Kidokoro,
1KR6	thermolysin with Z-D-glutamic acid (benzyloxycarbonyl-D-glutamic acid)	1.80	Senda, M., Senda, T., Kidokoro, S.
1KRO	thermolysin with Z-D-threonine (benzyloxycarbonyl-D-threonine)	1.70	Senda, M., Senda, T., Kidokoro, S.
1KS7	thermolysin with Z-D-aspartic acid (benzyloxycarbonyl-D-aspartic acid)	1.70	Senda, M., Senda, T., Kidokoro, S.
1KTO	thermolysin with Z-D-alanine (benzyloxycarbonyl-D-alanine)	1.90	Senda, M., Senda, T., Kidokoro, S.
1L3F	thermolysin substrate has an Open Conformation	2.30	Hausrath, A.C., Matthews, B.W.
1LNA	different metals in thermolysin	1.90	Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W.
1LNB	different metals in thermolysin	1.80	Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W.
1LNC	different metals in thermolysin	1.80	Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W.
1LND	different metals in thermolysin	1.70	Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W.
1LNE	different metals in thermolysin	1.70	Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W.
1LNF	different metals in thermolysin	1.70	Holland, D.R., Matthews, B.W.
1OS0	thermolysin with alpha-amino phosphinic inhibitop	2.10	Selkti, M., Tomas, A., Prange, T.
1PE5	thermolysin with tricyclic inhibitor	1.70	Juers, D., Holland, D., Morgan, B.P., Bartlett, P.A., Matthews, B.W.
1PE7	thermolysin with bicyclic inhibitor	1.82	Juers, D., Yusuff, N., Bartlett, P.A., Matthews, B.W.
1PE8	thermolysin with monocyclic inhibitor	1.80	Juers, D., Pyun, H.-J., Bartlett, P.A., Matthews, B.W.

<b>pdb ID</b>	<b>Description</b>	<b>Resolution</b>	<b>Authors</b>
1QF0	thermolysin with (2-sulphanyl-3-phenylpropanoyl)-phe-tyr. Parameters for Zn-bidentation	2.20	Gaucher, J.-F., Selkti, M., Tiraboschi, G., Prange, T., Roques, B.P., Tomas, A., Fournie-Zaluski, M.C.
1QF1	thermolysin thermolysin with (2-sulphanylheptanoyl)-phe-ala parameters for Zn-bidentation	2.00	Gaucher, J.-F., Selkti, M., Tiraboschi, G., Prange, T., Roques, B.P., Tomas, A., Fournie-Zaluski, M.C.
1QF2	thermolysin with (2-sulphanyl-3-phenylpropanoyl)-gly-(5-phenylproline). Parameters for Zn-monodentation	2.06	Gaucher, J.-F., Selkti, M., Tiraboschi, G., Prange, T., Roques, B.P., Tomas, A., Fournie-Zaluski, M.C.
1THL	Thermolysin complexed with a novel glutaramide derivative, n-(1-(2(r,s)-carboxy-4-phenylbutyl)cyclopentylcarbonyl)-(s)-tryptophan	1.70	Holland, D.R., Matthews, B.W.
1TLI	thermolysin (2% isopropanol soaked crystals)	2.05	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
1TLP	phosphoramidates as inhibitors and transition-state analogs of thermolysin	2.30	Tronrud, D.E., Monzingo, A.F., Matthews, B.W.
1TLX	thermolysin (native)	2.10	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
1TMN	N-carboxymethyl dipeptide inhibitors of thermolysin	1.90	Monzingo, A.F., Matthews, B.W.
1Y3G	structure of a slanediol potease inhibitor bound to thermolysin	2.10	Juers, D.H., Kim, J., Matthews, B.W., Sieburth, S.M.
1Z9G	structure analysis of thermolysin with the inhibitor ( <i>R</i> )-retro-thiorphan	1.70	Roderick, S.L., Fournie-Zaluski, M.C., Roques, B.P., Matthews, B.W.
1ZDP	structure of thermolysin with the inhibitor ( <i>S</i> )-thiorphan	1.70	Roderick, S.L., Fournie-Zaluski, M.C., Roques, B.P., Matthews, B.W.
2TLI	thermolysin (5% isopropanol soaked)	1.95	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.

<b>pdb ID</b>	<b>Description</b>	<b>Resolution</b>	<b>Authors</b>
2TLX	thermolysin (native)	1.65	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
2TMN	Phosphoramidates as inhibitors and transitionstate analogs of thermolysin	1.60	Tronrud, D.E., Monzingo, A.F., Matthews, B.W.
3TLI	thermolysin (10% isopropanol soaked)	1.95	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
3TMN	Binding of <i>L</i> -valyl- <i>L</i> -tryptophan to thermolysin: the mode of interaction of a product of peptide hydrolysis	1.70	Holden, H.M., Matthews, B.W.
4TLI	thermolysin (25% isopropanol soaked)	1.95	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
4TLN	Hydroxamic acid inhibitors of thermolysin suggest a pentacoordinated Zn intermediate in catalysis	2.30	Matthews, B.W., Holmes, M.A.
4TMN	Slow and fast binding inhibitors of thermolysin display different modes of binding	1.70	Holden, H.M., Tronrud, D.E., Monzingo, A.F., Weaver, L.H., Matthews, B.W.
5TLI	thermolysin (60%isopropanol soaked)	2.10	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
5TLN	Hydroxamic acid inhibitors of thermolysin suggest a pentacoordinated Zn intermediate in catalysis	2.30	Matthews, B.W., Holmes, M.A.
5TMN	Slow and fast binding inhibitors of thermolysin display different modes of binding	1.60	Holden, H.M., Tronrud, D.E., Monzingo, A.F., Weaver, L.H., Matthews, B.W.
6TLI	thermolysin (60% isopropanol soaked)	2.10	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
6TMN	Structures of thermolysin-inhibitors complexes that differ by a single hydrogen bond	1.6	Tronrud, D.E., Holden, H.M., Matthews, B.W.

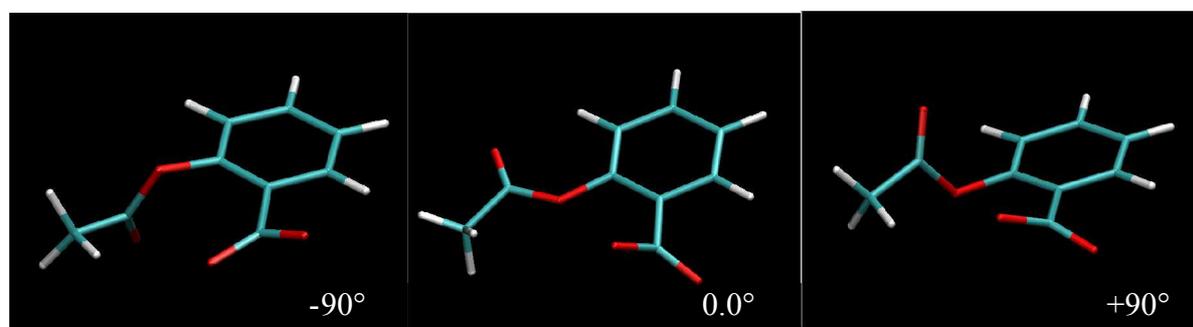
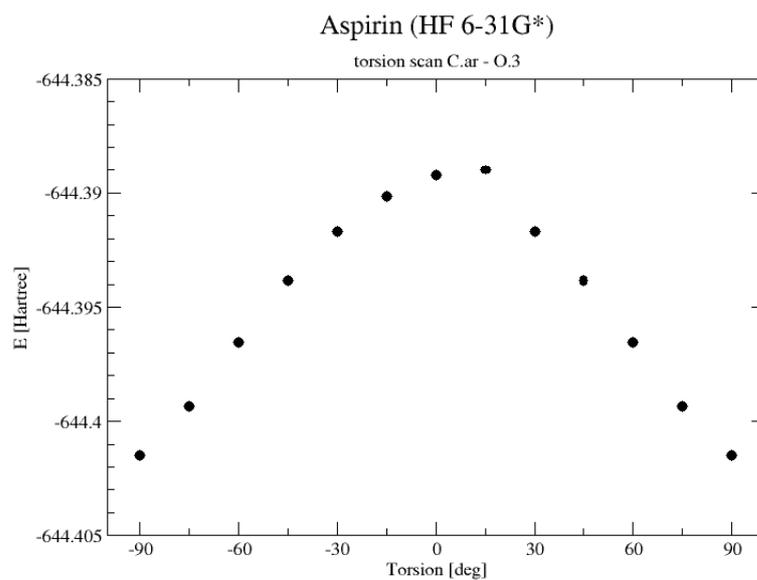
---

<b>pdb ID</b>	<b>Description</b>	<b>Resolution</b>	<b>Authors</b>
7TLI	thermolysin (90% isopropanol soaked)	1.95	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
7TLN	Inhibition of thermolysin by an active site directed irreversible inhibitor	2.3	Matthews, B.W., Holmes, M.A., Tronrud, D.E.
8TLI	thermolysin (100% isopropanol soaked)	2.2	English, A.C., Done, S.H., Groom, C.R., Hubbard, R.E.
8TLN	Comparison suggests that thermolysin and related neutral proteases undergo hinge-binding motion during catalysis	1.60	Tronrud, D., Matthews, B.W.

---

## Conformation search for aspirin

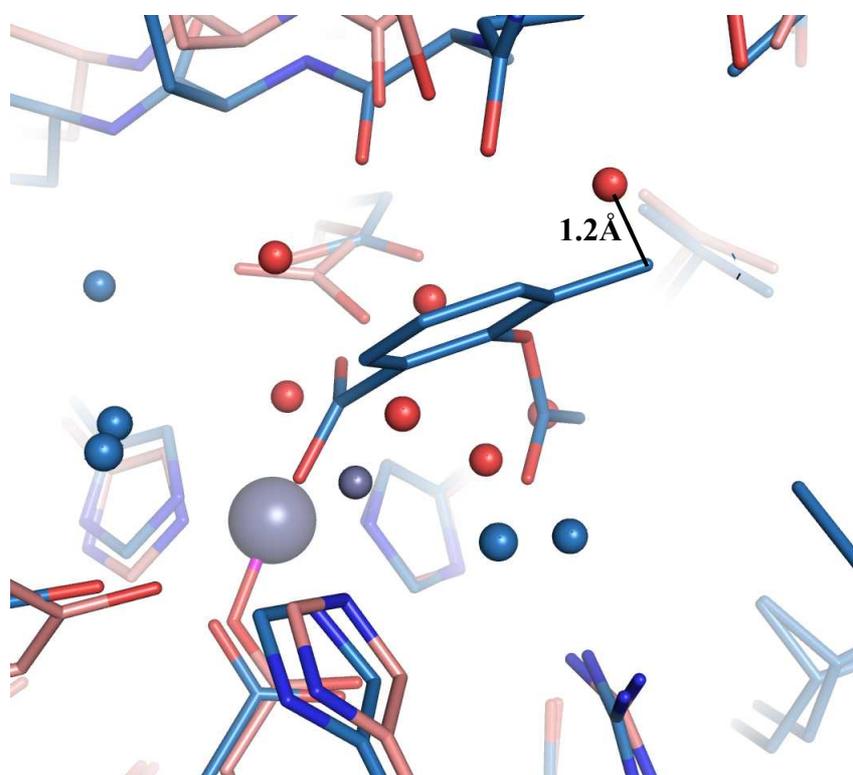
The phenyl carbon-phenol oxygen bond was turned in steps of 30°. Energy was determined using Gaussian.



**Figure s.1** Energy evaluation of different aspirin conformers

### ***Substrate free protein structure***

The herein determined protein-3-methylaspirin complex structure is superimposed onto the substrate free thermolysin protein structure (pdb ID 1L3F). The 3-methyl substituent occupies an area close to a structural water molecule position in 1L3F (distance =1.2Å). Direct comparison is hindered because in the case of a substrate free structure, the two proteins domains show a 5 degree rotation and a second zinc ion is bound into the active site.



**Figure s.2** Superimposition of 3-methylaspirin complex structure (blue: carbon atoms – blue spheres: water molecules) and pdb ID 1L3F (pink: carbon atoms – red spheres: water molecules). The relevant zinc ion is represented as a big purple sphere. The additionally bound zinc ion of the substrate free structure is shown as a little purple sphere.

**Refinement data****Table s.11:** refinement data for two 3-methylaspirin complex structures

	Complex structure at 1.95 Å resolution	Complex structure at 1.75 Å resolution
<b>Data collection and Processing</b>		
No. of crystals used	1	1
Wavelength [Å]	1.5418	0.91841
Space group	P6 <sub>1</sub> 22	P6 <sub>1</sub> 22
Unit cell parameters		
a, b, c [Å]	92.8, 92.8, 130.2	92.7, 92.7, 130.0
β [°]	90, 90, 120	90, 90, 120
Matthews coefficient [Å <sup>3</sup> /Da]	2.3	2.3
Solvent content	46.8	46.6
<b>Diffraction data*</b>		
Resolution range [Å]	50.0 – 1.95 (1.98-1.95)	50.0 – 1.75 (1.78 – 1.75)
Unique reflections	24 548 (1 178)	33 957 (1 669)
R(I) <sub>sym</sub> [%]	11.6 (50.7)	8.8 (49.5)
Completeness [%]	98.8 (97.8)	100 (100)
Redundancy	5.0 (5.0)	6.7 (7.1)
I/σ(I)	14.6 (3.2)	21.0 (3.9)
<b>Refinement</b>		
Resolution range [Å]	30.0 - 1.95	30.0 - 1.75
Reflections used in refinement (work/free)	22 493 / 1150	30957 / 1625
Final R values for all reflections (work/free) [%]	16.1 / 24.7	15.7 / 21.7
	14.5 / 23.1	14.5 / 19.9

	Complex structure at 1.95 Å resolution	Complex structure at 1.75 Å resolution
Final R values for reflections with $F > 4 \sigma$ (work/free) [%]	316	316
Protein residues	1 Zn <sup>2+</sup> , 4 Ca <sup>2+</sup>	1 Zn <sup>2+</sup> , 4 Ca <sup>2+</sup>
Metal ions	1	1
Inhibitor	308	328
Water molecules		
<b>RMSDs</b>		
Bonds [Å]	0.006	0.007
Angles [°]	1.7	1.9
<b>Ramachandran plot</b>		
Residues in most favored regions [%]	87.0	88.5
Residues in additional allowed regions [%]	12.6	10.0
Residues in generously allowed regions [%]	0.0	1.1
Residues in disallowed regions [%]	0.4 (Thr 26)	0.4 (Thr 26)
<b>Mean B factor [Å<sup>2</sup>]</b>		
Protein	16.3	13.6
Metal ions	15.4	11.4
Inhibitor	26.9	18.4
Water molecules	29.6	26.9

## SUMMARY & OUTLOOK

The present thesis demonstrates several examples of how to support a drug discovery project by computational methods. If some information about ligand and their binding properties is already available quantitative structure-activity relationships can be established to assist future ligand design.

In an example for antimalarial drugs, targeting the enzyme DOXP-reductoisomerase, three different 3D QSAR methods, CoMFA, CoMSIA and AFMoC have been compared. The training data set comprised 27 ligands and allowed to generate QSAR models based on leave-one-out analysis for all three methods. When applied to predict affinities for ligands, not used to calculate the model, AFMoC clearly outperformed the other two considered methods. CoMFA and CoMSIA models are solely based on ligand structures and thus more sensitive to size and composition of the training set data. AFMoC is a method which can be considered as either a target-tailored scoring function or a 3D-QSAR method which considers protein-ligand interactions. Using information from both, ligands and protein, it is less dependent on training data. Docking has been difficult because the protein comprises large flexible loops which interact with the potential ligands. Using all information available, a predictive AFMoC model could be established which can now be used to predict binding affinities of newly suggested inhibitors. However, as shown by synthesizing multiple derivatives, it has to be concluded that it is very difficult to optimize fosmidomycin. It is a very small ligand which mimics the natural substrate almost perfect while forming strong interactions. Further studies might include inhibitor design for DXR not only targeting malaria. DXR has been discovered in *Mycobacterium tuberculosis* as well as in *Pseudomonas aeruginosa*. The developed

AFMoC model could be applied not only to rationalize affinity but also to determine selectivity between targets.

A second example for lead optimization has been shown on benzophenone-based farnesyltransferase inhibitors. In previous studies the binding site of farnesyltransferase has been explored in a stepwise fashion. To further optimize affinity the so-called specificity pocket has been explored and additional derivatives have been prioritized using computational method. The newly synthesized inhibitors showed improved binding not only in vitro but in vivo as well.

Farnesyltransferase is a target not only for malaria but most prominent in the oncology field. While possible cell toxicity should be a concern, in general the double targeting of anti-infectives and cancer drugs is not a contradiction. Both are administered for short time and few dosages only, therefore another level of toxicity compared to e.g. cardiovascular drugs can be tolerated. Additionally, the binding site of *P. falciparum* and human ftase show some differences. This requires a high specificity in the ligands. The compounds investigated here are of a large size. A study made by Hann *et al.* [Hann, 2001] demonstrates the relationship between complexity of a ligand and its specificity towards a certain target. Therefore, the benzophenone-based inhibitors provide very promising ligand series to study affinity and selectivity for multiple farnesyltransferase targets.

Additionally to the computer-assisted lead optimization efforts, we applied a virtual screening protocol and docking for a fragment-based lead discovery project. Fragment-based screening has been applied recently successfully and has gained importance as an alternative to conventional HTS. However, most fragment-based approaches do not include computational support and if applied, mostly at the library design level or to predict substitutions for lead optimization. In contrast to those studies we applied virtual screening and docking as main screening tool. The decomposition of the substrate recognition site in proteases into multiple

sub-pockets makes the target enzyme thermolysin ideal for fragment-based screening. Furthermore, thermolysin is robust and allows access to crystallization and prior knowledge about fragment binding is available. Starting with 800,000 compounds several filter criteria led to 14 fragments purchased for experimental validation. 3-Methylaspirin could be determined as a thermolysin inhibitor and its binding geometry was identified crystallographically. In a subsequent study we analyzed binding affinities not only for 3-methylaspirin but aspirin and 3-chloroaspirin. The latter ligand was determined to inhibit thermolysin with 522  $\mu\text{M}$ .

Having successfully applied virtual screening to fragments several shortcomings became obvious. The small intersection between fragments of the World Drug Index and the commercial available molecules demonstrates the need for synthesis support already at the library design level. It additionally serves as evidence for fragment-based approaches which provide an easier coverage of chemical space. Compared to larger molecules fragments hardly address the binding site completely. The remaining area is filled with solvent and the fragment does interact not only with the protein but also with the residual water molecules occupying the partially filled binding pocket. While this might be true for any ligand, the reduced number of possible protein-fragment interactions enhances the importance of interactions to the water network as shown by several thermolysin-organic solvent crystal structures. Recently, some attempts have been made to include water in protein-ligand docking. However, to successfully apply automated docking to fragments a better consideration of the contribution of water in the protein-ligand interface is required.

In summary, several promising results have been obtained which demonstrate the capabilities of computational methods in the drug discovery process. They reduce time, cost and effort and provide guidance not only in the optimization but also in the discovery stage of lead compounds and series.

## REFERENCES

- Abad-Zapatero, C. and Metz, J. T. (2005) "Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today*, **10** (7): 464 - 469.
- Abagyan, R. and Totrov, M. (2001) "High-throughput docking for lead generation. *Curr Opin Chem Biol*, **5** (4): 375-82.
- Adnane, J., Bizouarn, F. A., Chen, Z., Ohkanda, J., Hamilton, A. D., Munoz-Antonia, T. and Sebti, S. M. (2000) "Inhibition of farnesyltransferase increases TGF beta type II receptor expression and enhances the responsiveness of human cancer cells to TGF beta. *Oncogene*, **19** (48): 5525 - 5533.
- Ajay Bemis, G. W. and Murcko, M. A. (1999) "Designing libraries with CNS activity. *Journal of Medicinal Chemistry*, **42** (24): 4942 - 4951.
- Alberts, I. L., Todorov, N. P. and Dean, P. M. (2005) "Receptor flexibility in de novo ligand design and docking. *J Med Chem*, **48** (21): 6585 - 6596.
- Andrews, P. R., Craik, D. J. and Martin, J. L. (1984) "Functional group contributions to drug-receptor interactions. *J. Med. Chem.*, **27** (12): 1648 - 1657.
- Austin, C. P. (2004) "The impact of the completed human genome sequence on the development of novel therapeutics for human disease. *Annu Rev Med*, **55** 1 - 13.
- Babaoglu, K. and Shoichet, B. K. (2006) "Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.*, **2** (12): 720 - 723.
- Barril, X. and Morley, S. D. (2005) "Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J Med Chem*, **48** (13): 4432 - 4443.
- Baurin, N., Baker, R., Richardson, C., Chen, I., Foloppe, N., Potter, A., Jordan, A., Roughley, S., Parratt, M., Greaney, P., Morley, D. and Hubbard, R. E. (2004) "Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *Journal of Chemical Information and Computer Sciences*, **44** (2): 643 - 651.

- Baxter, C. A., Murray, C. W., Waszkowycz, B., Li, J., Sykes, R. A., Bone, R. G., Perkins, T. D. and Wylie, W. (2000) "New approach to molecular docking and its application to virtual screening of chemical databases. "*J Chem Inf Comput Sci*, **40** (2): 254 - 262.
- Bemis, G. W. and Murcko, M. A. (1996) "The properties of known drugs.1. Molecular frameworks. "*Journal of Medicinal Chemistry*, **39** (15): 2887 – 2893.
- Bemis, G. W. and Murcko, M. A. (1999) "Properties of known drugs.2. Side Chains." *Journal of Medicinal Chemistry*, **42**: 5095 - 5099.
- Bergner, A., Gunther, J., Hendlich, M., Klebe, G. and Verdonk, M. (2001) "Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. "*Biopolymers*, **61** (2): 99-110.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) "The Protein Data Bank. "*Nucleic Acids Res*, **28** (1): 235 – 242.
- Bertini, I., Calderone, V., Fragai, M., Luchinat, C., Maletta, M., Kwon, J.Y. (2006) "Snapshots of the Reaction Mechanism of Matrix Metalloproteinases." *Angew. Chemie Int. Ed.*, **45** (47): 7952 – 7955.
- Bissantz, C., Folkers, G. and Rognan, D. (2000) "Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. "*J Med Chem*, **43** (25): 4759 - 4767.
- Blundell, T. L. (1996) "Structure-based drug design. "*Nature*, **384** (6604): 23 - 26.
- Blundell, T. L., Sibanda, B. L., Montalvao, R. W., Brewerton, S., Chelliah, V., Worth, C. L., Harmer, N. J., Davies, O. and Burke, D. (2006) "Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. "*Philos Trans R Soc Lond B Biol Sci*, **361** (1467): 413 - 423.
- Bohm, H. J. (1992a) "The computer program LUDI: a new method for the de novo design of enzyme inhibitors. "*J Comput Aided Mol Des*, **6** (1): 61 - 78.
- Bohm, H. J. (1992b) "LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. "*J Comput Aided Mol Des*, **6** (6): 593 - 606.

- Bohm, H. J. (1994a) "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. "*J Comput Aided Mol Des*, **8** (3): 243 - 256.
- Bohm, H. J. (1994b) "On the use of LUDI to search the Fine Chemicals Directory for ligands of proteins of known three-dimensional structure. "*J Comput Aided Mol Des*, **8** (5): 623-32.
- Bohm, H. J. (1998) "Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. "*Journal of Computer-Aided Molecular Design*, **12** (4): 309 - 323.
- Bohm, H. J. and Stahl, M. (1999) "Rapid empirical scoring functions in virtual screening applications. "*Medicinal Chemistry Research*, **9** (7-8): 445 - 462.
- Bohm, M. and Klebe, G. (2002) "Development of new hydrogen-bond descriptors and their application to comparative molecular field analyses. "*J Med Chem*, **45** (8): 1585-97.
- Bohm, M., Mitsch, A., Wissner, P., Sattler, I. and Schlitzer, M. (2001) "Exploration of novel aryl binding sites of farnesyltransferase using molecular modeling and benzophenone-based farnesyltransferase inhibitors. "*Journal of Medicinal Chemistry*, **44** (19): 3117 - 3124.
- Borrmann, S., Adegnika, A. A., Matsiegui, P. B., Issifou, S., Schindler, A., Mawili-Mboumba, D. P., Baranek, T., Wiesner, J., Jomaa, H. and Kremsner, P. G. (2004) "Fosmidomycin-clindamycin for Plasmodium falciparum Infections in African children. "*J Infect Dis*, **189** (5): 901 - 908.
- Bostrom, J., Bohm, M., Gundertofte, K. and Klebe, G. (2003) "A 3D QSAR study on a set of dopamine D4 receptor antagonists. "*J.Chem.Inf.Comput.Sci.*, **43** (3): 1020 - 1027.
- Botzki, A., Salmen, S., Bernhardt, G., Buschauer, A. and Dove, S. (2005) "Structure-based design of bacterial hyaluronan lyase inhibitors. "*Qsar und Combinatorial Science*, **24** (4): 458 - 469.
- Brenk, R., Naerum, L., Gradler, U., Gerber, H. D., Garcia, G. A., Reuter, K., Stubbs, M. T. and Klebe, G. (2003) "Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. "*J Med Chem*, **46** (7): 1133 - 1143.
- Breuker, K. (2004) "New mass spectrometric methods for the quantification of protein-ligand binding in solution. "*Angewandte Chemie-International Edition*, **43** (1): 22 - 25.

- Brooijmans, N. and Kuntz, I. D. (2003) "Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*, **32** 335 - 373.
- Brooks B. R. Bruccoleri, R., Olafsen, B., States, D. and M., K. (1983) *J comp chem*, **4** 187.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
- Bush, B. L. and Nachbar R. B., J. (1993) "Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J Comput Aided Mol Des*, **7** (5): 587 - 619.
- Böhm, H. J., Böhringer, M., Bur, D., Gmünder, H., Huber, W., Klaus, W., Kostrewa, D., Kühne, H., Lübbers, T., Meunier-Keller, N. and Müller, F. (2000) "Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J Med Chem*, **43** (14bk-md): 2664 - 2674.
- Carlson, H. A. (2002) "Protein flexibility and drug design: how to hit a moving target. *Curr Opin Chem Biol*, **6** (4): 447 - 452.
- Carlson, H. A. and McCammon, J. A. (2000a) "Accommodating protein flexibility in computational drug design. *Mol Pharmacol*, **57** (2): 213 - 218.
- Carlson, H. A., Masukawa, K. M., Rubins, K., Bushman, F. D., Jorgensen, W. L., Lins, R. D., Briggs, J. M. and McCammon, J. A. (2000b) "Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem*, **43** (11): 2100-14.
- Carr, R. and Jhoti, H. (2002) "Structure-based screening of low-affinity compounds. *Drug Discovery Today*, **7** (9): 522 - 527.
- Case, D. A., Cheatham, T. E., I. I. I., Darden, T., Gohlke, H., Luo, R., Merz, K. M., J., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. (2005) "The Amber biomolecular simulation programs. *J Comput Chem*, **26** (16): 1668 - 1688.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., I. I. I., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J.,

- Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. and Kollman, P. A. (2002). *AMBER 7*. San Francisco:.
- Chakrabarti, D.; Azam, T.; DelVecchio, C.; Qiu, L.; Park, Y.; Allen, C. M. Protein prenyl transferase activities of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **1998**, *94*, 175-184.
- Charifson, P. S., Corkery, J. J., Murcko, M. A. and Walters, W. P. (1999) "Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. "*J Med Chem*, **42** (25): 5100 – 5109.
- Chen, H., Lyne, P. D., Giordanetto, F., Lovell, T. and Li, J. (2006) "On evaluating molecular-docking methods for pose prediction and enrichment factors. "*J Chem Inf Model*, **46** (1): 401 – 415.
- Chen, J. M., Nelson, F. C., Levin, J. I., Mobilio, D., Moy, F. J., Nilakantan, R., Zask, A. and Powers, R. (2000) "Structure-based design of a novel, potent, and selective inhibitor for MMP-13 utilizing NMR spectroscopy and computer-aided molecular design. "*Journal of the American Chemical Society*, **122** (40): 9648 – 9654.
- Cherkasov, A., Lee, S., Nandan, D. and Reiner, N. E. (2006) "Large-scale survey for potentially targetable indels in bacterial and protozoan proteins. "*Proteins*, **62** (2): 371-80.
- Chin, D. N., Chuaqui, C. E. and Singh, J. (2004) "Integration of virtual screening into the drug discovery process. "*Mini Rev Med Chem*, **4** (10): 1053 – 1065. Clark, D. E. and Pickett, S. D. (2000) "Computational methods for the prediction of 'drug-likeness'. "*Drug Discov Today*, **5** (2): 49-58.
- Clark, G. J., Kinch, M. S., RogersGraham, K., Sebti, S. M., Hamilton, A. D. and Der, C. J. (1997) "The Ras-related protein Rheb is farnesylated and antagonizes Ras signaling and transformation. "*Journal of Biological Chemistry*, **272** (16): 10608 - 10615.

- Clark, R. D., Strizhev, A., Leonard, J. M., Blake, J. F. and Matthew, J. B. (2002) "Consensus scoring for ligand/protein interactions. "*J Mol Graph Model*, **20** (4): 281 - 295.
- Claussen, H., Buning, C., Rarey, M. and Lengauer, T. (2001) "FlexE: efficient molecular docking considering protein structure variations. "*J Mol Biol*, **308** (2): 377 - 395.
- Cole, J. C., Murray, C. W., Nissink, J. W., Taylor, R. D. and Taylor, R. (2005) "Comparing protein-ligand docking programs is difficult. "*Proteins*, **60** (3): 325 - 332.
- Congreve, M. S., Davis, D. J., Devine, L., Granata, C., O'Reilly, M., Wyatt, P. G. and Jhoti, H. (2003b) "Detection of ligands from a dynamic combinatorial library by X-ray crystallography. "*Angewandte Chemie-International Edition*, **42** (37): 4479 – 4482.
- Congreve, M., Carr, R., Murray, C. and Jhoti, H. (2003a) "A 'rule of three' for fragment-based lead discovery? "*Drug Discov.Today*, **8** (19): 876 – 877.
- Cornell, W. D., Cieplak, P., Merz, J. K., Ferguson, D. M. and Kollman, P. A. (1995), *J.Am.Chem.Soc.*, **117** 5179 – 5197.
- Cotesta, S., Giordanetto, F., Trosset, J., Crivori, P., Kroemer, R. T., Stouten, P. F. W. and Vulpetti, A. (2005) "Virtual screening to enrich a compound collection with CDK2 inhibitors using docking, scoring, and composite scoring models. "*Proteins*, **60** (4): 629-43.
- Coupez, B. and Lewis, R. A. (2006) "Docking and scoring--theoretically easy, practically impossible? "*Curr Med Chem*, **13** (25): 2995 – 3003.
- Cramer, R. D., I. I. I. De Priest, S. A. and Patterson D. E. Hecht, P. (1993). The Developing Practice of Comparative Molecular Field Analysis I. In H. Kubinyi (Ed.), *3D QSAR in Drug Design. Theory, Methods and Applications* Leiden:ESCOM.
- Cramer, R. D., I. I. I., Patterson, D. E. and Bunce, J. D. (1988) "Comparative Molecular Field Analysis (CoMFA). I. Effect of Shape on Binding of Steroids to Carrier Proteins. "*J Am.Chem.Soc.*, **110** 5959.
- Cramer, R. D., I. I. I., Patterson, D. E. and Bunce, J. D. (1989) "Recent advances in comparative molecular field analysis (CoMFA). "*Prog Clin Biol Res*, **291** 161 – 165.
- Crane, M., Kaiser, J., Ramsden, N., Lauw, S., Rohdich, F., Eisenreich, W., Hunter W. N., Bacher, A., Diederich, F., (2006) "Fluorescent Inhibitors for IspF, an Enzyme in the Non-

- Mevalonate Pathway for Isoprenoid Biosynthesis and a Potential Target for Antimalarial Therapy”, *Angewandte Chemie Int. Ed.*, **45** (7) 1069 - 1074
- Cruciani, G. and Watson, K. A. (1994) "Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. " *J Med Chem*, **37** (16): 2589 - 2601.
- Cruciani, G., Carosati, E., De Boeck, B., Ethirajulu, K., Mackie, C., Howe, T. and Vianello, R. (2005) "MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. " *J Med Chem*, **48** (22): 6970 - 6979.
- Cui, G., Wang, B. and Merz, K. M. J. (2005) "Computational studies of the farnesyltransferase ternary complex part I: substrate binding. " *Biochemistry*, **44** (50): 16513-23.
- Cummings, M. D., DesJarlais, R. L., Gibbs, A. C., Mohan, V. and Jaeger, E. P. (2005) "Comparison of automated docking programs as virtual screening tools. " *J Med Chem*, **48** (4): 962 – 976.
- Dayam, R., Sanchez, T., Clement, O., Shoemaker, R., Sei, S. and Neamati, N. (2005) "beta-Diketo acid pharmacophore hypothesis. 1. Discovery of a novel class of HIV-1 integrase inhibitors 3. " *Journal of Medicinal Chemistry*, **48** (1): 111 – 120.
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. <http://www.pymol.org>
- DeWitte, R. S. and Shakhnovich, E. (1997) "SMoG: De novo design method based on simple, fast and accurate free energy estimates. " *Abstracts of Papers of the American Chemical Society*, **214** 6 – Comp.
- DeWitte, R. S. and Shakhnovich, E. I. (1996) "SMoG: de Novo design method based on simple, fast, and accurate free energy estimates.1. Methodology and supporting evidence. " *Journal of the American Chemical Society*, **118** (47): 11733 – 11744.
- Degen, J. and Rarey, M. (2006) "FlexNovo: structure-based searching in large fragment spaces. " *ChemMedChem*, **1** (8): 854-68.
- DesJarlais, R. L. and Dixon, J. S. (1994) "A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. " *J Comput Aided Mol Des*, **8** (3): 231 – 242.

- Devreux, V., Wiesner, J., Goeman, J. L., der, V., Eycken, J., Jomaa, H. and Van Calenbergh, S. (2006) "Synthesis and biological evaluation of cyclopropyl analogues of fosmidomycin as potent *Plasmodium falciparum* growth inhibitors. " *Journal of Medicinal Chemistry*, **49** (8): 2656 – 2660.
- Dewar, M., Zoebisch, E., Healy, E. and Stewart, J. (1985) "Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. " *J.Am.Chem.Soc.*, **107** 3902 - 3909.
- Dinsmore, C. J. and Bell, I. M. (2003) "Inhibitors of farnesyltransferase and geranylgeranyltransferase-I for antitumor therapy: substrate-based design, conformational constraint and biological activity. " *Curr Top Med Chem*, **3** (10): 1075 - 1093.
- Dixon, M. (1953) "The determination of the enzyme inhibition constants. " *Biochemical Journal*, **55** 170-171.
- Dixon, M. (1972) "The graphical determination of  $K_m$  and  $K_i$ . " *Biochemical Journal*, **129** 197-202.
- Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Connolly, D. T. and Shoichet, B. K. (2002) "Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. " *J.Med.Chem.*, **45** (11): 2213 – 2221.
- Drews, J. (2000) "Drug discovery: a historical perspective. " *Science*, **287** (5460): 1960 – 1964.
- Egan, W. J. and Lauri, G. (2002) "Prediction of intestinal permeability. " *Adv Drug Deliv Rev*, **54** (3): 273 – 289.
- Eisenreich, W., Bacher, A., Arigoni, D. and Rohdich, F. (2004) "Biosynthesis of isoprenoids via the non-mevalonate pathway. " *Cell Mol Life Sci*, **61** (12): 1401 - 1426.
- Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M. H. and Bacher, A. (1998) "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. " *Chem Biol*, **5** (9): R221 - R233.
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. and Mee, R. P. (1997) "Empirical scoring functions: I. The development of a fast empirical scoring function to

- estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*, **11** (5): 425 – 445.
- Emsley, P. & Kowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta cryst. D* **60**, 2126-2132.
- English, A. C., Done, S. H., Caves, L. S., Groom, C. R. and Hubbard, R. E. (1999) "Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins*, **37** (4): 628-40.
- English, A. C., Groom, C. R. and Hubbard, R. E. (2001). *Experimental and computational mapping of the binding surface of a crystalline protein*.
- Erlanson, D. A. (2006) "Fragment-based lead discovery: a chemical update. *Curr.Opin.Biotechnol.*, **17** (6): 643 – 652.
- Erlanson, D. A. and Hansen, S. K. (2004b) "Making drugs on proteins: site-directed ligand discovery for fragment-based lead assembly. *Curr.Opin.Chem.Biol.*, **8** (4): 399 – 406.
- Erlanson, D. A., Braisted, A. C., Raphael, D. R., Randal, M., Stroud, R. M., Gordon, E. M. and Wells, J. A. (2000) "Site-directed ligand discovery. *Proceedings of the National Academy of Sciences of the United States of America*, **97** (17): 9367 – 9372.
- Erlanson, D. A., Lam, J. W., Wiesmann, C., Luong, T. N., Simmons, R. L., Delano, W. L., Choong, I. C., Burdett, M. T., Flanagan, W. M., Lee, D., Gordon, E. M. and O'Brien, T. (2003) "In situ assembly of enzyme inhibitors using extended tethering. *Nature Biotechnology*, **21** (3): 308 – 314.
- Erlanson, D. A., McDowell, R. S. and O'Brien, T. (2004a) "Fragment-based drug discovery. *J.Med.Chem.*, **47** (14): 3463 - 3482.
- Erlanson, D. A., Wells, J. A. and Braisted, A. C. (2004c) "Tethering: fragment-based drug discovery. *Annu.Rev.Biophys.Biomol.Struct.*, **33** 199 - 223.
- Esteva, M. I., Kettler, K., Maidana, C., Fichera, L., Ruiz, A. M., Bontempi, E. J., Andersson, B., Dahse, H., Haebel, P., Ortmann, R., Klebe, G. and Schlitzer, M. (2005) "Benzophenone-based farnesyltransferase inhibitors with high activity against *Trypanosoma cruzi*. *J Med Chem*, **48** (23): 7186-91.

- Evers, A. and Klabunde, T. (2005) "Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the Alpha1A adrenergic receptor. "*Journal of Medicinal Chemistry*, **48** (4): 1088 - 1097.
- Evers, A. and Klebe, G. (2004) "Ligand-supported homology modeling of g-protein-coupled receptor sites: models sufficient for successful virtual screening. "*Angew Chem Int Ed Engl*, **43** (2): 248 - 251.
- Ewing, T. J. A. and Kuntz, I. D. (1997) "Critical evaluation of search algorithms for automated molecular docking and database screening. "*Journal of Computational Chemistry*, **18** (9): 1175 - 1189.
- Ewing, T. J., Makino, S., Skillman, A. G. and Kuntz, I. D. (2001) "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. "*J Comput Aided Mol Des*, **15** (5): 411 – 428.
- Farid, R., Day, T., Friesner, R. A. and Pearlstein, R. A. (2006) "New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. "*Bioorg Med Chem*, **14** (9): 3160 – 3173.
- Feder, J., Brougham, L. R. and Wildi, B. S. (1974) "Inhibition of Thermolysin by Dipeptides. "*Biochemistry*, **13** (6): 1186 – 1189.
- Fejzo, J., Lepre, C. A., Peng, J. W., Bemis, G. W., Ajay Murcko, M. A. and Moore, J. M. (1999) "The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. "*Chemistry und Biology*, **6** (10): 755 – 769.
- Fernandes, R. P. M., Phaosiri, C. and Proteau, P. J. (2005) "Mutation in the flexible loop of 1-deoxy-D-xylulose 5-phosphate reductoisomerase broadens substrate utilization. "*Archives of Biochemistry and Biophysics*, **444** (2): 159 - 164.
- Ferrara, P., Gohlke, H., Price, D. J., Klebe, G. and Brooks, C. L. 3. (2004) "Assessing scoring functions for protein-ligand interactions. "*J Med Chem*, **47** (12): 3032-47.
- Ferrari, A. M., Wei, B. Q., Costantino, L. and Shoichet, B. K. (2004) "Soft docking and multiple receptor conformations in virtual screening. "*J Med Chem*, **47** (21): 5076 - 5084.
- Fielding, L., Fletcher, D., Rutherford, S., Kaur, J. and Mestres, J. (2003) "Exploring the active site of human factor Xa protein by NMR screening of small molecule probes. "*Organic und Biomolecular Chemistry*, **1** (23): 4235 - 4241.

- Finkelstein, A. V. and Janin, J. (1989) "The Price of Lost Freedom - Entropy of Bimolecular Complex-Formation. "*Protein Engineering*, **3** (1): 1 - 3.
- Forbes, I. T., Cooper, D. G., Dodds, E. K., Hickey, D. M., Ife, R. J., Meeson, M., Stockley, M., Berkhout, T. A., Gohil, J., Groot, P. H. and Moores, K. (2000) "CCR2B receptor antagonists: conversion of a weak HTS hit to a potent lead compound. "*Bioorg.Med.Chem.Lett.*, **10** (16): 1803 - 1806.
- Forino, M., Johnson, S., Wong, T. Y., Rozanov, D. V., Savinov, A. Y., Li, W., Fattorusso, R., Becattini, B., Orry, A. J., Jung, D. W., Abagyan, R. A., Smith, J. W., Alibek, K., Liddington, R. C., Strongin, A. Y. and Pellecchia, M. (2005) "Efficient synthetic inhibitors of anthrax lethal factor. "*Proceedings of the National Academy of Sciences of the United States of America*, **102** (27): 9499 - 9504.
- Fradera, X., Knegt, R. M. and Mestres, J. (2000) "Similarity-driven flexible ligand docking. "*Proteins*, **40** (4): 623 - 636.
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P. and Shenkin, P. S. (2004) "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. "*J Med Chem*, **47** (7): 1739 - 1749.
- Gad, S. (2005). Drug Discovery in the 21st Century. In S. Gad (Ed.), *Drug Discovery Handbook* (pp. 1 - 10). Hoboken:Wiley.
- Gane, P. J. and Dean, P. M. (2000) "Recent advances in structure-based rational drug design. "*Curr Opin Struct Biol*, **10** (4bk-md): 401 - 404.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M. and Subramanian, (2002a) "Genome sequence of the human malaria parasite *Plasmodium falciparum*. "*Nature*, **419** (6906): 498 - 511.
- Gardner, M. J., Shallom, S. J., Carlton, J. M., Salzberg, S. L., Nene, V., Shoabibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B., Jarrahi, B., Brenner, M., Parvizi, B., Tallon, L., Moazzez, A., Granger, D., Fujii, C., Hansen, C., Pederson, J., Feldblyum, T., Peterson, J.,

- Suh, B., Angiuoli, S., Perteau, M., Allen, J., Selengut, J., White, O., Cummings, L. M., Smith, H. O., Adams, M. D., Venter, J. C., Carucci, D. J., Hoffman, S. L. and Fraser, C. M. (2002b) "Sequence of Plasmodium falciparum chromosomes 2, 10, 11 and 14. " *Nature*, **419** (6906): 531 - 534.
- Gasteiger, J. and Marsili, M. (1980) "Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. " *Tetrahedron*, **36** (22): 3219 - 3228.
- Gaucher, J. F., Selkti, M., Tiraboschi, G., Prange, T., Roques, B. P., Tomas, A. and Fournie-Zaluski, M. C. (1999) "Crystal structures of alpha-mercaptoacyldipeptides in the thermolysin active site: structural parameters for a Zn monodentation or bidentation in metalloendopeptidases. " *Biochemistry*, **38** (39): 12569-76.
- Gerber, P. R. and Müller, K. (1995) "MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. " *J.Comput.Aided Mol.Des.*, **9** 251 - 268.
- Gleeson, M. P. (2007) "Plasma protein binding affinity and its relationship to molecular structure: an in-silico analysis. " *J Med Chem*, **50** (1): 101 - 112.
- Glenn, M. P., Chang, S., Horney, C., Rivas, K., Yokoyama, K., Pusateri, E. E., Fletcher, S., Cummings, C. G., Buckner, F. S., Pendyala, P. R., Chakrabarti, D., Sebt, S. M., Gelb, M. and Van Voorh, (2006) "Structurally simple, potent, Plasmodium selective farnesyltransferase inhibitors that arrest the growth of malaria parasites. " *J Med Chem*, **49** (19): 5710-27.
- Glenn, M. R., Chang, S. Y., Hucke, O., Verlinde, C. L. M. J., Rivas, K., Horney, C., Yokoyama, K., Buckner, F. S., Pendyala, P. R., Chakrabarti, D., Gelb, M., Van Voorhis, W. C., Sebt, S. M. and Hamilton, A. D. (2005) "Structurally simple farnesyltransferase inhibitors arrest the growth of malaria parasites. " *Angewandte Chemie-International Edition*, **44** (31): 4903 - 4906.
- Gohlke, H. and Klebe, G. (2001) "Statistical potentials and scoring functions applied to protein-ligand binding. " *Curr Opin Struct Biol*, **11** (2): 231 - 235.
- Gohlke, H. and Klebe, G. (2002a) "DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. " *J Med Chem*, **45** (19): 4153-70.

- Gohlke, H. and Klebe, G. (2002b) "Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl*, **41** (15): 2644-76.
- Gohlke, H., Hendlich, M. and Klebe, G. (2000a) "Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*, **295** (2): 337 - 356.
- Gohlke, H., Hendlich, M. and Klebe, G. (2000b) "Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. *Perspectives in Drug Discovery and Design*, **20** (1): 115 - 144.
- Gohlke, H., Schwarz, S., Gundisch, D., Tilotta, M. C., Weber, A., Wegge, T. and Seitz, G. (2003) "3D QSAR analyses-guided rational design of novel ligands for the (alpha4)2(beta2)3 nicotinic acetylcholine receptor. *J Med Chem*, **46** (11): 2031 - 2048.
- Good, A. C., Cheney, D. L., Sitkoff, D. F., Tokarski, J. S., Stouch, T. R., Bassolino, D. A., Krystek, S. R., Li, Y., Mason, J. S. and Perkins, T. D. (2003) "Analysis and optimization of structure-based virtual screening protocols. 2. Examination of docked ligand orientation sampling methodology: mapping a pharmacophore for success. *J Mol Graph Model*, **22** (1): 31 - 40.
- Good, A. C., Krystek, S. R. and Mason, J. S. (2000) "High-throughput and virtual screening: core lead discovery technologies move towards integration. *Drug Discov Today*, **5** (12 Suppl 1): 61 - 69.
- Goodsell, D. S., Morris, G. M. and Olson, A. J. (1996) "Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit*, **9** (1): 1 - 5.
- Green, D. V. (2003) "Virtual screening of virtual libraries. *Prog Med Chem*, **41** 61 - 97.
- Green, N. M. (1975) "Avidin. *Adv. Protein Chem*, **29** 85-133.
- Gruneberg, S., Stubbs, M. T. and Klebe, G. (2002) "Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem*, **45** (17): 3588-602.
- Gruneberg, S., Wendt, B. and Klebe, G. (2001) "Subnanomolar Inhibitors from Computer Screening: A Model Study Using Human Carbonic Anhydrase II. *Angew Chem Int Ed Engl*, **40** (2): 389-393.

- Guida, W. C., Hamilton, A. D., Crotty, J. W. and Sebt, S. M. (2005) "Protein farnesyltransferase: Flexible docking studies on inhibitors using computational modeling. *Journal of Computer-Aided Molecular Design*, **19** (12): 871 - 885.
- Gunther, J., Bergner, A., Hendlich, M. and Klebe, G. (2003) "Utilising structural knowledge in drug design strategies: applications using Relibase. *J Mol Biol*, **326** (2): 621-36.
- Hajduk, P. J., Dinges, J., Schkeryantz, J. M., Janowick, D., Kaminski, M., Tufano, M., Augeri, D. J., Petros, A., Nienaber, V., Zhong, P., Hammond, R., Coen, M., Beutel, B., Katz, L. and Fesik, S. W. (1999) "Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *Journal of Medicinal Chemistry*, **42** (19): 3852 - 3859.
- Hajduk, P. J., Gomtsyan, A., Didomenico, S., Cowart, M., Bayburt, E. K., Solomon, L., Severin, J., Smith, R., Walter, K., Holzman, T. F., Stewart, A., McGaraughty, S., Jarvis, M. F., Kowaluk, E. A. and Fesik, S. W. (2000) "Design of adenosine kinase inhibitors from the NMR-based screening of fragments. *Journal of Medicinal Chemistry*, **43** (25): 4781 - 4786.
- Hajduk, P. J., Huth, J. R. and Fesik, S. W. (2005a) "Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem*, **48** (7): 2518 - 2525.
- Hajduk, P. J., Huth, J. R. and Tse, C. (2005b) "Predicting protein druggability. *Drug Discov Today*, **10** (23-24): 1675 - 1682.
- Hajduk, P. J., Sheppard, G., Nettesheim, D. G., Olejniczak, E. T., Shuker, S. B., Meadows, R. P., Steinman, D. H., Carrera, G. M., Marcotte, P. A., Severin, J., Walter, K., Smith, H., Gubbins, E., Simmer, R., Holzman, T. F., Morgan, D. W., Davidsen, S. K., Summers, J. B. and Fesik, S. W. (1997) "Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *Journal of the American Chemical Society*, **119** (25): 5818 - 5827.
- Halperin, I., Ma, B., Wolfson, H. and Nussinov, R. (2002) "Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, **47** (4): 409 - 443.
- Haluska, P., Dy, G. K. and Adjei, A. A. (2002) "Farnesyl transferase inhibitors as anticancer agents. *Eur J Cancer*, **38** (13): 1685 - 1700.
- Hann, M. M., Leach, A. R. and Harper, G. (2001) "Molecular complexity and its impact on the probability of finding leads for drug discovery. *J.Chem.Inf.Comput.Sci.*, **41** (3): 856 - 864.

- Hansen, K. K., Hansen, H. C., Clark, R. C. and Bartlett, P. A. (2003) "Identification of novel macrocyclic peptidase substrates via on-bead enzymatic cyclization. "*J Org Chem*, **68** (22): 8459-64.
- Hausrath, A. C. and Matthews, B. W. (1994) "Redetermination and refinement of the complex of benzylsuccinic acid with thermolysin and its relation to the complex with carboxypeptidase A. "*J Biol Chem*, **269** (29): 18839-42.
- Hendlich, M., Bergner, A., Gunther, J. and Klebe, G. (2003) "Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. "*J Mol Biol*, **326** (2): 607-20.
- Henriksen, B. S., Zahn, T. J., Evanseck, J. D., Firestine, S. M. and Gibbs, R. A. (2005) "Computational and conformational evaluation of FTase alternative substrates: insight into a novel enzyme binding pocket. "*J Chem Inf Model*, **45** (4): 1047-52.
- Henriksson, L. M., Bjorkelid, C., Mowbray, S. L. and Unge, T. (2006) "The 1.9 Å resolution structure of Mycobacterium tuberculosis 1-deoxy-D-xylulose 5-phosphate reductoisomerase, a potential drug target. "*Acta Crystallogr D Biol Crystallogr*, **62** (Pt 7): 807-13.
- Hillisch, A., Pineda, L. F. and Hilgenfeld, R. (2004) "Utility of homology models in the drug discovery process. "*Drug Discov Today*, **9** (15): 659 - 669.
- Hindle, S. A., Rarey, M., Buning, C. and Lengau, T. (2002) "Flexible docking under pharmacophore type constraints. "*J Comput Aided Mol Des*, **16** (2): 129 - 149.
- Hochgurtel, M., Biesinger, R., Kroth, H., Piecha, D., Hofmann, M. W., Krause, S., Schaaf, O., Nicolau, C. and Eliseev, A. V. (2003) "Ketones as building blocks for dynamic combinatorial libraries: Highly active neuraminidase inhibitors generated via selection pressure of the biological target. "*Journal of Medicinal Chemistry*, **46** (3): 356 - 358.
- Holden, H. M., Tronrud, D. E., Monzingo, A. F., Weaver, L. H. and Matthews, B. W. (1987) "Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphoramidate transition-state analogues. "*Biochemistry*, **26** (26): 8542-53.
- Holloway, M. K., Wai, J. M., Halgren, T. A., Fitzgerald, P. M., Vacca, J. P., Dorsey, B. D., Levin, R. B., Thompson, W. J., Chen, L. J., de-Solms, S. J., Gaffin, N., Ghosh, A. K.,

- Giuliani, E. A., Graham, S. L., Guare, J. P., Hungate, R. W., Lyle, T. A., Sanders, W. M., Tucker, T. J., Wiggins, M., Wiscount, C. M., Woltersdorf, O. W., Young, S. D., Darke, P. L. and Zugay, J. A. (1885) "A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. "*J.Med.Chem.*, **38** 305 - 317.
- Holmes, M. A. and Matthews, B. W. (1982) "Structure of Thermolysin Refined at 1.6-Å Resolution. "*Journal of Molecular Biology*, **160** (4): 623 - 639.
- Honma, T. (2003) "Recent advances in de novo design strategy for practical lead identification. "*Med Res Rev*, **23** (5): 606 - 632.
- Hopkins, A. (2003) *Winning before the battle is fought: a priori target druggability assessment*, International Workshop New Approaches in Drug Design & Discovery.
- Hopkins, A. L., Groom, C. R. and Alex, A. (2004) "Ligand efficiency: a useful metric for lead selection. "*Drug Discov.Today*, **9** (10): 430 - 431.
- Hou, T. and Xu, X. (2004) "Recent development and application of virtual screening in drug discovery: an overview. "*Curr Pharm Des*, **10** (9): 1011 - 1033.
- Huc, I. and Lehn, J. M. (1997) "Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly. "*Proceedings of the National Academy of Sciences of the United States of America*, **94** (6): 2106 - 2110.
- Hunt, J. T., Lee, V. G., Leftheris, K., Seizinger, B., Carboni, J., Mabus, J., Ricca, C., Yan, N. and Manne, V. (1996) "Potent, cell active, non-thiol tetrapeptide inhibitors of farnesyltransferase. "*J Med Chem*, **39** (2): 353 – 358.
- Hunter, W.N., Bond, C.S., Gabrielsen, M., Kemp, L.E., "Structure and reactivity in the non mevalonate pathway of isoprenoid biosynthesis" (2003) *Biochem Soc Trans.*, **31**(3):537-42
- Hyde, J., Braisted, A. C., Randal, M. and Arkin, M. R. (2003) "Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. "*Biochemistry*, **42** (21): 6475 – 6483.
- Ishchenko, A.V., Shakhnovich, E.I., "SMall Molecule Growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein-ligand interactions." *J Med Chem.* (2002), **45** (13):2770-80.

- Jacobsson, M., Liden, P., Stjernschantz, E., Bostrom, H. and Norinder, U. (2003) "Improving structure-based virtual screening by multivariate analysis of scoring data. "*J Med Chem*, **46** (26): 5781 – 5789.
- Jacoby, E., Davies, J. and Blommers, M. J. (2003) "Design of small molecule libraries for NMR screening and other applications in drug discovery. "*Curr.Top.Med.Chem.*, **3** (1): 11 – 23.
- Jalaie, M. and Shanmugasundaram, V. (2006) "Virtual screening: are we there yet? "*Mini Rev Med Chem*, **6** (10): 1159 – 1167.
- Jencks, W. P. (1981) "On the Attribution and Additivity of Binding-Energies. "*Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **78** (7): 4046 - 4050.
- Jhoti, H. (2001) "High-throughput structural proteomics using x-rays. "*Trends in Biotechnology*, **19** (10): S67 - S71.
- Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D. and Beck, E. (1999) "Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. "*Science*, **285** (5433): 1573 - 1576.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R. and Taylor, R. (1997) "Development and validation of a genetic algorithm for flexible docking. "*J Mol Biol*, **267** (3): 727 – 748.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
- Jung, M., Kim, H., Nam, K. Y. and No, K. T. (2005) "Three-dimensional structure of Plasmodium falciparum Ca<sup>2+</sup>-ATPase(PfATP6) and docking of artemisinin derivatives to PfATP6. "*Bioorganic und Medicinal Chemistry Letters*, **15** (12): 2994 - 2997.
- Kairys, V., Fernandes, M. X. and Gilson, M. K. (2006) "Screening drug-like compounds by docking to homology models: a systematic study. "*J Chem Inf Model*, **46** (1): 365 - 379.
- Kalyanaraman, C., Bernacki, K. and Jacobson, M. P. (2005) "Virtual screening against highly charged active sites: Identifying substrates of alpha-beta barrel enzymes 1. "*Biochemistry*, **44** (6): 2059 – 2071.

- Kellenberger, E., Rodrigo, J., Muller, P. and Rognan, D. (2004) "Comparative evaluation of eight docking tools for docking and virtual screening accuracy. "*Proteins*, **57** (2): 225 – 242.
- Kemp, C. A., Flanagan, J. U., van Eldik, A. J., Marechal, J. D., Wolf, C. R., Roberts, G. C. K., Paine, M. J. I. and Sutcliffe, M. J. (2004) "Validation of model of cytochrome p450 2D6: An in silico tool for predicting metabolism and inhibition. "*Journal of Medicinal Chemistry*, **47** (22): 5340 – 5346.
- Keseru, G. M. and Makara, G. M. (2006) "Hit discovery and hit-to-lead approaches. "*Drug Discov Today*, **11** (15-16): 741 – 748.
- Kettler, K., Wiesner, J., Ortmann, R., Dahse, H. M., Jomaa, H. and Schlitzer, M. (2006) "Antimalarial activity of methylpiperazinyl-substituted benzophenone-based farnesyltransferase inhibitors. "*Pharmazie*, **61** (1): 63 – 65.
- Kettler, K., Wiesner, J., Silber, K., Haebel, P., Ortmann, R., Sattler, I., Dahse, H., Jomaa, H., Klebe, G. and Schlitzer, M. (2005) "Non-thiol farnesyltransferase inhibitors: N-(4-aminoacylamino-3-benzoylphenyl)-3-[5-(4-nitrophenyl)-2 furyl]acrylic acid amides and their antimalarial activity. "*Eur J Med Chem*, **40** (1): 93-101.
- Kirtay, C., Mitchell, J. and Lumley, J. (2005) "Knowledge BAsed Potentials: the REverse Boltzmann Methodology, Virtual Screening and Molecular Weight Dependence. "*QSAR Comb.Sci.*, **24** 527 - 536.
- Kirton, S. B., Murray, C. W., Verdonk, M. L. and Taylor, R. D. (2005) "Prediction of binding modes for ligands in the cytochromes p450 and other heme-containing proteins. "*Proteins-Structure Function and Bioinformatics*, **58** (4): 836 - 844.
- Kitchen, D. B., Decornez, H., Furr, J. R. and Bajorath, J. (2004) "Docking and scoring in virtual screening for drug discovery: methods and applications. "*Nat Rev Drug Discov*, **3** (11): 935 - 949.
- Klebe, G.; Mietzner, T.; Weber, F. (1999) "Methodological developments and strategies for a fast flexible superposition of drug-size molecules." *J.Comput.-Aided Mol. Des.* **13**, 35-49.
- Klebe, G. (1994) "The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. "*J Mol Biol*, **237** (2): 212-35.
- Klebe, G. (2000) "Recent developments in structure-based drug design. "*J Mol Med*, **78** (5): 269 - 281.

- Klebe, G. (2006) "Virtual ligand screening: strategies, perspectives and limitations. "*Drug Discov Today*, **11** (13-14): 580-94.
- Klebe, G., Abraham, U. and Mietzner, T. (1994) "Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. "*J Med Chem*, **37** (24): 4130 – 4146.
- Kolb, P. and Caflisch, A. (2006) "Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. "*Journal of Medicinal Chemistry*, **49** (25): 7384 – 7392.
- Koppisch, A. T., Fox, D. T., Blagg, B. S. and Poulter, C. D. (2002) "E. coli MEP synthase: steady-state kinetic analysis and substrate binding. "*Biochemistry*, **41** (1): 236 - 243.
- Kraemer, O., Hazemann, I., Podjarny, A. D. and Klebe, G. (2004) "Virtual screening for inhibitors of human aldose reductase. "*Proteins*, **55** (4): 814-23.
- Kramer, B., Rarey, M. and Lengauer, T. (1999) "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. "*Proteins*, **37** (2): 228 - 241.
- Kroemer, R. T., Vulpetti, A., McDonald, J. J., Rohrer, D. C., Trosset, J., Giordanetto, F., Cotesta, S., McMartin, C., Kihlen, M. and Stouten, P. F. W. (2004) "Assessment of docking poses: interactions-based accuracy classification (IBAC) versus crystal structure deviations. "*J Chem Inf Comput Sci*, **44** (3): 871-81.
- Krovat, E. M. and Langer, T. (2004) "Impact of scoring functions on enrichment in docking-based virtual screening: an application study on renin inhibitors. "*J Chem Inf Comput Sci*, **44** (3): 1123-9.
- Kubinyi, H. (2003) "Drug research: myths, hype and reality. "*Nat Rev Drug Discov*, **2** (8): 665 - 668.
- Kuhn, B., Gerber, P., Schulz-Gasch, T. and Stahl, M. (2005) "Validation and use of the MM-PBSA approach for drug discovery. "*J Med Chem*, **48** (12): 4040 - 4048.
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E. (1982) "A geometric approach to macromolecule-ligand interactions. "*J Mol Biol*, **161** (2): 269 - 288.
- Kuntz, I. D., Chen, K., Sharp, K. A. and Kollman, P. A. (1999) "The maximal affinity of ligands. "*Proc.Natl.Acad.Sci.U.S.A*, **96** (18): 9997 – 10002.

- Kuo, C. L., Assefa, H., Kamath, S., Brzozowski, Z., Slawinski, J., Saczewski, F., Buolamwini, J. K. and Neamati, N. (2004) "Application of CoMFA and CoMSIA 3D-QSAR and docking studies in optimization of mercaptobenzenesulfonamides as HIV-1 integrase inhibitors. *J Med Chem*, **47** (2): 385 - 399.
- Kuzuyama, T. (2002) "Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotechnol Biochem*, **66** (8): 1619 - 1627.
- Kuzuyama, T., Takagi, M., Takahashi, S. and Seto, H. (2000) "Cloning and characterization of 1-deoxy-D-xylulose 5-phosphate synthase from *Streptomyces* sp. Strain CL190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis. *J Bacteriol*, **182** (4): 891 - 897.
- Ladbury, J. E. (1996) "Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol*, **3** (12): 973 - 980.
- Lahana, R. (1999) "How many leads from HTS? *Drug Discov. Today*, **4** (10): 447 - 448.
- Lameijer, E. W., Kok, J. N., Back, T. and Ijzerman, A. P. (2006) "Mining a chemical database for fragment co-occurrence: Discovery of 'chemical cliches'. *Journal of Chemical Information and Modeling*, **46** (2): 553 - 562.
- Lange, G., Lesuisse, D., Deprez, P., Schoot, B., Loenze, P., Benard, D., Marquette, J. P., Broto, P., Sarubbi, E. and Mandine, E. (2003) "Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of (pp60)Src are identical to those for high affinity binding of full length inhibitors. *Journal of Medicinal Chemistry*, **46** (24): 5184 - 5195.
- Langer, T. and Krovat, E. M. (2003) "Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr Opin Drug Discov Devel*, **6** (3): 370 - 376.
- Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
- Leach, A. R., Hann, M. M., Burrows, J. N. and Griffen, E. J. (2006b) "Fragment screening: an introduction. *Mol. Biosyst.*, **2** (9): 430 - 446.
- Leach, A. R., Shoichet, B. K. and Peishoff, C. E. (2006a) "Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem*, **49** (20): 5851 - 5855.

- Lell, B., Ruangweerayut, R., Wiesner, J., Missinou, M. A., Schindler, A., Baranek, T., Hintz, M., Hutchinson, D., Jomaa, H. and Kremsner, P. G. (2003) "Fosmidomycin, a novel chemotherapeutic agent for malaria." *Antimicrob Agents Chemother*, **47** (2): 735 - 738.
- Lerner, E. C., Qian, Y. M., Blaskovich, M. A., Fossum, R. D., Vogt, A., Sun, J. Z., Cox, A. D., Der, C. J., Hamilton, A. D. and Sebti, S. M. (1995) "Ras Caax Peptidomimetic Fti-277 Selectively Blocks Oncogenic Ras Signaling by Inducing Cytoplasmic Accumulation of Inactive Ras-Raf Complexes." *Journal of Biological Chemistry*, **270** (45): 26802 - 26806.
- Lesuisse, D., Lange, G., Deprez, P., Benard, D., Schoot, B., Delettre, G., Marquette, J. P., Broto, P., Jean-Baptiste, V., Bichet, P., Sarubbi, E. and Mandine, E. (2002) "SAR and X-ray. A new approach combining fragment-based screening and rational drug design: Application to the discovery of nanomolar inhibitors of Src SH2." *Journal of Medicinal Chemistry*, **45** (12): 2379 - 2387.
- Lewell, X. Q., Judd, D. B., Watson, S. P. and Hann, M. M. (1998) "RECAP - Retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry." *Journal of Chemical Information and Computer Sciences*, **38** (3): 511 - 522.
- Lewis, W. G., Green, L. G., Grynszpan, F., Radic, Z., Carlier, P. R., Taylor, P., Finn, M. G. and Sharpless, K. B. (2002) "Click chemistry in situ: Acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks." *Angewandte Chemie-International Edition*, **41** (6): 1053.
- Lichtenthaler, H. K. (1999) "The 1-Deoxy-D-Xylulose-5-Phosphate Pathway of Isoprenoid Biosynthesis in Plants." *Annu Rev Plant Physiol Plant Mol Biol*, **50** 47 - 65.
- Liebeschuetz, J. W., Jones, S. D., Morgan, P. J., Murray, C. W., Rimmer, A. D., Roscoe, J. M. E., Waszkowycz, B., Welsh, P. M., Wylie, W. A., Young, S. C., Martin, H., Mahler, J., Brady, L. and Wilkinson, K. (2002) "PRO\_SELECT: Combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors." *Journal of Medicinal Chemistry*, **45** (6): 1221 - 1232.
- Limited, D. J. J. P. (1993). *MOPAC 6.0, Quantum Chemical Program Exchange*. Tokyo:.

- Lipinski, C. A., Lombardo, F., Dominy, B. W. and Feeney, P. J. (1997) "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." *Advanced Drug Delivery Reviews*, **23** (1-3): 3 - 25.
- Lipinski, C. A., Lombardo, F., Dominy, B. W. and Feeney, P. J. (2001) "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." *Adv Drug Deliv Rev*, **46** (1-3): 3 - 26.
- Liu, G., Huth, J. R., Olejniczak, E. T., Mendoza, R., DeVries, P., Leitz, S., Reilly, E. B., Okasinski, G. F., Fesik, S. W. and von Geldern, T. W. (2001) "Novel p-arylthio cinnamides as antagonists of leukocyte function-associated antigen-1/intracellular adhesion molecule-1 interaction. 2. Mechanism of inhibition and structure-based improvement of pharmaceutical properties." *Journal of Medicinal Chemistry*, **44** (8): 1202 - 1210.
- Liu, G., Xin, Z. L., Pei, Z. G., Hajduk, P. J., bad-Zapatero, C., Hutchins, C. W., Zhao, H. Y., Lubben, T. H., Ballaron, S. J., Haasch, D. L., Kaszubska, W., Rondinone, C. M., Trevillyan, J. M. and Jirousek, M. R. (2003) "Fragment screening and assembly: A highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor." *Journal of Medicinal Chemistry*, **46** (20): 4232 - 4235.
- Liu, Z. M., Huang, C. K., Fan, K. Q., Wei, P., Chen, H., Liu, S. Y., Pei, J. F., Shi, L., Li, B., Yang, K., Liu, Y. and Lai, L. H. (2005) "Virtual screening of novel noncovalent inhibitors for SARS-CoV 3C-like proteinase." *Journal of Chemical Information and Modeling*, **45** (1): 10 - 17.
- Long, S. B., Casey, P. J. and Beese, L. S. (2002) "Reaction path of protein farnesyltransferase at atomic resolution." *Nature*, **419** (6907): 645 - 650.
- Lundqvist, T. (2005) "The devil is still in the details--driving early drug discovery forward with biophysical experimental methods." *Curr Opin Drug Discov Devel*, **8** (4): 513 - 519.
- Lyne, P. D., Kenny, P. W., Cosgrove, D. A., Deng, C., Zabudoff, S., Wendoloski, J. J. and Ashwell, S. (2004) "Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening." *J Med Chem*, **47** (8): 1962 - 1968.
- Mac Sweeney, A., Lange, R., Fernandes, R. P., Schulz, H., Dale, G. E., Douangamath, A., Proteau, P. J. and Oefner, C. (2005) "The crystal structure of E.coli 1-deoxy-D-xylulose-5-phosphate reductoisomerase in a ternary complex with the antimalarial compound

- fosmidomycin and NADPH reveals a tight-binding closed enzyme conformation. "*J Mol Biol*, **345** (1): 115 - 127.
- MacKarell, A. D., D., B., Bellott, R. L. and K. Arplus, D. (1998), *J.Phys.Chem.B.*, 3586 - 3616.
- Machicado, C., Lopez-Llano, J., Cuesta-Lopez, S., Bueno, M. and Sancho, J. (2005) "Design of ligand binding to an engineered protein cavity using virtual screening and thermal up-shift evaluation. "*J Comput Aided Mol Des*, **19** (6): 421-43.
- Maly, D. J., Choong, I. C. and Ellman, J. A. (2000) "Combinatorial target-guided ligand assembly: Identification of potent subtype-selective c-Src inhibitors. "*Proceedings of the National Academy of Sciences of the United States of America*, **97** (6): 2419 - 2424.
- Marshall, C. J. (1993) "Protein prenylation: a mediator of protein-protein interactions. "*Science*, **259** (5103): 1865 - 1866.
- Matthews, B. W. (1988) "Structural Basis of the Action of Thermolysin and Related Zinc Peptidases. "*Accounts of Chemical Research*, **21** (9): 333 - 340.
- McGovern, S. L. and Shoichet, B. K. (2003b) "Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. "*J Med Chem*, **46** (14): 2895 - 2907.
- McGovern, S. L., Caselli, E., Grigorieff, N. and Shoichet, B. K. (2002) "A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. "*J.Med.Chem.*, **45** (8): 1712 - 1722.
- McGovern, S. L., Helfand, B. T., Feng, B. and Shoichet, B. K. (2003a) "A specific mechanism of nonspecific inhibition. "*Journal of Medicinal Chemistry*, **46** (20): 4265 - 4272.
- Medina-Franco, J. L., Rodriguez-Morales, S., Juarez-Gordiano, C., Hernandez-Campos, A. and Castillo, R. (2004) "Docking-based CoMFA and CoMSIA studies of non-nucleoside reverse transcriptase inhibitors of the pyridinone derivative type. "*J Comput Aided Mol Des*, **18** (5): 345 - 360.
- Merkle, L., de Andres-Gomez, A., Dick, B., Cox, R. J. and Godfrey, C. R. A. (2005) "A fragment-based approach to understanding inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase. "*Chembiochem*, **6** (10): 1866 - 1874.

- Mesa, R. A. (2006) "Tipifarnib: farnesyl transferase inhibition at a crossroads. " *Expert Rev Anticancer Ther*, **6** (3): 313 - 319.
- Meyer, B. and Peters, T. (2003) "NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. " *Angewandte Chemie-International Edition*, **42** (8): 864 - 890.
- Michel, J., Verdonk, M. L. and Essex, J. W. (2006) "Protein-ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization? " *J Med Chem*, **49** (25): 7427 - 7439.
- Missinou, M. A., Borrmann, S., Schindler, A., Issifou, S., Adegnika, A. A., Matsiegui, P. B., Binder, R., Lell, B., Wiesner, J., Baranek, T., Jomaa, H. and Kremsner, P. G. (2002) "Fosmidomycin for malaria. " *Lancet*, **360** (9349): 1941 – 1942.
- Mitchell, J.B.O., Laskowski, R.A., Alex, A., Thornton, J.M., "BLEEP – Potential of mean force describing protein-ligand interactions: I. Generating potential." *J Comput Chem*, (1999), **20**:1165-1176.
- Miteva, M. A., Lee, W. H., Montes, M. O. and Villoutreix, B. O. (2005) "Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. " *J Med Chem*, **48** (19): 6012-22.
- Mitsch, A., Wissner, P., Silber, K., Haebel, P., Sattler, I., Klebe, G. and Schlitzer, M. (2004) "Non-thiol farnesyltransferase inhibitors: N-(4-tolylacetyl-amino-3-benzoylphenyl)-3-arylfurylacrylic acid amides. " *Bioorganic und Medicinal Chemistry*, **12** (17): 4585 – 4600.
- Monzingo, A. F. and Matthews, B. W. (1984) "Binding of N-carboxymethyl dipeptide inhibitors to thermolysin determined by X-ray crystallography: a novel class of transition-state analogues for zinc peptidases. " *Biochemistry*, **23** (24): 5724-9.
- Mooij, W. T. M. and Verdonk, M. L. (2005) "General and targeted statistical potentials for protein-ligand interactions. " *Proteins-Structure Function and Bioinformatics*, **61** (2): 272 - 287.
- Mooij, W. T., Hartshorn, M. J., Tickle, I. J., Sharff, A. J., Verdonk, M. L. and Jhoti, H. (2006) "Automated protein-ligand crystallography for structure-based drug design. " *ChemMedChem*, **1** (8): 827 - 838.

- Moomaw, J. F. and Casey, P. J. (1992) "Mammalian protein geranylgeranyltransferase. Subunit composition and metal requirements. "*J Biol Chem*, **267** (24): 17438 - 17443.
- Morgan, G. and Fruton, J. S. (1978) "Kinetics of Action of Thermolysin on Peptide Substrates. "*Biochemistry*, **17** (17): 3562 - 3568.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998) "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. "*Journal of Computational Chemistry*, **19** (14): 1639 - 1662.
- Morris, G. M., Goodsell, D. S., Huey, R. and Olson, A. J. (1996) "Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. "*J Comput Aided Mol Des*, **10** (4): 293 - 304.
- Moy, F. J., Haraki, K., Mobilio, D., Walker, G., Powers, R., Tabei, K., Tong, H. and Siegel, M. M. (2001) "MS/NMR: A structure-based approach for discovering protein ligands and for drug design by coupling size exclusion chromatography, mass spectrometry, and nuclear magnetic resonance spectroscopy. "*Analytical Chemistry*, **73** (3): 571 – 581.
- Mozziconacci, J. C., Arnoult, E., Bernard, P., Do, Q. T., Marot, C. and Morin-Allory, L. (2005) "Optimization and validation of a docking-scoring protocol; Application to virtual screening for COX-2 inhibitors. "*Journal of Medicinal Chemistry*, **48** (4): 1055 – 1068.
- Mpamhanga, C. P., Chen, B. N., Mclay, I. M., Ormsby, D. L. and Lindvall, M. K. (2005) "Retrospective docking study of PDE4B ligands and an analysis of the behavior of selected scoring functions 1. "*Journal of Chemical Information and Modeling*, **45** (4): 1061 – 1074.
- Muegge, I. (1999) "Protein-ligand docking with a knowledge-based scoring function. "*Abstracts of Papers of the American Chemical Society*, **218** U493 – U493.
- Muegge, I. (2006) "PMF scoring revisited. "*J Med Chem*, **49** (20): 5895 - 5902.
- Muegge, I. and Martin, Y. C. (1999b) "A general and fast scoring function for protein-ligand interactions: A simplified potential approach. "*Journal of Medicinal Chemistry*, **42** (5): 791 - 804.
- Muegge, I., Martin, Y. C., Hajduk, P. J. and Fesik, S. W. (1999a) "Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. "*J Med Chem*, **42** (14): 2498 - 2503.

- Mueller, C., Schwender, J., Zeidler, J. and Lichtenthaler, H. K. (2000) "Properties and inhibition of the first two enzymes of the non-mevalonate pathway of isoprenoid biosynthesis. " *Biochem Soc Trans*, **28** (6): 792 - 793.
- Murray, C. W. and Verdonk, M. L. (2002) "The consequences of translational and rotational entropy lost by small molecules on binding to proteins. " *J.Comput.Aided Mol.Des*, **16** (10): 741 - 753.
- Navaza, J. (1994) "Amore - An Automated Package for Molecular Replacement. " *Acta Crystallographica Section A*, **50** 157 - 163.
- Nettles, J. H., Jenkins, J. L., Bender, A., Deng, Z., Davies, J. W. and Glick, M. (2006) "Bridging chemical and biological space: 'target fishing' using 2D and 3D molecular descriptors. " *J Med Chem*, **49** (23): 6802 - 6810.
- Neumann, T., Junker, H. D., Keil, O., Burkert, K., Otleben, H., Gamer, J., Sekul, R., Deppe, H., Feurer, A., Tomandl, D. and Metz, G. (2005) "Discovery of thrombin inhibitor fragments from chemical microarray screening. " *Letters in Drug Design und Discovery*, **2** (8): 590 - 594.
- O'Connor, S. J., Barr, K. J., Wang, L., Sorensen, B. K., Tasker, A. S., Sham, H., Ng, S. C., Cohen, J., Devine, E., Cherian, S., Saeed, B., Zhang, H., Lee, J. Y., Warner, R., Tahir, S., Kovar, P., Ewing, P., Alder, J., Mitten, M., Leal, J., Marsh, K., Bauch, J., Hoffman, D. J., Sebti, S. M. and Rosenberg, S. H. (1999) "Second-generation peptidomimetic inhibitors of protein farnesyltransferase demonstrating improved cellular potency and significant in vivo efficacy. " *J Med Chem*, **42** (18): 3701 - 3710.
- Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N. and Hirono, S. (2006) "Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. " *J Chem Inf Model*, **46** (1): 380 – 391.
- Ohkanda, J., Lockman, J.W., Yokoyama, K., Gelb, M.H., Croft, S.L., Sebti, S.M., Hamilton, A.D., (2001) "Peptidomimetic inhibitors of protein farnesyltransferase show poteiteal antimalarial activity", *Bioorg. Med. Chem. Lett.*, **11** (6): 761-764
- Olsen, L., Pettersson, I., Hemmingsen, L., Adolph, H. and Jorgensen, F. S. (2004) "Docking and scoring of metallo-beta-lactamases inhibitors. " *J Comput Aided Mol Des*, **18** (4): 287-302.

- Oprea, T. I. (2002) "Current trends in lead discovery: are we looking for the appropriate properties?" *Mol.Divers.*, **5** (4): 199 - 208.
- Oprea, T. I. and Matter, H. (2004) "Integrating virtual screening in lead discovery." *Curr Opin Chem Biol*, **8** (4): 349 - 358.
- Oprea, T. I., Blaney, J. M. (2006). Cheminformatics Approaches to Fragment-based Lead Discovery. In W. Jahnke, D. A. Erlanson and (Ed.), *Fragment-based Approaches in Lead Discovery* (pp. 91). Weinheim:Wiley-VCH.
- Oprea, T. I., Davis, A. M., Teague, S. J. and Leeson, P. D. (2001) "Is there a difference between leads and drugs? A historical perspective." *Journal of Chemical Information and Computer Sciences*, **41** (5): 1308 - 1315.
- Ortiz, A. R., Pisabarro, M. T., Gago, F. and Wade, R. C. (1995) "Prediction of drug binding affinities by comparative binding energy analysis." *J Med Chem*, **38** (14): 2681 - 2691.
- Osterberg, F., Morris, G. M., Sanner, M. F., Olson, A. J. and Goodsell, D. S. (2002) "Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock." *Proteins*, **46** (1): 34 - 40.
- Otwinowski, Z. and Minor, W. (1997) "Processing of X-ray diffraction data collected in oscillation mode." *Macromolecular Crystallography, Pt A*, **276** 307 - 326.
- Pang, Y. P., Quiram, P., Jelacic, T., Hong, F. and Brimijoin, S. (1996) "Highly potent, selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase - Steps toward novel drugs for treating Alzheimer's disease." *Journal of Biological Chemistry*, **271** (39): 23646 - 23649.
- Park, H. W., Boduluri, S. R., Moomaw, J. F., Casey, P. J. and Beese, L. S. (1997) "Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution." *Science*, **275** (5307): 1800 - 1804.
- Pastor, M., Cruciani, G. and Clementi, S. (1997) "Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships." *J Med Chem*, **40** (10): 1455 - 1464.
- Pearlstein, R. A., Vaz, R. J., Kang, J., Chen, X. L., Preobrazhenskaya, M., Shchekotikhin, A. E., Korolev, A. M., Lysenkova, L. N., Miroshnikova, O. V., Hendrix, J. and Rampe, D.

- (2003) "Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches. *Bioorg Med Chem Lett*, **13** (10): 1829 - 1835.
- Pedretti, A., Villa, L. and Vistoli, G. (2002) "Modeling of binding modes and inhibition mechanism of some natural ligands of farnesyl transferase using molecular docking. *J Med Chem*, **45** (7): 1460-5.
- Peng, Y., Keenan, S. M. and Welsh, W. J. (2005) "Structural model of the Plasmodium CDK, Pfmrk, a novel target for malaria therapeutics. *J Mol Graph Model*, **24** (1): 72-80.
- Perola, E., Walters, W. P. and Charifson, P. S. (2004) "A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins*, **56** (2): 235 - 249.
- Phaosiri, C. and Proteau, P. J. (2004) "Substrate analogs for the investigation of deoxyxylulose 5-phosphate reductoisomerase inhibition: synthesis and evaluation. *Bioorganic und Medicinal Chemistry Letters*, **14** (21): 5309 - 5312.
- Pierce, A. C., Rao, G. and Bemis, G. W. (2004) "BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38, and HIV protease. *Journal of Medicinal Chemistry*, **47** (11): 2768 - 2775.
- Pirard, B. (2005) "Knowledge-driven lead discovery. *Mini Rev Med Chem*, **5** (11): 1045 - 1052.
- Proteau, P. J. (2004) "1-Deoxy-D-xylulose 5-phosphate reductoisomerase: an overview. *Bioorganic Chemistry*, **32** (6): 483 - 493.
- Proteau, P. J., Woo, Y. H., Williamson, R. T. and Phaosiri, C. (1999) "Stereochemistry of the reduction step mediated by recombinant 1-deoxy-D-xylulose 5-phosphate isomeroreductase. *Organic Letters*, **1** (6): 921 - 923.
- Proudfoot, J. R. (2002) "Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg.Med.Chem.Lett.*, **12** (12): 1647 - 1650.
- Qian, Y. M., Blaskovich, M. A., Saleem, M., Seong, C. M., Wathen, S. P., Hamilton, A. D. and Sebt, S. M. (1994) "Design and Structural Requirements of Potent Peptidomimetic Inhibitors of P21(Ras) Farnesyltransferase. *Journal of Biological Chemistry*, **269** (17): 12410 - 12413.

- Qian, Y. M., Vogt, A., Sebti, S. M. and Hamilton, A. D. (1996) "Design and synthesis of non-peptide Ras CAAX mimetics as potent farnesyltransferase inhibitors. " *Journal of Medicinal Chemistry*, **39** (1): 217 - 223.
- Radestock, S., Bohm, M. and Gohlke, H. (2005) "Improving binding mode predictions by docking into protein-specifically adapted potential fields. " *J Med Chem*, **48** (17): 5466 - 5479.
- Rao, J. H. and Whitesides, G. M. (1997) "Tight binding of a dimeric derivative of vancomycin with dimeric L-Lys-D-Ala-D-Ala. " *Journal of the American Chemical Society*, **119** (43): 10286 - 10290.
- Rao, J. H., Lahiri, J., Weis, R. M. and Whitesides, G. M. (2000) "Design, synthesis, and characterization of a high-affinity trivalent system derived from vancomycin and L-Lys-D-Ala-D-Ala. " *Journal of the American Chemical Society*, **122** (12): 2698 - 2710.
- Rarey, M., Kramer, B. and Lengauer, T. (1995) "Time-efficient docking of flexible ligands into active sites of proteins. " *Proc Int Conf Intell Syst Mol Biol*, **3** 300 - 308.
- Rarey, M., Kramer, B. and Lengauer, T. (1997) "Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. " *J Comput Aided Mol Des*, **11** (4): 369 - 384.
- Rarey, M., Kramer, B. and Lengauer, T. (1999) "Docking of hydrophobic ligands with interaction-based matching algorithms. " *Bioinformatics*, **15** (3): 243 - 250.
- Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. (1996) "A fast flexible docking method using an incremental construction algorithm. " *J Mol Biol*, **261** (3): 470 - 489.
- Rees, D. C., Congreve, M., Murray, C. W. and Carr, R. (2004) "Fragment-based lead discovery. " *Nature Reviews Drug Discovery*, **3** (8): 660 - 672.
- Reese, J. H. and Maltese, W. A. (1991) "Post-translational modification of proteins by 15-carbon and 20-carbon isoprenoids in three mammalian cell lines. " *Mol Cell Biochem*, **104** (1-2): 109 - 116.
- Reichenberg, A., Wiesner, J., Weidemeyer, C., Dreiseidler, E., Sanderbrand, S., Altincicek, B., Beck, E., Schlitzer, M. and Jomaa, H. (2001) "Diaryl ester prodrugs of FR900098 with improved in vivo antimalarial activity. " *Bioorg Med Chem Lett*, **11** (6): 833 - 835.

- Reiss, Y., Goldstein, J. L., Seabra, M. C., Casey, P. J. and Brown, M. S. (1990) "Inhibition of purified p21ras farnesyl:protein transferase by Cys-AAX tetrapeptides. "*Cell*, **62** (1): 81 - 88.
- Reuter, K., Sanderbrand, S., Jomaa, H., Wiesner, J., Steinbrecher, I., Beck, E., Hintz, M., Klebe, G. and Stubbs, M. T. (2002) "Crystal structure of 1-deoxy-D-xylulose-5-phosphate reductoisomerase, a crucial enzyme in the non-mevalonate pathway of isoprenoid biosynthesis. "*J Biol Chem*, **277** (7): 5378-84.
- Ricagno, S., Grolle, S., Bringer-Meyer, S., Sahn, H., Lindqvist, Y. and Schneider, G. (2004) "Crystal structure of 1-deoxy-d-xylulose-5-phosphate reductoisomerase from *Zymomonas mobilis* at 1.9-Å resolution. "*Biochim Biophys Acta*, **1698** (1): 37 - 44.
- Rishton, G. M. (2003) "Nonleadlikeness and leadlikeness in biochemical screening. "*Drug Discov.Today*, **8** (2): 86 - 96.
- Roche, O., Schneider, P., Zuegge, J., Guba, W., Kansy, M., Alanine, A., Bleicher, K., Danel, F., Gutknecht, E. M., Rogers-Evans, M., Neidhart, W., Stalder, H., Dillon, M., Sjogren, E., Fotouhi, N., Gillespie, P., Goodnow, R., Harris, W., Jones, P., Taniguchi, M., Tsujii, S., von der, S. W., Zimmermann, G. and Schneider, G. (2002) "Development of a virtual screening method for identification of 'frequent hitters' in compound libraries. "*J.Med.Chem.*, **45** (1): 137 - 142.
- Rohdich, F., Eisenreich, W., Wungsintaweeikul, J., Hecht, S., Schuhr, C. A. and Bacher, A. (2001a) "Biosynthesis of terpenoids. 2C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) from *Plasmodium falciparum*. "*Eur J Biochem*, **268** (11): 3190 - 3197.
- Rohdich, F., Hecht, S., Gartner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A. and Eisenreich, W. (2002) "Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. "*Proc Natl Acad Sci U S A*, **99** (3): 1158 - 1163.
- Rohdich, F., Kis, K., Bacher, A. and Eisenreich, W. (2001b) "The non-mevalonate pathway of isoprenoids: genes, enzymes and intermediates. "*Curr Opin Chem Biol*, **5** (5): 535 - 540.
- Rong, S. B., Enyedy, I. J., Qiao, L., Zhao, L., Ma, D., Pearce, L. L., Lorenzo, P. S., Stone, J. C., Blumberg, P. M., Wang, S. and Kozikowski, A. P. (2002) "Structural basis of RasGRP binding to high-affinity PKC ligands. "*J Med Chem*, **45** (4): 853 - 860.

- Sadowski, J. and Bostrom, J. (2006) "MIMUMBA revisited: torsion angle rules for conformer generation derived from X-ray structures. *J Chem Inf Model*, **46** (6): 2305 - 2309.
- Sadowski, J., Schwab, C., Gasteiger, J. *CORINA, 3D Structure Generator* version 2.6, University of Erlangen, Germany.
- Sakowski, J., Bohm, M., Sattler, I., Dahse, H. M. and Schlitzer, M. (2001) "Synthesis, molecular modeling, and structure-activity relationship of benzophenone-based CAAX-peptidomimetic farnesyltransferase inhibitors. *Journal of Medicinal Chemistry*, **44** (18): 2886 - 2899.
- Salo, O. M. H., Raitio, K. H., Savinainen, J. R., Nevalainen, T., Lahtela-Kakkonen, M., Laitinen, J. T., Jarvinen, T. and Poso, A. (2005) "Virtual screening of novel CB2 ligands using a comparative model of the human cannabinoid CB2 receptor. *J Med Chem*, **48** (23): 7166-71.
- Schlechter, I. and Berger, A. (1967) "On the Size of the active Site in Proteases. *Biochem. Biophys. Res. Commun.*, **27** (2): 157 - 162.
- Schlitzer, M. (2005) "Inhibition of Farnesyltransferase as a Strategy for the Development of Novel Anti-Malarials. *Current Medicinal Chemistry - Anti-Infective Agents*, **4** (3): 277-286.
- Schlitzer, M. and Sattler, I. (1999) "Design, synthesis, and evaluation of novel modular bisubstrate analogue inhibitors of farnesyltransferase. *Angewandte Chemie-International Edition*, **38** (13-14): 2032 - 2034.
- Schlitzer, M. and Sattler, I. (2000) "Non-thiol farnesyltransferase inhibitors: the concept of benzophenone-based bisubstrate analogue farnesyltransferase inhibitors. *European Journal of Medicinal Chemistry*, **35** (7-8): 721 - 726.
- Schneider, G. and Bohm, H. J. (2002) "Virtual screening and fast automated docking methods. *Drug Discov Today*, **7** (1): 64 - 70.
- Schneider, G. and Fechner, U. (2005) "Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov*, **4** (8): 649 - 663.

- Schuffenhauer, A., Ruedisser, S., Marzinzik, A. L., Jahnke, W., Blommers, M., Selzer, P. and Jacoby, E. (2005) "Library design for fragment based screening. " *Current Topics in Medicinal Chemistry*, **5** (8): 751 - 762.
- Schwender, J., Seemann, M., Lichtenthaler, H. K. and Rohmer, M. (1996) "Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*. " *Biochem J*, **316** ( Pt 1) 73 - 80.
- Sebti, S. M. and Hamilton, A. D. (1998) "New approaches to anticancer drug design based on the inhibition of farnesyltransferase. " *Drug Discovery Today*, **3** (1): 26 - 33.
- Seto, H. (1997) "[Non-mevalonate pathway. A new pathway for the biosynthesis of isopentenyl diphosphate]. " *Tanpakushitsu Kakusan Koso*, **42** (16): 2590 - 2600.
- Sheldrick, G. M. and Schneider, T. R. (1997) "SHELXL: High-resolution refinement. " *Macromolecular Crystallography, Pt B*, **277** 319 - 343.
- Shoichet, B. K., Kuntz, I. D. and Bodian, D. L. (1992) "Molecular docking using shape descriptors. " *Journal of Computational Chemistry*, **13** (3): 380 - 397.
- Shoichet, B. K., McGovern, S. L., Wei, B. and Irwin, J. J. (2002) "Lead discovery using molecular docking. " *Curr Opin Chem Biol*, **6**: 439 - 446.
- Shoichet, B. K., Stroud, R. M., Santi, D. V., Kuntz, I. D. and Perry, K. M. (1993) "Structure-based discovery of inhibitors of thymidylate synthase. " *Science*, **259** (5100): 1445 - 1450.
- Shuker, S. B., Hajduk, P. J., Meadows, R. P. and Fesik, S. W. (1996) "Discovering high-affinity ligands for proteins: SAR by NMR. " *Science*, **274** (5292): 1531 - 1534.
- Silber, K. (2001). *Ligand- und Rezeptor-basierte Suchstrategien zum Auffinden neuer Proteininhibitoren*. Halle Wittenberg:.
- Singh, N., Cheve, G., Avery, M. A. and McCurdy, C. R. (2006) "Comparative protein modeling of 1-deoxy-D-xylulose-5-phosphate reductoisomerase enzyme from *Plasmodium falciparum*: A potential target for antimalarial drug discovery. " *Journal of Chemical Information and Modeling*, **46** (3): 1360 - 1370.
- Sippl, W. (2002) "Development of biologically active compounds by combining 3D QSAR and structure-based design methods. " *J Comput Aided Mol Des*, **16** (11): 825 - 830.

- Sivanesan, D., Rajnarayanan, R. V., Doherty, J. and Pattabiraman, N. (2005) "In-silico screening using flexible ligand binding pockets: a molecular dynamics-based approach. "*J Comput Aided Mol Des*, **19** (4): 213 - 228.
- Sotriffer, C. A., Sanschagrin Paul; Matter, Matter, H., Klebe, G. (2007) "SFCscore: Scoring functions for affinity prediction of protein-ligand complexes. "*submitted*.
- Sotriffer, C. A. and Dramburg, I. (2005) "In situ cross-docking to simultaneously address multiple targets. "*J Med Chem*, **48** (9): 3122-5.
- Sotriffer, C. A., Gohlke, H. and Klebe, G. (2002a) "Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. "*J Med Chem*, **45** (10): 1967-70.
- Sotriffer, C. and Klebe, G. (2002b) "Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. "*Il Farmaco*, **57** (3): 243-51.
- Sousa, S. F., Fernandes, P. A. and Ramos, M. J. (2006) "Protein-ligand docking: current status and future challenges. "*Proteins*, **65** (1): 15 - 26.
- Springer, C., Adalsteinsson, H., Young, M. M., Kegelmeyer, P. W. and Roe, D. C. (2005) "PostDOCK: a structural, empirical approach to scoring protein ligand complexes. "*J Med Chem*, **48** (22): 6821 - 6831.
- Stahl, M. and Rarey, M. (2001) "Detailed analysis of scoring functions for virtual screening. "*J Med Chem*, **44** (7): 1035 - 1042.
- Steinbacher, S., Kaiser, J., Eisenreich, W., Huber, R., Bacher, A. and Rohdich, F. (2003) "Structural basis of fosmidomycin action revealed by the complex with 2-C-methyl-D-erythritol 4-phosphate synthase (IspC). Implications for the catalytic mechanism and anti-malaria drug development. "*J Biol Chem*, **278** (20): 18401 - 18407.
- Stewart, J. (1990) "MOPAC: a semiempirical molecular orbital program. "*J Comput Aided Mol Des*, **4** (1): 1 - 105.
- Strickland, C. L., Windsor, W. T., Syto, R., Wang, L., Bond, R., Wu, Z., Schwartz, J., Le, H. V., Beese, L. S. and Weber, P. C. (1998) "Crystal structure of farnesyl protein transferase complexed with a CaaX peptide and farnesyl diphosphate analogue. "*Biochemistry*, **37** (47): 16601-11.

- Swayze, E. E., Jefferson, E. A., Sannes-Lowery, K. A., Blyn, L. B., Risen, L. M., Arakawa, S., Osgood, S. A., Hofstadler, S. A. and Griffey, R. H. (2002) "SAR by MS: A ligand based technique for drug lead discovery against structured RNA targets. *Journal of Medicinal Chemistry*, **45** (18): 3816 - 3819.
- Szczepankiewicz, B. G., Liu, G., Hajduk, P. J., bad-Zapatero, C., Pei, Z. H., Xin, Z. L., Lubben, T. H., Trevillyan, J. M., Stashko, M. A., Ballaron, S. J., Liang, H., Huang, F., Hutchins, C. W., Fesik, S. W. and Jirousek, M. R. (2003) "Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *Journal of the American Chemical Society*, **125** (14): 4087 - 4096.
- TRIPOS (2002). *SYBYL Molecular Modeling Software*. St. Louis, MO:.
- Takahashi, S., Kuzuyama, T., Watanabe, H. and Seto, H. (1998) "A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proc Natl Acad Sci U S A*, **95** (17): 9879 - 9884.
- Tame, J. R. (1999) "Scoring functions: a view from the bench. *J Comput Aided Mol Des*, **13** (2): 99 - 108.
- Tame, J. R. (2005) "Scoring functions--the first 100 years. *J Comput Aided Mol Des*, **19** (6): 445 - 451.
- Taylor, R. D., Jewsbury, P. J. and Essex, J. W. (2002) "A review of protein-small molecule docking methods. *J Comput Aided Mol Des*, **16** (3): 151 - 166.
- Teague, S. J., Davis, A. M., Leeson, P. D. and Oprea, T. (1999) "The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed Engl*, **38** (24): 3743-3748.
- Terp, G. E., Johansen, B. N., Christensen, I. T. and Jorgensen, F. S. (2001) "A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J Med Chem*, **44** (14): 2333 - 2343.
- Thibaut, U., Folkers, G., Klebe, G., Kubinyi, H., Merz, A. and Rognan, D. (1993). Recommendations to CoMFA Studies and 3D QSAR Publications. In H. Kubinyi (Ed.), *3D QSAR in Drug Design. Theory, Methods and Applications* (pp. 711 - 716). Leiden, The Netherlands:ESCOM.

- Tobin, D. A., Pickett, J. S., Hartman, H. L., Fierke, C. A. and Penner-Hahn, J. E. (2003) "Structural characterization of the zinc site in protein farnesyltransferase." *J Am Chem Soc*, **125** (33): 9962 - 9969.
- Toledo-Sherman, L., Deretey, E., Slon-Usakiewicz, J. J., Ng, W., Dai, J. R., Foster, J. E., Redden, P. R., Uger, M. D., Liao, L. C., Pasternak, A. and Reid, N. (2005) "Frontal affinity chromatography with MS detection of EphB2 tyrosine kinase receptor. 2. Identification of small-molecule inhibitors via coupling with virtual screening 1." *Journal of Medicinal Chemistry*, **48** (9): 3221 - 3230.
- Totrov, M. and Abagyan, R. (1997) "Flexible protein-ligand docking by global energy optimization in internal coordinates." *Proteins*, **Suppl 1** 215 - 220.
- Tronrud, D. E., Monzingo, A. F. and Matthews, B. W. (1986) "Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin." *Eur J Biochem*, **157** (2): 261-8.
- Vasudevan, A., Qian, Y. M., Vogt, A., Blaskovich, M. A., Ohkanda, J., Sebti, S. M. and Hamilton, A. D. (1999) "Potent, highly selective, and non-thiol inhibitors of protein geranylgeranyltransferase-I." *Journal of Medicinal Chemistry*, **42** (8): 1333 - 1340.
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W. and Kopple, K. D. (2002) "Molecular properties that influence the oral bioavailability of drug candidates." *J. Med. Chem.*, **45** (12): 2615 - 2623.
- Velec, H. F., Gohlke, H. and Klebe, G. (2005) "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*, **48** (20): 6296 - 6303.
- UNITY, Chemical information software* version 4.0, Tripos Inc., St.,Loius, MO USA.
- Verdonk, M. L., Berdini, V., Hartshorn, M. J., Mooij, W. T., Murray, C. W., Taylor, R. D. and Watson, P. (2004) "Virtual screening using protein-ligand docking: avoiding artificial enrichment." *J Chem Inf Comput Sci*, **44** (3): 793 - 806.
- Verdonk, M. L., Chessari, G., Cole, J. C., Hartshorn, M. J., Murray, C. W., Nissink, J. W., Taylor, R. D. and Taylor, R. (2005) "Modeling water molecules in protein-ligand docking using GOLD." *J Med Chem*, **48** (20): 6504 - 6515.

- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. and Taylor, R. D. (2003) "Improved protein-ligand docking using GOLD. "*Proteins*, **52** (4): 609 - 623.
- Vieth, M., Siegel, M. G., Higgs, R. E., Watson, I. A., Robertson, D. H., Savin, K. A., Durst, G. L. and Hipskind, P. A. (2004) "Characteristic physical properties and structural fragments of marketed oral drugs. "*Journal of Medicinal Chemistry*, **47** (1): 224 – 232.
- Villar, H.O., Koehler, R.T., (2000) "Comments on the design of chemical libraries for screening", *Mol. Divers.*, **5**, 13-24
- Vigers, G. P. and Rizzi, J. P. (2004) "Multiple active site corrections for docking and virtual screening. "*J Med Chem*, **47** (1): 80 - 89.
- Virag, I., Polgar, T. and Keseru, G. M. (2005) "Functional virtual screening of estrogen receptor alpha modulators by FlexX-Pharm. "*Journal of Molecular Structure-Theochem*, **725** (1-3): 239 - 242.
- WHO (2005). *The Africa Malaria Report*. Nairobi/Geneve/New York:.
- Walters, W., Stahl, M. and Murcko, M. (1998) "Virtual Screening - an overview. "*Drug Discovery Today*, **3** (4): 160 - 178.
- Wang, R. and Wang, S. (2001a) "How does consensus scoring work for virtual library screening? An idealized computer experiment. "*J Chem Inf Comput Sci*, **41** (5): 1422 - 1426.
- Wang, R., Lu, Y. and Wang, S. (2003) "Comparative evaluation of 11 scoring functions for molecular docking. "*J Med Chem*, **46** (12): 2287 - 2303.
- Wang, T. and Wade, R. C. (2001b) "Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. "*J Med Chem*, **44** (6): 961 - 971.
- Wang, T. and Wade, R. C. (2002) "Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. "*J Med Chem*, **45** (22): 4828 - 4837.
- Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E. and Head, M. S. (2006) "A critical assessment of docking programs and scoring functions. "*J Med Chem*, **49** (20): 5912 - 5931.

- Weber, A., Casini, A., Heine, A., Kuhn, D., Supuran, C. T., Scozzafava, A. and Klebe, G. (2004) "Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. "*J Med Chem*, **47** (3): 550-7.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) "A new force field for molecular mechanical simulation of nucleic acids and proteins. "*J.Am.Chem.Soc.*, **106** 765 - 784.
- Welch, W., Ruppert, J. and Jain, A. N. (1996) "Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. "*Chem Biol*, **3** (6): 449 - 462.
- Wendt, M. D., Rockway, T. W., Geyer, A., McClellan, W., Weitzberg, M., Zhao, X. M., Mantei, R., Nienaber, V. L., Stewart, K., Klinghofer, V. and Giranda, V. L. (2004) "Identification of novel binding interactions in the development of potent, selective 2-naphthamidine inhibitors of urokinase. Synthesis, structural analysis, and SAR of N-phenyl amide 6-substitution. "*Journal of Medicinal Chemistry*, **47** (2): 303 - 324.
- Wermuth, C. G. (2004) "Selective optimization of side activities: another way for drug discovery. "*J Med Chem*, **47** (6): 1303 - 1314.
- Wermuth, C. G. (2006a) "Selective optimization of side activities: the SOSA approach. "*Drug Discov Today*, **11** (3-4): 160 - 164.
- Wermuth, C. G. (2006b) "Similarity in drugs: reflections on analogue design. "*Drug Discov Today*, **11** (7-8): 348 - 354.
- White, R. J., Margolis, P. S., Trias, J. and Yuan, Z. Y. (2003) "Targeting metalloenzymes: a strategy that works. "*Current Opinion in Pharmacology*, **3** (5): 502 - 507.
- Wieman, H., Tondel, K., Anderssen, E. and Drablos, F. (2004) "Homology-based modelling of targets for rational drug design. "*Mini Rev Med Chem*, **4** (7): 793 - 804.
- Wiesner, J., Mitsch, A., Jomaa, H. and Schlitzer, M. (2003b) "Structure-activity relationships of novel anti-malarial agents. Part 7: N-(3-benzoyl-4-tolylacetylaminophenyl)-3-(5-aryl-2-furyl)acrylic acid amides with polar moieties. "*Bioorg Med Chem Lett*, **13** (13): 2159 - 2161.
- Wiesner, J., Ortmann, R., Jomaa, H. and Schlitzer, M. (2003a) "New antimalarial drugs. "*Angew Chem Int Ed Engl*, **42** (43): 5274 - 5293.

- Wold, S., Johansson, E. and Cocchi, M. (1993). PLS - Partial Least Squares Projections to Latent Structures I. In H. Kubinyi (Ed.), *3D QSAR in Drug Design. Theory, Methods and Applications* Leiden:ESCOM.
- Wold, S., Ruhe, A., Wold, H. and Dunn I. I. I, W. J. (1984) "The Collinearity Problem in Linear Regression. The Partial Least Squares Approach to Generalized Inverses. "*SIAM J.Sci.Stat.Comput.*, **5** 735 - 743.
- Woo, Y. H., Fernandes, R. P. M. and Proteau, P. J. (2006) "Evaluation of fosmidomycin analogs as inhibitors of the *Synechocystis* sp PCC6803 1-deoxy-D-xylulose 5-phosphate reductoisomerase. "*Bioorganic und Medicinal Chemistry*, **14** (7): 2375 - 2385.
- Wyss, D. F., Arasappan, A., Senior, M. M., Wang, Y. S., Beyer, B. M., Njoroge, F. G. and Mccoy, M. A. (2004) "Non-peptidic small-molecule inhibitors of the single-chain hepatitis C virus NS3 protease/NS4A cofactor complex discovered by structure-based NMR screening. "*Journal of Medicinal Chemistry*, **47** (10): 2486 - 2498.
- Yajima, S., Hara, K., Sanders, J. M., Yin, F. L., Ohsawa, K., Wiesner, J., Jomaa, H. and Oldfield, E. (2004) "Crystallographic structures of two bisphosphonate : 1-deoxyxylulose-5-phosphate reductoisomerase complexes. "*Journal of the American Chemical Society*, **126** (35): 10824 - 10825.
- Yajima, S., Nonaka, T., Kuzuyama, T., Seto, H. and Ohsawa, K. (2002) "Crystal structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase complexed with cofactors: implications of a flexible loop movement upon substrate binding. "*J Biochem (Tokyo)*, **131** (3): 313 - 317.
- Zamora, I., Afzelius, L. and Cruciani, G. (2003) "Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. "*J Med Chem*, **46** (12): 2313 - 2324.
- Zartler, E. R. and Shapiro, M. J. (2005) "Fragonomics: fragment-based drug discovery. "*Curr.Opin.Chem.Biol.*, **9** (4): 366 - 370.
- Zentgraf, M., Fokkens, J. and Sotriffer, C. A. (2006) "Addressing protein flexibility and ligand selectivity by in situ cross-docking. "*ChemMedChem*, **1** (12): 1355-9.
- Zentgraf, M. (2006) "Characterization of Binding Pocket Flexibility of Aldose Reductase" thesis, Philipps-Universitaet Marburg.

## APPENDIX

### ***Acknowledgment of work from collaboration partners***

**DXR project:** All compound synthesis was performed by members of the research groups of Prof. Martin Schlitzer (Marburg), Prof. A. Link (Greifswald) and Dr. Kurz (Hamburg). Biochemical assays were done by Drs. Jochen Wiesner and Hassan Jomaa (Giessen). Drs. Holger Gohlke and Christoph Sotriffer contributed significantly throughout constructive discussions. The DXR project was financed by the Graduiertenkolleg “Protein function on atomic level”, the BMBF Biochance program and the European Union “quality for life, QLK2-CT-2002-00887” project.

**Ftase project:** All compound synthesis was performed by Prof. Martin Schlitzer, Drs. Katja Kettler, Jacek Sakowski, Andreas Mitsch and coworkers. Biochemical assays were done by Drs. Jochen Wiesner and Hassan Jomaa, Justus Liebig Universität, Giessen, Germany. The *Plasmodium falciparum* homology model was done by Dr. Peter Haebel,

**Fragment project:** 3-Chloroaspirin was synthesized by Sascha Brass. Biochemical validation was supported by Christian Sohn and X-ray crystallography was supported by Dr. Holger Steuber and Dr. Andreas Heine. The fragmented structures of the world drug index were provided by Robert Jaeger, Sanofi-Aventis, Frankfurt.

## **Abbreviations**

ADME	administration, distribution, metabolism, excretion
BW	body weight
CAC	commercially available compounds data set
COX	cyclooxygenase
CSD	Cambridge structural database
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DXR	DOXP-Reductoisomerase
ED <sub>50/90</sub>	effective dose, that produces a response in 50/90% of the test species
EMA	European Agency for the Evaluations of Medicinal Products
Fagla	3-(2-furyl)acryloyl-glycyl-leucinamide
FBS	fragment-based screening
FDA	food and drug administration
FTASE	farnesyltransferase
HIV	human immunodeficiency virus
HTS	high throughput screening
K <sub>i</sub>	Inhibition constant
K <sub>m</sub>	michaelis-menten constant
MMP	matrix-metallo-proteinase
MOPS	3-(N-morpholino)propansulfonic acid
MW	molecular weight
NME	new molecular entity
NMR	nuclear magnetic resonance

---

<i>P. falc.</i>	Plasmodium falciparum
PDB	protein data bank
PDE	phosphodiesterase
PLS	Partial least-square
RECAP	retrosynthetic combinatorial analysis procedure
RMSD	(root mean square deviation)
SAS	Solvent Accessible Surface
SFC	Scoring function consortium
SPR	surface plasmon resonance
TLN	Thermolysin
TRIS	2-amino-2-hydroxymethyl-1,3-propanediol
WDI	World Drug Index resp. Drug fragments data set
XRAY	X-ray reference data set

## Overview over described scoring function:

### Boehm function as defined in FlexX:

$$\begin{aligned} \Delta G = & \Delta G_0 + \Delta G_{rot} \times N_{rot} \\ & + \Delta G_{hb} \sum_{neutral\_H-bonds} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{io} \sum_{ionic\_int.} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{aro} \sum_{arom\_int.} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{lipo} \sum_{lipo\_cont.} f^*(\Delta R) \end{aligned}$$

Here,  $f(\Delta R, \Delta \alpha)$  penalizes deviations from the ideal geometry.  $\Delta G$  Coefficients are derived from regression analysis of a training set.

### DrugScore

$$\Delta W = \gamma \sum_{l \in L_k} \sum_{p \in P_k} \Delta W_{T(l), T(p)}(r_i) + (1 - \gamma) \left[ \sum_{l \in L_k} \Delta W_{T(l)}^L(S_i^{Kpl}, S_j^{Frei}) + \sum_{p \in P_k} \Delta W_{T(p)}^P(S_k^{Kpl}, S_l^{Frei}) \right]$$

Interactions  $\Delta W_{i,j}$  between atoms of type  $i$  and  $j$ , located at a distance  $r$ , can be obtained from the normalized radial pair distribution function  $g_{i,j}(r)$ . The binding score between protein and ligand by DrugScore as the sum of all occurring atom-atom interactions. The second part is the solvent accessible surface term, for which a specific description can be found in Gohlke *et al.* (JMB 2000). DrugScore can be used with the PDB or the CSD as data base. The main difference are better statistics in the CSD for several atom types which are rarely found in the PDB (F, S, Cl, Br,)

## AutoDock

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}$$

The first four terms represent dispersion, repulsion, hydrogen bonding, electrostatics and deviations from ideal geometry, “tor” considers restriction of internal rotors “sol” desolvation.

## GOLD

GOLDScore function is a molecular mechanics-like function with four terms:

$$\text{GOLD Fitness} = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int}$$

$S_{hb\_ext}$  is the protein-ligand hydrogen-bond score,  $S_{vdw\_ext}$  is the protein-ligand van der Waals score.  $S_{hb\_int}$  is the contribution due to intramolecular hydrogen bonds in the ligand.  $S_{vdw\_int}$  is the contribution due to intramolecular strain in the ligand

## Chemscore

The Chemscore function estimates free energy of binding of a ligand to a protein as follows:

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} S_{hbond} + \Delta G_{metal} S_{metal} + \Delta G_{lipo} S_{lipo} + \Delta G_{rot} H_{rot}$$

where  $S_{hbond}$ ,  $S_{metal}$ , and  $S_{lipo}$  are scores for hydrogen-binding, acceptor-metal, and lipophilic interactions, respectively.  $H_{rot}$  is a score to compensate for conformational entropy loss. Coefficients are derived from regression analysis of a training set.

## SFC-Score

Regression functions have the general form:

$$pK_i = \left( \sum_k c_k \times desc_k \right) + const.$$

where  $k$  is the number of descriptor variables and  $c_k$  is the coefficient of the descriptor  $desc_k$  in the regression equation.

## Virtual Screening examples

**Table a.1:** Reported virtual screening examples (derived from Coupez, 2006 and extended):

Target	Procedure and methods	Results
T4-lysozyme and W66F/L44A double mutant [Machicado, 2005]	a) Physical-chemical filtering the ligand database. b) ligand flexible and fast docking (LigandFit) procedure + consensus scoring c) fine docking (Affinity) + free energy calculation	3 step Protocol set-up and validation of the protocol with known T4-lysozyme binder. Identification of binders on the W66F/L44A double mutant.
Cannabinoid CB2 receptor (GPCR) [Salo, 2005]	a) Ligand-based screening with pharmacophoric and surface volume constrains b) flexible ligand docking (GOLD) on the homology model of the receptor + consensus scoring	Protocol combining ligand-based approach to filter the database and docking on a homology model of the receptor. 1 hit found in G-protein activation assays
<i>P. falciparum</i> kinase (Pfmrk) [Peng, 2005]	a) homology model based on the hCDK7 X-ray structure refined by QM, energy minimization and molecular dynamics b) docking of oxindole-based Pfmrk inhibitors (GOLD) + scoring (GOLD score) to validate the model.	A 3D structural model of Pfmrk was constructed and refined using homology modeling, ab initio quantum mechanical calculations, and EM and MD simulations. The refined structural model was further evaluated by the molecular docking of a series of known Pfmrk inhibitors. Identification of new in vitro inhibitor hits against Pfmrk.
CDK2 [Cotesta, 2005]	a) Docking (Gold or QXP) b) Scoring functions (PLP, Ligscore, Ludi, Jain, ChemScore, PMF) and four composite scoring models	Using the QXP and GOLD programs, Comparison of the ability of six single scoring functions and four composite scoring models to separate compounds that are active against CDK2 from inactives.

Target	Procedure and methods	Results
Estrogen receptor, thymidine kinase, coagulation factor VIIa, and neuraminidase [Miteva, 2005]	a) Physical-chemical filtering of the ligand database b) Rigid body docking fitting (Fred) c) flexible ligand docking (Surflex or Dock)	Multistep VLS protocol set-up based on sequential docking/scoring steps using a consensus docking / scoring approach. Tuning docking/scoring parameters according to the pocket properties improved the performance of both docking methods.
Bacterial hyaluronan lyase [Botzki, 2005]	a) Homology model based on two streptococcal hyalronan lyases X-ray structures b) Docking and scoring (Ludi) c) Physical-chemical filtering of the docked ligand	Identification of Hits
type 4 phosphodiesterase (PDE4B) [Mpamhanga, 2005]	a) Docking (Ligandfit) b) Scoring functions (PMF, JAIN, PLP2, LigScore2, and DockScore)	Comparison and testing the combination of various scoring functions
Plasmodium falciparum Ca <sup>2+</sup> - ATPase (PfATP6) [Jung, 2005]	a) Homology model of PfATP6 b) Docking (Ligandfit) c) Scoring (LUDI)	Validation of a model of PfATP6 by docking known artemisinin derivatives
Erythropoietin- producing hepatocellular B2 (EphB2) kinase [Toledo-Sherman, 2005]	a) ligand-based pharmacophore search b) Docking (GOLD) + consensus scoring (C-score)	Ligand-based and docking approaches combined with a mass spectrometry screening approach enable the identification of hits as inhibitors for EphB2
cytochrome P450 [Kirton, 2005]	a) Docking (GOLD) b) Scoring (Goldscore, Chemscore, modified Goldscore, modified Chemscore)	Improvement of the scoring functions by providing a better treatment of the environment around the heme, which leads to higher success rates.

Target	Procedure and methods	Results
Alpha1A Receptor GPCR [Evers, 2005]	a) Homology model based on the bovine rhodopsin X-ray structure b) 2D and three dimensional pharmacophore filtering of the ligand database c) Docking (GOLD) d) Scoring (PMF)	The generation and validation of a homology model for the alpha1A receptor is described. The generated model was used to discover alpha1A antagonists hits which were validated in an experimental assay.
Cyclooxygenase-2 [Mozziconacci, 2005]	a) docking (DOCK) b) Consensus scoring (C-score+SCORE) c) geometric criteria filtering	Optimization of the DOCK parameters and validation study of this protocol shows that the SCORE and DOCK scoring functions are the best for this target. Screening retrieved a set of compounds bearing a sulfonyl group and compounds, that are fairly similar to known potent inhibitors. One-third of the dozen tested compounds ended up inhibiting the COX-2 enzyme.
alpha-beta barrel enzymes [Kalyanaraman, 2005]	a) Docking : Glide b) Rescoring: OPLS-AA + a generalized Born implicit solvent model	Presentation of physics-based method for rescoring protein-ligand complexes generated by a docking program in the case of a highly charged active site.
HIV-1 Integrase (IN) [Dayam, 2005]	a) 3D pharmacophore guided database search b) Docking + Scoring (GOLD) c) Selection based on structural novelty, ability to chelate Mg <sup>2+</sup> ion in the active site of IN, and satisfying Lipinski's rule-of-five	Discovery of a new class of HIV-1 Integrase inhibitor, by applying ligand-based screening methods, followed by docking.
Metallo-β-lactamases [Olsen, 2004]	a) Docking + scoring (AutoDock, Gold, FlexX) b) rescoring (GRID)	Evaluation the AutoDock, GOLD and FlexX docking programs for docking of dicarboxylic acid inhibitors into metallo-β-lactamases (MBLs). GOLD provides the best performance.

Target	Procedure and methods	Results
SARS coronavirus 3C-like proteinase [Liu, 2005]	a) model for the receptor built by homology modeling and multicanonical MD method b) Docking and scoring (DOCK) c) Pharmacophore model, consensus scoring, and "drug-like" filters	Discovery of hits that inhibiting the SARS coronavirus 3C-like proteinase, by using a homology model for docking and by applying Pharmacophore model, consensus scoring, and "drug-like" filters
Cytochrome P450 2D6 [Kemp, 2004]	a) Docking b) scoring (ChemScore)	Validation of Model of Cytochrome P450 2D6 by docking.
Aspartic proteinase renin [Krovat, 2004]	a) Docking (Ligandfit) b) Scoring functions (LigScore1, LigScore2, PLP1, PLP2, JAIN, PMF, LUDI)	Comparison and testing the combination of various scoring functions
Chk-1 kinase [Lyne, 2004]	a) Physical-chemical filtering of the ligand database b) 3-D pharmacophore filtering for compounds with kinase binding motifs c) docking (FlexX-Pharm) d) consensus scoring (C-Score)	Protocol based on knowledge-based strategy using ligandbased and docking approaches. 36 of 103 selected compounds for testing (corresponding to four chemical classes) were found to inhibit the enzyme.
Estrogen receptor $\alpha$ [Virag, 2005]	a) Docking (FlexX-Pharm, FlexX) b) Scoring functions (PMF, Chem, Dock, GOLD, FlexX)	The combination of FlexX-Pharm and ChemScore gave the best performance. Increased performance of FlexXPharm over FlexX was also demonstrated on the estrogen receptor.
Carbonic anhydrase II [Gruneberg, 2002]	a) pharmacophore search Unity b) ligand superimposition FlexS c) docking FlexX, DrugScore	Inhibitors with nanomolar affinity ( sulfonamides) including structural validation
tRNA-guanine transglycosylase [Brenk, 2003],	a) Hot spot analyses with SuperStar and DrugScore b) Pharmacophore 3D search Unity c) Docking FlexX, Scoring with Flex and DrugScore	Submicromolar inhibitors with different chemotypes. Extensive affinity and structural validation (multiple X-ray crystal structures)

**Publications arising from this work:****Articles**

Breu B., Silber K, Gohlke H.

*Consensus AFMoC models incorporate ligand and receptor conformational variability into tailor-made scoring functions*

*Submitted*

Silber K., Steuber H., Jaeger, R., Heine, A., Sohn C., Klebe G.

Successful Fragment-based Virtual Screening for Metalloproeinase Inhibitors

*Manuscript n preparation*

Kettler K, Wiesner J, Sakowski J, Silber K, Haebel P, Klebe G, Schlitzer M, *et al.*

Development of Farnesyltransferase Inhibitors as novel Anti-Malarials

*J. Med Chem*, submitted

Silber K, Heidler P, Kurz T, Klebe G

AFMoC enhances predictivity of 3D QSAR: A Case Study with DOXP-reductoisomerase

*J. Med Chem*, **2005** May 19;48(10):3547-63

Kettler K, Wiesner J, Silber K, Haebel P, Ortmann R, Sattler I, Dahse HM, Jomaa H, Klebe G, Schlitzer M.

Non-thiol farnesyltransferase inhibitors: N-(4-aminoacylamino-3-benzoylphenyl)-3-[5-(4-nitrophenyl)-2 furyl]acrylic acid amides and their antimalarial activity.

*Eur J Med Chem*. **2005** Jan;40(1):93-101

Mitsch A, Wissner P, Bohm M, Silber K, Klebe G, Sattler I, Schlitzer M.

Non-thiol farnesyltransferase inhibitors: utilization of the far aryl binding site by 5-cinnamoylaminobenzophenones.

*Arch Pharm (Weinheim)*. **2004** Sep;337(9):493-501

Mitsch A, Wissner P, Silber K, Haebel P, Sattler I, Klebe G, Schlitzer M.

Non-thiol farnesyltransferase inhibitors: N-(4-tolylacetyl-amino-3-benzoylphenyl)-3-arylfurylacrylic acid amides.

*Bioorg Med Chem.* **2004** Sep 1;12(17):4585-600.

Wiesner J, Kettler K, Sakowski J, Ortmann R, Katzin AM, Kimura EA, Silber K, Klebe G, Jomaa H, Schlitzer M.

Farnesyltransferase inhibitors inhibit the growth of malaria parasites in vitro and in vivo.

*Angew Chem Int Ed Engl.* **2004** Jan;43(2):251-4

Kettler K, Sakowski J, Silber K, Sattler I, Klebe G, Schlitzer M.

Non-thiol farnesyltransferase inhibitors N-(4-acylamino-3-benzoylphenyl)-3-[5-(4-nitrophenyl)-2-furyl]acrylic acid amides. *Bioorg Med Chem.* 2003 Apr 3;11(7):1521-30.

### **Oral communications**

Silber, K., Klebe, G.

Inhibitors of DOXP-Reductoisomerase - Structure-based drug design using AFMoC

DPhG Doktorandentagung, Freudenstadt-Lauterbad, Germany, **2004**

Silber, K., Klebe, G.

AFMoC: Combining Receptor- and Ligand-based Strategies for lead optimization *Frontiers in Medicinal Chemistry*, Fulda, Germany, **2003**

Silber, K., Klebe, G.

Combining Receptor-based and Ligand-based Searching Strategies for Virtual Screening.

16<sup>th</sup> Molecular Modeling Workshop, Darmstadt, Germany, **2002**

**Poster presentations**

Silber K., Steuber H., Klebe, G.:

Scoring Fragments for Metalloproteinase Inhibitors,

European Union- Asia link Sommer School of Medicinal Chemistry, Shanghai, China, **2005**

Silber K., Steuber H., Reinscheid U., Klebe, G.:

A Fragment-based Screening Approach for Metalloproteinase Inhibitors,

Gordon Research Conference on Computer Aided Drug Design, Tilton, NH, USA, **2005**

Silber K., Gohlke H., Wiesner J., Heidler P., Kurz T., Jomaa H., Klebe G.:

*Rationalizing Inhibitor Potency for New Antimalarials: 3D QSAR, Docking, and AFMoC for DOXP-Reductoisomerase Inhibitors*

15<sup>th</sup> European Symposium on Quantitative Structure-Activity Relationships and Molecular Modelling in Istanbul, Turkey, **2004**

*(awarded best poster)*

Silber K., Wiesner J., Jomaa H., Klebe, G.

*Rationalizing Inhibitor Potency using AFMoC*

Gordon Research Conference on Computer Aided Drug Design, Tilton, NH, USA, **2003**

Silber K., Wiesner J., Jomaa H., Klebe G.: *AFMoC, a Ligand- and Receptor-based 3D-QSAR Approach tested on Inhibitor for DOXP-Reductoisomerase*

Frontiers in Medicinal Chemistry, Fulda; Germany, **2003**

*(awarded best poster)*

Silber K., Wiesner J., Jomaa H. und Klebe G.:

*Inhibiting DOXP-Reductoisomerase: A new Approach to Combat Malaria*

17. Darmstädter "Molecular Modelling Workshop", Erlangen, Germany, **2003**

*(awarded best poster)*

## Acknowledgments

Numerous people have contributed to this work and I am grateful to all of them. Especially, I would like to thank the following people:

- *Prof. Dr. Gerhard Klebe* for the opportunity to do my PhD thesis under his supervision. He provided not only guidance but also allowed space for own creative ideas. He encouraged, supported and enabled my attendance to international conferences and I am very grateful for all his advice throughout my entire education in Marburg.
- *Prof. Christoph Sotriffer* for his support throughout the thesis and the great supervision and cooperation of the first semester lab. He is remarkably knowledgeable and patient and was very creative in lighting up meetings with English poems.
- *Dr. Nils Rackelmann* for listening whenever needed, spontaneous traveling, his permanent motivation and his efforts in getting me finalizing this thesis.
- My parents, *Lydia and Prof. Dieter Silber* for their never ending support, for convincing me that University is a special place and for being exceptional role models.
- The A203 gang, *Dr. Andreas Evers and Dr. Peter Haebel* for a great time, discussion on- and off-work and countless pizza ceremonies. *Dr. Matthias Zentgraf* for supporting my work, for pizza, movie and game nights, the autumn day photo sessions and most of all for being a great friend.
- *Nils Weskamp* for the great neighborhood relationship, Stella and more.
- *Holger Steuber and Dr. Andreas Heine* for the collaboration on crystal structures.
- *Drs. Matthias Zentgraf, Nils Weskamp, Peter Block, Christoph Sotriffer and Hans Velec* for dedicating a lot of time in keeping the Klebe-lab computer facilities running,
- *Prof. Holger Gohlke* for his support and the introduction into molecular modeling.
- The *Erstsemester Assistenten* for their commitment to teaching and supervision.

- 
- *Christian Sohn* for incredibly precise work on the fragment assay and for keeping the X-ray facilities running.
  - *Drs. Clayton Springer, Donovan Chin, William Egan and Nils Rackelmann* for proofreading and fruitful discussions on fragments, QSAR, docking and statistics.
  - *Angela Scholz* for support for meetings, orders, offices, application forms, etc...
  - The whole *AG Klebe* group for the nice and stimulating working atmosphere. The large diversity of topics and people, the group meetings, coffee room discussions and meeting organization provided an exceptional experience for the future.
  - The *AG Schlitzer* group for the nice collaboration on the Ftase project. It opened unexpected doors and taught me a lot.
  - The Graduiertenkolleg '*Proteinfunktion auf atomarer Ebene*' for funding parts of this work and for some nice seminar trips to Hirschegg.

**Curriculum vitae**

<b>Full name:</b>	<u>Katrin</u> Edith Elisabeth Silber
<b>Date of birth:</b>	August, 11 1976
<b>Place of birth:</b>	Frankfurt, Germany
<b>2006- 2007</b>	Research associate at the Novartis Institutes for Biomedical Research in Collaboration with Prof. Mike Yaffe, MIT  <i>⇒ Druggability assessment and computational characterization of protein-protein interfaces</i>
<b>2001-2006</b>	PhD student at the Institute of Pharmaceutical Chemistry, Philipps-University, Marburg, in the group of Prof. G. Klebe  <i>⇒ Fragment-, ligand- and receptor-based virtual screening strategies for new enzyme inhibitors</i>
<b>2002</b>	Additional certificate as “Diplom-Pharmazeut”
<b>2003- 2005</b>	Fellow of the Marburg Graduiertenkolleg “Protein function at the atomic level”
<b>2001</b>	Third state examination and licensure as pharmacist
<b>2000- 2001</b>	Obligatory practical training student in pharmacy (research and trade)  First half (research): Virtual screening approaches for new inhibitors of metalloproteinases (modelling and biochemical validation), G. Klebe, Philipps-University, Marburg Second half (public pharmacy): Obligatory practical training in public pharmacies in Munich
<b>1999- 2000</b>	Studies of pharmacy at the University Paris sud (XI) as participant of an Erasmus/Sokrates student exchange program
<b>1996- 2000</b>	Studies of pharmacy at Philipps-University, Marburg
<b>1996</b>	Abitur (majors: chemistry and maths)
<b>1987- 1996</b>	Marienschule der Ursulinen, Offenbach,
<b>1983- 1987</b>	Friedrich- Fröbel Schule, Obertshausen

## **Erklärung**

Ich versichere, dass ich meine Dissertation

LEAD OPTIMIZATION FOR NEW ANTIMALARIALS

AND

SUCCESSFUL LEAD IDENTIFICATION FOR METALLOPROTEINASES:

A FRAGMENT-BASED APPROACH USING VIRTUAL SCREENING

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 ähnlich Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg an der Lahn, den 08. August 2007