

Calculation of binding free energies including protein flexibility

Diplomarbeit

vorgelegt der Fakultät für Chemie der
Ruprecht-Karls-Universität Heidelberg

Frauke Meyer

2002

Die vorliegende Arbeit wurde von Prof. J. Warnatz begutachtet. Sie wurde in der Arbeitsgruppe von Prof. J. C. Smith am Interdisziplinären Zentrum für wissenschaftliches Rechnen im Zeitraum vom 20.08.2001 bis 31.05.2002 ausgeführt.

Besonderer Dank gebührt Dr. Stefan Fischer, der mich während der Diplomarbeit fachlich angeleitet und betreut hat.

Ich erkläre hiermit des Eides statt, dass ich die vorliegende Diplomarbeit selbständig unter Anleitung verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Datum

Unterschrift

Abstract

The prediction of binding affinities of ligands to a protein remains a major problem in computer-aided drug design. Fast methods based on empirical models are rather unreliable and assume protein rigidity, whereas accurate methods based on molecular dynamics simulations are computationally costly. This diploma thesis presents a continuum method to calculate binding free energies ΔG taking protein flexibility into account. ΔG is calculated from valence terms, electrostatic and van der Waals interactions, solvent effects and entropic contributions. A set of 47 benzamidine derivatives binding to trypsin, for which the experimental values were determined experimentally in the Hoffmann La Roche laboratories, were docked into the protein binding pocket and energy minimised using combined quantum/classical mechanics to generate a conformational ensemble of structures of the complex. The experimental range ($\Delta G_{exp} = -3.9$ to -7.6 kcal/mol) was reproduced in the calculation ($\Delta G_{calc} = -2.6$ to -9.2 kcal/mol) with a comparatively low root-mean-square error of 1.3 kcal/mol. Thus, including protein flexibility outperforms the corresponding rigid protein approach (RMSD = 3.2 kcal/mol for the same set of ligands). Hence, even in the case of a well defined binding pocket such as is the case for trypsin, conformational flexibility cannot be neglected.

Contents

1	Introduction	1
1.1	Background	1
1.1.1	The nature of binding	3
1.1.2	Current status of research	5
1.2	Aim	11
1.3	The model system	11
2	Theory and methods	17
2.1	Coordinate generation	18
2.1.1	Describing molecular systems: Quantum mechanics and molecular mechanics	18
2.1.2	Docking procedure	21
2.2	Binding free energy function	24
2.2.1	Van der Waals energy	26
2.2.2	Electrostatics	27
2.2.3	Solvation free energy	30
2.2.4	Valence energy	32
2.2.5	Quantum mechanical energy	32
2.2.6	Entropic contributions	34

3 Results	37
3.1 Rigid protein and ligand	38
3.2 Flexible protein and ligand	40
3.2.1 Conformational ensembles	40
3.2.2 Binding free energies	41
4 Discussion	43
4.1 Fixed protein and ligand	43
4.2 Flexible protein and ligand	46
4.2.1 Conformational ensembles and total complex energies . .	46
4.2.2 Binding free energies	49
4.2.3 Contributions to the binding free energy	51
4.2.4 Case studies: benzofurane-3-carboxamidine and deriva- tives	58
4.3 Comparison: flexible versus fixed protein and ligand	62
5 Conclusions and outlook	66
5.1 Conclusions	66
5.2 Outlook	68
6 Zusammenfassung	70
A Starting positions for ligand docking	72
B Van der Waals parameters	74
C Vibran modification	75
D Tables of results: fixed protein	76

<i>CONTENTS</i>	vi
E Tables of results: flexible protein	78
E.1 Boltzmann weighted ensemble averages	78
E.2 Binding free energies	83
F LUDI scores	85

Chapter 1

Introduction

1.1 Background

In order to act as an active agent, a drug must bind to a certain target molecule, usually a protein, in the body. By binding to the target the drug inhibits, causes or modifies a specific reaction. A condition for binding is that the drug sterically fits into the binding pocket and forms favorable local interactions with its environment. The binding free energy is a measure for the binding affinity of a potential drug to its target. A rational approach to develop drugs is structure-based drug design, which aims at predicting the pharmaceutical potency of a molecule using the structure of the complex, protein and ligand. In sharp contrast to this approach most of the currently available drugs have been discovered by pure chance due to the lack of detailed knowledge about the biological system [1]. This irrational way of developing pharmaceuticals shall be illustrated with the following historic example:

“Ein glücklicher Zufall hat uns in die Hände gespielt...” is the first sentence of the publication on acetanilide of A. Cahn and P. Hepp in the Centralblatt

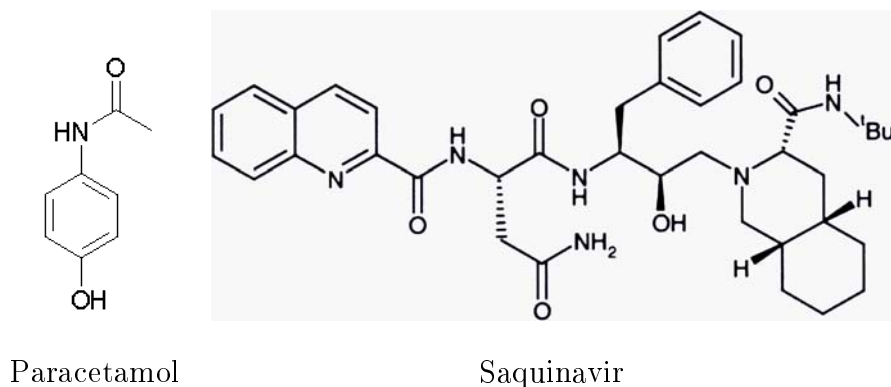


Figure 1.1: Paracetamol as a classical example of drug discovery by chance, and Saquinavir, an HIV-1 protease inhibitor, as an example of rational drug design.

für Klinische Medizin in 1886, saying that it was nothing but luck, that allowed them to discover a fever reducing drug [2]. Cahn and Hepp planned to give naphththaline to a dog to test this substance for its eventual antipyretic effect. What they accidentally gave to the dog without knowing it was acetanilid, which proved to be effective. Acetanilid was further modified to give the well-known pain killer and antipyretic paracetamol (Figure 1.1).

Since an increasing number of protein structures are becoming available, this structural information can be used for a rational design of pharmaceutical compounds. One famous example of successful rational drug design is the development of inhibitors for HIV-1 protease [3, 4]. The steps in the development process were: (a) solution of the X-Ray structure of the protease; (b) transition state mimic; (c) data base searching and de novo design resulting in a set of test ligands fitting in the protease binding pocket; and (d) estimation of binding affinities to rank the candidates and to give the most promising one (*e. g.* Saquinavir, developed by Hoffmann-La Roche, Fig. 1.1).

As shown in the example above, a key step in structure-based drug design is the prediction of binding affinities. It requires structural information as well as an understanding of the physical forces involved. Binding affinities are one of the main target properties of rational drug design, and also contribute to the general understanding of molecular recognition.

1.1.1 The nature of binding

A measure for the affinity of a ligand to bind to a protein is the binding constant K_{bind} . For a given complexation reaction $ligand + protein \rightleftharpoons complex$ K_{bind} is defined as

$$K_{bind} = \frac{k_1}{k_{-1}} = \frac{[complex]}{[protein][ligand]}, \quad (1.1)$$

where k_1 is the rate constant of association, k_{-1} is the rate constant of dissociation, and the brackets indicate concentrations. The dissociation constant K_{diss} is the reciprocal of K_{bind} . The binding constant can be experimentally determined. However, because it strongly depends on the assay conditions of the specific experiment, comparisons are possible only if measurements were carried out following the same protocol. The property often measured experimentally is the so called IC_{50} value, giving the concentration of the inhibitor required to reduce the binding of a ligand by half (see Eq. 1.3). The binding free energy is the main target value of theoretical calculations. It is connected to the binding constant as follows:

$$\Delta G_{bind} = -RT \ln K_{bind} = G_C - G_L - G_P, \quad (1.2)$$

where R is the gas constant, T the temperature, and G_C , G_L and G_P the free energy of the complex, ligand and protein, respectively. The binding free

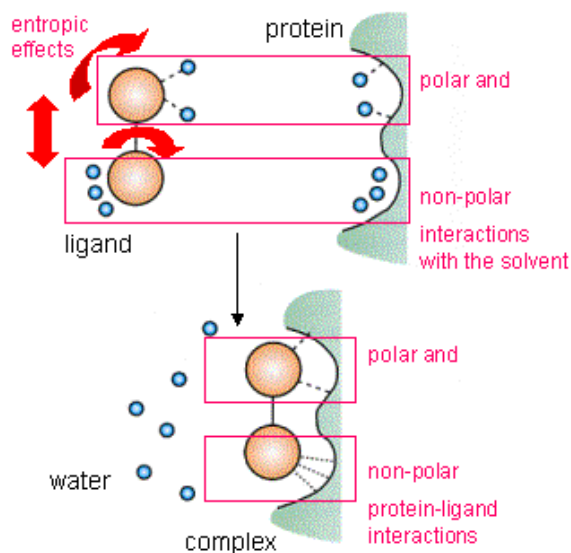


Figure 1.2: Schematical presentation of the effects determining the binding affinity of a ligand to a protein (bonded interactions are not included). Polar and non-polar interactions and entropic contributions (arrows) are depicted.

energy is the difference of the free energies of the protein and the ligand in their bound and unbound state. The change of the free energy of the system upon binding arises from the changes in bonded and non-bonded intramolecular interactions within protein and ligand, and from changes in non-bonded intermolecular interactions between protein, ligand and solvent. Non-bonded interactions at the protein:ligand interface can further be categorized into:

- polar interactions, comprising hydrogen bonds, salt bridges and other interactions between (partially) charged groups. These interactions are responsible for binding specificity.
- hydrophobic interactions of non-polar residues such as aliphatic and aromatic groups (van der Waals interactions). These interactions are the

main forces driving binding.

In addition, polar and non-polar contributions to the solvation free energies have to be considered (Figure 1.2). Binding requires the desolvation of the ligand and the protein binding pocket, and subsequent solvation of the complexed binding pocket. Hydrogen bonds between the solute and the solvent break upon binding, and are reformed between ligand and receptor and between water molecules. Consequently, their total number, and hence the hydrogen bond energy, does not change significantly. The release of water molecules from hydrophobic regions of the solute leads to an increase in van der Waals interactions in the formed complex and in hydrogen bonds of the displaced water molecules. This results in a favorable contribution to the binding free energy.

Beyond the effects mentioned above, entropic contributions to the binding free energy due to changes in translational, rotational and vibrational degrees of freedom are included in ΔG_{bind} . Six translational and rotational degrees of freedom of the unbound ligand as well as internal rotations of the receptor and the ligand freeze, while new vibrational modes are created upon complexation.

1.1.2 Current status of research

In spite of the complex nature of protein:ligand interactions, major advances have been made in understanding and predicting binding. Nevertheless, the prediction of binding affinities of ligands to a protein still remains a major problem in computer-aided drug design. Different methods have been developed to estimate relative and absolute binding free energies [5, 6]. Molecular dynamics (MD) simulations can provide an atom-detailed insight into the dy-

namics of complex biological processes such as ligand binding to a receptor. Methods based on MD, currently presenting the most accurate approaches, are free energy perturbation (FEP) and thermodynamic integration (TI) [7]. Less rigorous methods have been developed to reduce the computational expense: the Poisson Boltzmann / surface area (PBSA) method, the linear interaction energy (LIE) method, and empirical methods based on scoring functions [8]. The most important developments are reviewed briefly.

FEP and TI methods allow the relative binding free energy between two ligands bound to the same protein to be determined. Thus, they serve as tools for ranking two candidates, but do not give absolute binding free energies. Solvent molecules are explicitly included, thereby providing an accurate representation of electrostatic and solvation effects. Reddy et al applied MD and FEP to calculate the relative binding free energies of a set of 20 inhibitors to fructose-1,6-bisphosphatase and obtained good agreement with experiment [9]. The study of Essex et al is an example of combining Monte Carlo and free energy perturbation to calculate relative binding free energies [10]. Their simulation was able to predict the strongest among the four benzamidine-like inhibitors binding to trypsin. Van Gunsteren and coworkers calculated the relative binding free energies of inhibitors to the estrogen receptor ligand-binding domain using thermodynamic integration [11]. In this study they exploited the method to a one-step perturbation, allowing them to derive the binding affinity of several similar ligands from one single MD simulation.

Recently, Ota et al have developed a non-Boltzmann thermodynamic integration (NBTI) method, which overcomes the limitations of sampling conformational space for complex systems [12]. The calculated binding free energy

difference of benzamidine and benzylamine associated with trypsin using this method differs from the experimental value ($\Delta\Delta G_{exp} = 2.6$ kcal) by 0.4 kcal instead of 0.8 kcal in the case of conventional TI. Erion and Reddy combined free energy perturbation with quantum mechanics to study the energetics of the hydration of organic compounds [13], an example, that also shows, that FEP can in principle be used to calculate any free energy differences. However, as FEP requires MD simulations including explicit water molecules, it is not suitable for virtual data base screening.

More computationally efficient is the linear interaction energy method, originally proposed by Åqvist [14]. It is based on estimations of the protein ligand interaction energies derived from empirical force fields. Only intermolecular interactions are taken into account. Electrostatic and van der Waals interaction energies are estimated for an ensemble average from an MD trajectory, and are scaled with empirical parameters. Essex et al applied this method to the calculation of binding free energies of 15 inhibitors binding to the enzyme neuraminidase. They could reproduce the experimental results with an error of 1.51 kcal/mol [15].

Scoring functions can be evaluated very quickly, and hence are especially suitable for use in database searching. The local ligand-receptor interactions are counted and weighted. The scoring function is then parametrized empirically as a sum of these terms. Böhm presents such an empirical function, in which the different protein:ligand contacts, such as hydrogen bonds, electrostatic interactions and hydrophobic contacts are parametrized [16]. The application to 82 complexes of known crystal structure resulted in a standard deviation of 1.8 kcal/mol from experiment. While conventional scoring

functions failed in reproducing binding free energies for a p38 MAB Kinase protein system determined with TI simulation methods, the so-called one-window free energy grid (OWFEG) method performed well in this test case [17]. The approximate local free energy on each grid point in the binding region is determined, and ligand flexibility is considered. Scoring functions give a rather rough estimation of the binding affinities, and consequently serve as drug design tools for discriminating ligands of low affinity from potential drug candidates.

In terms of accuracy and computational expense, continuum methods, also denoted as Poisson Boltzmann / surface area (PBSA) methods, are situated in-between the most rigorous perturbation methods based on MD, and the empirical scoring functions. In contrast to FEP simulations, these methods model the solvent as a continuum of a certain dielectric constant. They thereby take solvent effects into account at lower computational cost. The solvation free energy is partitioned into the polar contribution, calculated by solving the Poisson-Boltzmann equation, and the non-polar contribution, estimated as linearly dependent on the solvent accessible surface area.

Several recent studies successfully applied continuum methods to protein:ligand binding and demonstrate the advancement in this field. These include the binding free energy calculation of peptides to the chaperone DnaK [18]; of antibiotics to a ribosomal unit [19]; of TIBO-like inhibitors to HIV-1 reverse transcriptase [20] and of cathepsinD-like inhibitors to aspartyl protease [21]. All studies mentioned used snapshots of molecular dynamics simulations with an explicit solvent model to generate an ensemble of complex conformations. This set of coordinates is subsequently applied to the free energy calculation.

In combination with MD, the continuum method approach lacks its advantage of not requiring extensive computational recourses. In most studies, the calculated free energies deviate from the experimental results by approx. 1 kcal/mol [20, 21]. Thus, the currently available PBSA methods outperform the scoring functions, but can not reach the accuracy of MD based free energy perturbation methods. The PBSA approach has been further exploited to study free energy changes due to protein:protein interactions, as shown in the study of Wang et al on the HIV protease [22]. The broad applicability of the continuum method could be further demonstrated by applying the PBSA method to the calculation of binding free energy differences between protein mutants complexed with the same ligand [23].

Entropic contributions to the binding free energy are either neglected or taken into account as a constant value. Gehring et al included the change in conformational entropy of only the peptide chains upon complexation, but thereby could not improve their calculation [18]. The study of Schwarzl et al aimed at improving the prediction of binding free energies by introducing an ideal-gas entropy correction into the energy function. The entropic contributions due to changes of the translational, rotational and vibrational degrees of freedom are taken into account. For six small and rigid benzamidine-like ligands binding to trypsin, ΔG_{calc} deviated by 0.55 kcal/mol on average relative to the experimentally determined values [24]. Their approach gave less accurate results for sterically more demanding ligands, for which the docking into the binding pocket of trypsin is not uniquely defined [25]. This led to the assumption, that in the case of sterically demanding ligands, conformational changes upon binding should be considered, when generating three-

dimensional complex structures.

For the work presented here, we chose the continuum method for the following reasons:

- Partitioning the free energy into several contributing terms allows one to analyse the effects driving binding in detail.
- It requires reasonable computational power.
- Preceding studies proved its potential to accurately rank ligand candidates according to their binding free energy. Entropy contributions can be taken into account and improve the prediction.
- The results for the calculations with a rigid system encourage one to try to improve the method by including protein flexibility and considering an ensemble of complex conformations.
- The molecular mechanics method to generate the coordinates by energy minimization can be combined with quantum mechanics as a further attempt to increase the reliability of the calculation method.

1.2 Aim

The aim of this diploma thesis is to develop a method to reliably predict free energy changes upon ligand binding to a protein at reasonable computational expense.

For this purpose a previously developed method based on continuum electrostatics and an entropy correction shall be improved. Whereas the scoring method of this approach is in principle adopted, the generation of the three-dimensional structure of the complex used here is more sophisticated. Introducing protein and ligand flexibility, using a combined quantum mechanical and molecular mechanical approach, and considering an ensemble of probable complex conformations shall allow us to accurately predict binding free energies, also for ligands of ambiguous conformations.

1.3 The model system

This thesis aims at a calculation method for binding free energies, which is applicable to any protein-ligand-system of interest. Hence, the energy function should not depend on the model system, that has been chosen to derive it. A main source of error in the prediction of binding affinities is the generation of the three-dimensional structures. To reduce the uncertainties of the structure prediction, the criteria for the choice of a model system are:

- a protein or a complex, of which a crystal structure of high resolution is available;
- a rather inflexible binding pocket;

- a set of ligands with little conformational freedom.

In addition, a comparison of the calculated and the measured binding affinities requires a consistent data set of the experimental values. For this thesis, trypsin as the receptor and a set of benzamidine-like inhibitors have been chosen. This model system meets the criteria and has already served as test system for the preceding study [25].

Trypsin is a digestive enzyme and belongs to the class of serine proteases, like chymotrypsin or thrombin [26]. Those proteins cut peptide bonds of different substrates and with certain specificity. Serine proteases have the so-called catalytic triade in common, which consists of Ser 195, His 57 and Asp 102 and is shown in figure 1.3. The serine residue features an especially high chemical reactivity, and plays a key role in the catalytic mechanism of the peptide bond hydrolysis.

A hydroxyl group is usually not nucleophilic enough to attack an amide group. The Ser 195 is highly activated by the neighboring His 57 that forms a hydrogen bond to the serine residue. The histidine accepts the proton and thereby enables the serine anion to form a tetraedric transition state with the substrate, as shown in figure 1.4. The vicinal aspartate can accept one proton from the histidine to return it later. The following steps (not shown in figure 1.4) are the breakup of the transition state into an amine and a carboxylate group by an attacking water molecule and the subsequent reprotonation of Ser 195.

While all serine proteases have this mechanistic pattern in common, they differ from each other in their substrate specificity. This specificity is mainly achieved by the molecular recognition of the substrate in the P1 binding pocket

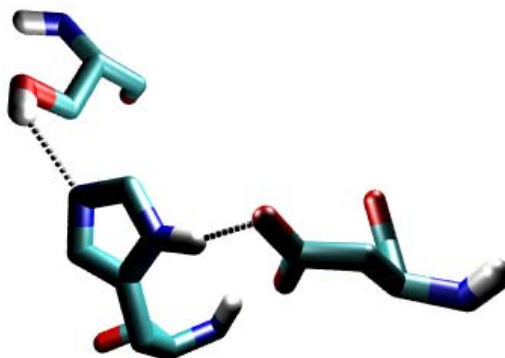


Figure 1.3: The geometry of the catalytic triade in trypsin. The initial catalytic step is a proton transfer from Ser 195 to His 57 followed by the attack of the substrate by the unprotonated serine. The illustration was generated with VMD [27] and rendered with POV-Ray[28].

of the protease. Only those peptide bonds with adjacent amino acid side chains, that fit into the P1 pocket, are cut by the enzyme. The crystal structure of trypsin shows an asparagine (Asp 189) pointing into the binding pocket. This negatively charged residue is responsible for the lysine and arginine selectivity of trypsin.

Trypsin-like serine proteases are well-studied in terms of structural and binding properties, as they are of special interest as potential drug targets in blood coagulation. The characteristic of trypsin is to bury the substrate side chain in the P1 pocket, so that an effective inhibition can be achieved by blocking this pocket. A well-studied class of non-covalent trypsin inhibitors

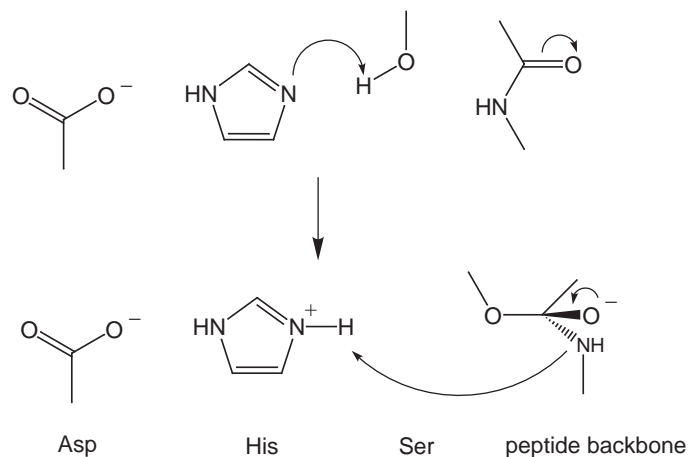


Figure 1.4: Catalytic mechanism of the transition state formation in serine proteases. The nucleophilic attack of the serine is promoted by the histidine side chain, which itself is further stabilized by the asparagine.

are benzamidine derivatives mimicking arginine and lysine. These substrate analogs show binding constants in the micro-molar range, corresponding to binding free energies of -4 to -8 kcal/mol [29].

Among the crystal structures solved for trypsin complexed with an inhibitor, the structure of the benzamidine-trypsin complex 1bty [30] has the highest resolution (1.5 Å). This structure has also been used in the diploma thesis of S. M. Schwarzl [25]. Several modifications, such as the removal of crystal water molecules and slight side chain adjustments in the region of the P1 pocket, were made. The resulting prepared structure served as a model system in the present work.

One cannot conclude from experiment, whether His 57 is protonated or not. Binding free energy calculations based on the rigid protein approach were performed for both cases [25]. We used the protein structure with an unprotonated His 57.

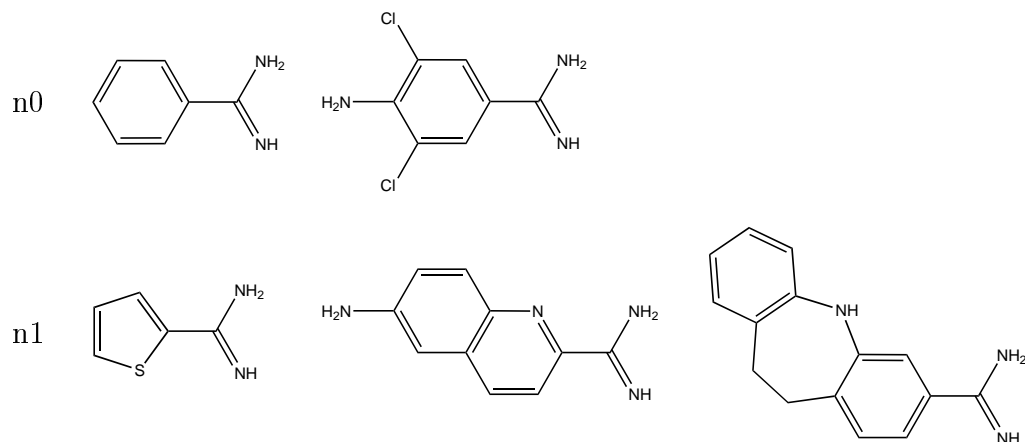


Figure 1.5: Examples of ligands in set n0 and n1.

The criteria for the choice of a set of trypsin inhibitors were an amidinium group and maximal one degree of freedom during docking. Only those molecules were chosen, that do not show any protonation state variability in the bound of unbound state. As substrate analogs they form a salt bridge from the positively charged amidinium group to the carboxylate group of Asp 189. One subset (here denoted as set n0) comprises six rigid inhibitors of C2-symmetry, and thus ligands without conformational variability. The other subset (here denoted as set n1) comprises 42 ligands featuring one variability in the docking conformation due to their asymmetry (C1) or to an internal rotational degree of freedom (in the case of $-\text{CH}_3$, $-\text{NH}_2$ and others). Some representative structures are shown in figure 1.5.

The measurement of the IC_{50} value of these ligands have been carried out in the Hoffmann-LaRoche laboratories [31]. With UV spectroscopy the cleavage of para-nitroanilide covalently bound to a peptide by trypsin was monitored in the presence of the inhibitor in varying concentrations. For these experiments, the ionic strength was $IS = 92 \text{ mM}$ and the pH value set to 7.8. The IC_{50} values

were derived from the inhibitor concentration response curves and converted into the dissociation constant using the following formula [32],

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_M}}, \quad (1.3)$$

where K_i is the dissociation constant, $[S]$ is the substrate concentration and K_M the Michaelis-Menten constant. The experimental error in the dissociation constant is appr. 20%, corresponding to an error of 0.1 kcal/mol for the binding free energy. Because the experimental values are not yet published by Hoffmann-LaRoche for all of the ligands investigated here, they can only be listed in the Appendices D and E without the corresponding structure.

When docking the ligands of set n0 into the binding pocket of rigidly fixed trypsin the binding free energies calculated with the continuum method correlates well with the experimental values (see Section 1.1.2). Using this approach, ΔG_{bind} could not be reliably predicted for set n1. The calculated values deviated from the experimental results by 3.20 kcal/mol. For the work presented here, the calculations were carried out with both sets n0 and n1 in order to compare the performance of the new approach, that includes conformational flexibility, with the rigid protein approach.

Chapter 2

Theory and methods

The calculation of ΔG_{bind} requires two steps. First, the structures have to be generated, of which the modeling of the complex is the most crucial part. The docking of the ligand into the binding pocket is followed by an energy minimisation to give the three-dimensional coordinates of the complex. For this minimisation, an energy function to describe the molecular system is needed. Second, the function to calculate the binding free energy of the complex is developed and applied. So this step in turn requires the three-dimensional structure. The interdependence of coordinate generation and scoring can not be circumvented and should be kept in mind when deriving methods to predict binding free energies.

The following chapter describes the computational procedures and underlying theory of the calculations. Section 2.1 deals with the first step, the docking procedure to generate the coordinates. In Section 2.2, the function for the binding free energy calculations and the contributing terms are presented.

2.1 Coordinate generation

2.1.1 Describing molecular systems: Quantum mechanics and molecular mechanics

The input for the function to calculate the binding free energies are the structures of the ligand and receptor in the bound and unbound state. These structures are generated by minimising the potential energy as a function of atomic coordinates. Various methods differing in the degree of approximation and computational cost have been developed to give the energy of a molecular system using an atomic detail description.

With molecular mechanics (MM) methods, molecules are described by parametrized potential functions. The variables are the coordinates of the atoms. Atoms are treated as spheres of a certain mass and charge, linked with each other by bonds, modeled as springs. The so-called empirical force field gives the potential energy of the system, which is the sum of the interactions between the atoms. The simulations of the present work are carried out with the CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field and software [33]. The potential energy of the CHARMM force field is

$$E_{MM} = E_{val} + E_{coul} + E_{vdw} \quad (2.1)$$

The valence energy accounts for valence interactions between bonded atoms, in other words for the structural strain in the molecule, and will here be indicated as E_{val} .

E_{vdw} denotes the van der Waals interaction energy between non-bonded atoms. The Lennard-Jones 12-6 potential (Equation 2.2) is often used to calcu-

late the van der Waals energy. It assumes a long range attraction proportional to r^{-6} due to a dipole-induced dipole interaction, and a short range repulsion proportional to r^{-12} resulting from the overlap of electron clouds.

$$E_{vdw} = 4 \sum_{i>j} \epsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}^{12}} - \frac{\sigma_{ij}}{r_{ij}^6} \right) \quad (2.2)$$

where r_{ij} is the distance between atoms i and j , ϵ_{ij} is the van der Waals energy minimum, and σ_{ij} the distance at which E_{vdw} is zero.

E_{coul} is the electrostatic interaction energy, and is calculated as the sum of Coulomb interactions between the partial atomic charges:

$$E_{coul} = \sum_{i>j} \frac{q_i q_j}{\epsilon r_{ij}} \quad (2.3)$$

where q_i and q_j are the atomic charges of atoms i and j respectively, ϵ is the dielectric constant of the medium, and r_{ij} is the atom distance. The electrostatic energy of a molecule is strongly influenced by the dielectric medium around it as well as the inner dielectric properties. Without explicitly including water molecules into the simulation, the Coulomb law does not account for solvent effects. Hence, simulations based on the CHARMM force field are vacuum simulations, as long as water molecules are not explicitly included. A way to account for solvent screening effects - charge reparametrisation - is described in Section 2.1.2.

The CHARMM program will here be used for energy minimisations. The aim of energy minimisations is to locally explore the potential energy surface in order to find the local energy minima on the surface. Different iterative procedures have been developed for this purpose. The steepest descent, the conjugate gradient and the Newton-Raphson method are minimisation algo-

rithms implemented into CHARMM, that have been used in this work. For further details, see Ref. [34].

In contrast to molecular mechanics (MM), that models the molecule as masses (atoms) interacting with each other according to classical mechanics, quantum mechanics (QM) uses wave functions to describe the electron density around the nuclei. Semi-empirical quantum mechanics methods have been developed to increase the computational efficiency of pure so-called *ab initio* methods. This is achieved by calculating only the wave functions of the valence electrons and empirical parametrisation of the remaining integrals to account for the neglect of core and other orbitals. The approaches developed so far differ in the number of neglected orbitals and of the parameters introduced. Presently, AM1 (Austin Model 1) [35] developed by Dewar et al is the most commonly used semi-empirical method.

The interface of CHARMM and MOPAC 4.0 [36] allows to perform combined AM1 and molecular mechanical calculations. A small part of the system, *e. g.* the reacting part in a chemical reaction or a ligand bound to a receptor, is treated semi-empirically, while the remaining part, mostly the protein, is modeled by the force field. The potential energy becomes $E = E_{MM} + E_{QM} + E_{QM/MM}$. E_{MM} is the energy of the part treated with MM, E_{QM} is the energy of the purely quantum mechanically treated part, and $E_{QM/MM}$ is the interaction energy between those two regions. The latter is of exclusively non-bonded nature in the case of a trypsin (MM part) - ligand (QM part) system.

A main advantage of treating the ligand quantum mechanically is that no parametrization of the bonded interactions is needed. Since CHARMM is

set up for simulations of biological macromolecules, parameters for the small chemical molecules investigated here are not available. For the parameters of the non-bonded interactions, see Section 2.1.2.

2.1.2 Docking procedure

The docking procedure to generate the coordinates of the trypsin-inhibitor system were performed with the program CHARMM using the MM force field to treat the protein and AM1 to treat the ligand. In the preceding study assuming a rigid protein and ligand, only one possible conformation, that one of the lowest energy found, was taken into account for the binding free energy calculation [24]. However, it is well known that proteins with a certain conformational flexibility undergo an induced fit upon ligand binding. They can adopt different structures with an energy in-essentially higher than the minimal energy accessible. Recent docking studies have shown, that methods using only the minimum structure perform poorly, whereas ensemble average methods permit more accurate docking [37, 38]. Here, within the flexible protein approach, an ensemble of possible conformations for one complex is considered. Each conformation of the ensemble represents a local minimum of the potential energy surface of the complex.

In order to obtain a set of possible structures of one complex the following docking procedure has been developed: One initial protein structure is used. The ligand is transferred to a certain starting position in the binding pocket. An energy minimisation is carried out to give a structure of the bound state of certain potential energy. The docking into the initial protein structure and subsequent minimisation is done for 70 different starting positions (see

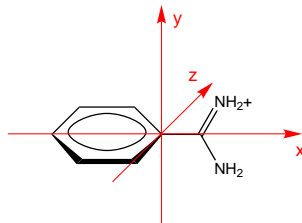


Figure 2.1: Illustration of the axes, along which the ligands are translated and rotated to give a set of 70 protein-ligand-structures for minimisation.

Appendix A). As shown in figure 2.1, three axes are defined for translations and rotations. The ligand is translated along these axes. At seven different resulting positions (including the origin), the ligand is rotated around them.

Each structure is energy minimized using the combined QM/MM method in CHARMM. Most atoms of the protein are kept fixed during the minimisation. The flexible region should include enough atoms, so that backbone and side chain atoms can adjust themselves upon complexation. At the same time, it should be as small as possible to reduce uncertainties in the structural prediction. Here, it comprises the binding pocket including the catalytic triade, the ligand, and five crystal water molecules located in the binding region. Approx. all atoms within a distance of 9 Å from the aromatic carbon atom of the ligand bound to the amidinium group are included.

The minimisation is performed using the steepest descent, the conjugate gradient and the Newton-Raphson algorithms. During one cycle of minimisation a self-consistent field (SCF) calculation is run. The electronic solutions are gradually refined, until the change in energy on two successive iterations is less than 10^{-8} kcal/mol.

Electrostatic interactions between charges are weakened in the presence of a

solvent of a high dielectric constant. If no explicit water molecules are included, this screening effect of the solvent is disregarded. To take solvent effects into account, nonuniform charge scaling, as implemented into CHARMM is used [24].

The protein is partitioned into groups I , each of them spanning a part of a side chain or backbone (I denotes a group, i an atom). By dividing each charge q_i by a factor $\lambda_I > 1$, that depends on the position of the group I in the protein, the absolute value of the vacuum interaction energy is scaled down to the interaction energy in solution. The scaling factors are calculated as $\lambda_I = \sqrt{\epsilon_I}$, where ϵ_I is the effective dielectric constant for the interaction of the group I with the other groups of the system. ϵ_I is derived from the ratio of the interaction energy of group I with the other groups J in vacuum E_{IJ}^{vac} and the same interaction energy in solution E_{IJ}^{solv} (Eq. 2.4).

$$\epsilon_I = \frac{\sum_{J, \epsilon_{IJ} > 0} |E_{IJ}^{vac}|}{\sum_{J, \epsilon_{IJ} > 0} |E_{IJ}^{solv}|} \quad (2.4)$$

E_{IJ}^{vac} is calculated with the Coulomb law (Eq. 2.3), E_{IJ}^{solv} is derived from the electrostatic potential in solution, which in turn is calculated using the Poisson-Boltzmann equation (see Section 2.2.2). For the energy minimisations in the present study, the scaling factors were calculated such that a solvation energy for a system with a dielectric constant of 4 for the protein and of 80 for the solvent is reproduced as accurately as possible by the vacuum interaction energy.

Among the 70 structures obtained after minimisation, those with a total potential energy not more than 25 kcal/mol higher than the minimum are chosen for the free energy calculation. Out of a group of similar structures with

a deviation of the total energy of less than 0.002 kcal/mol, one representative candidate is chosen. The resulting conformational ensemble is used for the free energy calculation, using the equation derived below (Eq.2.11, P:L interactions and $\Delta G_{QM}(L)$ excluded). Each structure is weighted by a Boltzmann factor f_i so as to account for its probability:

$$f_i = \frac{N_i}{N_{tot}} = \frac{e^{-\frac{G_{tot,i}(C)}{RT}}}{\sum_i e^{-\frac{G_{tot,i}(C)}{RT}}}, \quad (2.5)$$

where N_i is the number of molecules in the state i , N_{tot} the total number of molecules, $G_{tot,i}(C)$ is the total free energy of state i as described in Eq. 2.11, R is the gas constant and T the temperature. The average complex free energy then is $\langle G_{tot}(C) \rangle = \sum_i f_i \cdot G_{tot,i}(C)$.

2.2 Binding free energy function

This thesis aims at the calculation of the binding free energy $\Delta G_{bind} = G_C - G_L - G_P$ for the complex (C) formation from a protein (P) and a ligand (L).

The following terms contribute to ΔG_{bind} :

- ΔG_{int} intramolecular and intermolecular interactions, comprising van der Waals, Coulombic and valence energy terms

$$\Delta G_{int} = \Delta G_{val} + \Delta G_{coul} + \Delta G_{vdw}$$
- ΔG_{solv} solvation free energy, which includes polar and non-polar contributions: $\Delta G_{solv} = \Delta G_{solv,p} + \Delta G_{solv,np}$
- $\Delta G_{entropy}$ entropic contribution due to changes in the translational, rotational and vibrational degrees of freedom.

The binding free energy then is

$$\Delta G_{bind} = \Delta G_{int} + \Delta G_{solv} + \Delta G_{entropy} \quad (2.6)$$

The sum of the van der Waals, the Coulombic and the valence energy terms equals to the total potential energy of the CHARMM force field E_{MM} . Because the ligand is treated quantum mechanically, these terms can not be derived from E_{MM} directly. Instead, ΔG_{int} has to be further partitioned into the contributions due to conformational changes of the protein $\Delta G(P)$, the ligand $\Delta G(L)$, and the direct interaction between protein and ligand in the bound state $\Delta G(P : L)$:

$$\Delta G_{int} = \Delta G_{int}(P) + \Delta G_{int}(L) + \Delta G_{int}(P : L) \quad (2.7)$$

The difference of the quantum mechanical AM1 energy of the ligand in the bound and unbound states contains all three terms of interest, and hence can be understood as

$$\begin{aligned} \Delta G_{int}(L) &= \Delta E_{QM}(L) \\ &= \Delta G_{vdw}(L) + \Delta G_{coul}(L) + \Delta G_{val}(L) \end{aligned} \quad (2.8)$$

Terms such as van der Waals or valence energies are not defined in quantum mechanics. Equation 2.8 transfers a force field description to a QM description in order to result in a consistent terminology and to get a concrete picture of the energy function and its terms.

The respective energies for the protein and protein:ligand interaction can be derived from the CHARMM force field and are given by

$$\begin{aligned} \Delta G_{int}(P) &= \Delta E_{MM}(P) \\ &= \Delta G_{vdw}(P) + \Delta G_{coul}(P) + \Delta G_{val}(P) \end{aligned} \quad (2.9)$$

and

$$\begin{aligned}\Delta G_{int}(P : L) &= \Delta E_{MM}(P : L) \\ &= \Delta G_{vdw}(P : L) + \Delta G_{coul}(P : L)\end{aligned}\quad (2.10)$$

The Coulombic interactions were calculated with the program `coulomb` written by S. M. Schwarzl, but in principle could also be derived with CHARMM. Because only non-covalently binding inhibitors are chosen, the valence energy term for the protein:ligand interaction $\Delta G_{val}(P : L)$ equals 0. Equations 2.6 to 2.10 lead to the following 'Master equation' [5] used for the free energy calculations.

$$\begin{aligned}\Delta G_{bind} &= \Delta G_{val}(P) + \Delta G_{coul}(P) + \Delta G_{coul}(P : L) + \Delta G_{vdw}(P) \\ &\quad + \Delta G_{vdw}(P : L) + \Delta G_{QM}(L) + \Delta G_{solv} + \Delta G_{entropy}\end{aligned}\quad (2.11)$$

In the case of the method assuming a rigid protein and ligand, $\Delta G_{int}(P)$ as well as $\Delta G_{int}(L)$ cancels out, as no structural changes upon complexation are allowed. When introducing flexibility, both terms have to be taken into account. The equation also reflects the introduction of quantum mechanics for the ligand ($\Delta G_{QM}(L)$).

2.2.1 Van der Waals energy

The van der Waals energy contribution is modeled by the Lennard-Jones-Potential (see Eq. 2.2). The change of the van der Waals energy within the protein and between protein and ligand is taken from the CHARMM potential energy.

The calculation requires parameters for the van der Waals radius, which is defined as the radius at the energy minimum of the Lennard-Jones-Potential

between two atoms of the same species, and the value of the minimal potential energy. For some of the atoms, namely the chlorine atom bound to benzene and the sulfur atom in thiophene, these van der Waals parameters are not defined in the CHARMM force field. Parameters of the MAB force field [39] as implemented in the modeling program MOLOC[40] were used here. The MAB force field can generally be applied to a broad range of structures, because parameters for various molecules and atoms can be generated from a small set of basic parameters. The van der Waals parameters taken from MAB are listed in Appendix B.

2.2.2 Electrostatics

The electrostatic contribution to the binding free energy ΔG_{elec} includes two terms: the Coulombic interactions between the partial charges of the system, ΔG_{coul} , and the polar fraction of the solvation free energy, $\Delta G_{solv,p}$. In the continuum solvent model, the system under investigation is modeled as consisting of two kinds of continuum dielectric. The solvent is water with a dielectric constant of $\epsilon(solv) = 80$. In the microscopic picture of the protein, a dielectric constant is actually not defined, but experimental data shows, that a dielectric constant in the range of 1 to 4 is justified. Here, $\epsilon(prot) = 4$ is used.

The total electrostatic energy can be calculated from the electrostatic potential ϕ_i at the positions of the charges q_i with

$$G_{elec} = \sum_{i=1}^n q_i \phi_i, \quad (2.12)$$

where n is the number of charges. When mobile ions are present in the system, the electrostatic potential is accurately described by the Poisson-Boltzmann

equation

$$\nabla \cdot \epsilon_i \nabla \Phi_i - \kappa' \sinh \Phi_i = -4\pi \rho_i, \quad (2.13)$$

where ρ_i is the charge distribution, and κ' is related to the Debye-Hückel inverse length, which in turn is a function of the ionic strength and the temperature. Eq. 2.13 is a non-linear second order differential equation, that can be linearized *via* a Taylor series expansion of the hyperbolic sine function and truncation after the first term to give the linear Poisson-Boltzmann equation

$$\nabla \cdot \epsilon_i \nabla \Phi_i - \kappa' \Phi_i = -4\pi \rho_i. \quad (2.14)$$

Using finite differences methods, the linear Poisson-Boltzmann equation can be solved numerically. Here, it was solved using the software package UHBD (University of Houston Brownian Dynamics)[41, 42]. A three-dimensional grid is applied to a box containing the molecule, which is surrounded by the continuum solvent. The charges are assigned to the grid points using a trilinear interpolation scheme. By using a finite differences scheme the derivatives of the linear Poisson-Boltzmann equation are calculated. The electrostatic potential of the grid points is obtained and interpolated to the atoms. The accuracy of this scheme is higher with lower grid spacing.

An artefactual Coulombic interaction of the charge with itself arises from the assignment of charges to grid points. It results in an artefactual grid energy, that is included in G_{elec} and can not be separated. The artefactual grid energy depends on the charges, their coordinates and the dielectric constant. Thus, it does not cancel out when calculating the change of the electrostatic energy upon ligand binding unless when using the rigid protein approach. The atoms within the flexible region are allowed to move, and consequently cause a grid energy, which is different for the bound and unbound state.

This technical problem due to the grid based implementation can be circumvented using a thermodynamic cycle as shown in Figure 2.2. According to the Hess's law of constant heat summation, the value of $\Delta X = X(B) - X(A)$, where X is a state function, only depends on the initial state A and the final state B, and does not depend on the path leading from A to B. Instead of a direct path, which is the complexation in water, the system is desolvated, the complexation then takes place in vacuum, and the system is re-solvated to end up in a solvated bound state. The electrostatic contribution to the binding free energy can be calculated as

$$\begin{aligned}\Delta G_{elec} &= -\Delta G_{solv,p}(P) - \Delta G_{solv,p}(L) + \Delta G_{coul}(\epsilon = 4) + \Delta G_{solv,p}(C) \\ &= \Delta G_{solv,p} + \Delta G_{coul}(\epsilon = 4)\end{aligned}\quad (2.15)$$

with

$$\Delta G_{solv,p}(X) = G_{elec}(X, \epsilon = 80) - G_{elec}(X, \epsilon = 4), \quad (2.16)$$

where p denotes, that the solvation free energy calculated in this way only takes into account the polar contributions, and $\Delta G_{solv,p}$ the overall change in the polar part of the solvation free energy $\Delta G_{solv,p}(C) - \Delta G_{solv,p}(L) - \Delta G_{solv,p}(P)$. Polar solvation effects arise from the electrostatic intermolecular interaction between the solvent and the solute. Non-polar contributions to the solvation free energy are discussed in Section 2.2.3.

The binding reaction takes place in a dielectric medium of $\epsilon = 4$, which equals the dielectric constant of the protein. In this case of an unchanging dielectric medium, the Poisson-Boltzmann equation becomes the Coulomb law, and $\Delta G_{elec}(\epsilon = 4)$ can be calculated as the the Coulombic energy $\Delta G_{elec} = \Delta G_{coul} = \sum_{i>j} \frac{q_i q_j}{4\pi \epsilon_{ij}}$. It comprises the intramolecular as well as the intermolecular Coulombic interactions. Eq. 2.8 shows, that the potential energy of the

quantum mechanically treated ligand contains the intramolecular Coulombic interactions $\Delta G_{coul}(L)$, which therefore is not included in ΔG_{coul} in the free energy calculations (see Section 2.2.5 for details).

The other three terms in Eq. 2.15 are solvation free energies, that can be evaluated by solving the Poisson-Boltzmann equation as implemented in UHBD. The desolvation of protein and ligand and the solvation of the complex are reactions without any changes of the atom coordinates. Thus, the artefactual grid error cancels out, when subtracting the electrostatic energies of the solvated and non-solvated system.

Thus, *via* the thermodynamic cycle, the Coulombic interaction and also the polar contribution to the solvation free energy are calculated (see 'Master equation' 2.11). The parameters required are the partial atomic charges and the atomic radii. Ligand and protein radii and the protein charges were taken from the CHARMM force field. *Ab initio* quantum mechanics simulations were performed to calculate the Electrostatic Surface Potential (ESP) of the ligands. The charges were obtained by fitting their potential to the ESP. The module CHELPG in the Gaussian94 software [43], a 3-21G* basis set for prior ligand geometry optimization, and a HF/6-31G* basis set for the ESP calculations were used.

2.2.3 Solvation free energy

The binding of the inhibitor to its receptor requires the removal of the water molecules solvating the inhibitor and the binding pocket, and the addition of water molecules to solve the receptor-inhibitor complex. Modeling the solvent implicitly by a continuum enables to predict the solvation energetics by us-

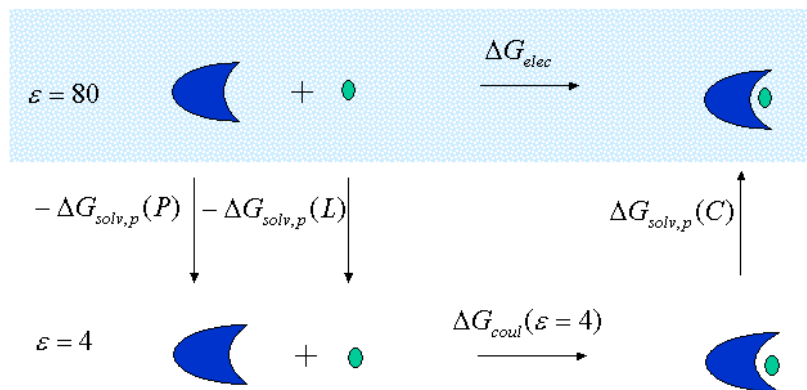


Figure 2.2: The thermodynamic cycle for the calculation of the electrostatic contributions to the solvation free energy.

ing the PBSA approach: Poisson-Boltzmann (PB) calculations for the polar contributions as described in Section 2.2.2 and an estimation of the non-polar contribution, based on surface area (SA) calculations.

The non-polar contribution accounts for the hydrophobic effect due to the solvent cavity formation and van der Waals interactions between solvent and solute. It linearly depends on the change of the solvent-accessible surface area ΔA :

$$\Delta G_{solv,np} = \gamma \Delta A, \quad (2.17)$$

where γ is an empirical coefficient and equals 25 cal/mol/\AA^2 [44, 45]. The surface area of the solutes was computed with the program codisp [46], which is based on the Connolly algorithm [47].

2.2.4 Valence energy

The valence energy accounts for the intramolecular bonded interactions. In the empirical CHARMM force field, it is defined as

$$E_{val} = E_{bond} + E_{\Theta} + E_{\Phi} + E_{impr} + E_{urey}. \quad (2.18)$$

where E_{bond} is the bond stretching energy, E_{Θ} the bond-angle bending energy, E_{Φ} the dihedral angle bending energy, E_{impr} the out-of-plane bending energy, and E_{urey} the Urey-Bradley term, introducing a coupling of bond stretching and angle bending. The mathematical form of one of the addends is a sum of parabolic functions. Each of these functions gives the energy of a certain bonded interaction. It depends on the deviation from the equilibrium value of the bond length or angle, respectively, and is parametrized together with a force constant reflecting the stiffness.

The valence energy is only defined for the molecular mechanics region, i.e. the protein, and is computed as the difference between the protein in the bound and unbound state.

$$\Delta G_{val}(P) = E_{val}(P, bound) - E_{val}(P, unbound) \quad (2.19)$$

The quantum mechanical energy accounts for the change of the ligand conformational strain and is discussed in the following section.

2.2.5 Quantum mechanical energy

The quantum mechanical energy of the system is the eigenvalue obtained from solving the Schrödinger equation $\hat{H}\psi = E\psi$. The Born Oppenheimer approximation is introduced to separate the motions of the nuclei and the electrons.

Multiplication with the conjugated complex wave function and integration leads to

$$E = \frac{\int \psi^* \hat{H} \psi d\tau}{\int \psi^* \psi d\tau}. \quad (2.20)$$

According to the variation theorem, every wave function representing an approximate solution is of a higher energy than the exact energy E . The expansion of Eq. 2.20 in a given basis set gives the secular determinant $|H_{rs} - E \cdot S_{rs}|$, in which H is the Hartree matrix, S the overlap matrix, r and s are denoting one-electron-wave functions, and E is the eigenvalue or energy. In order to find the energy minimum, the secular determinant is set to 0. The solution of the resulting equations and evaluation of the integrals is simplified by introducing the approximations made in the AM1 method.

In order to distinguish the quantum terms from molecular mechanical energy terms, E will henceforth be denoted by E_{QM} . The change in the quantum energy then is

$$\Delta G_{QM}(L) = E_{QM}(L, bound) - E_{QM}(L, unbound). \quad (2.21)$$

$E_{QM}(L, bound)$ and $E_{QM}(L, unbound)$ are the quantum energies of the ligand after geometry optimization in the bound and unbound states, respectively. The critical value of this equation is $E_{QM}(L, bound)$, because it contains not only intramolecular interactions within the ligand, but also electrostatic protein:ligand interactions when taking E_{QM} from CHARMM. To obtain a term, that exclusively accounts for the ligand energy, the protein charges are set to 0 prior to the energy computation using the previously determined geometry of the ligand in the bound state.

The polarisation energy is the change of electrostatic energy due to a change in the electron density of a system in an external electric field. In consequence

of switching all the charges of the environment off, the wave function is not distorted by an external field. In other words, a favorable polarisation energy term is not included in $\Delta G_{QM}(L)$, and hence in the calculated binding free energy values presented here.

2.2.6 Entropic contributions

The Gibbs-Helmholtz equation

$$\Delta G = \Delta H - T\Delta S \quad (2.22)$$

connects the free energy (also called Gibbs energy) to the state functions enthalpy H and entropy S for a given temperature T . While a free energy term can only comprise enthalpic contributions, in general both enthalpy and entropy can contribute to the free energy. The solvation free energy for example partly arises from an unfavorable entropy loss because of the increase of water molecules oriented at the solute surface.

In this section, enthalpic and entropic contributions from translational (t), rotational (r) and vibrational (v) degrees of freedom are discussed. They can be calculated as

$$\Delta G_{entropy} = \Delta G^v + \Delta G^t + \Delta G^r \quad (2.23)$$

A system of N atoms and $3N$ total degrees of freedom has 3 translational, 3 rotational and $3N - 6$ vibrational degrees of freedom. Upon complex formation, mobility of the unbound ligand is lost, whereby the translational and rotational free energy change is unfavorable. 3 translational and 3 rotational degrees of freedom are converted to 6 additional normal modes, resulting in a favorable vibrational free energy change.

Vibration

Vibrational entropy and enthalpy effects in association reactions have been studied by simulations in detail [48, 49]. The thermodynamic quantities Helmholtz free energy A and entropy S and can be derived using statistical thermodynamics. In the case of the vibration [49], they are given by

$$S^v = R \sum_{i=1}^{3N-6} \left[\frac{h\nu_i}{k_B T} \frac{1}{e^{h\nu_i/k_B T} - 1} - \ln \left(1 - e^{-h\nu_i/k_B T} \right) \right] \quad (2.24)$$

and

$$A^v = \sum_{i=1}^{3N-6} \left[\frac{1}{2} h\nu_i + \frac{h\nu_i}{e^{h\nu_i/k_B T} - 1} \right] - S^v T, \quad (2.25)$$

in which h is the Planck constant, ν_i the normal mode frequency, k_B the Boltzmann constant, and $\sum_{i=1}^{3N-6} \frac{1}{2} h\nu_i$ is the zero point energy. Here, the Gibbs free energy equals the Helmholtz free energy, because the reaction takes place in solution, and thus volume changes can be neglected. Hence, the vibrational free energy change can be expressed as $\Delta G^v = A^v(C) + A^v(L) + A^v(P)$.

ΔG^v depends on the vibrational frequencies of protein, ligand and complex. For the normal mode analysis in CHARMM, the Hessian matrix, which is the matrix of second derivatives of the energy with respect to displacements of all pairs of atoms in x, y, and z directions, is calculated. This is done for both the flexible protein and the ligand atoms, using the MM and QM potential energy function, respectively. The current CHARMM version does not allow a combined QM/MM normal mode analysis with atoms kept fixed, so that the code of the `vibrant` module in CHARMM was modified to establish this interface (see Appendix C). The Hessian matrix is diagonalized to give the force constants for the molecule. The mass-weighted force matrix is then used for calculating the vibrational frequencies.

Translation and Rotation

The translational contribution is given by the Sackur-Tetrode equation [50]

$$S^t = R \cdot \left[\frac{5}{2} + \frac{3}{2} \ln \left(\frac{2\pi m k_B T}{h^2} \right) - \ln(\rho) \right] \quad (2.26)$$

and

$$A^t = \frac{3}{2} RT - S^t T. \quad (2.27)$$

ρ donates the number density, which is defined as the product of the concentration and the Avogadro number. The translational free energy only depends on the system's mass.

Accordingly, the rotational enthalpy and entropy are defined as

$$S^r = R \cdot \left[\frac{3}{2} + \frac{1}{2} \ln(\pi I_A I_B I_C) + \frac{3}{2} \ln \left(\frac{8\pi^2 k_B T}{h^2} \right) - \ln(\sigma) \right] \quad (2.28)$$

and

$$A^r = \frac{3}{2} RT - S^r T, \quad (2.29)$$

where I_X with X denoting A, B or C respectively, are the principal moments of inertia. The symmetry factor σ is the number of equivalent positions, that can be transformed into one another by rotation. For ligands of set n0 in the unbound state, σ equals two, for n1 ligands, the protein and all complexes, σ is one. The moments of inertia and thus the rotational free energy depend on atom masses and distances.

As for the vibrational terms, $\Delta G^{t,r}$ is approx. the Helmholtz free energy $\Delta A^{t,r}$, so that the translational and rotational contributions to the binding free energy can be expressed by

$$\Delta G^{t,r} = A^t(C) - A^t(L) - A^t(P) + A^r(C) - A^r(L) - A^r(P). \quad (2.30)$$

Chapter 3

Results

The results of the free energy calculations are listed in Appendices D and E. Because the experimental values are not yet published by Hoffmann-LaRoche for all ligands investigated here, they can only be listed in the Appendix without the corresponding structure.

The aim of the thesis is to show that binding free energy calculations can be improved by taking protein flexibility into account. For this purpose, the results obtained from treating the system flexibly have to be compared to results obtained for a rigid protein. However, the data for the rigid protein approach had been already derived by using the MAB force field, whereas here in the flexible protein approach a combination of the CHARMM force field and the semi-empirical AM1 description are used. In order to discriminate between differences in the results due to another molecular description and differences due to the introduction of flexibility, calculations were performed for the ligands of set n0 using the CHARMM-AM1 method and a rigid protein and ligand (Section 3.1). As a next step, the same protocol, including flexibility, was applied to the set n0 and n1 (Section 3.2).

The binding free energy was calculated as the sum of its contributing energy terms according to Equation 2.11. In empirical force fields, the different terms are balanced in order to fit to experimental data. Van der Waals parameters, partial atomic charges and valence parameters are adjusted properly for a certain force field, and hence are interdependent. Here, however, the system is considered to be surrounded by water instead of vacuum, and Coulomb energies are substituted by solvation free energies. To re-establish the balance of the terms summed-up to give the binding free energy, a factor α was introduced to scale the van der Waals energy contribution. It was obtained by fitting the calculated to the experimental results using the method of 'least-squares'. In the tables of results in the Appendix, the weighted values are listed.

3.1 Rigid protein and ligand

The computational procedures for the calculations for a fixed system correspond to those described in Section 2.2.6, but differ from them in the following respects. During the minimization, all protein atoms are kept fixed and only ligand atoms are allowed to move. The obtained ligand coordinates are assumed to represent the ligand conformation in both the bound and unbound states. For the normal mode analysis, the energy minimized free ligand was used. As a consequence of the rigidity, the intramolecular interactions including the quantum mechanical energy do not change during binding, and thus cancel out in the 'Master equation' 2.11. The binding free energy then becomes

$$\Delta G_{bind} = \Delta G_{coul}(P : L) + \Delta G_{vdw}(P : L) + \Delta G_{solv} + \Delta G_{entropy}. \quad (3.1)$$

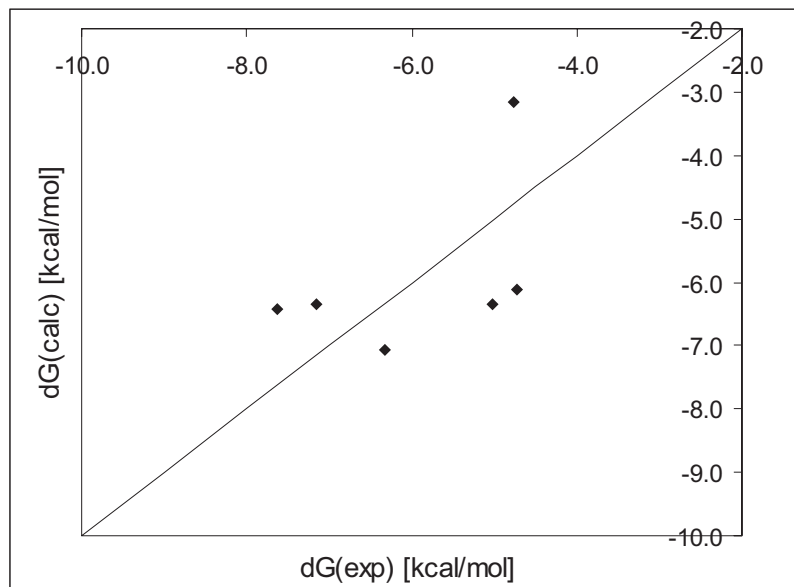


Figure 3.1: Calculated versus experimental binding free energies for the ligands of set n0 assuming a rigid system.

Here, the atoms do not move upon complex formation. Thus, the artefactual grid error cancels out when calculating the electrostatic energy by solving the Poisson-Boltzmann equation with the finite differences method. The electrostatic energy ΔG_{elec} accounts for the protein:ligand Coulombic interactions $\Delta G_{coul}(P : L)$ as well as the polar contributions to the solvation free energy $\Delta G_{solv,p}$. Thus, the binding free energy can be calculated as,

$$\Delta G_{bind} = \Delta G_{vdw}(P : L) + \Delta G_{elec} + \Delta G_{solv,np} + \Delta G_{entropy}. \quad (3.2)$$

The docking protocol resulted in only one conformation of reasonable energy per complex, so that no ensemble was considered. The results for set n0 are listed in table D.1. Figure 3.1 shows a plot of calculated binding free energies against experimental values. With a root mean square deviation of $\text{RMSD} = 1.22$ kcal/mol of the calculated binding free energies relative to the

experimental values, the CHARMM-AM1 method is less accurate than the method based on the MAB force field (RMSD = 0.55 kcal/mol [24]).

3.2 Flexible protein and ligand

3.2.1 Conformational ensembles

For each ligand, 70 different structures of its complex with trypsin were generated using the docking protocol described in Section 2.1.2. After removing very energetically unfavorable structures, the others were clustered into groups of similar energies, out of which one candidate was taken up into the ensemble. This procedure gave ensembles of 2 to 6 complex structures per ligand, or 3.5 on average. For the set n0, the mean number of structures in an ensemble is only 2.3.

A measure for the structural deviation between complexes of one ensemble is the root mean square deviation of their atomic coordinates. It was calculated from the deviation of the coordinates of the flexible atoms of only those structures significantly contributing to the Boltzmann weighted free energy (with $f_i > 0.1$) The RMSD then ranges from 0.1 Å to 1.5 Å.

The energy function of the CHARMM force field (Eq. 2.1) was used for the energy minimisation and the selection of the structures representing an ensemble. The free energy calculations were then carried out with the free energy function ('Master equation' 2.11), followed by the Boltzmann weighting and ensemble averaging. To examine to which degree they correlate with each other, the results of the complex energy calculations based on both functions are compared (see Table E.1 and E.2 in App. E). Only for few ensembles, the

structures are ranked similarly according to the CHARMM energy function and the free energy function. If not the same ranking, then often a similar tendency can be observed for many ensembles. On average the energy deviation within one ensemble is higher for the calculations with the CHARMM force field (4.5 kcal/mol) than for the calculations with the free energy function (1.0 kcal/mol).

According to the relative energies the complexes of one ensemble were Boltzmann weighted. Tables E.1 and E.2 in Appendix E shows the total energies and resulting weighting factors. All contributions to the complex free energy are listed, even if not varying among the conformations, as it is the case for the translational and rotational contributions. At room temperature $R \cdot T$ equals 0.59 kcal/mol. Thus, conformations with an energy difference of this order of magnitude are equally occupied, while conformations of an energy approx. 3 kcal higher than the lowest local minimum found contribute less than 1% to the average energy. Consequently, for several ligands, the simulations gave only a single structure that is occupied at room temperature.

3.2.2 Binding free energies

Tables E.3 and E.4 gives the results for ΔG_{val} , ΔG_{coul} , ΔG_{vdw} , ΔG_{QM} , ΔG_{solv} , ΔG_v and ΔG_{tr} for set n0 and n1. The root mean square deviation (RMSD) between ΔG_{calc} and ΔG_{exp} is 1.41 kcal/mol. The fitting of the calculated binding free energies to the experimental values led to a scaling factor $\alpha = 0.78$.

For one ligand, the binding free energy is predicted too unfavorable by 4.6 kcal/mol relative to the experimental value. However, the error $|\Delta G_{exp} -$

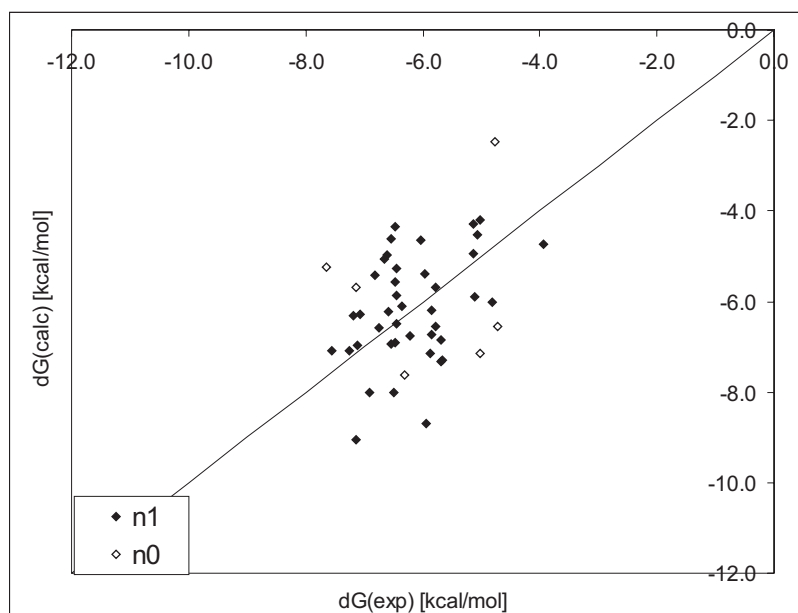


Figure 3.2: Calculated versus experimental binding free energies for the sets n0 and n1, one outlier excluded.

$|\Delta G_{calc}|$ for all other ligands is not no greater than 2.8 kcal/mol. This finding indicates that the outlier has some unique reason, which is different from other uncertainties causing deviations. A root mean square deviation of 1.26 kcal/mol of the calculated relative to the experimental values is obtained, when neglecting the outlier ($\alpha = 0.78$). The calculated binding free energies without the outlier are plotted versus the experimental binding free energies in Figure 3.2. The experimental binding free energies of set n0 was reproduced with an RMSD of 1.94 kcal/mol ($\alpha = 0.79$), so on average less accurately than with the same CHARMM and AM1 method not including flexibility (RMSD = 1.22 kcal/mol). In contrast, a comparatively small RMSD of 1.12 kcal/mol is observed for the ligands of set n1 ($\alpha = 0.78$).

Chapter 4

Discussion

4.1 Fixed protein and ligand

For the ligands of set n0, the pure molecular mechanical (MM) method using the MAB force field was able to predict the binding free energies with a higher accuracy than the combined quantum mechanical and molecular mechanical (QM/MM) method using the CHARMM force field and AM1. The results obtained using MAB are presented in Ref. [24]. Figure 4.1 shows the binding free energies calculated with the QM/MM method in comparison to those obtained with the MM method

In Table 4.1 the range of the energy terms contributing to the binding free energies are listed for comparison. The main difference between the MAB force field and the CHARMM force field is the van der Waals energy term. This term is bigger in the case of the CHARMM force field, and therefore is scaled down by $\alpha = 0.51$ instead of $\alpha = 0.85$ for the MAB force field. The scaled values vary among the ligands by $\Delta\Delta G_{vdw} = 1.7$ kcal/mol (CHARMM) and $\Delta\Delta G_{vdw} = 3.2$ kcal/mol (MAB). Thus, van der Waals energy contributions obtained from

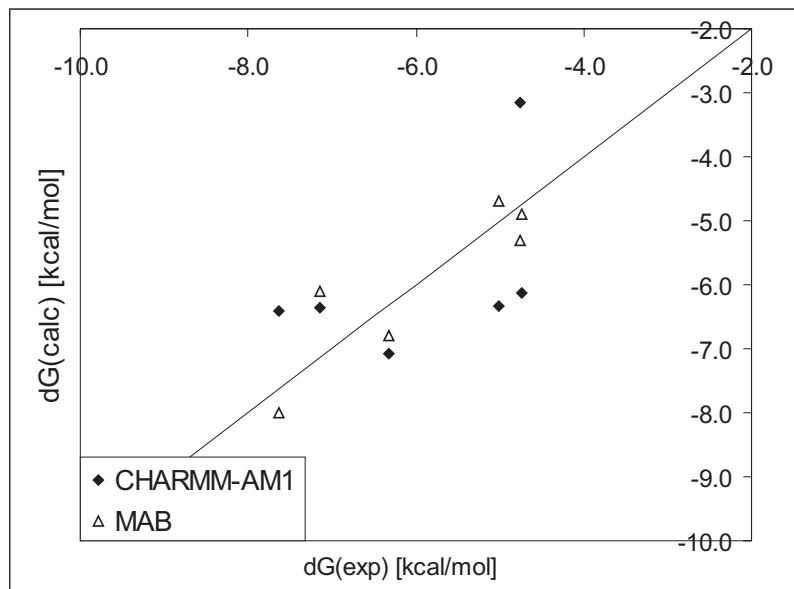


Figure 4.1: Calculated versus experimental binding free energies for the set n0 assuming a rigid system and using the MAB force field and the CHARMM force field combined with AM1.

CHARMM influence the ranking of the ligands according to their calculated binding free energy less than those from MAB. The range of the electrostatic contributions $\Delta\Delta G_{elec}$ is similar for both cases, but they differ significantly in the absolute values of ΔG_{elec} for one ligand. The non-polar contribution to the solvation free energy as well as the entropic terms are similar and give approximately the same ranking for the ligands.

Differences in the complex structures are mainly due to a rotation of the aromatic ring out of the plane of the amidinium group by approx. 15 degrees. The amidinium group itself is positioned very similarly and at the same distance from the Asp 189 carboxylate group, while the benzene ring position in the structures after the CHARMM-AM1 minimisation differs from its position

Table 4.1: Comparison of the contributions to the binding free energy using CHARMM-AM1 with the contributions using MAB for the ligands of set n0 assuming a rigid system. All energies are in kcal/mol.

	ΔG_{vdw}		ΔG_{elec}		$\Delta G_{solv,np}$		ΔG^v		ΔG^{tr}	
	min	max	min	max	min	max	min	max	min	max
MM	-16.1	-11.9	-1.2	0.8	-6.9	-5.7	-3.8	-2.5	16.5	18.0
QM/MM	-14.5	-12.8	-0.8	1.3	-6.9	-5.8	-3.5	-2.7	16.5	17.9

in the MAB energy minimized structures.

It can be concluded, that for this special case, the procedure based on the MAB force field performs better than the combined QM/MM method using CHARMM and AM1. Possible reasons might be the inconsistency in the CHARMM van der Waals parameters, which were partly taken from the MAB force field, or an unbalanced energy function when introducing quantum mechanics. For the energy minimisation, the ligand is treated quantum mechanically, allowing for changes of the charge density in the aromatic system upon ligand binding. Nevertheless, for the ΔG_{bind} calculations, the ESP partial charges derived by *ab initio* simulations in vacuum have been used for the ligand in the bound and unbound states, and polarization is neglected. When introducing a polarisation term into the energy function, which then accounts for a favorable energy due to a distortion of the electron density in the ligand as response to the protein dielectric field, a deviation of RMSD = 1.00 kcal/mol was found. Thus, here the accuracy can be increased by including polarisation effects into the calculation of binding free energies. However, the fact that the data set used contains only six points must be kept under consideration.

The test calculations with a fixed protein serve as a comparison for the calculations including flexibility as discussed in the following section. In general, only minor changes in the terms of the energy function were observed when using quantum mechanics for the ligand. The accuracy was satisfactory, so that the QM/MM method was used to introduce protein flexibility into the calculation.

4.2 Flexible protein and ligand

4.2.1 Conformational ensembles and total complex energies

One of the reasons why trypsin and benzaminidine-like inhibitors have been chosen for this study is their conformational rigidity. However, previous studies have shown that considering the system as partly flexible might improve the prediction of binding affinities [25]. The results from the previously described docking procedure indicate that several conformations instead of only the one of lowest energy should be considered. An ensemble, which consists of up to six structures, represents the conformational states that the complex can occupy. The fact that ligands of set n0 form a smaller number of complexes than those of set n1 reflects well their lower conformational variability.

While the docking of set n0 ligands gave structures with an RMSD of 0.1 to 0.3 Å, set n1 ligands can form complexes deviating by 1.0 to 1.5 Å from each other. Such RMSD values are typically found in the case of ligands of high variability, or sterically not very demanding ligands, for which two favorable positions are differing in a 180 degree rotation but are identical otherwise. An

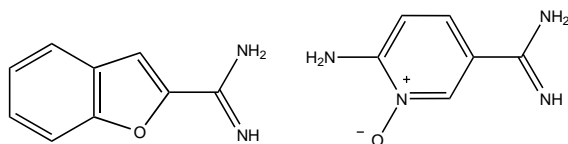


Figure 4.2: Structures of benzofurane-3-carboxamidine and N-oxo-pyridine-2-carboxamidine

example is benzofurane-3-carboxamidine (see Fig. 4.2), of which the complex can be described by an ensemble of two structures with an RMSD of 0.9 Å. The deviation arises mainly from the rotation of the ligand, whereas the protein structures are very similar.

By way of contrast, ligands with an aromatic nitrogen atom at the meta-position to the amidinium group are found to form one complex, in which the catalytic triade shows the same hydrogen bond pattern as in the uncomplexed trypsin (see Figure 1.3), and another one with a ligand position rotated by 180 degrees, in which the hydroxyl group in Ser 195 turns around to the aromatic nitrogen atom. The corresponding N-oxide, N-oxo-pyridine-2-carboxamidine (see Fig. 4.2) also has such a binding behavior. A hydrogen bond from Ser 195 to the oxygen atom of the ligand is of optimal geometry. The two possible states are both significantly occupied ($\Delta\Delta G_C = 0.8$ kcal/mol, $f_i/f_j = 4 : 1$) and are shown in Figure 4.3.

It should be noted that in the docking and minimisation method the conformational space is not explored completely and systematically. Hence, it is only an assumption that the minimised structures represent the lowest local minima of the trypsin-ligand complex. Furthermore, only the ligand position was varied, whereas the starting coordinates of the protein was the same for each minimisation. It remains to be seen if an extension of the conforma-

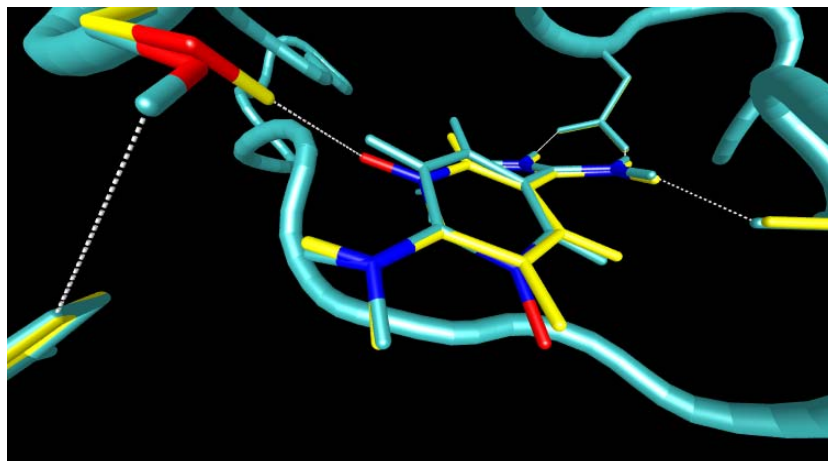


Figure 4.3: Overlap of two representative structures of N-oxo-pyridine-2-carboxamide inhibited trypsin, showing an intact catalytic triad (cyan) or a ligand-Ser 195 hydrogen bond (yellow), respectively.

tional sampling by rotations of protein side chains in the binding pocket or internal rotations of ligand residues or by Monte Carlo sampling leads to an improvement in the calculations.

For the energy minimisation, the recently developed charge reparametrisation method was used to account for solvent screening effects. In order to examine the influence of using scaled charges for the coordinate generation and subsequent free energy calculation, the energy minimisation was carried out also with the original charges for the ligands of set n0 and two ligands of set n1 (results not shown). The binding free energies were then predicted with an RMSD of 1.89 kcal/mol for these 8 ligands. A main difference in the contributing energy terms was found for the valence energy $\Delta G_{val}(P)$, which was favorable. The conformational strain in the unbound protein was bigger than in the bound protein by up to 6 kcal/mol (unscaled charges), which is not the case when using scaled charges. This result is rather surprising, as the

conformational adjustment of the protein to the ligand is expected to lead to an increase in the valence energy upon binding.

While the CHARMM potential energy function was used for the energy minimisation and to roughly discriminate between favorable and unfavorable conformations of inhibited trypsin, the Boltzmann weighted average was calculated with the free energy function taking solvation and entropic effects into account. The comparison of the ranking according to these two functions shows, that in principle a similar tendency is obtained. Because for the calculations presented here, ensembles contain conformations differing by up to 25 kcal/mol in their CHARMM potential energy, it can be assumed that all significantly contributing structures are considered, and hence the ensemble is representative. The structural deviations were not quantitatively examined for the ensemble generation, so that from a number of candidates of different conformations but with a free energy difference of less than 0.002 kcal/mol only one representative is chosen. In order to overcome an eventual neglect of conformations in addition to the energy cut off an RMSD calculation or some geometric criteria would have to be introduced into the ensemble generation protocol.

4.2.2 Binding free energies

For the total set of 48 n0 and n1 ligands, the experimental binding free energies could be reproduced by calculations including protein and ligand flexibility with a deviation of $\text{RMSD} = 1.41$ kcal/mol. The binding free energies were predicted to range from -9.2 to -2.6 kcal/mol with one outlier at -1.1 kcal/mol ($\Delta G_{exp} = -5.7$ kcal/mol), while the experimentally determined values range

from -7.6 to -3.9 kcal/mol. The outlier a ligand with a rather rigid and sterically demanding residue consisting of three condensed five-membered rings containing hetero-atoms, located in para-position to the amidinium group. The local minima used for the calculations were found only after an extended manual docking procedure including the exploration of internal rotational degrees of freedom. The inaccurate prediction can be traced back to a particularly high increase in the structural strain of the bound ligand ($\Delta G_{QM} = 3.9$ kcal/mol), which is even higher in other complex conformations obtained from the docking. It can be concluded that the calculation failed in reproducing the experimental binding free energy in this case, because the actual local minimum was not found. An extension of the flexible region might overcome this problem.

However, the procedure performed well in other cases of rather big and structurally not well defined ligands with internal degrees of freedom. When excluding the outlier discussed above, the average error between calculated and experimental binding free energies is $\text{RMSD} = 1.26$ kcal/mol. Thus, the accuracy obtained when taking conformational variations into account is approximately the same as for calculations treating the system as rigid and using the same CHARMM-AM1 molecular model ($\text{RMSD} = 1.22$ kcal/mol for set n0). The method is capable of identifying inhibitors for which docking does not lead to a reasonable protein:ligand structure by a remarkably high conformational strain.

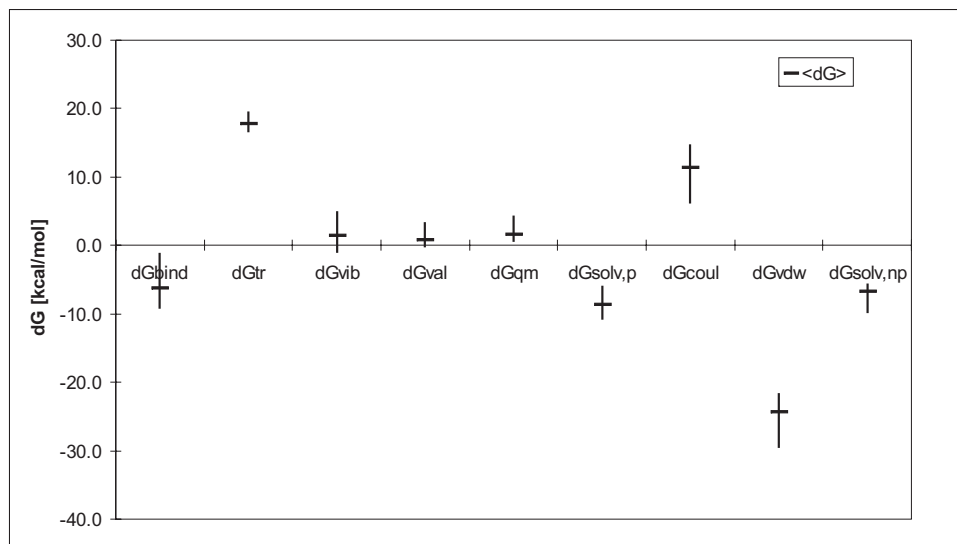


Figure 4.4: The range of the calculated binding free energies and the contributing terms for the ligands of set n0 and n1 (including the outlier). $\langle \Delta G \rangle$ denotes the average value over all ligands.

4.2.3 Contributions to the binding free energy

Partitioning of the total binding free energy into the contributing terms allows an analysis of the underlying effects. The deviation of the calculated binding free energies from the experimental values of more than 1 kcal/mol is an indication of the errors in the calculated energy terms. Nevertheless, conclusions concerning the effects driving binding can be drawn from the calculations.

Figure 4.4 shows the average, maximum and minimum energy values obtained for the sets n0 and n1 (including the outlier) of the calculated binding free energies and their contributing terms. The entropic term is partitioned into vibrational and the sum of translational and rotational contributions, and the electrostatic energy into the Coulombic interaction energy and the polar contribution to the solvation free energy. In general, the changes in the van

der Waals interactions and in the polar and non-polar solvation free energy favour ligand binding in general, whereas changes in the Coulombic interactions, in the intramolecular MM and QM energy and entropic effects disfavor binding. The comparison with the terms obtained for a rigid system and set n_0 shows that the energies lie in a similar range when introducing flexibility. The free energy function derived for a flexible system takes additional energy terms for intramolecular interactions into account, which are on average unfavorable. Besides, enthalpic and entropic effects arising from changes in the vibrational modes of the system are slightly unfavorable, which is in contrast to simulations assuming a rigid protein .

The van der Waals energy $\Delta G_{vdw}(P, P : L)$ is the most favorable energy contribution, and ranges from -21.6 to -29.6 kcal/mol (already scaled by $\alpha = 0.78$), therefore influencing the ranking of the ligands according to their binding free energies significantly. It is mainly caused by a negative protein:ligand energy term $\Delta G_{vdw}(P : L)$ due to dispersion effects between them. The larger the contact area between the ligand atoms and the protein atoms in the binding pocket, the higher the absolute value of the van der Waals contribution. On average, the intramolecular protein van der Waals energy $\Delta G_{vdw}(P)$ is unfavorable by 1.5 kcal/mol.

The valence energy term and the quantum mechanical energy term account for the increase in conformational strain in the protein and ligand, respectively, upon complex formation. As expected both terms are mostly unfavorable ($\Delta G_{val}(P) = -0.3$ to $+3.5$ kcal/mol and $\Delta G_{QM}(L) = +0.6$ to $+3.9$ kcal/mol). The variations from ligand to ligand are smaller for both terms than the variations of the van der Waals energy. Nevertheless, valence and quantum me-

chanical energy contributions are influencing the relative binding free energies ($\Delta\Delta G_{exp} = 3.7$ kcal/mol). Strictly speaking, $\Delta G_{QM}(L)$ also contains van der Waals and electrostatic interactions within the ligand, but cannot be further decomposed.

The favorable non-polar contribution to the solvation free energy is calculated as linearly dependent on the solvent accessible surface area. Thus, the more the ligands under consideration differ in their size, the more important is the role of $\Delta G_{solv,np}$ for the correct ranking of inhibitors. Here, it varies by up to 4.2 kcal/mol among the ligands.

The entropic and enthalpic contributions due to changes in the translational and rotational degrees of freedom lie within a range of 16.5 to 19.4 kcal/mol, thus are highly unfavorable. Although the number of vibrational degrees of freedom increases upon binding, the vibrational effects are found to mostly disfavor the association reaction. This finding is in sharp contrast to the results of the calculations assuming the protein to be rigid [24], that gave a favorable vibrational free energy ($\Delta G^v < 0$ and $|\Delta H^v| < |\Delta S^v \cdot T|$). When assuming the protein as rigid, the vibrational entropy is favorable ($\Delta S^v \gg 0$) and exceeds the unfavorable vibrational enthalpy including the zero point energy ($\Delta H^v > 0$). Including flexibility leads to a smaller increase in the entropy ($\Delta S^v > 0$), what can be interpreted as a stiffening of the system. Together with a higher increase in the enthalpy ($\Delta H^v \gg 0$) relative to the case of a fixed protein, the enthalpic effects exceed the entropic effect ($|\Delta H^v| > |\Delta S^v \cdot T|$), resulting in a positive ΔG^v .

The calculation of ΔG_{elec} via the thermodynamic cycle allows to examine the competing electrostatic effects. The total electrostatic energy, comprising

polar solvation energies and Coulombic energies, is on average unfavorable and ranges from -0.1 kcal/mol to 6 kcal/mol. The negative polar contribution to the solvation free energy ($\Delta G_{solv,p} = -5.9$ to -10.8 kcal/mol) and the positive change in the Coulombic energy ($\Delta G_{coul} = 6.1$ to 14.8 kcal/mol) almost cancel each other out. However, the electrostatic energy plays an important role in the correct ranking of the ligands. The polar part of the solvation free energy is favorable, because water molecules solvating the hydrophobic regions of the ligand and protein binding pocket are released upon binding to form favorable hydrogen bonds between each other. Despite the formation of the salt bridge between the positively charged amidinium group and the negatively charged Asp 189 residue, the Coulombic interactions highly disfavor the complexation of trypsin, a protein with a total charge of +8, by an inhibitor with a charge of +1.

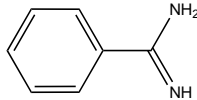
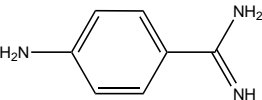
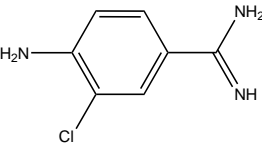
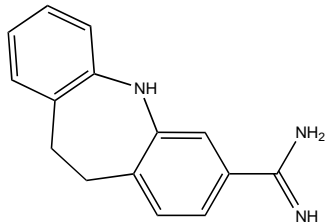
Using AM1 for the minimisation of the ligand makes it possible to take polarisation effects into account. Instead of using partial charges, the ligand electrons are modeled as an electron density delocalized in the conjugated system of the inhibitor. The interaction with the electric field of the protein results in a favorable polarisation energy. However, in the present computational procedure, the protein charges are set to 0 for the calculation of the quantum mechanical energy, so that the polarisation effect is neglected in the free energy calculation but not in the coordinate generation. The recently implemented decomposition module in CHARMM allows to compute the energy of an undistorted wave function in the presence of an electric field. The difference between the resulting quantum mechanical energy and the quantum mechanical energy of the distorted wave function equals the polarisation

energy, which is given by,

$$\begin{aligned} E_{pol} &= G_{QM}(\Psi) - G_{QM}(\Psi^0) \\ &= \langle \Psi | \hat{H}^0 + \hat{H}^{QM/MM} | \Psi \rangle - \langle \Psi^0 | \hat{H}^0 + \hat{H}^{QM/MM} | \Psi^0 \rangle, \end{aligned} \quad (4.1)$$

where Ψ is the distorted wave function, Ψ^0 the undistorted wave function, \hat{H}^0 the Hamiltonian of the system in vacuum, and $\hat{H}^{QM/MM}$ the Hamiltonian for the electrostatic interaction with the MM region [51]. The polarisation energy can be further decomposed into the energy needed to distort the electron density in vacuum, $E_{dist} = \langle \Psi | \hat{H}^0 | \Psi \rangle - \langle \Psi^0 | \hat{H}^0 | \Psi^0 \rangle$, and the energy gain in the QM/MM electrostatic interactions due to the distortion, $E_{stab} = \langle \Psi | \hat{H}^{QM/MM} | \Psi \rangle - \langle \Psi^0 | \hat{H}^{QM/MM} | \Psi^0 \rangle$. For four example ligands the results are listed in Table 4.2. They have been calculated for complexes of all ligands of set n0 and n1 using scaled protein charges to account for the screening of the polarisation effect by the solvent. The favorable stabilisation energy lies in the range of -3.0 to -5.7 kcal/mol and compensates the unfavorable distortion energy, which varies from 1.5 to 2.8 kcal/mol in the sets n0 and n1. The lowest absolute value of the polarisation energy is observed for benzamidine (-1.5 kcal/mol). The increase of the calculated polarisation energy correlates well with the increase in polarizable rests and the size of the aromatic system. When introducing the polarisation energy into the binding free energy calculation for the ligands of set n0 assuming a rigid protein, the accuracy was increased (RMSD = 1.00 kcal/mol). However, in the case of a flexible system, the calculated binding free energies including E_{pol} deviated from the experimental results by an RMSD of 1.42 kcal/mol for the ligands of set n0 and n1. Thus, by considering the polarisation effect in this way, the prediction of binding free energies cannot be improved.

Table 4.2: Results of the polarisation energy calculations for some example ligands using scaled charges on the protein atoms. $G_{QM}(\Psi)$ denotes the energy of the distorted ligand electrons, $G_{QM}(\Psi^0)$ the energy of the undistorted ligand electrons of the bound ligand. All energies are in kcal/mol.

	$G_{QM}(\Psi)$	$G_{QM}(\Psi^0)$	E_{pol}	E_{stab}	E_{dist}
	-80.4	-78.9	-1.5	-3.0	1.5
	-86.6	-84.7	-2.0	-3.8	1.9
	-89.5	-87.3	-2.2	-4.3	2.1
	-89.9	-87.1	-2.8	-5.6	2.8

A crucial parameter for the estimation of polarisation effects is the dielectric constant of the protein, because the electron density particularly sensitively responds to the electric field. Hence, the polarisation energy strongly depends on the strength of the external field. Here, the screening effect onto the electrostatic interactions is modeled by reparametrising the protein charges, resulting in a dielectric constant of $\epsilon(solv) = 80$ for the solvent, $\epsilon(solv) = 4$ and $\epsilon(lig) = 1$ for the ligand, as the latter is treated quantum mechanically. A straightforward method using a dielectric constant of $\epsilon = 4$ for the calculations of all electrostatic interaction energies is necessary. The modelling of electrostatic interactions used so far is inconsistent also for the following reason: Partial charges representing the distorted electron wave function differ from the ESP charges, which were derived for the free ligand and were used for energy calculation of the ligand in the unbound as well as in the bound state.

In order to calculate the electrostatic interactions of the protein partial charges with the polarized ligand instead of those with the ESP charges of the undistorted ligand, $\Delta G_{QM}(L)$ could be calculated without switching the partial charges of the protein off. It would also contain the Coulombic protein:ligand interaction energy, so that the modified 'Master equation' is then given by

$$\begin{aligned} \Delta G_{bind} = & \Delta G_{val}(P) + \Delta G_{coul}(P) + \Delta G_{vdw}(P, P : L) \\ & + \Delta G_{QM}(L, P : L) + \Delta G_{solv} + \Delta G_{entropy}. \end{aligned} \quad (4.2)$$

The terms containing Coulombic energy, $\Delta G_{coul}(P)$ and $\Delta G_{QM}(L, P : L)$, have to be calculated for $\epsilon = 4$ because they are accounting for the complexation reaction in the thermodynamic cycle. Since in a quantum mechanical AM1

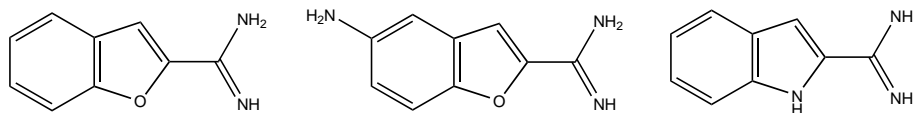


Figure 4.5: Structures of benzofurane-3-carboxamidine, 5-amino-benzofurane-3-carboxamidine, and indole-3-carboxamidine.

description, only vacuum simulations can be performed, a screening of the electrostatic interactions between the quantum region and the MM region by a factor of 4 can only be achieved by dividing the protein partial charges by 4. In this way, the polarisation effect should also be scaled down to a reasonable level. Using this approach, ligand partial charges are only used for the calculation of the solvation free energy. The otherwise consistent treatment of the electrostatics should result in an adequate inclusion of the polarisation energy and a proper balance of the contributing energy terms. It remains to be seen, if the described modification of the Master equation leads to an improvement in the prediction of binding free energies.

4.2.4 Case studies: benzofurane-3-carboxamidine and derivatives

What are the effects increasing the binding affinity of an inhibitor by changing a particular atom or adding a certain residue? This question shall be answered by comparing three ligands with a noticeable difference in their binding free energies in spite of their structural similarity: benzofurane-3-carboxamidine, its amino-substituted analogue 5-amino-benzofurane-3-carboxamidine, and its indole analogue indole-3-carboxamidine (Figure 4.5).

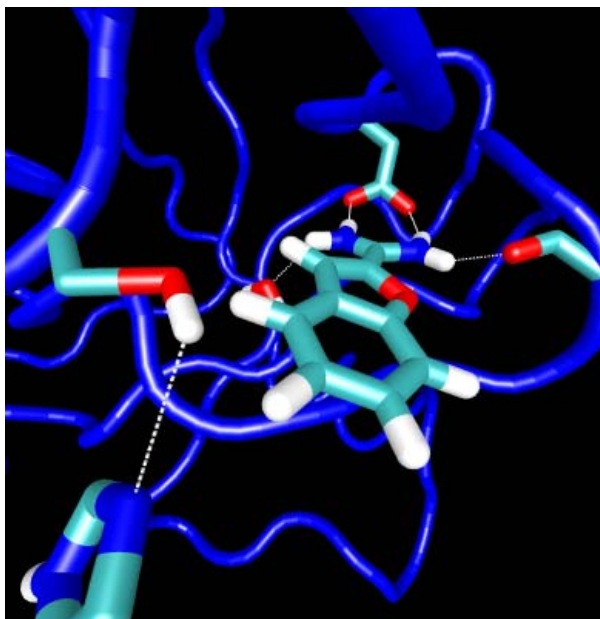


Figure 4.6: Illustration of benzofurane-3-carboxamide inhibited trypsin.

The experimentally determined higher binding affinity of the indole-like ligand relative to benzofurane-3-carboxamide could be reproduced. The calculations predict a difference in the binding free energy of $\Delta\Delta G_{bind} = 0.3$ kcal/mol. Representative complex structures are shown in Figures 4.6 and 4.7.

Benzofurane-3-carboxamide shows two different docking positions, which are identical except for the position of the aromatic oxygen atom. They can be transferred into each other by a 180 degree rotation. In both complex structures the protein conformation corresponds to the structure before minimisation, and thus is practically not affected by the inhibition. The two states differ in their binding free energy by only 0.3 kcal/mol, and are consequently similarly occupied.

In sharp contrast, the corresponding two possible structures of indole-

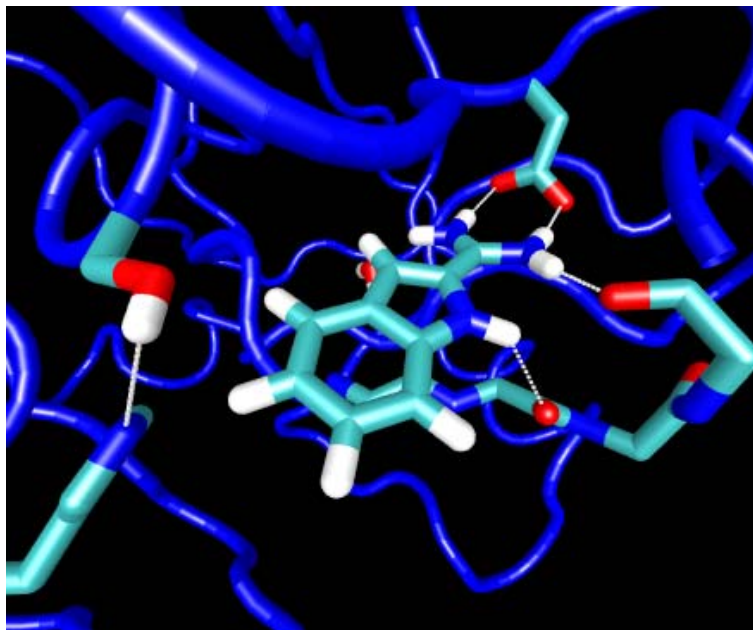


Figure 4.7: Illustration of indole-3-carboxamide inhibited trypsin.

3-carboxamide inhibited trypsin differ in their energies by more than 4 kcal/mol, so that only one structure has to be considered. A remarkable feature of the complex is the hydrogen bond from the backbone carboxyl group of Gly 211 to the H-atom bound to the indole nitrogen of the ligand. In order to achieve a favorable hydrogen bond geometry with a hydrogen-oxygen distance of 2.3 pm, the backbone of Gly 211 and its neighbouring amino acids is distorted. The ligand deviates from its optimally flat geometry as well. This structural characteristic is reflected by an increase in the quantum mechanical energy of the ligand, that is larger than that observed for the benzofurane ligand by $\Delta\Delta G_{QM} = 0.7$ kcal/mol.

The two different ligands are very similar with regard to their shape in the unbound as well as in the bound state. Therefore, the van der Waals energy contribution to the binding free energy can be estimated from their

difference in the ligand surface. The hydrogen atom of the indole ring leads to a bigger favorable van der Waals energy in comparison to the benzofurane ring ($\Delta\Delta G_{vdw} = 0.7$ kcal/mol).

A significant difference in the binding effects of the two ligands is also observed for the electrostatic interactions. The binding of benzofurane-3-carboxamide to trypsin leads to an unfavorable change in the Coulombic energy, that is higher than in the case of the corresponding indole ligand ($\Delta\Delta G_{coul} = 1.2$ kcal/mol). The latter forms an additional hydrogen bond resulting in a more favorable electrostatic protein:ligand interactions. The high tendency of this ligand to form strong hydrogen bonds leads on the other hand to a more negative solvation free energy $\Delta G_{solv,p}(L)$, which results in a less favorable change in the solvation free energy upon binding ($\Delta\Delta G_{solv,p} = -0.9$ kcal/mol). Because these two counter effects almost cancel each other out, the electrostatic energy contribution differs by only 0.3 kcal/mol. The other contributions do not show remarkable differences.

When adding an amino group to the benzofurane-like ligand in 5-position, a less favorable binding free energy is observed both in the experiment and in the calculations. It is mainly caused by a more positive vibrational free energy contribution ΔG^v . This finding could be interpreted as follows: the amino-substituted inhibitor is bound to trypsin by an additional hydrogen bond from the amino group to the aromatic nitrogen atom of His 57 in the catalytic triad (Figure 4.8). This is true for both possible rotations of the ligand in the docked position. In comparison to the ligand without amino substituent, the ligand is conformationally fixed to a higher extent in the binding pocket. The vibrational flexibility and therefore the increase in vibrational entropy is lower,

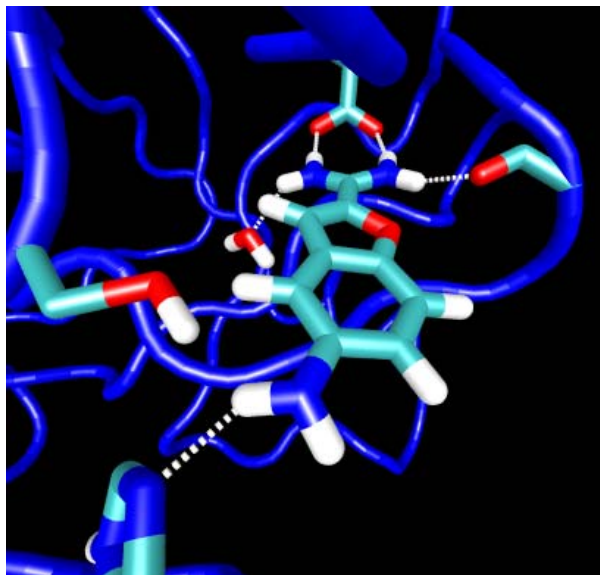


Figure 4.8: Illustration of 5-amino-benzofurane-3-carboxamide inhibited trypsin.

resulting in a less favorable free energy due to vibrational effects.

The additional hydrogen bond the amine-substituted inhibitor can form leads to less unfavorable ΔG_{coul} and to a less favorable $\Delta G_{solv,p}$, resulting in a total electrostatic contribution to the binding free energy very similar to benzofurane-3-carboxamide. In spite of the additional amino group the two ligands hardly differ in their van der Waals energy, because the NH_2 group shows very little hydrophobic interactions with the binding pocket.

4.3 Comparison: flexible versus fixed protein and ligand

Allowing the protein and ligand atoms to move in order to take up an energetically favorable conformation leads to the generation of complex structures

with eye-catching differences from those obtained by modelling the system as rigid:

- More than one structurally and energetically reasonable complex is obtained and has to be considered.
- Hydrogen bonds can be formed by movements of polar groups of the protein side chain or backbones in the direction of the ligand.
- Slight adjustments of protein residues at a certain cost of conformational energy can take place if compensated by favorable van der Waals and polar interactions.
- The catalytic triade might change its geometry, and thereby loose its function for sterical reasons or because of the possibility of more favorable interactions with the inhibitor. Those cases are rather rare and demand optimal substitutes for the destroyed hydrogen bonds in order to result in low free energies.

The prediction of binding free energies could be improved by the introduction of protein and ligand flexibility. Assuming the system to be rigid resulted in calculated binding free energies deviating from the experimental values by $\text{RMSD} = 3.20 \text{ kcal/mol}$ for the ligands of set n0 and n1 (for MAB, $\epsilon = 1$). The improvement of the energy calculation by allowing ligand and protein atoms to move is reflected by an RMSD of 1.26 kcal/mol obtained when using CHARMM and AM1. In Figure 4.9, the binding free energies calculated with: the MAB force field within the rigid protein approach; and the CHARMM-AM1 method and considering the protein as flexible, are plotted against the experimental

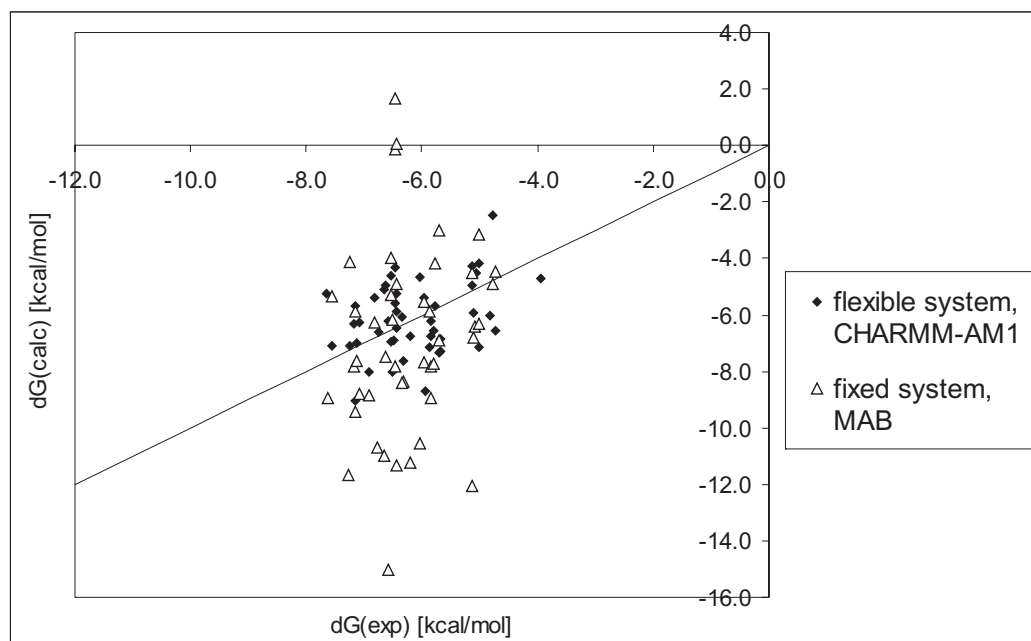


Figure 4.9: Calculated versus experimental binding free energies for the sets n0 and n1 using the MAB force field and assuming the system as rigid, and using the CHARMM force field combined with AM1 and assuming a flexible system.

energies. A much stronger correlation of the calculated and experimental values is achieved with the combined QM/MM computational method including flexibility.

The introduction of protein flexibility does not improve the prediction of binding affinities for the set n0 ligands. In this particular case of six ligands, accuracy is even lost to some extent (RMSD = 1.94 kcal/mol instead of RMSD = 0.55 kcal/mol for MAB, $\epsilon = 4$, or RMSD = 1.4 kcal/mol for MAB, $\epsilon = 1$). The ligands of set n0 show less conformational variability. Accordingly, the conformations of trypsin inhibited by those ligands can be assumed to

be very similar to the crystal structure of the benzamidine-trypsin complex. However, allowing coordinate changes during the minimisation might lead to structurally different local minima. In the case of ligands with unambiguous docking structures, the rigid protein approach is preferable in order to reduce any uncertainties concerning the generation of the complex coordinates.

In contrast, a significant improvement is achieved for ligands of set n1, for which the docking positions are comparatively ambiguous (RMSD = 1.12 kcal/mol instead of RMSD = 3.40 kcal/mol for MAB and a fixed system). This opposite effect of flexibility onto the prediction of binding affinities indicates, that a flexible region of the protein should only be defined if required by the ligands. It should be chosen such that reasonable docking positions can be found for the inhibitors under examination, but at the same time should be restricted to a number of amino acid residues as small as possible. It can also be concluded that the improvement is not due to the use of a combined quantum mechanical and molecular mechanical instead of a pure molecular mechanical method. Modelling the ligand quantum mechanically should change the accuracy of the binding free energy calculation for all ligands to the same degree.

For comparison, the binding free energies have also been calculated using the empirical scoring function of the program LUDI. The results are shown in App. F.

Chapter 5

Conclusions and outlook

5.1 Conclusions

The continuum method to predict absolute binding free energies can be improved by including protein and ligand flexibility into the computational procedure. The new method clearly outperforms the former rigid protein approach in the case of ligands with a certain variability in their docking position and conformation. The accuracy obtained approaches the accuracy when using computationally expensive molecular dynamics simulations for the coordinate generation (approx. 1 kcal/mol) [20, 21].

The complex of a flexible ligand bound to trypsin cannot be described by a single structure. Instead, an ensemble of conformations, representing local minima on the energy landscape of the complex, must be used. Conformations different from the one of lowest energy cannot be neglected when an accurate calculation of binding free energies is aimed at. A flexible region should only be introduced if required by the ligands and should be defined as small as possible. Otherwise, flexibility leads to structural uncertainties rather than a

gain in accuracy by exploring the conformational space of the system, resulting in a less reliable prediction of binding free energies.

A combined method of semi-empirical quantum mechanics for the ligand and molecular mechanics for the protein proved to be suitable for the generation of the complex structures. The advantage, that no parametrization of the ligand is required, is paid for by a higher computational cost. An increase in the accuracy of the structural prediction by using QM/MM instead of MM for the whole system is not observed.

The results for the flexible protein support the conclusion of previous studies, that entropic and enthalpic contributions due to changes in the translational, rotational and vibrational degrees of freedom have to be taken into account in binding free energy calculations. The energy contributions of bonded intramolecular interactions are slightly unfavorable, and influence the ranking of the ligands according to their binding affinity ($\Delta\Delta G_{val} = -0.3$ to 3.5 kcal/mol and $\Delta\Delta G_{QM} = 0.6$ to 3.9 kcal/mol). They can serve as criterion for identifying negligible structures with high conformational strain. Using a thermodynamic cycle gives a detailed insight into the electrostatic effects driving binding. In the case of the cationic inhibitors, a favorable polar solvation free energy of -8.6 kcal/mol on average is exceeded by an unfavorable Coulombic energy of 11.5 kcal/mol. In the present energy function, accounting for polarisation effects in the calculation of electrostatic energy changes does not improve the calculation of binding free energies.

5.2 Outlook

The continuum method including an ideal-gas entropy correction could be improved by introducing protein and ligand flexibility. However, the accuracy is not sufficient for a correct ranking of ligands showing small relative binding affinities, as is the case here. Nevertheless, the potential of continuum methods in discriminating between inhibitors of different affinity could be stressed by this study. The results are encouraging to identify possible sources of errors and to extend the method to other free energy calculations.

A source of error may lie in the inconsistency in the calculation of the electrostatic contributions. The protein:ligand Coulombic interactions are calculated assuming a charge distribution that is identical to that of the free ligand, whereas the underlying complex structures have been obtained by an energy minimisation allowing for a charge polarisation in the ligand. The next step in the improvement of the computational procedure will be the calculation of the electrostatic interactions between the protein partial charges and a distorted charge density of the polarized ligand. An expansion of the quantum mechanically treated region to certain protein residues might also lead to an increase in the reliability of the method. The electron density of the Asp 189 carboxylate group can be expected to be remarkably affected by the presence of a positively charged ligand.

For the calculations assuming protein and ligand as rigid, a protein dielectric constant of $\epsilon = 4$ proved to be adequate, and was therefore also chosen for the calculations presented here. However, allowing the atoms to orientate according to an electric field lowers the effective dielectric constant. Hence, a smaller dielectric constant might be more suitable to account for the lower

screening effect, and might yield a more reliable calculation of the binding free energies.

To date, the docking procedure only samples ligand positions relative to the protein coordinates. In order to overcome the limitations of the conformational search, internal rotations of the ligand as well as of crucial amino acid side chains in the binding pocket have to be explored. Such an expanded docking protocol is particularly necessary for ligands with higher variability than those examined here, as they can be found in the set n2 of the Hoffmann-LaRoche set of benzamidine-like inhibitors.

The results obtained so far encourage us to apply the developed docking procedure and free energy function to systems beyond the model system of trypsin and benzamidine. The method is expected to be applicable to any system undergoing a conformational transition or an association reaction. Examples are the inhibition of the serine protease thrombin by benzamidine derivatives, a very similar test case, or the binding of the natural substrate ATP to the protein myosin. For this purpose the procedure will be further automated aiming at a protocol that requires only some minor modifications from the user and can be applied to any system of interest.

Chapter 6

Zusammenfassung

Die Vorhersage der Bindungsaffinitäten von Liganden an Proteine ist trotz intensiver Bemühungen weiterhin ein Hauptproblem im computergestützten Wirkstoffdesign. Während die relativ genauen auf Molekulardynamik beruhenden Verfahren extrem rechenaufwendig sind, liefern empirische Modelle, die die Konformationsänderungen von Proteinen bei Inhibition zumeist vernachlässigen, eher unzuverlässige Ergebnisse.

Das Ziel dieser Diplomarbeit ist es, eine höhere Genauigkeit in der Vorhersage der freien Bindungsenergie durch die Einführung der Flexibilität von Protein und Ligand in die Berechnung zu erreichen. Dazu wird eine Kontinuumsmethode verwendet, die Solvatationseffekte durch Modellierung des Lösungsmittels Wasser als kontinuierliches Medium und die entropischen Beiträge zur freien Bindungsenergie durch Korrekturterme für Änderungen in Translations-, Rotations- und Schwingungsfreiheitsgraden berücksichtigt. Die freie Energie der Assoziationsreaktion von Ligand und Protein wird in polare und unpolare Wechselwirkungen zwischen Protein, Ligand und Lösungsmittel, in Beiträge kovalenter Bindungen und entropische Anteile aufgeteilt. Dies er-

laubt eine detaillierte Analyse der die Bindung herbeiführenden Effekte.

Das Protein Trypsin und ein Reihe Benzamidin-ähnlicher Liganden, deren Bindungsaffinitäten bereits experimentell in den Laboratorien der Hoffmann-La Roche AG, Basel, bestimmt wurden, dienten als Testsystem für die Anwendung des oben beschriebenen Verfahrens zur Berechnung der Bindungsenergien. Ausgehend von der Röntgenkristallstruktur des Trypsin-Benzamidin-Komplexes wurden die Liganden virtuell in die flexible Bindungstasche des Trypsins gedockt und die resultierende Struktur einer Energieminimierung unterzogen. Dazu wurde eine Kombination quantenmechanischer und molekularmechanischer Methoden verwendet. Ein Docking-Protokoll wurde zur Generierung eines Ensembles möglicher Komplexkonformationen entwickelt, die entsprechend ihrer Boltzmann-Wahrscheinlichkeit in der Bindungsenergieberechnung berücksichtigt wurden.

Die experimentell ermittelten freien Bindungsenergien ($\Delta G_{exp} = -3.9$ bis -7.6 kcal/mol) konnten rechnerisch gut reproduziert werden ($\Delta G_{calc} = -2.6$ bis -9.2 kcal/mol). Eine geringe mittlere quadratische Abweichung von 1.3 kcal/mol im Vergleich zu dem Fehler von Berechnungen unter Annahme eines rigiden Systems (3.2 kcal/mol) zeigt, dass die Vorhersage der Bindungsaffinität von Liganden mit variabler Position und Konformation im Komplex durch die Berücksichtigung von Flexibilität verbessert werden kann. Folglich sind strukturelle Änderungen von Protein und Ligand bei Komplexierung nicht zu vernachlässigen. Es bleibt zu untersuchen, ob eine weitere Verbesserung der entwickelten Methode durch die Einbeziehung der Polarisation des konjugierten Elektronensystems des Liganden im elektrischen Feld des Proteins erreicht werden kann.

Appendix A

Starting positions for ligand docking

The file initialize.str defines the axes, distances and angles for the docking procedure. The definition of the axes (see Figure 2.1) is not shown here.

```
* initialize.str, Frauke Meyer 15.01.02
*

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! distances for translation in Angstrom

SET disforw    1.5 ! along z-axis
SET disbackw  -1.5
SET disright   1.2 ! along y-axis
SET disleft    -1.2
SET disdown    1.2 ! along x-axis
SET disup      -1.2

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! angles for rotation
```

SET angZ1 30

SET angZ2 150

SET angZ3 180

SET angZ4 210

SET angZ5 330

SET angY1 22.5

SET angY2 -22.5

SET angX1 22.5

SET angX2 -22.5

Appendix B

Van der Waals parameters

Table B.1: Minimum van der Waals radii and well depths taken from the MAB force field.

	(minimum radius)/2 [Å]	well depth [kcal/mol]
sulphur atom in thiophene	2.05	0.30
chlorine in chlorobenzene	1.75	0.20

Appendix C

Vibran modification

In the standard CHARMM vibran module [52], possible commands are

REDUce FIX [CMPAct] to perform a normal mode analysis, that takes the presence of fixed atoms, which are not allowed to move, into account.

DIAG FINite to perform a combined QM/MM normal mode analysis. The Hessian matrix is generated from the finite differences of the forces.

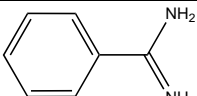
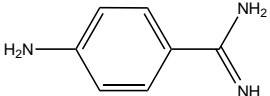
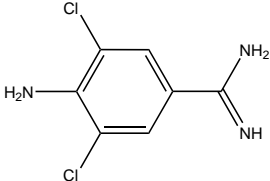
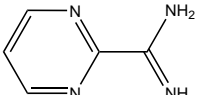
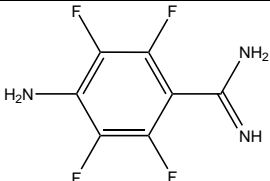
For an analysis with fixed atoms as well as quantum mechanically treated atoms in the system, a combination of the commands is needed, but has not been implemented in the vibran module. In the clf-finite version used for the calculations presented here, the DIAG FINITE algorithm is added to the REDUce FIX command such as the following command is possible.

REDUce FIX FINite [CMPAct] allowing a combined QM/MM analysis of a system containing atoms, which are constrained fix.

Appendix D

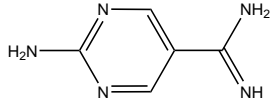
Tables of results: fixed protein

Table D.1: Results of the energy calculations for set n0 assuming the system as fixed. (The van der Waals energy contribution is scaled with $\alpha = 0.51$, all energies are in kcal/mol.)

No.	formula	ΔG_{vdw}	ΔG_{elec}	$\Delta G_{solv,np}$	ΔG^v	ΔG^{tr}	ΔG_{exp}	ΔG_{calc}
1		-13.69	-0.78	-6.04	-3.07	16.50	-6.32	-7.08
2		-14.22	0.00	-6.25	-2.73	16.84	-7.15	-6.36
3		-13.93	-0.70	-6.87	-2.86	17.94	-7.64	-6.42
4		-13.94	0.62	-5.76	-3.52	16.49	-4.73	-6.12
5		-12.82	1.33	-6.77	-2.76	17.86	-4.77	-3.16

continued on next page

continued from previous page

No.	formula	ΔG_{vdw}	ΔG_{elec}	$\Delta G_{solv,np}$	ΔG^v	ΔG^{tr}	ΔG_{exp}	ΔG_{calc}
6	 <chem>Nc1cc(C=O)ncn1</chem>	-14.52	0.46	-6.08	-3.03	16.83	-5.01	-6.34

Appendix E

Tables of results: flexible protein

E.1 Boltzmann weighted ensemble averages

Table E.1: Complex energies, Boltzmann factors, and total CHARMM free energies for set n1 assuming the system as flexible. (The van der Waals energy contribution is scaled by $\alpha = 0.78$, all energies are in kcal/mol.)

No.	G_{val}	G_{vdw}	G_{coul}	G_{QM}	$G_{solv,p}$	$G_{solv,np}$	G^v	G^{tr}	$G_{tot}(C)$	f_i	G_{MM}
1	45.55	-195.63	-12528.9	173.68	-548.55	189.10	731.20	-29.92	-12163.5	0.96	-549.55
1	45.29	-195.78	-12526.3	174.01	-549.23	189.01	731.38	-29.92	-12161.6	0.04	-540.27
2	45.50	-196.14	-12527.6	166.05	-548.69	189.03	742.03	-29.92	-12159.8	0.52	-569.72
2	46.28	-197.45	-12526.1	165.46	-548.70	189.16	741.50	-29.92	-12159.7	0.48	-562.92
3	46.81	-197.39	-12527.3	156.84	-548.47	189.17	728.68	-29.92	-12181.6	0.14	-581.77
3	46.40	-198.03	-12525.6	156.89	-548.44	189.14	727.26	-29.92	-12182.3	0.45	-575.57
3	46.36	-197.31	-12527.2	156.71	-548.28	189.12	728.30	-29.92	-12182.2	0.41	-581.27
4	45.64	-196.00	-12526.7	200.32	-548.75	189.13	715.20	-29.92	-12151.1	0.59	-528.23
4	45.46	-196.70	-12524.8	200.22	-548.74	189.09	714.47	-29.92	-12150.9	0.41	-521.87
5	47.16	-196.78	-12523.7	2.66	-548.67	188.97	719.24	-29.92	-12341.0	0.27	-728.19
5	47.78	-196.79	-12524.3	2.63	-548.69	188.92	719.23	-29.92	-12341.1	0.32	-729.05
5	47.14	-196.39	-12523.5	2.60	-548.73	188.95	718.61	-29.92	-12341.3	0.41	-726.20
6	45.43	-197.16	-12528.0	196.66	-549.49	189.12	726.70	-29.92	-12146.7	0.57	-538.95
6	45.43	-197.81	-12526.4	196.71	-549.52	189.02	726.00	-29.92	-12146.5	0.43	-533.51

Table E.2: Complex energies, Boltzmann factors, and total CHARMM free energies for set n1 assuming the system as flexible. (The van der Waals energy contribution is scaled by $\alpha = 0.78$, all energies are in kcal/mol.)

G_{val}	G_{vdw}	G_{coul}	G_{QM}	$G_{solv,p}$	$G_{solv,np}$	G^v	G^{tr}	$G_{tot}(C)$	f_i	G_{MM}
45.64	-195.24	-12529.2	187.83	-548.88	189.09	723.58	-29.92	-12157.1	0.28	-548.03
45.47	-195.97	-12526.7	188.02	-548.98	189.10	721.81	-29.92	-12157.2	0.31	-537.58
45.26	-196.31	-12526.6	187.94	-549.67	189.14	722.96	-29.92	-12157.2	0.33	-537.92
45.62	-195.99	-12527.6	187.93	-549.08	189.08	723.59	-29.92	-12156.4	0.08	-542.31
45.43	-196.69	-12526.8	195.48	-548.95	189.09	709.08	-29.92	-12163.3	0.28	-528.60
45.24	-195.82	-12527.3	195.78	-549.23	189.04	708.51	-29.92	-12163.7	0.54	-532.94
45.04	-196.29	-12525.8	195.60	-549.34	189.11	708.60	-29.92	-12163.0	0.18	-527.81
47.81	-197.07	-12526.5	224.69	-548.12	189.06	743.73	-29.92	-12096.4	0.34	-509.99
45.60	-197.47	-12523.6	224.83	-548.95	189.41	743.62	-29.92	-12096.5	0.43	-498.73
48.32	-197.43	-12525.2	224.65	-548.16	189.34	742.94	-29.92	-12095.4	0.07	-507.37
47.39	-197.85	-12523.8	224.59	-548.13	189.36	742.45	-29.92	-12095.9	0.16	-504.28
45.46	-196.77	-12528.0	129.00	-548.88	189.15	733.38	-29.92	-12206.6	0.17	-603.95
45.42	-195.92	-12529.8	128.81	-548.77	189.15	734.27	-29.92	-12206.8	0.24	-609.85
45.80	-196.57	-12526.6	128.45	-550.34	189.13	732.70	-29.92	-12207.3	0.59	-599.65
45.82	-199.06	-12531.6	141.44	-549.45	188.89	747.34	-29.92	-12186.6	0.29	-607.32
46.29	-198.27	-12532.3	139.49	-549.11	189.04	747.81	-29.92	-12186.9	0.54	-614.37
46.32	-198.74	-12532.4	140.46	-548.63	189.09	747.80	-29.92	-12186.0	0.12	-614.09
47.14	-197.52	-12531.2	140.17	-549.33	189.02	748.19	-29.92	-12183.5	0.00	-611.68
45.81	-199.42	-12528.4	140.93	-549.48	188.97	745.97	-29.92	-12185.5	0.05	-596.30
46.16	-196.05	-12526.4	170.09	-548.66	189.23	742.21	-29.92	-12153.4	0.01	-566.45
45.56	-196.57	-12524.8	170.68	-549.70	189.26	740.91	-29.92	-12154.6	0.11	-559.77
45.74	-196.80	-12524.7	170.18	-549.28	189.17	739.79	-29.92	-12155.8	0.81	-558.02
45.13	-196.79	-12523.0	171.27	-549.29	188.98	739.39	-29.92	-12154.2	0.06	-547.95
46.17	-196.72	-12524.7	170.27	-548.94	189.07	741.42	-29.92	-12153.3	0.01	-560.66
45.16	-194.15	-12521.5	169.87	-548.50	189.00	737.91	-29.92	-12152.1	0.00	-543.17
45.63	-197.67	-12527.5	201.23	-547.87	189.12	750.04	-29.92	-12117.0	0.70	-541.01
47.39	-197.01	-12527.1	200.72	-549.07	189.04	751.01	-29.92	-12115.0	0.02	-535.31
46.39	-198.08	-12524.6	200.72	-548.98	189.08	751.23	-29.92	-12114.2	0.01	-530.94
46.25	-199.03	-12525.7	200.87	-547.63	189.11	749.65	-29.92	-12116.4	0.27	-534.33
45.64	-195.25	-12527.1	185.95	-548.80	189.14	723.22	-29.92	-12157.1	0.14	-545.11
45.66	-196.54	-12525.5	186.00	-548.50	189.12	722.20	-29.92	-12157.5	0.25	-537.90
45.48	-195.89	-12525.3	185.91	-548.89	189.06	722.70	-29.92	-12156.8	0.09	-538.91
44.86	-195.95	-12524.1	185.94	-548.94	189.12	721.06	-29.92	-12157.9	0.52	-535.94
49.29	-198.15	-12527.9	208.27	-547.13	188.99	749.49	-29.92	-12107.1	0.00	-534.97
46.12	-199.19	-12526.6	208.05	-548.50	189.12	749.30	-29.92	-12111.6	1.00	-532.74
46.77	-197.73	-12525.6	176.18	-548.50	189.18	741.08	-29.92	-12148.6	0.04	-558.10
46.21	-198.68	-12525.5	176.12	-548.74	189.20	741.20	-29.92	-12150.2	0.65	-558.94
45.63	-197.86	-12525.3	176.38	-548.53	189.02	740.85	-29.92	-12149.7	0.30	-560.07
45.98	-199.54	-12523.4	187.55	-548.87	188.86	770.71	-29.92	-12108.6	1.00	-532.64
45.68	-197.74	-12527.0	202.76	-548.38	189.05	749.44	-29.92	-12116.1	0.68	-536.13
47.44	-197.15	-12527.5	202.65	-548.52	189.11	750.70	-29.92	-12113.2	0.01	-538.75
46.75	-198.68	-12524.5	202.86	-548.25	189.16	751.04	-29.92	-12111.5	0.00	-531.93

continued on next page

continued from previous page

G_{val}	G_{vdw}	G_{coul}	G_{QM}	$G_{solv,p}$	$G_{solv,np}$	G^v	G^{tr}	$G_{tot}(C)$	f_i	G_{MM}
46.18	-198.90	-12525.0	202.84	-548.75	189.14	748.78	-29.92	-12115.6	0.31	-529.96
49.79	-201.99	-12522.7	174.98	-552.20	189.24	745.45	-29.92	-12147.3	0.00	-553.02
45.89	-198.20	-12529.0	175.93	-548.91	188.99	743.71	-29.92	-12151.5	0.53	-574.09
45.95	-199.10	-12527.4	176.18	-548.95	189.01	742.81	-29.92	-12151.4	0.47	-567.92
45.67	-196.06	-12527.7	202.56	-548.50	189.10	715.75	-29.92	-12149.1	0.31	-530.81
44.83	-195.99	-12524.3	202.58	-549.07	189.07	714.51	-29.92	-12148.3	0.08	-518.91
45.13	-196.33	-12524.7	202.58	-549.62	189.14	715.37	-29.92	-12148.4	0.09	-521.59
45.66	-196.91	-12526.2	202.61	-548.63	189.11	714.88	-29.92	-12149.4	0.52	-524.95
45.96	-198.46	-12528.5	185.83	-548.20	188.90	771.93	-29.92	-12112.5	0.66	-560.82
48.27	-196.36	-12529.1	185.83	-548.20	189.02	772.38	-29.92	-12108.1	0.00	-559.22
48.14	-196.92	-12528.8	185.76	-548.22	189.10	773.70	-29.92	-12107.2	0.00	-560.33
46.21	-199.53	-12525.7	185.63	-548.46	189.16	770.58	-29.92	-12112.1	0.34	-551.11
46.85	-198.99	-12526.7	208.59	-548.52	189.13	739.49	-29.92	-12120.1	0.05	-534.08
45.88	-199.69	-12525.1	208.41	-548.59	189.14	739.15	-29.92	-12120.7	0.14	-529.62
46.00	-199.57	-12525.4	208.23	-548.27	189.05	739.62	-29.92	-12120.3	0.07	-529.45
45.66	-200.40	-12525.3	208.20	-548.42	189.06	739.38	-29.92	-12121.7	0.74	-529.47
46.84	-198.94	-12527.2	181.50	-547.56	189.26	767.30	-29.92	-12118.8	0.24	-560.65
46.53	-199.64	-12525.1	181.71	-550.69	189.12	768.66	-29.92	-12119.4	0.67	-550.95
48.09	-199.36	-12527.6	181.46	-547.62	189.24	767.61	-29.92	-12118.1	0.08	-560.99
45.97	-197.73	-12525.8	205.24	-547.98	189.25	753.29	-29.92	-12107.7	0.47	-537.87
47.44	-200.28	-12523.1	205.89	-550.02	189.13	753.96	-29.92	-12106.9	0.11	-523.94
45.94	-198.54	-12523.9	205.41	-548.21	189.04	752.54	-29.92	-12107.6	0.41	-531.01
45.45	-196.42	-12528.1	181.00	-548.94	189.05	734.26	-29.92	-12153.6	0.36	-557.22
45.53	-197.38	-12525.4	180.40	-549.26	188.96	733.46	-29.92	-12153.6	0.35	-547.38
45.48	-197.23	-12526.2	181.14	-549.04	188.92	733.36	-29.92	-12153.5	0.29	-551.32
45.51	-197.15	-12527.6	194.27	-548.64	188.99	726.20	-29.92	-12148.3	0.22	-541.47
45.64	-197.89	-12525.9	194.35	-548.78	189.01	724.62	-29.92	-12148.9	0.53	-535.80
45.57	-196.39	-12526.9	194.33	-549.35	189.07	725.68	-29.92	-12147.9	0.10	-538.18
45.54	-196.94	-12525.2	194.29	-549.39	188.99	725.33	-29.92	-12147.3	0.04	-532.69
45.41	-197.62	-12524.8	193.77	-549.39	189.12	725.45	-29.92	-12148.0	0.12	-532.87
45.52	-196.27	-12526.7	177.84	-549.04	189.06	733.80	-29.92	-12155.8	0.24	-557.13
45.78	-197.32	-12525.0	177.68	-548.72	188.95	732.68	-29.92	-12155.8	0.27	-550.49
46.27	-197.15	-12526.7	177.28	-548.95	189.13	734.02	-29.92	-12156.0	0.37	-555.81
45.46	-197.01	-12524.9	177.88	-549.05	189.01	733.21	-29.92	-12155.3	0.11	-550.98
46.11	-197.33	-12528.3	183.45	-548.60	188.96	743.95	-29.92	-12141.7	0.09	-559.48
46.08	-197.80	-12528.0	183.01	-548.88	189.21	743.35	-29.92	-12142.9	0.74	-558.09
46.14	-198.78	-12527.8	183.75	-549.64	189.24	745.10	-29.92	-12141.9	0.13	-551.32
46.18	-198.03	-12526.5	183.59	-548.69	189.16	743.00	-29.92	-12141.2	0.04	-553.32
45.68	-196.61	-12529.7	197.76	-549.75	188.99	737.04	-29.92	-12136.6	0.48	-550.27
46.31	-198.43	-12527.1	197.53	-549.83	189.07	736.55	-29.92	-12135.8	0.13	-539.55
45.72	-197.47	-12527.9	197.93	-549.96	189.05	736.07	-29.92	-12136.5	0.40	-544.08
45.81	-198.60	-12527.0	200.14	-548.28	189.15	763.38	-29.92	-12105.3	0.49	-544.52
45.94	-198.39	-12527.7	200.20	-548.33	189.02	763.99	-29.92	-12105.2	0.42	-546.21

continued on next page

continued from previous page

G_{val}	G_{vdw}	G_{coul}	G_{QM}	$G_{solv,p}$	$G_{solv,np}$	G^v	G^{tr}	$G_{tot}(C)$	f_i	G_{MM}
47.88	-200.42	-12526.0	199.87	-549.24	189.06	764.40	-29.92	-12104.3	0.09	-531.85
47.72	-199.76	-12524.7	204.72	-550.33	189.13	764.49	-29.92	-12098.7	0.11	-524.84
47.20	-199.60	-12525.1	204.88	-550.32	189.09	763.84	-29.92	-12099.9	0.89	-526.77
46.65	-199.24	-12526.3	174.14	-548.98	189.19	751.53	-29.92	-12142.9	0.29	-565.29
46.18	-198.80	-12525.6	174.10	-549.05	189.11	751.29	-29.92	-12142.7	0.19	-564.33
45.47	-198.20	-12527.3	174.40	-548.34	189.04	751.98	-29.92	-12142.9	0.30	-568.96
46.31	-198.48	-12526.4	174.15	-548.72	189.09	751.28	-29.92	-12142.7	0.22	-566.34
46.81	-200.31	-12527.1	138.02	-549.56	189.06	740.24	-29.92	-12192.7	0.16	-608.20
46.05	-199.27	-12528.8	138.49	-549.85	189.11	740.61	-29.92	-12193.6	0.64	-615.33
46.81	-199.61	-12528.5	137.96	-549.38	189.14	740.83	-29.92	-12192.7	0.14	-613.94
47.84	-201.20	-12526.2	139.04	-552.03	189.17	741.11	-29.92	-12192.2	0.06	-599.01
45.01	-195.84	-12524.3	169.03	-549.00	189.26	740.75	-29.92	-12155.0	0.27	-556.24
45.14	-193.59	-12525.0	169.32	-548.29	189.14	738.46	-29.92	-12154.8	0.18	-559.60
45.27	-195.64	-12524.0	170.40	-549.96	189.04	740.87	-29.92	-12153.9	0.04	-546.14
45.62	-198.31	-12521.8	168.82	-550.40	189.12	742.16	-29.92	-12154.7	0.15	-551.08
44.75	-196.23	-12522.1	168.76	-549.18	189.13	739.61	-29.92	-12155.2	0.35	-551.61
45.73	-199.44	-12527.8	159.34	-547.52	188.78	782.59	-29.92	-12128.2	0.37	-584.49
46.88	-199.71	-12525.9	159.16	-549.79	189.12	782.62	-29.92	-12127.5	0.10	-569.60
45.74	-200.31	-12525.9	159.46	-547.49	189.08	781.17	-29.92	-12128.1	0.31	-577.96
46.41	-200.60	-12525.3	158.93	-547.51	189.13	780.98	-29.92	-12127.9	0.22	-574.55
46.88	-198.66	-12526.8	170.94	-549.03	189.00	750.96	-29.92	-12146.6	0.36	-567.47
46.12	-199.11	-12525.5	170.87	-548.42	188.96	751.79	-29.92	-12145.2	0.03	-570.41
48.69	-197.58	-12527.3	170.80	-547.23	189.26	751.81	-29.92	-12141.4	0.00	-568.68
46.49	-199.56	-12525.4	170.74	-549.22	189.27	750.65	-29.92	-12146.9	0.57	-564.13
46.55	-199.47	-12525.4	171.48	-549.23	189.24	751.45	-29.92	-12145.3	0.04	-566.06
46.34	-198.13	-12527.8	204.47	-547.71	188.85	753.61	-29.92	-12110.2	0.39	-540.10
47.74	-197.05	-12525.7	204.64	-548.27	189.31	752.77	-29.92	-12106.4	0.00	-533.60
47.09	-200.67	-12524.9	205.32	-550.04	189.22	754.73	-29.92	-12109.2	0.06	-531.49
46.40	-198.95	-12526.0	204.57	-547.87	189.20	752.45	-29.92	-12110.1	0.32	-533.93
46.73	-199.07	-12525.9	204.53	-548.30	189.10	752.92	-29.92	-12109.9	0.23	-534.41
46.17	-199.46	-12527.2	274.90	-549.75	190.16	793.02	-29.92	-12002.1	0.90	-450.45
46.08	-201.05	-12527.9	277.94	-548.02	189.09	794.26	-29.92	-11999.5	0.01	-455.72
46.99	-200.20	-12526.9	275.09	-550.56	190.36	794.44	-29.92	-12000.7	0.09	-449.89
46.18	-197.64	-12526.7	201.38	-549.30	189.35	763.31	-29.92	-12103.3	0.58	-542.15
46.19	-198.45	-12524.6	201.55	-549.42	189.32	762.62	-29.92	-12102.7	0.22	-535.26
46.14	-198.65	-12524.5	201.70	-549.50	189.20	762.83	-29.92	-12102.7	0.21	-535.10
46.05	-197.59	-12527.7	203.92	-548.34	189.02	753.79	-29.92	-12110.8	0.48	-541.51
47.44	-196.82	-12528.3	204.35	-548.35	188.96	755.31	-29.92	-12107.3	0.00	-537.33
47.06	-199.30	-12526.3	205.30	-549.40	189.14	754.30	-29.92	-12109.1	0.03	-529.10
46.13	-198.50	-12525.8	204.12	-548.43	189.06	752.59	-29.92	-12110.8	0.49	-534.84
46.07	-198.80	-12526.8	162.70	-549.28	188.96	781.30	-29.92	-12125.8	0.59	-581.31
49.08	-196.50	-12529.1	163.56	-547.91	189.16	782.22	-29.92	-12119.4	0.00	-575.43
46.09	-199.31	-12524.7	163.25	-549.96	189.23	781.72	-29.92	-12123.6	0.01	-577.95

continued on next page

continued from previous page

G_{val}	G_{vdw}	G_{coul}	G_{QM}	$G_{solv,p}$	$G_{solv,np}$	G^v	G^{tr}	$G_{tot}(C)$	f_i	G_{MM}
46.09	-199.55	-12524.7	162.80	-549.48	189.13	780.02	-29.92	-12125.6	0.40	-574.89
48.19	-203.10	-12528.3	124.60	-547.84	189.12	864.84	-29.94	-12082.4	1.00	-608.09
48.44	-200.87	-12528.7	127.10	-548.60	189.21	865.59	-29.94	-12077.7	0.00	-609.10
49.44	-199.41	-12527.1	124.13	-548.11	188.98	865.99	-29.94	-12076.0	0.00	-611.46
49.55	-201.12	-12528.6	127.35	-548.40	189.07	864.56	-29.94	-12077.5	0.00	-609.42
49.84	-202.33	-12528.5	127.32	-548.52	189.14	864.68	-29.94	-12078.3	0.00	-609.47
45.62	-196.74	-12528.2	161.30	-548.46	188.97	735.00	-29.92	-12172.5	0.51	-578.33
46.67	-197.49	-12525.3	161.37	-549.01	189.20	734.47	-29.92	-12170.1	0.01	-568.47
45.67	-197.57	-12526.6	161.44	-548.57	189.17	733.94	-29.92	-12172.4	0.48	-572.07
51.13	-200.55	-12526.7	123.56	-546.70	188.84	863.99	-29.94	-12076.3	0.00	-606.79
51.01	-200.30	-12527.3	124.68	-546.30	189.09	866.18	-29.94	-12072.9	0.00	-609.23
49.66	-201.70	-12524.9	125.48	-548.01	189.53	867.16	-29.94	-12072.7	0.00	-598.73
48.62	-201.91	-12528.9	125.24	-547.12	190.09	864.17	-29.94	-12079.7	0.30	-616.70
47.72	-202.77	-12526.2	125.32	-547.23	189.53	864.07	-29.94	-12079.5	0.21	-606.34
47.48	-201.21	-12526.6	124.49	-547.05	189.47	864.79	-29.94	-12078.6	0.04	-609.11
47.25	-202.41	-12525.5	124.29	-547.82	189.87	864.29	-29.94	-12080.0	0.45	-608.02
48.74	-203.56	-12528.5	200.38	-546.76	189.16	815.35	-29.94	-12055.1	0.83	-553.26
49.94	-199.51	-12523.8	200.79	-546.94	189.28	814.42	-29.94	-12045.7	0.00	-526.69
48.98	-204.51	-12526.5	200.25	-547.24	189.23	815.54	-29.94	-12054.1	0.17	-546.30
47.14	-198.15	-12526.4	166.81	-549.17	189.05	751.74	-29.92	-12148.9	0.20	-567.47
46.29	-197.99	-12526.0	166.42	-548.71	188.98	752.02	-29.92	-12148.9	0.18	-569.75
45.69	-197.78	-12526.6	166.59	-548.91	189.12	752.20	-29.92	-12149.6	0.62	-569.98
47.86	-196.62	-12525.0	189.35	-548.71	189.08	772.42	-29.92	-12101.5	0.00	-538.05
48.42	-195.37	-12525.1	188.96	-549.44	188.92	774.92	-29.92	-12098.6	0.00	-538.06
46.11	-198.80	-12527.0	187.73	-547.35	189.28	772.04	-29.92	-12107.9	0.37	-560.56
46.09	-199.61	-12525.5	187.77	-547.53	189.32	771.14	-29.92	-12108.3	0.63	-554.05
50.09	-199.71	-12526.6	117.04	-550.87	188.88	768.50	-29.92	-12182.6	0.00	-625.66
46.40	-201.62	-12529.9	117.61	-549.93	188.86	769.37	-29.92	-12189.1	0.55	-639.44
46.27	-202.37	-12527.6	117.59	-549.98	188.83	768.14	-29.92	-12189.0	0.45	-632.50

E.2 Binding free energies

Table E.3: Results of the energy calculations for set n0 assuming the system as flexible. (The van der Waals energy contribution is scaled by $\alpha = 0.78$, all energies are in kcal/mol.)

No.	ΔG_{val}	ΔG_{coul}	ΔG_{vdw}	ΔG_{QM}	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$	ΔG_v	ΔG_{tr}	ΔG_{calc}	ΔG_{exp}
1	0.24	9.23	-21.07	1.19	-9.19	-5.96	1.45	16.50	-7.72	-6.32
2	0.56	11.20	-22.20	1.66	-9.24	-6.56	1.61	16.84	-5.80	-7.15
3	1.13	11.58	-23.09	1.62	-8.61	-6.54	0.77	17.95	-5.37	-7.64
4	0.26	12.15	-21.72	0.59	-8.89	-6.59	0.22	16.49	-6.64	-4.73
5	2.05	14.29	-22.06	1.56	-8.23	-6.08	-1.06	17.86	-2.57	-4.77
6	0.12	10.75	-22.87	1.11	-8.40	-6.05	1.25	16.83	-7.26	-5.01

Table E.4: Results of the energy calculations for set n1 assuming the system as flexible. (The van der Waals energy contribution is scaled by $\alpha = 0.78$, all energies are in kcal/mol.)

ΔG_{val}	ΔG_{coul}	ΔG_{vdw}	ΔG_{QM}	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$	ΔG_v	ΔG_{tr}	ΔG_{calc}	ΔG_{exp}
0.16	10.63	-21.31	1.22	-8.51	-5.69	0.70	16.91	-6.01	-5.11
-0.04	11.20	-21.58	0.84	-10.15	-5.56	0.39	16.89	-8.12	-6.91
1.53	13.30	-22.83	1.14	-9.96	-6.10	0.89	17.31	-4.84	-3.94
0.34	10.46	-21.88	1.37	-8.37	-5.95	0.93	17.25	-5.97	-6.44
0.83	6.18	-24.04	1.92	-6.12	-6.64	2.76	17.79	-7.45	-5.7
0.39	13.47	-22.19	1.41	-8.96	-6.00	0.27	17.28	-4.45	-6.47
0.54	11.06	-23.45	1.41	-8.62	-6.11	1.62	17.68	-6.40	-7.08
-0.08	13.12	-21.44	0.92	-9.45	-6.51	-0.50	16.90	-6.30	-5.85
0.81	11.46	-24.63	1.46	-8.74	-5.67	1.47	17.68	-7.20	-7.25
0.75	12.61	-23.83	0.75	-9.69	-6.59	1.34	17.68	-7.06	-6.54
0.68	14.70	-24.98	3.03	-9.97	-6.61	1.31	18.11	-4.65	-5.07
0.54	11.74	-23.54	1.17	-9.05	-7.38	1.08	17.68	-7.09	-7.12
0.61	9.85	-24.06	2.62	-7.32	-5.62	1.62	18.04	-5.53	-6.81
0.24	11.68	-21.96	0.89	-8.72	-6.78	0.56	16.91	-6.13	-4.82
0.74	10.52	-24.26	1.89	-9.12	-5.62	2.35	18.14	-7.20	-7.55
0.47	12.76	-25.61	0.75	-10.19	-7.33	1.69	17.86	-9.18	-7.15
1.43	12.22	-24.88	1.43	-10.85	-6.77	2.42	17.93	-7.27	-5.88
0.82	13.36	-23.79	1.07	-9.03	-6.85	1.30	17.89	-5.37	-6.44
0.18	11.45	-22.43	1.67	-8.79	-6.87	1.03	17.24	-5.79	-5.77

continued on next page

continued from previous page

ΔG_{val}	ΔG_{coul}	ΔG_{vdw}	ΔG_{QM}	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$	ΔG_v	ΔG_{tr}	ΔG_{calc}	ΔG_{exp}
0.27	11.87	-22.95	1.51	-8.88	-7.24	0.07	17.24	-6.96	-5.69
0.56	12.05	-22.41	1.66	-9.30	-5.96	0.86	17.24	-5.51	-5.97
0.79	10.16	-23.33	1.56	-8.37	-6.85	1.10	17.95	-6.85	-5.84
0.47	9.42	-22.62	2.17	-6.73	-6.59	1.77	17.54	-4.39	-5.14
0.75	10.87	-24.12	1.89	-8.93	-5.92	2.01	18.13	-6.70	-6.76
1.95	13.02	-25.05	1.92	-10.14	-6.28	2.34	18.14	-5.09	-6.62
0.83	11.59	-24.12	1.18	-9.05	-7.19	1.97	17.95	-6.60	-6.44
1.09	9.75	-25.04	1.87	-6.64	-7.14	1.99	18.16	-5.69	-6.47
-0.26	14.82	-21.38	1.13	-9.50	-6.83	-0.25	17.22	-4.29	-5.01
0.70	11.62	-25.43	1.94	-9.34	-6.94	1.74	17.97	-8.13	-6.5
1.32	12.18	-24.65	1.37	-9.52	-6.74	1.21	17.96	-7.02	-6.48
1.19	11.48	-24.21	1.01	-8.42	-5.97	1.44	17.90	-6.66	-5.79
0.94	10.90	-24.98	2.42	-7.93	-7.19	2.67	19.04	-4.76	-6.04
0.87	12.32	-23.46	1.43	-9.26	-6.74	1.82	18.16	-5.19	-6.66
0.82	11.32	-23.52	1.14	-8.94	-6.93	1.45	17.90	-6.89	-6.21
0.77	12.14	-24.54	0.10	-7.21	-7.69	1.19	17.71	-6.21	-6.35
2.89	9.76	-28.53	2.05	-5.91	-6.94	5.08	19.54	-5.08	-5.14
0.34	10.67	-22.58	1.91	-8.90	-6.93	0.80	17.85	-6.44	-7.18
2.47	11.35	-27.72	4.34	-6.40	-7.14	4.65	19.50	-1.02	-5.67
3.48	9.97	-29.16	2.24	-7.19	-9.82	3.03	18.82	-7.44	-5.67
0.78	11.62	-23.33	2.04	-8.61	-6.41	1.70	17.56	-4.74	-6.54
0.79	11.99	-24.75	1.50	-8.49	-9.06	1.75	18.49	-6.33	-6.58
1.04	9.22	-27.40	1.48	-5.95	-8.48	2.06	18.55	-8.83	-5.95

Appendix F

LUDI scores

The program LUDI has been developed by Böhm [53] and uses a simple empirical scoring function that accounts for hydrogen bonds and hydrophobic interactions to estimate binding affinities of protein-ligand complexes. The LUDI scores are a measure for the binding affinity and have been calculated for the ligands of set n0 and n1. In order to compare the accuracy of LUDI and of the free energy function presented in this work, the LUDI scores have been converted to binding free energies using the following expression [54]:

$$\text{LUDI score} = k \cdot \Delta G \text{ with } k = -73.33 \text{ mol/kcal.}$$

In Figure F.1 the binding free energies obtained from LUDI and those derived by the free energy function considering protein flexibility are plotted versus the experimental values. LUDI predicts ΔG_{bind} with an RMSD of 1.5 kcal/mol, and hence is less accurate than our continuum method. This is also underlined by a correlation coefficient of $\sigma = 0.21$ for LUDI in comparison to $\sigma = 0.34$ for our method. It should be noted that LUDI is a ligand design software tool that has already been developed and improved for more than 10 years. By way of contrast, the method used in this work is still under development, and thus has the potential to achieve an even higher accuracy.

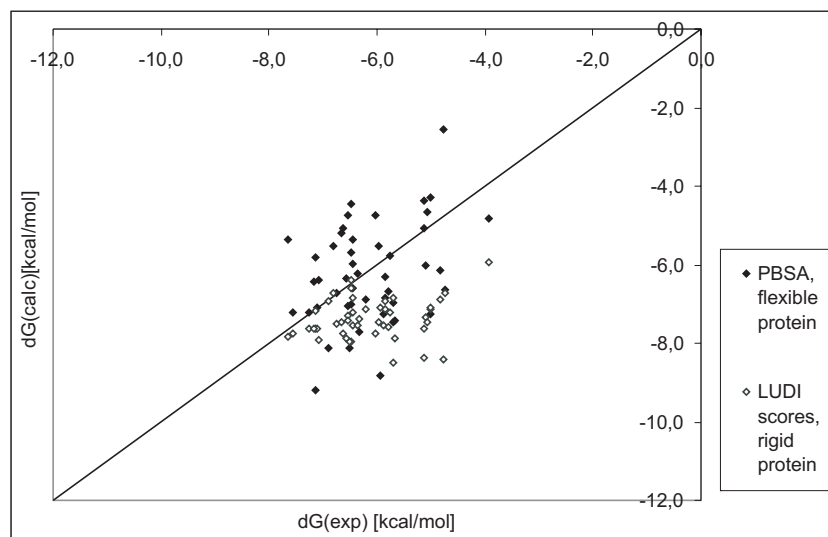


Figure F.1: Calculated versus experimental binding free energies for the ligands of set n0 and n1 using LUDI and the binding free energy function including protein flexibility.

Bibliography

- [1] Roberts R. M. Serendipity. Accidental Discoveries in Science. John Wiley & Sons, New York, 1998.
- [2] Reddy M. R. and Erion M. D. Das Antifebrin, ein neues Fiebermittel. Centralblatt für Klinische Medizin, 7:561 – 564, 1886.
- [3] Roberts N.A., Martin J.A., Kinchington D., Broadhurst A.V., Craig J.C., Duncan I.B., Galpin S.A., Handa B.K., Kay J., and Krohn A. et al. Rational design of peptide-based HIV proteinase inhibitors. Science, 248:358 – 361, 1990.
- [4] Jadhav P.K. Lam P.Y., Eyermann C.J.and Hodge C. N., Ru Y., Bachelier L.T., Meek J.L.and Otto M.J., Rayner M.M., and Wong Y.N. et al. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science, 263:380 – 384, 1994.
- [5] Murcko A. and M. A. Murcko. Computational methods to predict binding free energy in ligand-receptor complexes. J. Med. Chem., 38:4953 – 4967, 1995.
- [6] Wang W., Donini O., Reyes C.M., and Kollman P.A. Biomolecular simulations: recent developments in force fields, simulations of enzyme catal-

- ysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annu. Rev. Biophys. Biomol. Struct., 30:211 – 243, 2001.
- [7] Åqvist J. and Marelus J. The linear interaction energy method for predicting ligand binding free energies. Comb Chem High Throughput Screen, 4(8):613 – 26, 2001.
- [8] Gohlke H. and Klebe G. Statistical potentials and scoring functions applied to protein-ligand binding. Curr. Opin. Struct. Biol., 11(2):231 – 235, 2001.
- [9] Reddy M. R. and Erion M. D. Calculation of relative binding free energy differences for fructose 1,6-bisphosphatase inhibitors using the thermodynamic cycle perturbation approach. J. Am. Chem. Soc., 123:6246 – 6252, 2001.
- [10] Essex J.W., Severance D.L., Tirado-Rives J., and Jorgenson W.L. Monte Carlo simulations for proteins: binding affinities for trypsin-benzamidine complexes via free-energy perturbations. J. Phys. Chem. B, 123:9663 – 9669, 1997.
- [11] Oostenbrink B.C., Pitera J.W., van Lipzig M.M., Meerman J.H., and van Gunsteren W.F. Simulations of the estrogen receptor ligand-binding domain: affinity of natural ligands and xenoestrogens. J. Med. Chem., 43(24):4594 – 4605, 2000.
- [12] Ota N., Stroupe C., Ferreira da Silva J.M.S., Shah S.A., Mares-Guia M., and Brunger A.T. Non-Boltzmann thermodynamic integration (NBTI)

- for macromolecular systems: Relative free energy of binding of trypsin to benzamidine and benzylamine. Proteins, 37:641 – 653, 1999.
- [13] Erion M. D. and Reddy M. R. Calculation of relative free energy differences for the covalent hydration of organic compounds - a combined quantum mechanical and free energy perturbation study. J. Comput. Chem., 16:1513 – 1521, 1995.
- [14] Åqvist J., Medina C., and Samuelsson J. E. A new method for predicting binding affinity in computer-aided drug design. Protein Eng., 7:385 – 391, 1994.
- [15] Wall I.D., Leach A.R., Salt D.W., Ford M.G., and Essex J.W. Binding constants of neuraminidase inhibitors: An investigation of the linear interaction energy method . J. Med. Chem., 42:5142 – 52, 1999.
- [16] Böhm H. J. 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. J. Comput. Aided Mol. Des., 12:309 – 323, 1998.
- [17] Pearlman D.A. and Charifson P.S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. J. Med. Chem., 44:3417 – 3423, 2001.
- [18] Kasper P., Christen P., and Gehring H. Empirical calculation of the relative free energies of peptide binding to the molecular chaperone DnaK. Proteins, 40(2):185 – 192, 2000.

- [19] Ma C., Baker N.A., Joseph S., and McCammon J.A. Binding of aminoglycoside antibiotics to the small ribosomal subunit: a continuum electrostatics investigation. J. Am. Chem. Soc., 124:1438 – 1442, 2002.
- [20] Wang J., Morin P., Wang W., and Kollman P.A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc., 123:5221 – 5230, 2001.
- [21] Huo S., Wang J., Cieplak P., Kollman P.A., and Kuntz I.D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. J. Med. Chem., 45:1412 – 1419, 2002.
- [22] Wang W. and Kollman P.A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. J. Mol. Biol., 303:567 – 582, 2000.
- [23] Huo S., Massova I., and Kollman P.A. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J. Comput. Chem., 23:15 – 27, 2002.
- [24] Schwarzl S.M., Tschopp T.B., Smith J.C., and Fischer S. Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? J. Comput. Chem., 2002. in press.
- [25] Schwarzl S.M. Calculation of binding free energies, March 2000. Diplomarbeit, Universität Kaiserslautern.

- [26] Stryer L. Biochemie. Spektrum Akademischer Verlag, Heidelberg, 1996.
- [27] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. J. Molec. Graphics, 14:33 – 38, 1996.
- [28] Hallam Oaks Pty. Ltd. POV-Ray Version 3.1g, 1995 – 2000. <http://www.povray.org/>.
- [29] Hans-Joachim Böhm, Gerhard Klebe, and Hugo Kubinyi. Wirkstoffdesign. Spektrum Akademischer Verlag, Heidelberg, 1996.
- [30] B. A. Katz, J. Finer-Moore, R. Mortezaei, D.H. Rich, and R.M. Stroud. Episelection: Novel K_i nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface. Biochemistry, 34:8264, 1995.
- [31] F. Hoffmann-LaRoche AG. Inhibition of trypsin. unpublished.
- [32] Y.-C. Cheng and W. H. Prusoff. Relationship between the inhibition constant (K_i) and the concentration of the inhibitor that causes 50 percent inhibition (IC_{50}) of an enzyme reaction. Biochem. Pharmacol., 22:3099 – 3108, 1973.
- [33] Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S., and Karplus M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. J. Comput. Chem., 4:187 – 217, 1983.
- [34] Andrew R. Leach. Molecular Modelling Principles And Applications, chapter 4.4 and 4.5. Addison Wesley Longman Ltd., Harlow, 1996.

- [35] Dewar M.J.S., Zoebisch E.G., and Healy E.F. Stewart J.P. AM1: A new general purpose quantum mechanical molecular model. J. Am. Chem. Soc., 107:3902 – 3909, 1985.
- [36] Stewart J.J.P. Calculation of the geometry of a small protein using semiempirical methods. J. of Mol. Struct. (Theochem), 401:195 – 205, 1997.
- [37] Claussen H., Buning C., Rarey M., and Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J. Mol. Biol., 308:377 – 395, 2001.
- [38] Osterberg F., Morris G.M., Sanner M.F., Olson A.J., and Goodsell D.S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins, 46:34 – 40, 2002.
- [39] Paul R. Gerber and K. Müller. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. J. Comput. Aided Mol. Des., 6:251 – 268, 1995.
- [40] Paul R. Gerber. MOLOC. Roche inhouse modelling program.
- [41] M. E. Davis, J.D. Madura, B.A. Luty, and J.A. McCammon. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. Comp. Phys. Comm., 62:187–197, 1991.
- [42] J. D. Madura, J. M. Briggs, R.C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCam-

- mon. Electrostatics and diffusion in solution: Simulation with the University of Houston Brownian Dynamics program. Comp. Phys. Comm., 91:57 – 95, 1995.
- [43] M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople. Gaussian 94, Revision D.4. Gaussian, Inc., Pittsburgh PA, 1995.
- [44] Kim A. Sharp, Antony Nicholls, Richard F. Fine, and Barry Honig. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effect. Science, 252:106 – 109, 1991.
- [45] I. Tuñón and J.L. Pascual-Ahuir. Molecular surface area and hydrophobic effect. Protein Engineering, 5:715 – 716, 1992.
- [46] Stefan Fischer. CODISP, a program for computing continuum dispersion interactions. unpublished.
- [47] M.L. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. Science, 221:709 – 713, 1983.
- [48] Fischer S., Smith J. C., and Verma C. Dissecting the vibrational entropy change on protein/ligand binding: burial of a water molecule in bovine pancreatic trypsin inhibitor. J. Phys. Chem. B, 105:8050 – 8055, 2001.

- [49] Bruce Tidor and Martin Karplus. The contribution of vibrational entropy to molecular association: The dimerization of insulin. J. Mol. Biol., 238:405–414, 1994.
- [50] Wedler G. Lehrbuch der physikalischen Chemie. Wiley VCH Verlag, Weinheim, 1997. Chapter 4.2.
- [51] Gao J. and Xia X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. Science, 1992.
- [52] vibran.doc, c28b1 CHARMM documentation, 1995.
- [53] Böhm H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J. of Comput. Aided Mol. Des., 1994.
- [54] Ligand_Design, Version 95.0, Molecular Simulations Inc., San Diego, CA.

Danksagung

Mein Dank gilt Prof. J. Warnatz und Prof. J. Smith, die mir die Durchführung dieser Arbeit am IWR Heidelberg ermöglicht haben. Dr. Stefan Fischer sei für die Überlassung des interessanten und lehrreichen Themas und die durchweg erstklassige Betreuung gedankt. Ich danke Sonja M. Schwarzl für ihre Anleitung, ihre konstruktiven Vorschläge und Ideen für meine Arbeit. Sonja, ich habe viel von Dir gelernt! Bogdan möchte ich herzlich für die Änderung des CHARMM codes für die Normalmoden-Analyse danken. Bei allen Mitgliedern der Biocomputing Arbeitsgruppe und im besonderen beim unschlagbaren girls' room team möchte ich mich für die anregenden Diskussionen und stets gute Arbeitsatmosphäre bedanken.