Prediction of protonation states in ligand-protein complexes upon ligand binding

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie

der Philipps-Universität Marburg

vorgelegt von

Paul Czodrowski

aus Olsztyn

Marburg/Lahn 2006

Vom Fachbereich Pharmazie der Philipps-Universität Marburg
als Dissertation angenommen am: 23. November 2006

Erstgutachter: Prof. Dr. G. Klebe
Zweitgutachter: Dr. J. E. Nielsen

Tag der mündlichen Prüfung: 24. November 2006

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Februar 2003 bis September 2006 durchgeführt.

Pure Vernunft darf niemals siegen.
Dirk von Lowtzow

Contents

1 MOTIVATION **1**

1.1 Protein-Ligand Interactions 1

1.2 Scope of this thesis 2

2 DEVELOPMENT, VALIDATION AND APPLICATION OF ADAPTED PEOE CHARGES TO ESTIMATE PKA VALUES OF FUNCTIONAL GROUPS IN PROTEIN-LIGAND COMPLEXES **4**

2.1 INTRODUCTION 4

2.2 THEORY AND ALGORITHM 6

2.2.1 The PEOE procedure 6

2.2.2 Performance of the original PEOE charges in PB calculations 7

2.2.3 Adaptation of the PEOE procedure 8

2.2.4 Testing against 80 small organic molecules 10

2.2.5 Ionized molecules 11

2.2.6 Comparison to PARSE and CHARMM charges 12

2.2.7 Protein pKa calculations 14

2.3 MATERIALS AND METHODS 15

2.3.1 Solvation free energies 15

2.3.2 pKa calculations 15

2.3.3 X-ray structures 16

2.4 RESULTS AND DISCUSSION 16

2.4.1 pKa calculations of protein residues 16

2.4.2 pKa calculations on two complexes of thrombin 24

2.4.3 pKa calculations on a dihydrofolate reductase complex 27

2.5 Conclusions 28

3 PROTONATION CHANGES UPON LIGAND BINDING TO TRYPSIN AND THROMBIN: STRUCTURAL INTERPRETATION BASED ON PKA CALCULATIONS AND ITC EXPERIMENTS **30**

3.1 INTRODUCTION 30

3.2 RESULTS 33

3.2.1 Estimating the accuracy of calculated protonation changes 33

3.2.2 pKa Calculations on the complexes 35

3.3 DISCUSSION 38

3.3.1 Influence of conformational variations 40

3.3.2 pKa values of histidine residues . . . 40
3.4 Conclusions . . . 41
3.5 Materials and Methods . . . 42

4 ATYPICAL PROTONATION STATES IN THE ACTIVE SITE OF HIV-1 PROTEASE: A COMPUTATIONAL STUDY 46
4.1 INTRODUCTION . . . 46
4.1.1 Setting of the dielectric constant and handling of coupled systems . . . 48
4.2 MATERIALS AND METHODS . . . 49
4.3 RESULTS . . . 52
4.3.1 Calculations on two apo structures . . . 52
4.3.2 Calculations for the complexes . . . 54
4.4 DISCUSSION . . . 65
4.5 CONCLUSIONS . . . 66

5 PROTONATION EFFECTS IN HUMAN ALDOSE REDUCTASE 68
5.1 INTRODUCTION . . . 68
5.2 RESULTS . . . 69
5.2.1 Calculations for five holo structures . . . 69
5.2.2 Calculations for compounds with carboxylic head groups . . . 74
5.2.3 Calculations for compounds containing spiro-hydatoins . . . 74
5.3 DISCUSSION . . . 76
5.3.1 Holo structures . . . 76
5.3.2 Carboxylic head groups . . . 79
5.3.3 Spiro-hydatoins . . . 79
5.4 CONCLUSIONS . . . 80
5.5 MATERIALS AND METHODS . . . 82

6 PDB2PQR AS A NEW TOOL FOR THE SETUP OF PKA CALCULATIONS ON PROTEIN-LIGAND COMPLEXES 84
6.1 INTRODUCTION . . . 84
6.2 PDB2PQR PRINCIPLES . . . 84
6.3 SCRIPTS . . . 87
6.3.1 Substructure matching . . . 88
6.3.2 pka_lig_tool . . . 89

7 SUMMARY,ZUSAMMENFASSUNG 91
7.1 Summary . . . 91
7.2 Zusammenfassung . . . 93

Bibliography **97**

1 MOTIVATION

1.1 Protein-Ligand Interactions

Protein-ligand complexation involves the formation of intra-molecular interactions such as hydrogen-bonds, aromatic stacking, salt bridges and hydrophobic complementarity. A large variety of different computational approaches exist for modelling these processes, ranging from *ab initio*, semi-empirical, classical mechanics to empirical methods. These methods are capable of computing individual aspects of the binding process; e.g., calculating the free energy of binding, simulating the reaction path for a substrate, examining the conformational flexibility, investigating the role of water, to mention only a few. One effect that these methods usually neglect concerns the protonation change upon ligand binding. Possible explanations for this deficiency are the high computational burden and the lack of experimental data.

Protonation changes are omnipresent in the field of enzyme catalysis, because many catalyzed processes involve an acid-base reaction. Hence, a thorough understanding of such processes is of utmost importance being confronted with aspects of computational drug design such as ligand optimization, docking or molecular dynamics simulations. Protein X-ray crystallography represents an experimental method heavily applied in all stages of drug discovery, and we will use many such structures as reference points for our pKa calculations, but most often their resolution is not sufficient to draw conclusions about protonation states.

One impressive example for the thermodynamic contributions arising from protonation effects is pepstatin binding to plasmepsin II: it was experimentally shown that the proton transfer involved contributes almost 40 % of the total binding free energy change [1]. The following examples illustrate that atypical protonation states for active site residues are not as rare as one might anticipate: lysozyme is a very well studied enzyme with a glutamic acid having a pKa value of 6.2 [2]. In thioredoxin, an aspartic acid shows a highly shifted pKa value of 7.4 [3]. The pKa values of two histidine residues in a protein tyrosine phosphatase are 8.3 and 9.2, respectively [4]. For dehydratase and epimerase, the pKa values of central tyrosine residues are 6.4 [5] and 6.1, respectively [6]. In acetoacetate decarboxylase, a central lysine residue shows a pKa value of 6.0 [7]. In all these cases, the protonation pattern is different in the binding pocket, and it is crucial to obtain such knowledge when setting up a virtual screening or docking campaign, because the physicochemical properties of one key residue might have changed remarkably.

The experimental effort required to detect perturbed pKa values is rather high. Furthermore, applicable methods such as NMR impose limits with respect to protein size (NMR) or the reaction being catalyzed (kinetic measurements). One method which does not suffer from these deficiencies is isothermal titration calorimetry (ITC). It represents an experimental technique enabling a full thermodynamic characterization of the protein-ligand binding process. Furthermore, it allows to detect the net protonation change upon ligand binding, but no microscopic picture is obtained showing where the protonation changes actually occur. Such a structural interpretation is of utmost importance, in order to perform a factorisation of thE ITC data into enthalpic and entropic contributions with respect to certain residues in the binding pocket. For this purpose, theoretical approaches based on available structural data become attractive.

A calulation method with moderate computational demands is given by the Poisson-Boltzmann (PB) theory, which is a *continuum electrostatics* procedure. In general, inter- and intra-molecular *electrostatic* interactions are among the key factors determining the functional properties and structural stability of proteins [8, 9, 10, 11, 12, 13, 14, 15] and these interactions are directly related to charges of titratable groups, which in turn can have modulated protonation states. There is strong evidence for this, e.g., the pH-dependence of enzyme activity [16]. The ubiquitously present water molecules in proteins and protein-ligand complexes are simulated as a dielectric sphere, and this is the *continuum* part in the PB model.

1.2 Scope of this thesis

Several studies have proved the predictive power of PB calculations for an accurate pKa estimation in proteins [17, 18, 19, 20, 21]. However, to the best of our knowledge, no large-scale study of protonation states in protein-ligand complexes has been performed. One reason might be the lack of a generally applicable charge assignment procedure that works reliably with both, proteins and small organic molecules based on a consistent approach. In the first chapter of this thesis, we describe the development of such a charge model and its validation.

In the following chapters, three application studies for protein-ligand complexes with atypical protonation states are described. For the serine proteases trypsin and thrombin and human aldose reductase, in house and public ITC data are available, which reveal protonation changes upon ligand binding [22, 23]. For the human immunodeficiency virus 1 protease, NMR and kinetic measurements [24, 25, 26] detected perturbed pKa values of the catalytic

dyad for the complexed and uncomplexed state of this enzyme.

In the last chapter, we outline our efforts towards an open source PDB2PQR program [27] for a facilitated setup of pKa calculations on protein-ligand complexes. It includes a substructure search for the ligand based on a small molecule database with experimentally determined pKa values.

2 DEVELOPMENT, VALIDATION AND APPLICATION OF ADAPTED PEOE CHARGES TO ESTIMATE PKA VALUES OF FUNCTIONAL GROUPS IN PROTEIN-LIGAND COMPLEXES

2.1 INTRODUCTION

The first (purely theoretical, because no protein x-ray structures were available at that time) approach towards the handling of electrostatics in macromolecules was introduced by Linderstrom-Lang[28]. In their model, the charges (within a protein) are assumed to be spread uniformly over the surface of a spherical model. Three decades later, Tanford and Kirkwood presented an improved model in which they consider Coulombic interactions between the charges[29]; this model was further enhanced (obtaining a good fit to experimental data) by placing all charges onto the surface of the protein [30]. A consistent microscopic treatment of the dielectric effect in proteins based on the protein dipole Langevin dipole (PDLD) method was introduced by Warshel et al. [31][10][11][32]. Warshel also introduced the microscopic free energy pertubation (FEP) [33] and the microscopic linear response approximation (LRA) [32]. A (semi-macroscopic) combination of PDLD and LRA, called PDLD/S-LRA, has been introduced [32][34]. Besides these microscopic and semi-macroscopic electrostatic models, a wide range of different computational techniques has become available for pKa calculations in proteins, including, for example, empirical models [35][36], simplified electrostatic models using sigmoidally-screened Coulomb potentials [37], continuum electrostatic models such as Poisson-Boltzmann calculations [17]. Additionally, more sophisticated methods such as pH-dependent MD simulations [38][39] or QM/MM approaches [40] have been applied to protein pKa calculations. However, these approaches are computationally rather demanding and hamper routine application to large protein data sets.

One popular method for pKa calculations in proteins is based on the Poisson-Boltzmann (PB)[41] theory, which uses an implicit solvent model. Several investigations show the suitability of the PB theory to estimate pKa values in proteins [17][42][43][18][19][20][21]. In a PB calculation, the electrostatic potential of a solute molecule is calculated based on the charge distribution of its atoms. Atomic partial charges are normally used to describe such

charge distributions and they are, accordingly, key parameters in pKa calculations. Usually, partial charges applied in force fields such as CHARMM [44], Amber [45], or OPLS [46] are used for this purpose. Bashford et al. [47] analyzed the influence of different charge models on pKa calculations and observed performance differences. Warshel studied the influence that different charge models take in the PDLD method and they found that it is not crucial [48]. Warshel and Chu also established a new charge model for pKa calculations within the framework of the PDLD methodolgy and detected that the actual choice of the charges is far less crucial than the choice/adjustment of the radii [49]. Sitkoff et al. developed an optimized charge and radius set (PARSE) which is also often used for PB-based protein pKa calculations [50]. However, only few studies have been developed which perform PB-based pKa calculations on protein-ligand complexes [51][52]; the Åqvist group performed absolute binding free energies calculations considering the effect of charged ligands on proteins [53][54]. However, a systematic and large-scale study considering pKa shifts in protein-ligand complexes is still missing. In part, this may be due to the lack of sufficient experimental data to validate calculations on such systems; however, another possible explanation might be that at present no routine is available to assign generic charges reliably to both, proteins and small organic molecules, in one consistent approach.

Many different computational techniques can be used for assigning partial charges. They range from empirical to semi-empirical [55], including RESP methods (Restrained Electrostatic Potential) [56], and differ significantly with respect to the required computational resources. In the present contribution, a rather simple but generically applicable method was chosen which is well-established in the drug-design field for large scale charge assignments in database studies of organic molecules: the PEOE procedure, originally developed by Gasteiger and Marsili [57]. Another straightforward method for assigning charges based on a linear combination of atomic orbitals has been introduced by Del Re [58]. The PEOE procedure attributes initial charge values based on orbital electronegativities and redistributes these initial assignments iteratively until convergence is achieved. Interestingly, PEOE charges have not yet been used for PB-based pKa calculations. Therefore, we have tested the PEOE charges for the simultaneous charge assignment of protein and ligand functional groups. We re-parameterized them based on solvation free energies and developed an adapted PEOE procedure to calculate modified charges. Their broad applicability is based on the fact that only hybridization states of the atoms (as defined by the atom types) are required. Along with the development of a procedure to compute modified PEOE charges, pKa calculations on protein-ligand complexes can be performed in automated fashion for large-scale application studies.

The present contribution is structured as follows: First, the PEOE procedure is intro-

duced and its performance on proteinogenic functional groups is assessed. Subsequently, the re-parameterization for the calculation of solvation free energies is presented, first based on amino acid side chains and then expanded to a dataset of small organic molecules. A modification of the method to handle charged residues is discussed in the following. Using a data set of nine proteins with 132 residues of experimentally determined pKa values, the novel PEOE procedure is then tested retrospectively for its predictive power to reproduce pKa values of amino acid residues in proteins. Finally, three applications with respect to protein-ligand complexes are presented.

2.2 THEORY AND ALGORITHM

2.2.1 The PEOE procedure

Electronegativity is a chemically intuitive model descriptor for the charge distribution in covalent bonds. Mulliken [59] defined the electronegativity χ as the arithmetic mean of the ionization potential IP and the electron affinity EA [59]:

$$\chi = \frac{IP + EA}{2} \tag{2.1}$$

Hinze [60][61] expanded this concept and introduced the orbital electronegativity χ_v, i.e. the electronegativity of a specific orbital in a given valence state. The orbital electronegativity is dependent on the type of orbital along with its occupation (and hence the charge). The difference of the orbital charge density of an atom in its unperturbed state and in a molecular environment can be interpreted as partial atomic charge q_{iv} residing on the orbital v of atom A_i. Accordingly, a functional dependence of the orbital electronegativity and the charge can be established. In their original work [57], Gasteiger and Marsili used a second-order polynomial to express the dependence of the orbital electronegativity on the atomic charge q_i. In a more recent version it was replaced by a third-order polynomial:

$$\chi_v = a_v + b_v q + c_v q^2 + dq^3 \tag{2.2}$$

The coefficients av to dv could be derived from valence-state ionization potentials and electron affinities, with the coefficient av being equal to the orbital electronegativity of the uncharged state [60].

To calculate the charge distribution in molecules, Gasteiger and Marsili developed an iterative procedure corresponding to a "partial equalization of orbital electronegativities" (PEOE). The inductive effect is considered by allowing transfer of charge between two bonded atoms and directly bonded neighbors proportional to the difference of their orbital

electronegativities. Thus, PEOE distributes the charges rather locally. To avoid total equalization of electronegativity, the effect of the electrostatic field generated upon charge transfer is considered via a damping factor which is exponentially reduced from one iteration to the next. After an initial charge transfer, the new charges are used to update the electronegativties according to equation 2. The new electronegativities are then applied in the next iteration to redistribute the charges, which in turn serve to update the electronegativities and so forth. Due to the form of the damping factor, rapid convergence is achieved and the procedure is usually terminated after six iterations, making the entire approach very fast.

2.2.2 Performance of the original PEOE charges in PB calculations

To assess the quality of the original PEOE charges (computed with the PETRA program [62]) for PB calculations, solvation free energies were calculated using model compounds for the side chains of polar amino acids for which experimental values are available [63]. Solvation free energies were computed by solving the PB equation, using the program MEAD [64] with the settings described in section Materials and Methods. Sitkoff et al. have carried out a similar study to derive optimized parameters (charges and radii) for the estimation of solvation free energies (the PARSE parameters) [65].

Using the original PEOE charges, the calculated solvation free energies show a correlation of $r^2 = 0.79$ with the experimental values. PARSE parameters give a correlation of r^2 =1.0 for the same data set [65]. In the calculations with the PEOE charges, the solvation free energies are generally underestimated. This might result from the fact that PEOE charges were parameterized on gas phase data. Thus, these do not take solvation effects into account. Although the PEOE procedure yields good dipole moments for gas phase data, this dipole moment seems to be too small for dealing with solvation effects.

Therefore, in a first instance, the PEOE charges were uniformly scaled by a factor of 1.4 to enhance the polarization in the molecule's interior and to lower the shielding effects. This factor was adjusted by a least-squares fit of the calculated to the experimentally given solvation free energies. The RMSD (root mean square deviation) between calculated and experimental free energy of solvation amounts to 2.23 kcal/mol, $r^2 = 0.82$. Investigation of the revealed partial charges and comparison with the PARSE charges showed that the values of our computed partial charges for nitrogen and oxygen were significantly lower.

This observation led to a manual redistribution of the partial charges on oxygen and nitrogen atoms to achieve a comparable polarization as suggested by the PARSE charges. A very good correlation of calculated vs. experimental values resulted in $r^2 = 0.99$. However,

Table 2.1: Values of coefficient a for the PEOE procedure by Gasteiger et al. [16] and modifications of a in the new approach. In column 3, the original coefficients are listed. In columns 4 and 5, the modifications from the sections "Adaptation of the PEOE procedure" and "Testing against 80 small organic molecules" are listed.

Atom	SYBYL atom type	a_{orig}	Changes on a	Changes on a
H	H	7.17	0	0
C	C.3	7.98	0	0
	C.2	8.79	-0.4	0.5
	C.ar	7.98	0.5	0.55
	C.1	10.39	0	0
N	N.ar	12.87	0	-1.29
	N.3	11.54	2.5	6
	N.2	12.87	-0.1	0
	N.pl3	12.87	2	0.5
	N.am	12.87	2.2	3.5
	N.1	15.68	0	0
O	O.3	14.18	0	-3.1
	O.OH	14.18	-	0.8
	O.2	17.07	-2.2	-2.2
	O.co2	17.07	0	-1.82
F	F	14.66	0	0
Cl	Cl	11	0	1
Br	Br	10.08	0	0.8
I	I	9.9	0	1
S	S.3	10.14	0.12	0.12

the performed modifications have the disadvantage that rather arbitrary, purely result-oriented group-dependent corrections are required, which contradicts our initial idea of developing a generally applicable partial charge model. Therefore, the PEOE procedure was re-implemented in order to perform extensive parameter studies.

2.2.3 Adaptation of the PEOE procedure

Modifications were performed exclusively with respect to parameter a (see equation 2). The original coefficients a, b, c, and d were taken from the VEGA software [66]. The modulated coefficients a are shown in Table 2.1. The orbital electronegativities assigned to different valence states are expressed in terms of SYBYL atom types. A slight modification of the sp^2 hybridized carbon atoms (scaling down 4% for C.2, scaling up 6% for C.ar) yielded better

Table 2.2: Calculated and experimental solvation free energies for the polar amino acids in kcal/mol. In the middle columns, the results with the charges from the section Adaptation of the PEOE procedure shown (r^2 = 0.94, RMSD = 0.84). In the right-most columns, the results with the charges from the section Testing against 80 small organic molecules are shown (r^2 = 0.99, RMSD = 0.54).

AS	model compound	ΔG_{exp}	ΔG_{calc}	$\Delta\Delta G_{exp}$	ΔG_{calc}	$\Delta\Delta G_{exp}$
ARG	N-Propylguanidine	-10.92	-10.9	0.02	-9.85	1.07
ASN	Acetamide	-9.72	-9.44	0.28	-8.98	0.74
ASP	Acetic acid	-6.7	-7.02	-0.32	-6.57	0.13
CYS	Methylthiol	-1.24	-1.48	-0.24	-1.02	0.22
GLN	Propionamide	-9.42	-8.77	0.65	-8.5	0.92
GLU	Propionic acid	-6.47	-6.47	0	-6.04	0.43
HIS	Methylimidazole	-10.25	-10.45	-0.2	-10.81	-0.56
LYS	N-Butylamine	-4.38	-1.42	2.96	-3.67	0.71
MET	Methylethylsufide	-1.49	-1.18	0.31	-1.04	0.45
PHE	Toluene	-0.76	-1.13	-0.37	-0.54	0.22
SER	Methanol	-5.08	-4.82	0.26	-4.83	0.25
THR	Ethanol	-4.9	-4.68	0.22	-4.53	0.37
TRP	Methylindole	-5.91	-6.14	-0.23	-5.72	0.19
TYR	p-Cresole	-6.13	-6.09	0.04	-5.94	0.19

agreement with the experimental values. Most modifications were required for nitrogen atom types. The SYBYL atom type notation provides several descriptions for a sp^2 nitrogen atom. The practical application showed that such multiplicity is required for an optimal description.

Another modification concerns the damping factor: It was modulated from its original value of 0.5 to finally 0.778. Additionally, the scaling factor which was adjusted by a least-squares fit, converged to 1.56. This combination provide the best correlation between experimental and calculated solvation free energies for the amino acid model compounds. Subsequent to the PEOE procedure, this correction factor is multiplied uniformly to all partial charges, scaling them to larger absolute values.

The calculated solvation free energies of the polar amino acid side chains based on the modified charges are shown in Table 2.2 in comparison to the experimentally derived values. Figure 2.1 depicts the deviation between computational prediction and experiment. The adaptation of the PEOE charge assignment shows that through minor modifications of the original parameter set a convincing correlation between calculated and experimentally determined solvation free energies can be obtained. Using the original PEOE charges, the

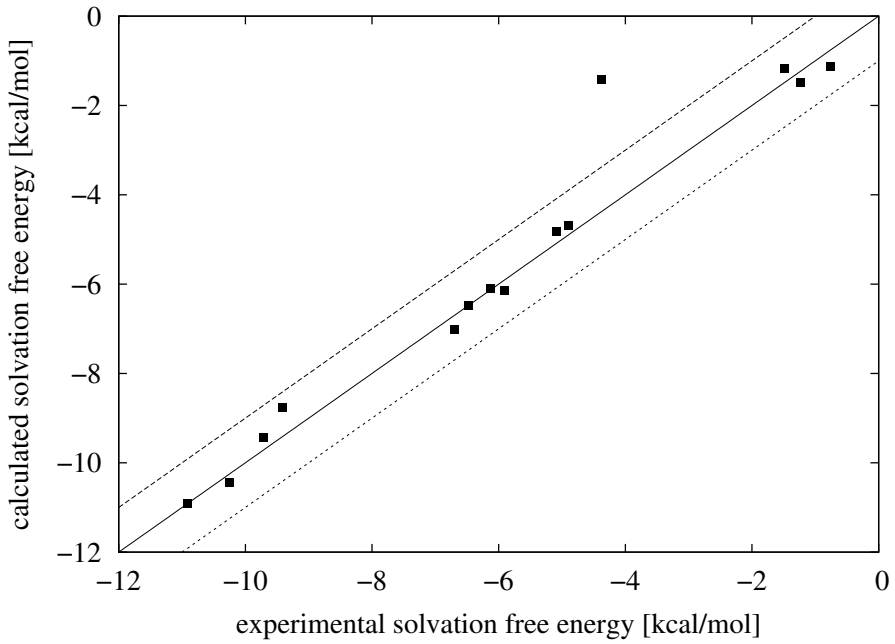

Figure 2.1: Calculated vs. experimental solvation free energies for the polar amino acids model compounds with parameters from the section **Adaptation of the PEOE procedure**. The correlation coefficient r^2 is 0.94, the RMSD is 1.32. The outlier is N-butylamine.

correlation coefficient r^2 is 0.79, whereas with the modified charge set r^2 amounts to 0.94. The applied modifications are purely heuristic. Testing was performed exclusively using polar amino acids as model compounds. To assess the scope and limitations for broad scale applications, 80 small organic molecules for which experimental solvation free energies have been reported were selected and a further cycle of parameter optimization was performed.

2.2.4 Testing against 80 small organic molecules

A set of 80 small organic molecules similarly used by Viswanadhan et al. [67] was taken from the GALAXY program [68]. The original data set consisting of 292 compounds was reduced to 80 representative compounds to avoid unnecessary redundancies displayed by very similar compounds.

Directly applying the modified PEOE parameter set described in section Adaptation of the PEOE procedure resulted in solvation free energies for the 80 small organic molecules which deviated significantly from the experimental reference values. Ethers and alcohols were particularly poorly predicted. Accordingly, we decided to introduce the new atom type O.OH, which corresponds to a hydroxyl sp^3 oxygen, leaving atom type O.3 for all remaining non-hydroxy sp3 oxygens. In addition, the solvation free energies of amines were also rather inadequately predicted. These observations stimulated a further cycle of iterative optimiza-

tion of the electronegativity parameter a. The newly assigned values for a are listed in Table 2.1, in comparison to the parameters obtained from the first optimization cycle (section Adaptation of the PEOE procedure). The most pronounced changes concerned the nitrogen atoms.

The calculated solvation free energies for the model compounds of the polar amino acids with these newly derived a coefficients are listed in Table 2.2. Compared to the first cycle of the PEOE procedure, r^2 improves from 0.94 to 0.99 (cf. Table 2.2). The calculated solvation free energies for the 80 small organic molecules reveal an RMSD of 1.57 and a correlation coefficient r^2 of 0.78 (Figure 2.2).

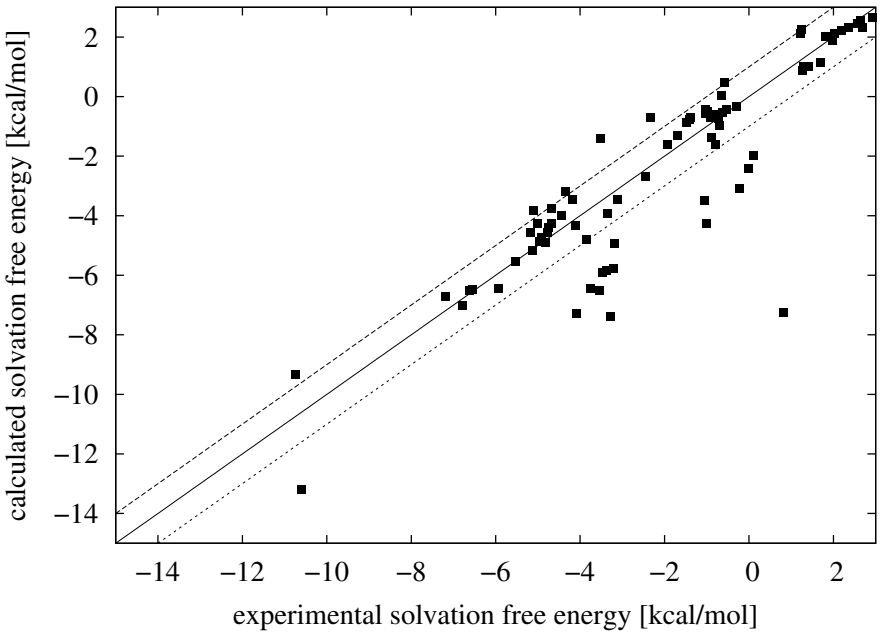

Figure 2.2: Calculated vs. experimental solvation free energies for the validation set with 80 small organic molecules with parameters from the section **Testing against 80 small organic molecules**. The correlation coefficient r^2 is 0.78, the RMSD is 1.57.

2.2.5 Ionized molecules

All molecules considered in the data set so far were handled as neutral species. The amount of experimental values for solvation free energies of ionized molecules is very limited. However, ionized functional groups are ubiquitously present in proteins: the amino acids arginine, aspartate, glutamate, lysine normally bear a formal charge under physiological conditions. Furthermore, cysteine, histidine, and tyrosine can possess a formal charge due to an altered protonation state.

Formal charges can be treated in our adapted version of the PEOE procedure; they are

Table 2.3: Comparison of experimental and calculated solvation free energies of ionized molecules.

ionized system	corresponding amino acid	experimental value	calculated value
N-butylammonium	Lysine	-66.35	-87
Methylimidazolium	Histidine	-59.14	-61.85
Acetate ion	Aspartate	-82.42	-91.97
Propionate ion	Glutamate	-80.63	-90.51

redistributed over several atoms. A fraction of the formal charge is added in each iteration cycle of the PEOE procedure to finally achieve the total charge. In each iteration cycle, the added fraction amounts to the formal charge divided by the number of applied iteration cycles (default is 6), accordingly, after completing all iteration cycles the full amount of formal charges has been assigned. The assignment of formal charges has to be performed prior to the application of the PEOE procedure. For titratable residues, the initially assigned formal charges are as follows: +0.5 (His, both ring nitrogens), -0.5 (Asp and Glu, both carboxylic oxygens), +1.0 (Lys, terminal nitrogen), -1.0 (Tyr, hydroxylic oxygen without proton). Cys and Ser are considered as too basic.

Experimental solvation energies of the charged form of the amino acids were taken from Smith [69], Table 2.3 shows a comparison of the experimental and calculated values. The best estimate is achieved for the model compound imidazole. Both acidic model compounds, aspartate and glutamate, and the model compound for lysine are predicted significantly worse. However, it must be noted that a statistically robust analysis would require many more data points. Furthermore, the reported experimental values (-60 to -80 kcal/mol) fall into a different range compared to the experimental values found for the neutral species (+3 to -10 kcal/mol) which were exclusively used for parameterization.

2.2.6 Comparison to PARSE and CHARMM charges

The resulting PEOE_PB charges for the most essential atoms of the protein titratable groups are listed in Table 2.4. Directly comparing charge models is complicated by the fact that charges are not a physical observable. They are assessed indirectly by reproducing properties accessible by experiment, e.g. dipole moment, electrostatic potential, or solvation energy. PARSE charges [65] were optimized with the aim to reproduce solvation energies, but only polar atoms bear charges unequal to 0, so that a direct comparison to our PEOE_PB is difficult. Major deviations are observed for the backbone carbonyl group, where our PEOE_PB charges adopt smaller absolute values than in the CHARMM and PARSE model. In contrast, for the backbone NH group, our approach converges to larger absolute values. For Asp and

Table 2.4: Comparison of the different charges which were used for the pKa calculations.

		CHARMM	PARSE	PEOE_PB
bb CO	C	0.51	0.55	0.14
	O	-0.51	-0.55	-0.3
bb NH	N	-0.47	-0.4	-0.73
	HN	0.31	0.4	0.49
Asp/Glu	C	0.62	0.1	0.11
(charged)	O	-0.76	-0.55	-0.7
Lys	N	-0.3	-0.32	-0.64
(charged)	H	0.3	0.33	0.56
His	$N_{\delta 1}$	-0.51	-0.35	-0.33
(charged)	$C_{\delta 2}$	0.19	0.14	-0.01
	$C_{\epsilon 1}$	0.32	0.14	-0.07
	$N_{\epsilon 2}$	-0.51	-0.35	-0.34
	$H_{\delta 1}$	0.44	0.45	0.51
	$H_{\delta 2}$	0.13	0.13	0.28
	$H_{\epsilon 1}$	0.18	0.13	0.33
	$H_{\epsilon 2}$	0.44	0.45	0.5
Tyr	OH	-0.54	-.49	-0.6
(neutral)	HH	0.43	0.44	0.44
	CZ	0.11	0.06	0

Glu, the carboxylate oxygens are handled by similar charges, whereas the CHARMM model assigns a more positive value to the carboxylate carbon. For Lys, our PEOE_PB charge model assigns larger absolute values, whereas for His and Tyr all three models agree moderately.

2.2.7 Protein pKa calculations

Since the pKa value of a titratable residue (with all other titratable sites fixed in their neutral state) cannot be directly calculated, a thermodynamic cycle must be setup 9, resulting in the following expression:

$$pK_a^{int} = pK_a^{mod} + \frac{1}{2.303 k_B T}[\Delta G_{s,p}(A) - \Delta_{s,p}(AH)] \tag{2.3}$$

This gives the intrinsic pKa value pK_a^{int}: It is composed of the model pKa value pK_a^{mod}, which corresponds to the pKa value of the isolated residue in aqueous solution and the difference of the free energies of transfer from aqueous solution (s) to the protein environment (p) for the protonated (AH) as well as the deprotonated form (A): $[\Delta G_{s,p}(A) - \Delta_{s,p}(AH)]$. $\Delta G_{s,p}$ is composed of two terms: the Born term, accounting for the desolvation energy, and the background term, which considers the electrostatic interaction with the permanent partial charges of the protein. For the calculation of the intrinsic pKa value, all titratable groups are treated in their neutral state. k_B is the Boltzmann constant, and T is the temperature. A comprehensive summary of how to calculate the required free energies and other competitive methods for pKa calculations, the reader is referred to Warshel's recent review [34] and the references therein.

The interaction with the titratable groups adopting all possible protonation states is evaluated via the site-site interactions energy term. Approximations are required to compute this term since for m titratable sites 2^m possible interactions must be considered. Bashford [70] developed the "reduced site" approach: It neglects those groups in the calculation of site-site interactions that are either very close to full deprotonation or full protonation. This method will be used throughout all pKa calculations in this study. The finally calculated pKa value for each titratable residue is the sum of its intrinsic pKa value and the corresponding site-site interaction energy term.

Table 2.5: Parameter settings for ΔG_{solv} calculations with MEAD.

Parameter	Value
Grid size	65x65x65Å
Grid spacing	0.25 Å
Temperature	300 K
Ionic strength	0.1 mol/l
$\epsilon_{Solvent}$	80
$\epsilon_{Molecule}$	2

Table 2.6: Parameter settings for pKa calculations with MEAD.

Parameter	Value
Temperature	300.0 K
Ionic strength	0.1 mol/l
$\epsilon_{Solvent}$	80
$\epsilon_{Molecule}$	20
r_{H_2O}	1.4 Å
Stern layer	2.0Å
Grid margin	20 Å

2.3 MATERIALS AND METHODS

2.3.1 Solvation free energies

A three-dimensional structure for all amino acid model compounds was generated using the program CORINA [71] and minimized with the program MOE [72]. MEAD [64] was used for the calculation of ΔG_{solv}, with the settings listed in Table 2.5. The surface term was calculated by MSMS [73], using Bondi's radii [74]. The validation set of 80 small organic molecules was taken from Viswanadhan et al. [67]; 3D coordinates were generated using CORINA [71] and further applied without additional minimization.

2.3.2 pKa calculations

MEAD was applied for all pKa calculations [64] with the parameter settings listed in Table 2.6. All water molecules in the crystal structures were removed prior to the pKa calculation. Site-site interactions were treated by the MEAD reduced-site approximation [70]. If not

otherwise stated, $\epsilon_{Protein}$ was set to 20. A grid spacing of 1.0 Å was applied; next to the titratable groups a focussed grid of 0.25 Å was utilized. The program REDUCE [75] was used for adding hydrogen atoms. The following residues were considered as titratable (model pKa value in parentheses): aspartic acid (4.0), C-terminus (3.8), glutamic acid (4.4), histidine (6.3; only N_ϵ was titrated), lysine (10.4), N-terminus (7.5), and tyrosine (9.6). No explicit hydrogens were set at the carboxylic functions of glutamate and aspartate. All titratable residues of the nine proteins listed in the next section were considered for the pKa calculations.

To assign titratable residues in the napsagatran/thrombin complex, all groups within a 12 Å sphere around napsagatran were considered. The model pKa values of the carboxylic functions of napsagatran and CRC220 were set to the experimentally determined values of 3.4 and 3.8, respectively [22].

2.3.3 X-ray structures

In total, nine proteins were used for testing. In the following, they are referred to by their PDB codes: 1A2P (barnase), 4PTI (bovine pancreatic trypsin inhibitor), 3ICB (calbindin), 2LZT (lysozyme), 1PPF (turkey ovomucoid third domain), 1PGA (protein G), 3RN3 (ribonuclease A), 2RN2 (ribonuclease H), and 1XNB (xylanase). The test data set thus comprised 132 experimentally determined pKa values [37, 18] (cases where only upper or lower bounds of the experimental pKa were given are not considered).

The structure of the napsagatran/thrombin complex was kindly provided by Roche, Basel. The coordinates of the CRC220/thrombin complex were taken from Fokkens [76].

2.4 RESULTS AND DISCUSSION

2.4.1 pKa calculations of protein residues

The test set consists of 132 experimental pKa values from a total of nine (uncomplexed) proteins: barnase, bovine pancreatic trypsin inhibitor, calbindin, hen egg white lysozyme, turkey ovomucoid third domain, protein G, ribonuclease A, ribonuclease H, and xylanase. This data set was previoulsy used by Nielsen et al. [20] and we will refer to their results for comparison.

The calculated pKa values for all 132 residues show an overall RMSD of 0.88 with respect

Table 2.7: The overall RMSD (calculated versus experimental pKa values) listed for the nine proteins of the validation set.

protein	number of exp. values	RMSD PEOE_PB	RMSD PARSE	RMSD CHARMM	RMSD reference [20]	RMSD Null model
Lysozyme	19	0.77	1	0.91	0.83	1.02
BPTI	10	0.48	0.52	0.45	0.37	0.6
Ovomucoid	15	0.67	0.55	0.61	0.8	1.08
Protein G	13	0.88	0.77	0.86	0.87	0.62
Ribonuclease A	16	0.91	1.34	1.16	1.37	0.86
Ribonuclease H	25	0.8	1.11	1.08	1.18	0.89
Barnase	12	1.12	1.09	1.03	1.29	1.37
Calbindin	10	0.66	0.66	0.72	0.39	0.71
Xylanase	12	1.35	1.4	1.55	1.18	1.74
total	132	0.88	1.01	1.01	1.02	1.04

to the experimentally determined pKa values (Figure 3). In Table 2.7, the average RMSD

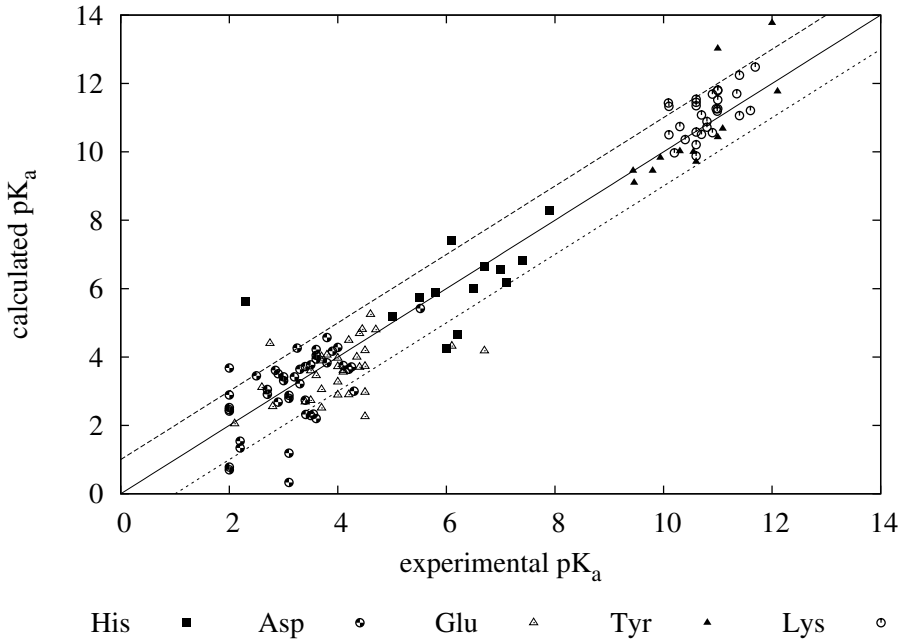

Figure 2.3: Comparison of calculated and experimental pKa values for the validation set consisting of nine proteins. The overall RMSD is 0.88 log units.

for each protein is listed. Substitution of the newly adapted PEOE charges (PEOE_PB) by CHARMM22 [77] charges and the corresponding radii reveals a less satisfactory RMSD of 1.01, substitution by PARSE charges gives also a RMSD of 1.01. Using an optimized hydrogen bond network, Nielsen et al. achieved an RMSD of 1.02 [20] (cf. Table 2.7). Splitting up the 132 experimentally determinded pKa values into the different residue types results in

the following RMSDs: Asp (44 values, RMSD: 0.88), Glu (33 values, RMSD: 0.94), His (13 values, RMSD: 1.24), Lys (29 values, RMSD: 0.61), Tyr (13 values, RMSD: 0.95). The overall RMSD only measures the overall quality of pKa calculations. However, the experimental values of a large percentage of residues in the data set shows only slight deviations from the actual model values: Among the 132 experimental pKa values, 62 residues deviate by less than 0.5 pKa units from the model pKa value. Such values have to be faced with a null model which assumes no pKa shifts for any titratable residue. Overall, such a model reveals an RMSD of 1.04 (for all 132 residues with experimental pKa values).

A more indicative analysis can be achieved by comparing only those titratable residues that deviate experimentally by more than one pKa unit with respect to the model pKa. Overall, 37 residues obey this criterion, and the resulting RMSD_l (l for large) with the PEOE_PB charges amounts to 1.04. By using an optimized hydrogen bond network, Nielsen et al. obtain an RMSD_l of 1.12 for these residues. Considering the null model, the RMSD_l is 1.78. The experimental pKas of 21 residues deviate even more than 1.5 log units from the model pKa, as listed in Table 2.8. In this table, not the absolute pKa values are listed, but the differences $\Delta(\mathrm{pKa}^{exp}\text{-}\mathrm{pKa}^{mod})$ for the deviation from experiment to the model pKa and $\Delta(\mathrm{pKa}^{calc}\text{-}\mathrm{pKa}^{mod})$ for the different charge sets. In Figure 2.4, a discriminative benchmark which plots these differences is displayed - such a benchmark was suggested by Warshel et al. [34]. Interestingly, some residues show large shifts away from the model compound in the opposite direction of the solvation free energy contribution (ΔG_{solv} favours the neutral state of a titratable residue in the protein). This indicates that factors other than solvation free energy must be responsible for these shifts. As a structural analysis shows, in most of the cases, these residues are involved in salt bridges. For a number of residues, the difference between $\Delta(\mathrm{pKa}^{exp}\text{-}\mathrm{pKa}^{mod})$ and $\Delta(\mathrm{pKa}^{calc}\text{-}\mathrm{pKa}^{mod})$ is larger than one pKa unit. The pKa shift of Asp83 in xylanase is strongly influenced by a salt bridge to Arg136. A similar explanation holds for Asp148 in ribonuclease H which is involved in a salt bridge to Arg46. Tyr31 in turkey ovomucoid third domain forms a hydrogen bond to Asp27. Asp101 in barnase forms three hydrogen bonds to Thr105 (backbone amide and side chain hydroxyl) and Tyr103 (backbone amide). Asp14 in ribonuclease A is involved in a network of three hydrogen bonds (side chains of Ser16, Thr17, Tyr25). His149 in xylanase is discussed in the subsequent paragraph. The overestimation of the pKa shifts of the mentioned residues could possibly be circumvented by the usage of a higher dielectric constant for the charge-charge interactions (ϵ_{eff} in PDLD/S terminology) compared to $\epsilon_{Protein}$ (as performed by Warshel et al. 13), but this is not possible in our PB setup. Alternatively, a further optimization of the PEOE_PB charges with respect to the interaction energies could yield an improvement. The same may hold for the other charge sets which show a similar trend in the calculated pKa shifts (and contributions from the site-site interactions) as the calculations with the

Table 2.8: Residues which show a pKa shift of more than 1.5 log units from the model pKa, listed as Δ(pKa^{exp}-pKa^{mod}). The differences for Δ(pKa^{calc}-pKa^{mod}) with the PEOE_PB, PARSE and CHARMM charges are also listed. Figure 2.4 shows the corresponding diagram.

PDB code	residue	Δ(pKa^{exp} -pKa^{mod})	Δ(pKa^{calc} -pKa^{mod})	Δ(pKa^{PARSE} -pKa^{mod})	Δ(pKa^{CHARMM} -pKa^{mod})
2LZT	7-GLU	-1.8	-1.29	-1.72	-1.53
2LZT	35-GLU	1.7	-0.08	-0.29	-0.29
2LZT	53-TYR	2.5	2.17	1.92	1.97
2LZT	66-ASP	-2	-1.11	-2.31	-2.22
3RN3	2-GLU	-1.6	-1.85	-2.48	-2.68
3RN3	14-ASP	-2	-2.46	-3.49	-3.28
1PPF	31-TYR	2.9	4.5	3.13	4.28
2RN2	10-ASP	1.52	1.42	1.35	2.04
2RN2	102-ASP	-2	-1.48	-2.25	-2.57
2RN2	127-HIS	1.6	2	1.49	2.02
2RN2	148-ASP	-2	-3.22	-4.58	-4.1
1A2P	54-ASP	-1.8	-2.66	-2.7	-2.48
1A2P	73-GLU	-2.3	-2.35	-2.48	-2.54
1A2P	93-ASP	-2	-1.58	-2.81	-2.07
1A2P	101-ASP	-2	-0.32	-2.48	-1.08
1XNB	11-ASP	-1.5	-0.55	-1.17	-0.67
1XNB	83-ASP	-2	-3.3	-3.54	-4.71
1XNB	101-ASP	-2	-1.54	-2.63	-2.49
1XNB	149-HIS	-4	-0.69	-0.89	-0.84
1XNB	172-GLU	2.3	-0.22	-0.73	-0.74

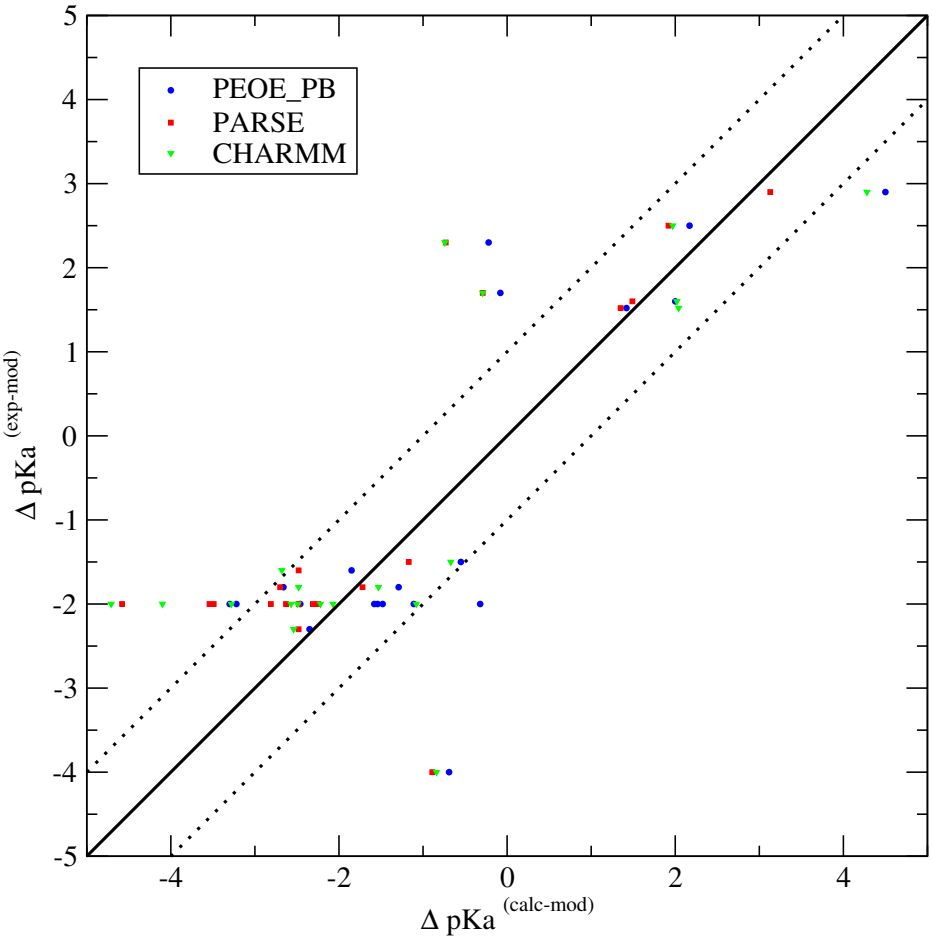

Figure 2.4: Discriminative benchmark for the residues in Table VIII (according to Warshel's discriminative benchmark [34]). Residues deviating more than one pKa unit are discussed in the text.

PEOE_PB charges.

The RMSD between experimental and calculated pKa values for all residues in Table 2.8 is 1.28; CHARMM charges give an RMSD of 1.29, while PARSE charges result in an RMSD of 1.37. Nielsen et al. obtain an RMSD of 1.38, the null model would give an RMSD of 2.15. It is worthwile noting, however, that the only two residues for xwhich the calculated *protonation state* at physiological pH differs significantly from experiment are Glu35 in lysozyme and Glu172 in xylanase. These two residues will be discussed below.

Xylanase is the protein which shows the largest overall RMSD (1.35, cf. Table 2.7). Only a limited number of residues is responsible for this deviation: the active site residues Glu78 and Glu172 (their titration behaviour will be explored in detail in a subsequent paragraph), Asp83 with a calculated pKa of 0.7 (pKa^{exp} = 2.0; pKa^{calc} = -0.7 using CHARMM charges), and His149 with a calculated pKa of 5.6 (pKa^{exp} = 2.3; pKa^{calc} = 5.5 using CHARMM charges). The pKa shift of Asp83 can be explained by the formation of a buried ion pair with Arg136. The reason for the lowered pKa value of the buried His149 (see also Table 2.8)

Table 2.9: Influence of the orientation of the OH-group of Tyr80 on the pKa values of Glu78 and Glu172. *std* corresponds to the standard hydrogen setting with the program REDUCE, *rotated* corresponds to a change in the dihedral angle of the hydroxyl group of Tyr80 by 180°. For comparison, the results with CHARMM charges are also shown.

$\epsilon_{Protein} = 4$	exp	*std* PEOE_PB	*rotated* PEOE_PB	*std* CHARMM	*rotated* CHARMM
Glu78	4.6	8.2	5.4	7.2	3.8
Glu172	6.7	4.1	9.7	2.9	10.1

is a hydrogen bond to a water molecule in close proximity. Although the deviation (pKacalc - pKaexp) of both, Asp83 and His149, exceeds one log unit, the predicted protonation state at physiological pH based on the calculated pKa values yields the correct sign. The same conclusion applies to the results obtained with CHARMM charges.

In continuum electrostatics calculations, the choice of the dielectric constant of the protein is crucial. For microscopic solvent simulations, where all interactions are considered explicitly, the dielectric constant of the protein is set to 1. Replacing the explicit solvent molecules by an implicit solvation model, the dielectric constant of the protein has to be assigned a value greater than 1. There is no universal dielectric constant to be assigned to proteins [15]. The standard definition of a dielectric constant would be the relationship between dipole fluctuations and the dielectric constant [34]. King et al. were able to show that the (macroscopic) dielectric constant depends locally on the protein region and that the reaction field (which is the effect of the missing solvent around the given region) has to be taken into account [78]. In general, $\epsilon_{Protein}$ is simply an empirical parameter that tries to represent contributions not included explicitly in the model [32, 15], such as protein reorganization [34]. Warshel concluded that the actual value of the dielectric constant entirely depends on the model and assumptions used to define this constant [11]. An overview on how optimally selected values for $\epsilon_{Protein}$ depend on the selected model (Born, TK, MTK, PDLD/S, PDLD/S-LRA) is summarized in Warshel's review from 2001 [34]. Antosiewicz et al. [18] obtained best agreement (concerning the overall RMSD) with experiment when using a dielectric constant of 20. Also in our present study, this value has been used throughout all performed pKa calculations, if not stated otherwise. Antosiewicz et al. rationalize this setting by the fact that a rather high dielectric constant compensates for conformational flexibility which is not taken into account explicitly. It was argued that this statement does not hold, since the actual choice of the dielectric constant of protein depends on the benchmark used for validation; furthermore, a rather high dielectric constant is caused by the ambition to calculate self-energies and charge-charge interactions with the same dielectric constant [34]. Nevertheless, the selection of the dielectric constant is crucial and this entity has to be

regarded in the present calculations as an additional adjustable parameter.

For two enzymes in the validation set, experimental pKa values have been determined for active site residues. Such residues are of special interest, because they are directly involved in ligand binding. However, their prediction is not trivial due to a high degree of burial and/or strong electrostatic interactions. In addition, the prediction of such residues is challenging, because the previously described overall RMSD (0.88 with PEOE_PB charges) mostly considers surface exposed residues with small pKa shifts and every model using a large dielectric constant (for the charge-charge interaction) will be accurate enough [79]. These two examples will show how cruical the choice of the dielectric constant is.

In lysozyme, glutamic acid 35 shows a pKa value of 6.2. The other catalytic residue, aspartic acid 52, exhibits an experimental pKa of 3.6. The spatial arrangement of these residues is shown in Figure 2.5. Warshel et al. investigated the influence of different dielectric constants

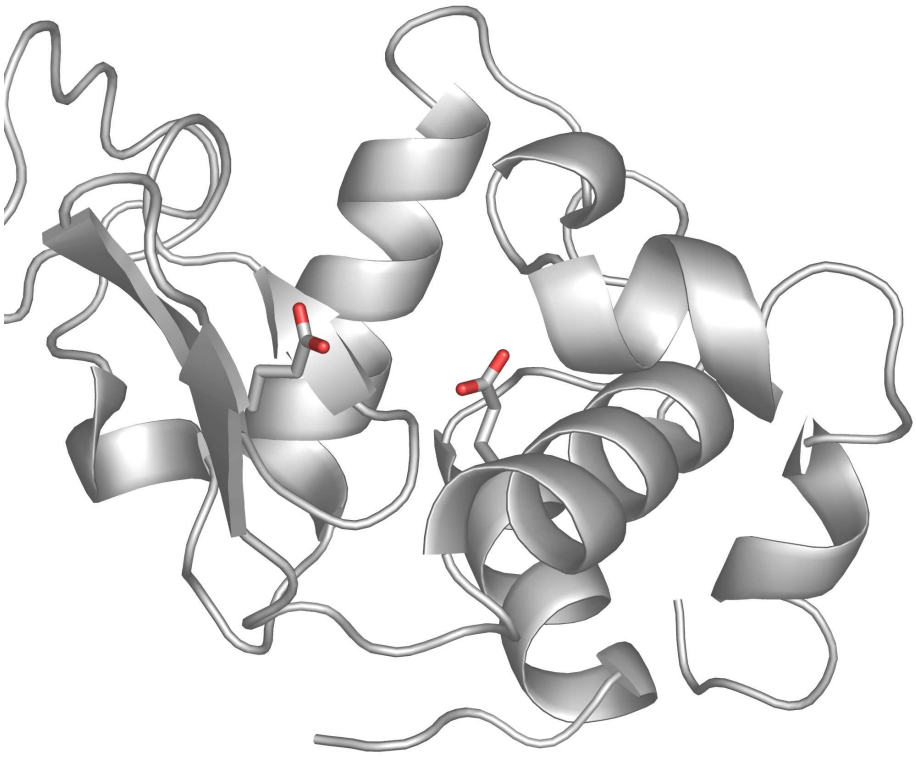

Figure 2.5: Active site residues Glu35 (pKa=6.2) and Asp52 (pKa=3.6) of lysozyme.

for $\epsilon_{Protein}$ and ϵ_{eff} (for the charge-charge interactions) and obtained best agreement with experiment for Glu35 when setting $\epsilon_{eff} = 80$ and $\epsilon_{Protein} = 4$. Their results suggest to use rather large values for ϵ_{eff} [34]. Such a treatment is not possible with our protocol, but

it can be realized in the PDLD/S-LRA methodology. The latter approach refrains from using a rather *artificially* high dielectric constant of the protein. Nevertheless, the present pKa calculations on lysozyme show that a simple adjustment of $\epsilon_{Protein}$ (no differentiation between ϵ_{eff} and ϵ_P as done by Warshel [34]) is able to produce correct trends: When setting $\epsilon_{Protein}$ to 4, the calculated pKa of Glu35 is 6.0 (with $\epsilon_{Protein}$ =20, pKa^{calc} = 4.3) and 3.4 for Asp52 (with $\epsilon_{Protein}$=20, pKa^{calc} = 3.8). Substituting our PEOE_PB by CHARMM charges results in similar values: With $\epsilon_{Protein}$ = 4, the calculated pKa of Glu35 is 5.1 (with $\epsilon_{Protein}$=20, pKacalc = 4.1) and 2.8 for Asp52 (with $\epsilon_{Protein}$=20, pKacalc = 3.4). The same trends are observed for PARSE charges: With $\epsilon_{Protein}$= 4, the calculated pKa of Glu35 is 5.1 (with $\epsilon_{Protein}$=20, pKa^{calc} = 4.1) and 2.9 for Asp52 (with $\epsilon_{Protein}$=20, pKa^{calc} = 3.3). As mentioned, in these calculations the actual value of $\epsilon_{Protein}$cannot be really justified by physical reasons (e.g. a dielectric constant derived from changes in the average dipole moment [34]), which is a known problem of the PB method. Based on an optimized hydrogen-bond network, Nielsen et al. [20] obtained a pKa of 5.7 for Glu35 and 1.8 for Asp52. Antosiewicz et al. [18] calculated a pKa of 5.9 for Glu35 and 2.8 for Asp52 ($\epsilon_{Protein}$=4). Demchuk and Wade [19] did not achieve satisfactory prediction of the pKa value of Glu35 (4.5; pKa for Asp52: 3.0), although their model uses two dielectric constants for the protein which would suggest a better spatial adaptation of the dielectric conditions.

In xylanase, the active site comprises two glutamic acid residues (Glu172 and Glu78), with pKa values of 6.7 and 4.6 respectively. Figure 2.6 depicts the overall geometry of the active site. The experimentally determined pKa values (especially of Glu78) are not well reproduced applying $\epsilon_{Protein}$ =20 (Glu172: pKa^{calc}=4.2; Glu78: pKa^{calc}=5.3). When decreasing $\epsilon_{Protein}$ to 4, the prediction becomes even worse, because a dramatic upward shift in pKa results for Glu78. A similar trend is recognized using CHARMM charges (see Table 2.9). Nielsen et al. [20] obtained a pKa of 5.0 for Glu172 and 3.3 for Glu78 (based on an optimized hydrogen bond network). Calculations based on mutants for which experimental values are also available [80] demonstrate the importance of the hydroxyl group of Tyr80 which is in close proximity to both active site glutamic acids. Upon mutation of this residue to Phe, the experimentally determined pKa value of Glu172, which is also shifted upwards, can be reproduced: In the Y80F mutant, the experimental pKa values of Glu172 and Glu78 are 7.9 and 5.0, respectively; the calculations yield 7.9 (Glu172) and 5.3 (Glu78) with an $\epsilon_{Protein}$=4. By variation of the dihedral angle of the hydroxyl group in the wild type (rotating off from the standard geometry suggested by the program REDUCE [75]), the computed pKa of Glu172 experiences the expected upward shift ($\epsilon_{Protein}$= 4). The same trend is suggested using CHARMM charges (see Table 2.9). Increasing $\epsilon_{Protein}$ from 4 to 20 reveals results that are in less satisfactory agreement with experiment. Obviously in the present case, the usage of a rather low value for the dielectric constant permits better pKa prediction of the

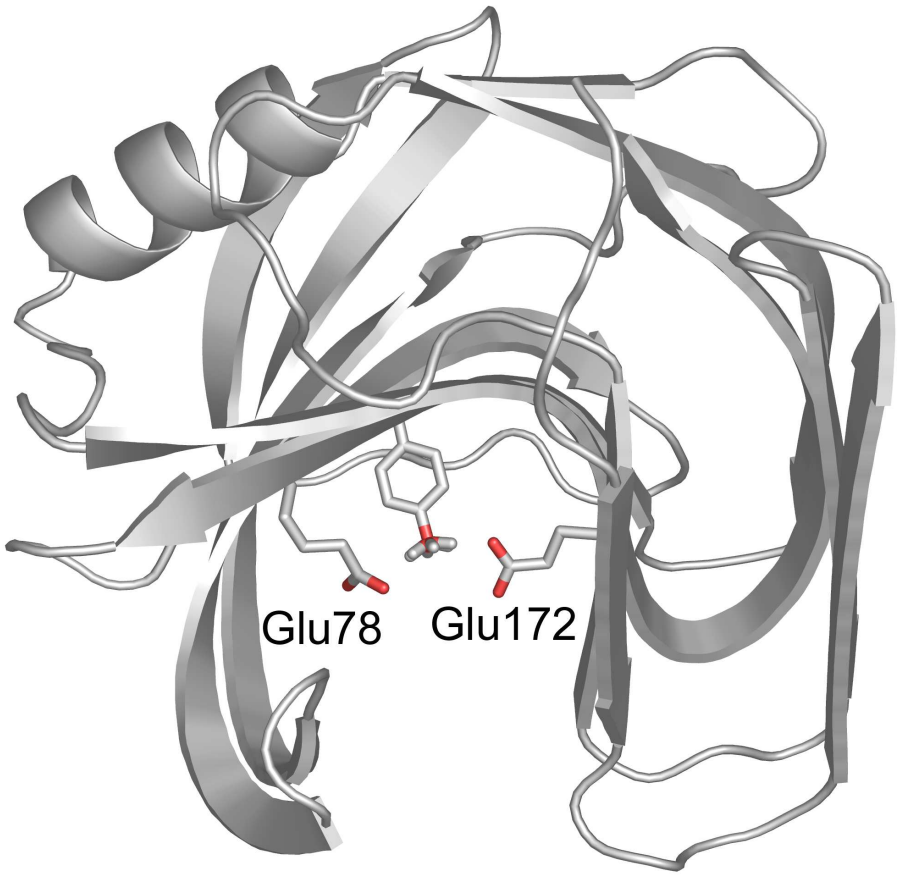

Figure 2.6: Rotation of the OH-group of Tyr80 (of xylanase) in steps of 60 degrees. The catalytically relevant glutamates 78 and 172 are displayed as capped sticks.

active site residues. In summary, our PEOE_PB charges yield quantitatively reliable pKa predictions of comparable or even better quality to competitive PB calculations in literature, thus justifying their application in continuum electrostatics calculations. As major advantage compared to other methods, the PEOE protocol can be easily extended to ligands covering a much broader range of functional groups compared to protein residues.

2.4.2 pKa calculations on two complexes of thrombin

Isothermal titration calorimetry (ITC) directly records the thermodynamic contributions of ligand binding to a protein. If ITC experiments are performed in different buffers, it is possible to discover protonation changes upon ligand binding. However, no microscopic picture is obtained where the protonation changes actually occur, because only the net changes in the protonation inventory are recorded. Dullweber et al. [22] performed ITC measurements on a series of serine protease complexes. Two examples from this study (one with an observed overall change of protonation and the other without any change) are studied computationally here and the PEOE_PB charges are applied for pKa calculation.

Napsagatran and CRC220 are inhibitors of the serine protease thrombin (the chemical formulae are given in Figure 2.7a and 2.7b). They share in common very similar functional

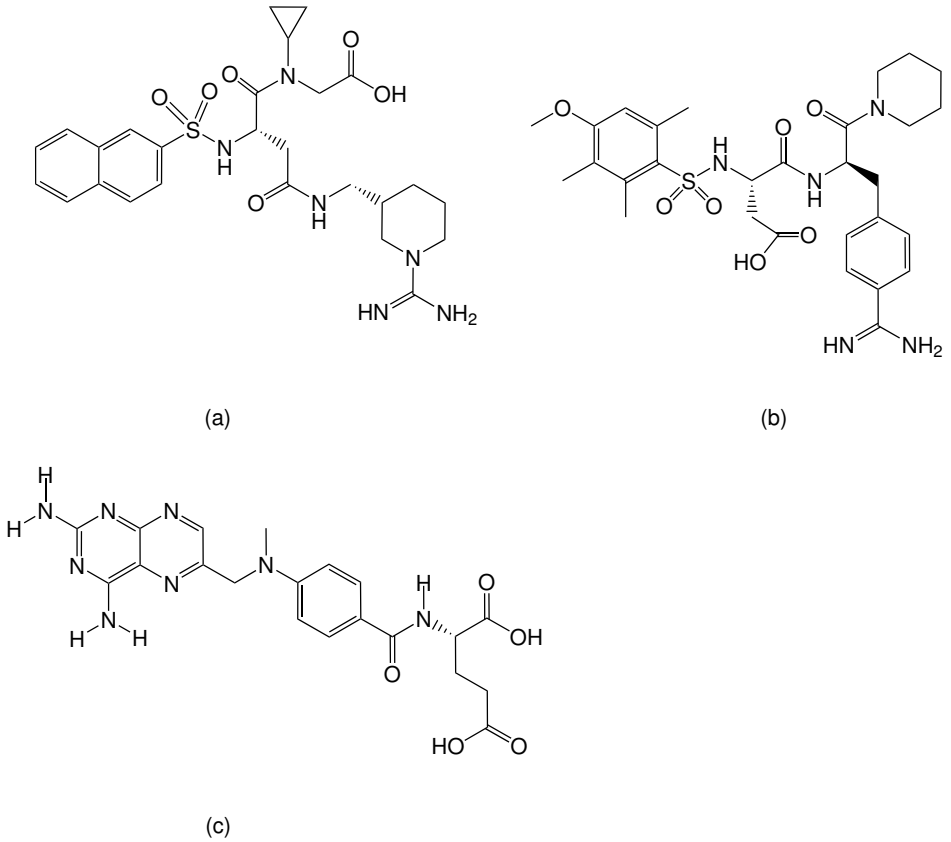

Figure 2.7: Napsagatran (a) and CRC220 (b) - inhibitors of the serine protease thrombin. The studied inhibitor of dihydrofolate reductase, Methotrexate (c).

groups that could be classified as potentially titratable. Nevertheless, ITC indicates that only Napsagatran binding induces a change in protonation state. Figure 2.8 shows the complexes between thrombin, napsagatran and CRC220. An overall protonation change of +0.9 moles of protons is observed upon formation of the thrombin-napsagatran complex [22]; thus, the dissolved protein-ligand complex captures 0.9 moles of protons from the buffer. To obtain a first clue on possible changes in protonation, Dullweber et al. also measured the napsagatran ethylester. Interestingly enough upon complex formation no change in protonation states could be observed. This suggests on first glance that the carboxylic function of napsagatran picks up the proton. More striking is the fact that CRC220 does not provoke such changes even though it has a very similar carboxylic function. Both ligands exhibit similar pKa values for this functional group. ITC indicates only overall changes but does not answer the question where the proton actually goes to, thus pKa calculations are consulted to provide an explanation.

pKa calculations were performed for uncomplexed thrombin and for the thrombin-inhibitor

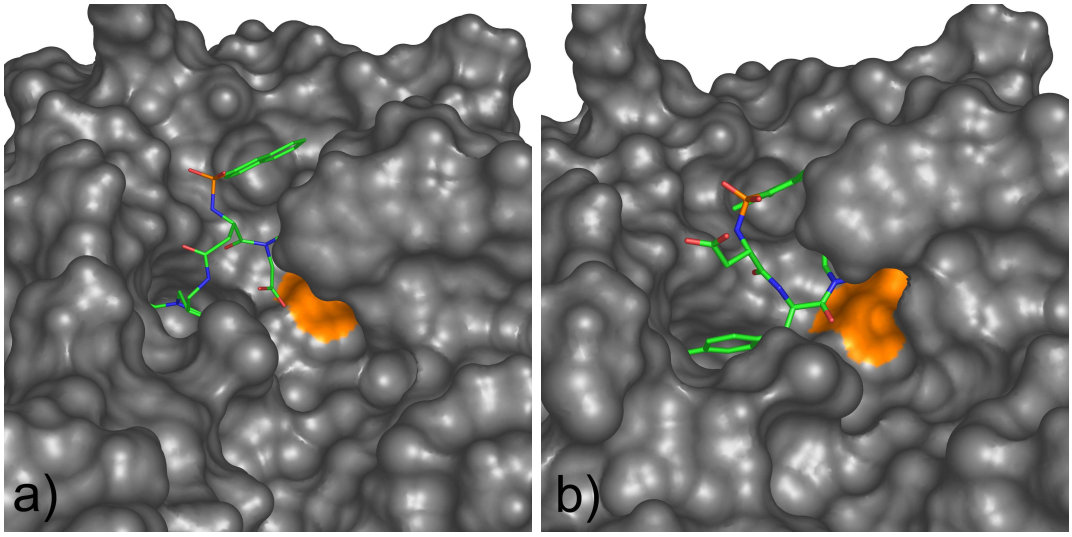

Figure 2.8: Thrombin in complex with napsagatran (a) and CRC220 (b) - the surface of His57 is colored in orange.

complexes. The uncomplexed form was generated using the coordinates of the complex by removing the ligand. This rather pragmatic strategy is born out of the desire to achieve general applicability, such that a structure of the uncomplexed protein does not have to be necessarily available to perform such calculations. Admittedly, thrombin is an easy case since only minor conformational adaptations are generally observed upon ligand binding. Thus, the uncomplexed structure is indeed not required in this case.

The pKa calculations ($\epsilon_{Protein} = 20$) show a pKa shift of histidine 57: its pKa value switches from 7.3 in the uncomplexed structure to 8.5 in the complex with napsagtran. This pKa shift can be considered significant, recalling the range defined in section pKa calculations of protein residues. Furthermore, an experimental value for the pKa of His57 has been determined for the homologous serine proteases trypsin and chymotrpysin, and it was found to be close to 7.0, well in agreement with our calculations for the uncomplexed thrombin. No other residues of the protein show significant pKa shifts upon complex formation, nor do functional groups of the ligand - the carboxylic group of napsagatran definitely remains deprotonated.

Surprisingly, ITC reveals no change of protonation upon formation of the CRC220/thrombin complex, even though very similar functional groups are involved. However, crystallography has revealed a different binding mode with respect of the placement of the carboxylic group in a different area of the binding site. It is pointing towards the adjacent solvent environment. The calculations with the new PEOE procedure show no significant pKa shifts of any residues or ligand functional groups when comparing the complexed and uncomplexed form. The pKa values for His57 move only slightly from 6.1 (uncomplexed) to 5.7 (complexed), which results in no significant change of protonation. The carboxylic group of the CRC220

also remains in its original protonation state (i.e., deprotonated). No other titratable residue shows significant pKa shift which could lead to a change of protonation. Accordingly, the calculations are in agreement with the ITC observations for the CRC220/Thrombin complex.

2.4.3 pKa calculations on a dihydrofolate reductase complex

DHFR (dihydrofolate reductase) belongs to the enzyme class of oxido-reductases. MTX (methotrexate, for chemical formula see Figure 2.7c) is a small molecule inhibitor of DHFR which resembles the structure of the natural substrate, folate. The binding of MTX to DHFR is known to be strongly pH-dependent [81, 82].

Early binding studies suggested that Asp27 is protonated in the free enzyme with a pKa of 6.6 [82]. NMR studies [81, 83, 84] showed that in the DHFR/MTX complex, the N1 nitrogen of the pteridine ring is protonated even at a pH greater than 10, although the pKa value in solution is significantly lower (pKa=5.7). It is believed that Asp27 and the N1 nitrogen form a salt bridge. As a result, Asp27 would be no longer protonated.

To study the crucial protonation states and the question of a potentially formed salt bridge, pKa calculations with the PEOE_PB charges were carried out, in similar fashion as described by Cannon et al. [51]. Two calculations were performed for the crystal structure with the PDB code 4DFR, since two independent molecules are present in the unit cell (referred to as MTX-A and MTX-B), due to higher completeness the usage of MTX-B for PB calculations was recommended by Cannon et al. [51].

The calculations ($\epsilon_{Protein}$=20) for MTX-B show that a significant pKa shift occurs for nitrogen N1 of MTX, namely from 5.7 (apo) to 7.6 (complexed). Asp27 shows a pKa shift in the opposite direction, from 3.4 to 0.5. The same trend can be observed when using MTX-A as model: The pKa of N1 shifts from 5.7 (apo) to 7.9 (complexed), and the pKa of Asp27 is lowered from 2.9 to 0.0. This result suggests the formation of a salt bridge between the protonated N1 of MTX and the deprotonated Asp27 of DHFR. Our results are in disagreement with the calculations of Cannon et al. [51], but correspond to the experimental results [81, 82, 83, 84]. Cannon at al. applied the same methodology (PB-based pKa calculations), but they used OPLS charges for the protein and ab initio charges (HF/3-21G) for the ligand. The application of charges derived from quantum chemical calculations is one of Cannon's explanations for the failure of their pKa calculations [51]. With our adapted PEOE charges, we are able to produce consistently charges using one computational method for both, protein and ligand. Furthermore, the computational burden for the charge calculation (involving ab initio methods) is reduced to a minimum, and the setup of the pKa calculations is simplified.

Most important, our results are in agreement with experiment.

2.5 Conclusions

The adaptation of the well-established PEOE formalism [57] to charge calculations in aqueous phase and of proteins, as well as its application to compute solvation free energies and pKa values in proteins is described in this contribution. Experimental solvation free energies of small organic molecules are well reproduced (see section Testing against 80 small organic molecules). The application of our adapted PEOE charges in pKa calculations shows convincing agreement with experiment, and equally good or better results are obtained compared to other PB-based studies [17][18][19][20] (see section pKa calculations of protein residues). Even though a competitive approach, the PDLD/S-LRA method, yields more accurate results for lysozyme [34], the pKa values of enzyme active site residues are still satisfactorily reproduced with our approach, suggesting the application of the method to pKa calculations of active site residues and protein-ligand complexes. Good agreement with experiment could be achieved for three protein-ligand complexes. As a major advantage compared to any other presently known charge assignment procedure used for continuum electrostatic calculations of macromolecular systems, the approach is parameterized for proteinogenic groups and a broad range of organic molecules.

Many protein-ligand complexes are known to change protonation states upon ligand binding. With the newly introduced charge model, we propose a simple protocol useful for performing pKa calculations on such complexes. This suggests incorporating such calculations in early steps of drug design, in particular to assume correct protonation states of the residues involved in ligand binding. For docking and scoring it is of utmost importance to consider the binding site residues and the functional groups of the docked ligands with appropriate protonation.

Furthermore, such calculations are an essential prerequisite to correctly interpret thermodynamic measurements. The complex formation of the above-mentioned thrombin inhibitors napsagatran, its ethylester and CRC220 has been studied by isothermal titration calorimetry [22]. This method provides experimental access to the binding free energy ΔG and the heat produced or absorbed during the binding process. After correction of all contributions involving changes in protonation states, the latter entity allows estimating the enthalpy of binding. By calculating the difference between ΔG and ΔH the entropic contribution to binding ($-T\Delta S$) can be determined. Correct factorization of ΔG into enthalpy and entropy provides important insight into the thermodynamic driving forces of binding. CRC220 and

the ethylester of napsagatran bind to thrombin without change of protonation state. Accordingly, the recorded heat can be directly translated into ΔH and, subsequently, ΔG can be factorized in ΔH and TΔS (see Dullweber et al. [22]). The situation is very different for napsagatran. Here a proton is transferred and the present calculations suggest that the proton is picked up by His57 and not, as originally assumed by Dullweber et al. [22], by the carboxylic group of napsagatran. If a proton is transferred from the buffer, corrections for the heat of ionization of the buffer functional group has to be performed. But this correction covers only part of the superimposed heat contributions. Furthermore, the heat of ionization of the group picking up the proton has to be corrected. At this point, it is very important whether the proton goes - as originally assumed - to the carboxylic group of the inhibitor or - as the calculations suggest - to the nitrogen atom of His57. This is simply due to the fact that the heat of ionization of an oxygen functional group (e.g. COO^-/COOH $\approx$ 1-2 kJ/mol) [22] is much lower than that of nitrogen functional group (e.g. imidazole/imidazolium $\approx$ 30 kJ/mol)(http://www.lsbu.ac.uk/biology/enztech/ph.html). Since the binding of napsagatran involves the transfer of one proton, the difference in correction for either COO^- or His57 as acceptor amounts to about 28 kJ/mol, which would suggest very different factorizations of ΔG into ΔH and -TΔS. Deviating interpretations of the thermodynamic properties of napsagatran binding will be the consequence. Perhaps the last example emphasizes the utmost importance of pKa calculations and the correct assignment of protonation states for the interpretation of ITC data, in particular considering that this method become more and more popular in recent times.

SUPPLEMENTARY MATERIAL AVAILABLE

The adapted PEOE_PB charges will be made available in a future release of the open source project PDB2PQR [27].

3 PROTONATION CHANGES UPON LIGAND BINDING TO TRYPSIN AND THROMBIN: STRUCTURAL INTERPRETATION BASED ON PKA CALCULATIONS AND ITC EXPERIMENTS

3.1 INTRODUCTION

The protein-ligand binding process is governed by several types of interactions, such as hydrogen bonds, charge interactions, aromatic stacking, and hydrophobic complementarity. One feature which is often neglected for the sake of simplicity is the change of protonation of ligand and protein functional groups upon complex formation. Systematic studies considering such protonation effects have been rarely performed due to their experimental complexity. Isothermal titration calorimetry (ITC) represents one experimental method for the detection of overall changes in protonation. However, ITC reveals only the net stoichiometry for the exchange of protons and does not provide structural evidence which groups release or pick up the protons being transferred. More detailed insights on the molecular level can be gained computationally by calculating pKa values of functional groups and their changes upon complex formation. Calculations based on the Poisson-Boltzmann (PB) have achieved convincing agreement with experimental pKa values observed for protein residues [17, 18, 20, 21]. Other competitive approaches such as the protein-dipole Langevin-dipole method (PDLD) [10, 11, 31, 32], empirical models based on rules for hydrogen-bonding contributions and solvation effects [36, 35] or simplified electrostatic models using sigmoidally-screened Coulomb potentials [37] have also shown convincing agreement with experimental data. We have recently developed a generally applicable charge model to be used in PB-based pKa calculations of protein-ligand complexes [85]. The present contribution describes the application of this method on trypsin and thrombin complexes (Figure **REF**), which have been experimentally studied in our laboratory by ITC [22]. Evidence from the computational evaluations is used to interpret and factorize experimental microcalorimetric data recorded for a series of structurally related complexes.

Trypsin and thrombin belong to the family of trypsin-like serine proteases. They are well-established model systems to study fundamental concepts of protein-ligand interactions.

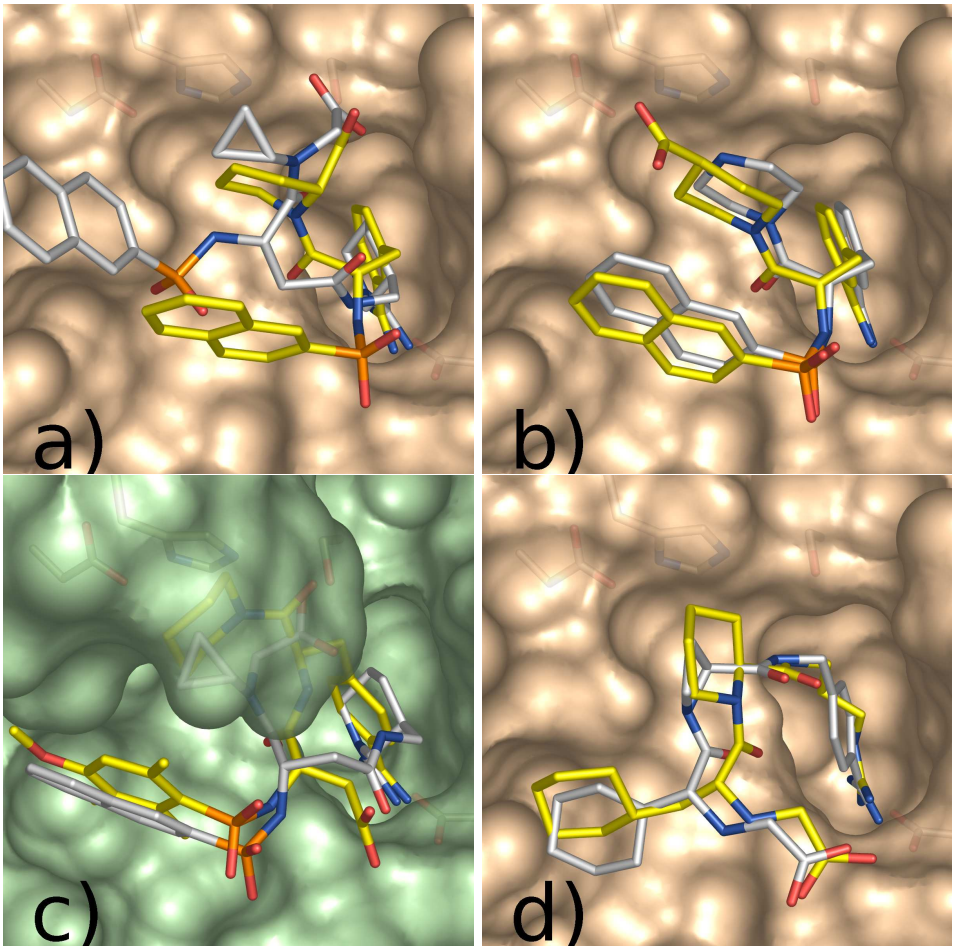

Figure 3.1: Trysin (colored pale yellow) and thrombin (colored green) in complex with the following ligands: a) **1b** (yellow) and **2** (white), b) **1c** (yellow) and **1d** (white), c) **2** (white) and **3** (yelow), d) **4** (yellow) and **5** (white). The catalytic triad and Asp189 of the S_1 pocket are shown in stick representation. The figure was generated with PyMOL [86]

Both enzymes catalyze the peptide-bond cleavage by means of their active-site residues His57, Asp102, and Ser195 which together compose the catalytic triad. Although the catalyzed reaction is identical, trypsin and thrombin are parts of different biological pathways: Trypsin is involved in digestion, whereas thrombin is engaged in the blood clotting cascade. Thrombin has been a target of pharmaceutical relevance for a long time, since its inhibition seems promising for the treatment of thrombosis via orally available drugs. However, such drugs should be selective for thrombin avoiding undesired crossreactivity with the structurally very similar trypsin. Therefore, the two proteins represent an ideal system to elucidate the selectivity determining features. In Figure 3.1, the key characteristics of the binding pocket of trypsin/thrombin are shown: The S_1 pocket can accommodate a positively charged functional group, normally forming a salt bridge to Asp189. Aromatic interactions are possible on the residues flanking the S_3 pocket. In thrombin, different to trypsin, the extended 60-loop is present and closes the S_2 pocket from the surrounding solvent environment. In both enzymes,

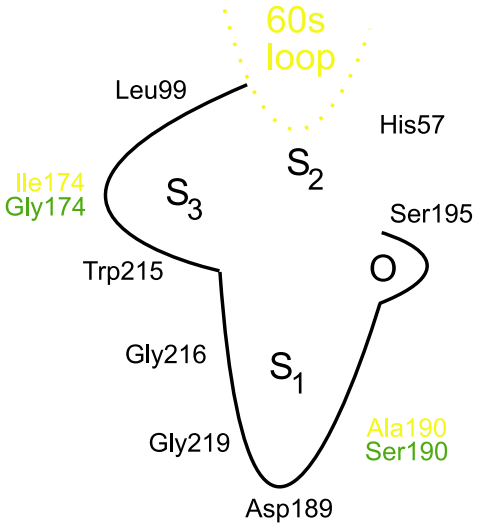

Figure 3.2: Schematic representation of the trypsin/thrombin active site. The following color code is used for the amino acids: green for trypsin, yellow for thrombin, grey for trypsin and thrombin. The 60s loop is only present in thrombin.

adjacent to the catalytic triad the oxy-anion hole is found.

One prerequisite for endowing ligands with enhanced target selectivity by rational concepts is the experimental factorization of affinity into entropic and enthalpic contributions, e.g. by means of isothermal titration calorimetry (ITC). In addition, this method provides indirectly the overall change in protonation when a ligand binds to a protein. In the present study, we show results from a computational study which help to understand and interpret unexpected protonation effects upon ligand binding to trypsin and thrombin. Such results are essential for correctly factorizing ITC data, because the heat of ionization required for this factorization depends on the titratable group. Only if this heat effect is appropriately accounted for, a meaningful partitioning in ΔH and -ΔTS can be performed.

In a previous communication, the experimental characterization of the thermodynamic and structural properties for a series of protein-ligand complexes of trypsin and thrombin has been described [22]. For one thrombin and three trypsin complexes, a change of protonation has been detected during the ligand binding process (Table 3.1). However, ITC and X-ray crystallography did not allow to unequivocally assign the proton donor or acceptor. This assignment, which is crucial for the thermodynamic interpretation, will be the main subject of this contribution.

Table 3.1: Net change of protonation (measured by ITC) upon ligand binding to trypsin or thrombin [22]: a positive sign corresponds to a proton uptake upon complexation, whereas a negative sign symbolizes a release of protons.

complex	Δn_{exp}
1b.Try	+0.90
1d.Try	-0.53
2.Try	+0.93
2.Thr	+0.88

3.2 RESULTS

The results of the pKa calculations will be discussed in light of the experimentally determined protonation changes Δn_{exp} (a positive sign corresponds to a proton uptake from the buffer upon complexation, whereas a negative sign symbolizes a release of protons); it is given in moles of protons per mole titratable group. Only pKa shifts between the apo and the complexed form which result in a protonation change will be described. Subsequently, the corresponding degrees of protonation and the resulting change in protonation .ncalc upon ligand binding will be portrayed based on the Henderson-Hasselbalch equation and a pH of 7.8 as used in all ITC experiments. Besides the overall Δn_{calc}, the protonation change $\Delta n_{calc/HL}$ (based solely on pKa shifts of His57 and the ligand) will be evaluated. All results are summarized in Table 3.2.

3.2.1 Estimating the accuracy of calculated protonation changes

Before describing any individual results, we would like to point out some general concepts of protonation states and their possible changes combined with an attempt to assess the significance and accuracy of protonation-change calculations.

In the context of the present study, we aim at the prediction of protonation changes occurring upon protein-ligand binding. Using a model screnario, we try to clarify what is a significant pKa shift with respect to protonation changes. As an example, consider a titratable group with a pKa of 7.0. In a solution of pH 7.0 (i.e, the same value as the pKa), the group is 50 protonated. If in this group a pKa shift of one unit is induced (e.g., by the ligand-binding process), the protonation changes by ± 41%, i.e. to 91% protonation if a shift of +1 pKa unit is induced or to 9% protonation if the pKa is shifted by -1. This essentially results in a fully protonated or deprotonated group, respectively. In contrast, if the same shift of one unit is taking place at a group with a pKa of 5.0, the effect is much

Table 3.2: Calculated pKa values for His57 and the ligand titratable groups and the corresponding protonation degrees at a pH of 7.8. Δn_{calc} corresponds to the protonation effect considering all pKa shifts observed in the 12 Å sphere; $\Delta n_{calc/HL}$ corresponds to the protonation effects considering the pKa shifts of His57 and the ligand; n.a. = not applicable.

					apo			complexed		
	compound	Δn_{exp}	Δn_{calc}	$\Delta n_{calc/HL}$	His57	ligCOO	ligAMINO	His57	ligCOO	ligAMINO
Trypsin	1b	0.9	0.51	0.49	0.43 (7.68)	0.0 (3.21)	n.a.	0.92 (8.84)	0.0 (2.69)	n.a.
	1bMe	0	-0.18	-0.14	0.20 (7.20)	n.a.	n.a.	0.06 (6.62)	n.a.	n.a.
	1c	0	-0.03	-0.01	0.20 (7.21)	0.0 (4.17)	n.a.	0.21 (7.23)	0.0 (3.69)	n.a.
	1cMe	0	-0.29	-0.21	0.30 (7.44)	n.a.	n.a.	0.09 (6.80)	n.a.	n.a.
	1d	-0.53	-0.14	-0.06	0.29 (7.42)	n.a.	0.33 (7.49)	0.00 (5.50)	n.a.	0.56 (7.90)
	1dAc	0	-0.21	-0.12	0.24 (7.29)	n.a.	n.a.	0.12 (6.92)	n.a.	n.a.
	2	0.93	0.57	0.54	0.46 (7.73)	0.00 (3.40)	n.a.	1.00 (10.92)	0.00 (3.30)	n.a.
	3	0	0.14	0.15	0.38 (7.59)	0.00 (3.84)	n.a.	0.53 (7.85)	0.00 (3.40)	n.a.
	4	0	0.34	0.36	0.38 (7.58)	0.00 (2.65)	0.32 (7.48)	0.11 (6.91)	0.00 (0.77)	0.95 (9.07)
	5	0	0.08	0.1	0.46 (7.73)	0.00 (2.51)	0.59 (7.95)	0.18 (7.13)	0.00 (0.46)	0.97 (9.28)
Thrombin	4	0	0.2	0.26	0.77 (8.32)	0.00 (2.65)	0.32 (7.48)	0.37 (7.56)	0.00 (-0.68)	0.98 (9.56)
	5	0	0.15	0.01	0.76 (8.31)	0.00 (2.51)	0.59 (7.95)	0.36 (7.55)	0.00 (-0.85)	1.00 (10.18)

less pronounced and virtually irrelevant for the protonation inventory (e.g., for pKa 5.0 → 6.0 at pH 7.0 the protonation degree rises from 1% to 8%, for pKa 5.0 → 4.0 at pH 7.0 it decreases from 1% to virtually 0%). Thus, the maximum effect on the protonation inventory is experienced at pKa values near the pH.

A pKa shift of one unit is close to the estimated minimum deviation of predicted pKa values from experimental data based on our calculation method (a root-mean-square deviation of 0.88 pKa units was obtained for a data set of 132 experimentally determined pKa values [85]). Therefore, with respect to estimates of protonation changes, caution is warranted if pKa shifts of less than one unit near the pH are observed. This also means that protonation changes of less than 40% (or 0.4 mole transferred protons) are of limited significance if they are based on a pKa shift of less than one unit (which is possible only at the critical pH).

In the results section, we evaluate the protonation change from the computed pKa shifts for the buffer pH of the ITC experiments. Rather small pKa shifts near this pH for complexes with no measured protonation change might lead to the impression that a proton transfer is falsely estimated. However, such small shifts must be considered in the context of the limited significance mentioned above. E.g., a protonation change of 0.33 mol based on a pKa shift from 7.4 to 6.8 at pH 7.0 is not sufficient to postulate a true proton transfer from the calculations.

Finally, also the experimental error of the ITC measurement should be mentioned, which is estimated to be ±0.12 mol [22]. Such a protonation change corresponds to a pKa shift of at least ±0.20 units (if occurring around the solution pH; larger at other pH values). This needs to be considered for the interpretation of the following results.

3.2.2 pKa Calculations on the complexes

For the binding of **1b** (Figure 3.3) to trypsin, ITC detected a proton uptake by the system of $\Delta n_{exp} = +0.90$. Visual inspection of the protein-ligand complex reveals either His57 or the carboxylate group of the ligand as likely candidates for picking up a proton. The distance between the N_{delta} of His57 and the Asp102 carboxy oxygen OD2 amounts to 2.7 Å, whereas $N_{epsilon}$ is 3.7 Å distant from the ligand's carboxy oxygen (**Figure 1a**). Furthermore, it must be noted that the ligand's carboxylic function is in hydrogen-bond distance (2.9 Å) to Ser195. In an earlier contribution, we assumed that the carboxy function of ligand **1b** would pick up the proton, simply because the corresponding methylester of **1b** did not show a similar protonation change [22]. The computed pKa values of complexed and uncomplexed trypsin show a significant shift only for His57. For the apo structure, a pKa value of 7.68

is calculated, whereas for the complex a pKa of 8.84 is suggested. Taking all apparent pKa shifts into account a net uptake of $\Delta n_{calc/all} = +0.51$ moles of protons (96 % of this proton uptake results from the His57 pKa shift) is computed. In contrast, the ligand's carboxylic group shows no pKa increase and, obviously, remains deprotonated.

In order to study whether the pKa calculations would suggest unchanged protonation states for the binding of the ester **1bMe** to trypsin a reasonable binding geometry of the ester had to be generated based on the crystal structure of the acid **1b**. A crystal structure could not be obtained. As both possibilities to attach the methyl group to the carboxylate oxygens of **1b** appeared feasible, we performed calculations on the basis of both binding orientations. However, the two possibilities did not show major differences in the resulting pKa values. The net change of protonation $\Delta n_{calc/all}$ is predicted to be -0.18.

Shifting the carboxylic group at the piperidine ring (as in **1b**) from the 2- to the 4- position gives **1c**. At first glance, this appears to be a minor variation. Both ligands **1b** and **1c** place their carboxylic functions in related regions of the binding pocket and, thus, similar protonation effects might be expected (**Figure 1b**). However, ITC revealed that no exchange of protons occurs upon complexation of **1c** (we will call this a zero effect). Based on the calculations, the His57 pKa in the uncomplexed form is 7.21 and remains at 7.23 in the complex. No further titratable group shows any significant pKa shift giving rise to a protonation change ($\Delta n_{calc/all}$ is -0.03).

Esterification of the ligand's 4-carboxylic function (**1cMe**) showed no protonation effect in the ITC experiment: Complexation of **1cMe** with trypsin results neither in proton uptake nor in proton release. The tendency for a *zero effect* is predicted correctly by our computations, although a net value $\Delta n_{call/all}$ of -0.29 is suggested.

Upon binding of trypsin to **1d**, the complex shows an overall release of Δn_{exp} = -0.53 moles of protons in the ITC experiment. Visual inspection of the complex geometry reveals a similar arrangement of the **1b.Try**, **1bMe.Try**, **1c.Try** and **1cMe.Try** complexes: The piperazine ring of **1d** occupies the same region of trypsin as in **1b**, **1bMe**, **1c** and **1cMe** (**Figure 1**). However, a basic instead of an acidic function is now accommodated in the S_2 pocket. Although the conserved binding mode might imply a very similar situation in the **1c.Try** and **1d.Try** complexes, the altered functional group has dramatic influence on the pKa calculation: It shows irregular (non-Henderson Hasselbalch/non-sigmoidal) titration curves for His57 and the ligand. Such curves have been observed experimentally [87] as well as theoretically (mostly in PB calculations) [43, 88, 89, 90]. Two requirements are necessary to produce irregular curves [91]: (1) The electrostatic interaction between the two

groups has to be strong, and (2) they must titrate in the same pH range. In the case of the **1d.Try** complex, the two requirements are fulfilled: The strong interaction is caused by close proximity of the titrating groups; their mutual distance measured between the titrating nitrogen atoms amounts to 4.1 Å. Furthermore, the two nitrogen functionalities titrate in the same pH range since they are both bases and experience similar intrinsic pKa values (the intrinsic pKa value of a titratable group corresponds to its pKa value when all other titratable groups are in the neutral state): the intrinsic pKas of His57 and the ligand are 5.38 and 7.07, respectively. In analogy to Trylska's treatment of the coupled system of the two aspartates in the catalytic dyad of HIV protease [52], the two titration curves (His57/ligand) are superimposed and the His57/ligand piperazine system is considered as one coupled base system. Evaluating this titration curve, the following pKa values can be extracted: His57 experiences a downward pKa shift from 7.42 to 5.59, whereas the pKa value of the ligand's secondary amino group changes from 7.49 (in aqueous solution) to 7.90 (in the complex). The resulting net protonation change (upon ligand binding) is estimated to be $\Delta n_{call/all} = -0.14$.

The experimentally observed deprotonation occurring upon **1d** complexation by trypsin can be prevented by acetylating the secondary nitrogen of the piperazine ring. This is also evidenced by ITC measurement of the corresponding **1dAc.Try** complex. Our pKa calculations yielded a $\Delta n_{call/all}$ of -0.21.

Ligand **2** differs structurally from the 1b-d series (**Figure 3**). The pKa calculation for the **2.Try** complex yields a protonation change of $\Delta n_{call/all} = +0.57$ (95 % of this Δn result from an upward pKa shift of His57). The ITC experiment showed a net proton uptake of $\Delta n_{exp} = +0.93$.

Inhibitor **3** also bears a carboxylic function, but it differs structurally from the **1b-d** series (**Figure 3**). However, as the crystal structure **3.Try** shows, its carboxylate is hydrogen-bonded to Gln174, which is more than 7 Å apart from Ser195 and more than 8 Å of from His57. Furthermore, the acidic group is strongly exposed to the adjacent solvent environment. ITC reveals no protonation effect upon binding of **3** to trypsin, the calculations give $\Delta n_{calc/all} = +0.14$.

The remaining two ligands for which pKa calculation were performed also differ structurally from the previous ligands. For both ligands, complex crystal stuctures with trypsin and thrombin have been determined. Interestingly, the ligands do not bear any aromatic side chain to address the S_3 pocket, and they bind with their guanidino (**4**) or benzamidino group (**5**) into the S_1 pocket (**Figure 1d**). The ligands have two titratable groups in common: a terminal carboxylic group and a secondary amine, separated by two bonds. In the crystal

structure, these groups are more than 9 Å remote from His57 and, thus, only marginal influence on its pKa value is expected. Indeed, rather small pKa shifts for His57 are observed in the pKa calculations (see Table 3.2). The ITC experiment showed no proton exchange when **4** or **5** bind to trypsin or thrombin. The following overall changes in protonation were obtained (the values for thrombin are given in parenthesis): $\Delta n_{calc/all}$ is +0.34 for **4.Try** (+0.20 for **4.Thr**), while for **5.Try** $\Delta n_{calc/all}$ is 0.08 (+0.15 for **5.Thr**).

Besides the above-described pKa shifts, there are further titratable groups that experience changes in pKa values upon complex formation in all considered complexes. For example, Asp189 in the S_1 pocket of trypsin shows a downward pKa shift (mean pKa value for apo trypsin: 4.3, mean pKa value for complexed trypsin: 2.6). Tyr228, also contributing to the S_1 pocket, is involved in another downward pKa shift: it changes from 12.5 (mean pKa value in apo trypsin) to 11.5 (mean pKa value in complexed trypsin). However, these shifts do not provoke any change in overall protonation. Therefore, they can be neglected for the interpretation of the observed net release or uptake of protons.

3.3 DISCUSSION

Our pKa calculations evidence that the protonated His57 and the deprotonated ligand **1b** form a salt bridge-like electrostatic interaction. At first glance, this suggests that the placement of a charged carboxylate group next to the catalytic triad would generally induce a protonation change of His57. However, as **1c.Try** shows, this is not the case. Closer inspection of the **1c.Try** complex shows that the ligand's carboxy oxygen (O55) is 5.1 Å apart from N_ϵ of His57, which corresponds to an expansion of 1.4 Å compared to **1b.Try**. In addition, the distance between Ser195 oxygen O_γ and the ligand's oxygen (O55) is increased from 2.9 to 7.5 Å. Most likely, these expanded distances alter the electrostatic properties of the complex (**1c.Try**) significantly. Accordingly, the pKa calculation suggests no change in protonation for His57 which is in agreement with experiment. Purely based on structural terms (without any pKa calculation), it would be difficult to predict such a *zero effect* for the latter complex.

For the **1d.Try** complex, the results deviate from the experimental finding. Originally, it was assumed that a downward pKa shift would only occur at the ligand's amino group upon trypsin binding [22]. The pKa calculation suggests His57 to undergo a significant downward pKa shift, along with a rather marginal upward pKa shift for the ligand. Although the ligand's pKa value changes by less than 0.5 log units, it almost outweights the effect on His57. This example shows that a small pKa shift in the physiological pH range may result

in a very complex picture of mutually compensating effects which prevent a straight-forward interpretation of experimentally observed protonation changes. The calculations suggest a correct trend for the net change in protonation of **1b.Try**, the quantitative agreement, however, is unsatisfactory. Nevertheless, assessing the relevance of such calculations, one has to bear in mind the range of experimental error and the reduced accuracy in pKa estimations that result for the present example from coupled titration curves. These are much more difficult to handle and allow only for crude pKa predictions.

The positions of the carboxylate of ligand **2** in trypsin and thrombin are similar. Both complexes show almost equal protonation effects in the ITC experiment (**2.Thr** $\Delta n_{calc/all} = +0.88$; **2.Try** $\Delta n_{calc/all} = +0.93$). For both ligands, the esterified form has also been complexed with trypsin and thrombin and measured by ITC, and it was observed that a zero effect results. Accordingly, it was originally assumed that the pKa shift in **2.Thr** and **2.Try** occurs at the carboxylic group of the ligand [22]. In a previous computational study [85], we could correctly predict the effect for **2.Thr** using the available crystal structure. For **2.Try**, only a docked structure served as starting geometry for the pKa calculation since no corresponding crystal structure is available. Possibly, this explains the somewhat less satisfactory quantitative agreement of experiment and computer prediction for **2.Try** ($\Delta\Delta n = 0.36$) compared to **2.Thr** ($\Delta\Delta n = 0.28$).

The computer simulations suggest that the two ligand titratable groups of **4** and **5** tend to adopt a zwitterionic state when complexed to the protein. The amino group and the carboxylate form a strongly coupled electrostatic system over short distance which induces a downward pKa shift for the carboxylate and an upward shift for the amino group. While the downward pKa shift does not alter the protonation state for the carboxylic group, the upward pKa shift of the amino function suggests a protonation change from the neutral to the positively charged state. Rather small pKa shifts of the amino function are suffcient to provoke this net change; e.g. for **4** ($pKa_{(aq)} = 7.48$), a pKa shift of +0.5 units induces a protonation change of +28 % (-0.5 units: -19 %) referring to a buffered solution of pH 7.8. The net protonation inventory resulting from the pKa shift of the amino groups in **4** and **5** might be partially compensated by a downward pKa shift of His57, but still a value deviating from zero results as Δn_{calc}. Interestingly, if one of the two ligand titratable groups is considered as permanently charged and, thus, not involved in protonation changes, the measured zero effect is much better predicted by the simulations. This may indicate that the site-site interaction between the ligand titratable groups is overestimated in the pKa calculation and prevents quantitative agreement. Nevertheless, for three of the four complexes the zero effect is correctly indicated if we consider all Δn_{calc} values smaller than 0.2 as negligible (**cf. Section 2.1 and 4**).

3.3.1 Influence of conformational variations

To assess the influence of slightly different conformations on the computed pKa values, we used the crystallographically determined apo structure of trypsin (1TGN) along with nine artificially generated uncomplexed structures ("*ligand-deleted* trypsin", obtained by removing the bound ligand) to determine the pKa values of uncomplexed trypsin. The pKa of His57 in 1TGN is 7.88. Analyzing all calculations on the nine *ligand-deleted* trypsin structures (cf. Table 3.2) unveils a slight variation of the pKa for the crucial residues His57 (7.49 $\pm$0.19) and Asp102 (1.99 $\pm$ 0.36). This suggests that the uncertainties in our calculations will be at least in the range of the estimated experimental error of ITC measurements and provides an additional approximation about the significance of the computed pKa shifts. The strategy of using the *ligand-deleted* trypsin structures appeared justified, as Guvench et al. [92] had previously shown by MD simulations that a high structural conservation of the trypsin binding pocket is given. The overall RMSD based on a C_α-fit between 1TGN and our trypsin complexes ranges between 0.28 and 0.42 Å, which illustrates the low degree of flexibility of the trypsin structures in response to ligand binding.

3.3.2 pKa values of histidine residues

Finally, it may be instructive to examine the spread of experimentally observed pKa shifts of histidine residues in proteins as an example to estimate the impact of the protein environment on pKa values. While in our calculations for the different complexes the pKa of His57 varies over a range of approximately 5 units, the experimental pKa values of histidines in various protein structures are reported to be distributed over more than seven pH units [93]. His149 in Bacillus circulans shows a value below 2.3 [94]. In contrast, the pKa values of His66 and His72 in a protein tyrosine phosphatase are 8.3 and 9.2, respectively [4]. The pKa value of His31 in T4 lysozyme is 9.1, caused by the formation of a salt bridge to Asp70 (pKa = 0.5) [95]. In a recently published data base [96], more than 300 pKa values for histidine residues are available: The mean value is 6.33 (standard deviation of 1.35); 35 deposited histidine residues exhibit a pKa value greater than 7.5. The experimental pKa for His57 in rat trypsin is 6.7 (NMR- measurement for a D189S mutant) [97], whereas the pKa value of the imidazole moiety of the free amino acid in water has been determined to 6.3 [98].

3.4 Conclusions

In this study, systematic investigations about protonation changes in due course of ligand binding to trypsin and thrombin have been performed. The trend for the nine complexes which show a *zero effect* in ITC could be predicted correctly. This conclusion is based on the following trends in the calculations of $\Delta n_{calc/all}$: -0.18 (**1bMe.Try**), -0.03 (**1c.Try**), -0.29 (**1cMe.Try**), -0.21 (**1dAc.Try**), +0.14 (**3.Try**), +0.34 (**4.Try**), +0.08 (**5.Try**), +0.20 (**4.Thr**), +0.15 (**5.Thr**). On average, a net change of ± 0.18 moles of protons (with a standard deviation of 0.09) is suggested. This has to be faced with the experimental accuracy which is estimated to be ± 0.12 [22]. Larger deviations are observed for those complexes where most likely a coupled system of titratable groups exists. For the trypsin and thrombin complexes where ITC suggests a net change in the protonation inventory also the calculations indicate the correct trends. For one complex (**1d.Try**) only the correct tendency is suggested, again obviously a case with strongly coupled titratable groups. Such groups deviate from simple Henderson-Hasselbalch behaviour and complicate straightforward handling.

The present calculations allow structural interpretation of protonation states in conjunction with ITC measurements. ITC experiments give access to the Gibbs free energy ΔG along with the heat absorbed or produced during complex formation. This heat contains the binding enthalpy, but changes in protonation states may be superimposed. This is due to the fact that any release or pick-up of protons involves a heat of ionization of the functional groups, which concerns not only the functional groups of the buffer molecules, but also the groups of the protein residues or the bound ligand that change their protonation state. In this context, it matters dramatically whether the groups involved in the protonation reaction exhibit oxygen or nitrogen functionalities.

In our previous communication [22] we falsely assumed that the carboxylate function of **1b** and **2** would undergo protonation upon trypsin and thrombin binding. This interpretation was born out of the simple conclusion that the corresponding esters of **1b** and **2** did not show any protonation changes. Accordingly, we corrected for a heat of ionization contribution of 2.26 or 0.84 kJ/mol [22], resp., for the acidic oxygen functions. After substraction of this contribution from the measured heat, a net ΔH_{bind} of -45.6 kJ/mol for **1b.Try** was assumed. The measured ΔG_{bind} was then factorized into enthalpy and entropy. The calculations presented in this study suggest that the carboxylic function of the ligand remains deprotonated, whereas His57 picks up a proton. This now involves ionization of a nitrogen functionality which produces a much higher heat contribution (ΔH_{ion} of His: 30 kJ/mol; taken from http://www.lsbu.ac.uk/biology/enztech/ph.html). Accordingly, ionization of His57 upon complex formation of **1b** with trypsin requires a correction of 27 kJ/mol

(0.90×30 kJ/mol). After substraction of this value from the measured heat, a quite different value for $\Delta H_{bind} = -20.6$ kJ/mol remains. In consequence, this suggests a strongly deviating factorization in enthalpic and entropic contributions. For the complexes of **2**, similar corrections have to be performed (Table 3.3). Accordingly, the interpretation of the driving forces responsible for binding of **1b** and **2** to the considered serine proteases has to be modified (Table 3.3). Binding is now characterized by a significantly smaller enthalpic contribution and a more favorable entropic contribution. The originally suggested pronounced enthalpic binding of **2** to thrombin is now reduced and thus better comparable with data of the other ligands in the series. Furthermore, the formerly contrasting partitioning for the complexes of acid **1b** and ester **1bMe** with trypsin is now suggested to be of equal size for both inhibitors. Even though the stronger binding of napsagatran (**2**) to thrombin results (still) from a stronger exothermic binding compared to the ester, the entropic advantage of the ester compared to the acid is much less pronounced according to the new partitioning.

For **1d.Try** a more complex protonation transfer is suggested by the computer simulations. However, with respect to the partitioning of enthalpy and entropy the assignment of heat of ionization corrections remains virtually unchanged because the amount to be corrected for the histidine residue is very similar to that of the ligand's piperazine moiety. Accordingly, no significant changes with respect to our previous interpretation and factorization for **1d.Try** are required.

In summary, protonation changes can lead to very different factorization into enthalpy and entropy depending on the involved ionizable groups. The pKa calculations presented in this communication can support the thermodynamic characterization and interpretation of the protein-ligand binding process, if protonation effects are superimposed. The experimental efforts to fully characterize protonation changes are quite elaborate and consume large amounts of expensive protein material. In contrast, the presented calculations are cheap and fast to perform. As such they can provide valuable assistance in properly planning and analyzing ITC experiments.

3.5 Materials and Methods

The investigated ligands are shown in **Figure 3**. Most of them contain three titratable groups: a guanidino group, a sulfonamide or secondary amino group, and a carboxylic acid group. For all ligands, experimental pKa values have been determined [22]. In the present pKa calculations, only the carboxylic acid and secondary amino group were considered as titratable, since their pKa values fall into the physiological pH range. A 3D representation of

Table 3.3: Corrected factorization of enthalpic and entropic binding contributions from ITC experiments. All energy values are given in kJ/mol.

complex	Δn_{obs}	Δ G	ΔH_x	ΔH^{COO}	ΔH^{His}	ΔH^{old}_{bind}	ΔH^{new}_{bind}	$-T\Delta S^{old}$	$-T\Delta S^{new}$
1b	0.9 ±0.6	-36.4	-47.6 ±1.9	2	30	-45.6	-20.6	9.2	-15.8
1bMe.Try	0	-37.0	-16.9 ±0.8	not necessary	not necessary	-16.9	-16.9	-20.6	-20.6
2.Try	0.93 ±0.12	-32.0	-32.3 ±3.6	0.8	30	-31.5	-4.4	-0.5	-27.6
2.Thr	0.88 ±0.08	-52.2	-75.1 ±1.8	0.8	30	-74.4	-48.7	22.2	-3.5
2Et.Thr	0	-40.3	-24.4 ±0.9	not necessary	not necessary	-24.4	-24.4	-15.9	-15.9

the complexes is shown in **Figure 1**. The following complex crystal structures (referred to by their PDB code) were used as starting geometries: 1K1I (**1b.Try**), 1K1J (**1cMe.Try**), 1K1L (**1d.Try**), 1K1M (**1dAc.Try**), 1K1N (**3.Try**), 1K1O (**4.Try**), 1K1P (**5.Try**); for thrombin, the complexes 1K21 (**4.Thr**) and 1K22 (**5.Thr**) were evaluated. A geometry of the trypsin complex with **1c** was generated by replacing the methylester of **1cMe.Try** with a carboxylic acid group. Subsequently, the modelled complex was energy minimized with the program MOLOC using the MAB force field [99]. The geometry of the complex **1bMe.Try** was generated similarly in reverse fashion by attaching a methyl group to either oxygens of the carboxylic function of **1b**; thus, both possibilities for ester formation have been explored. A 3D-structure of the **2.Try** complex was obtained by docking using GOLD [100]. 1K1P was chosen as template structure, and 20 docking poses were generated and evaluated. The rank 3 solution was selected for the pKa calculation based on its convincing agreement with the thrombin/napsagatran crystal structure [22].

For the pKa calculations, the same methodology as recently described by us was applied [85]. To study the protonation effects upon ligand binding, two pKa calculations are required: One without ligand and a second with the bound ligand. MEAD was employed as PB solver (using a dielectric constant of 20 for the protein) [64], REDUCE was used to generate hydrogens for the residues of the protein [75]. SYBYL was applied to add hydrogens to the ligands [101]. Only the titratable residues within a 12 Å sphere around the active site were selected for evaluation of the site-site interactions in order to reduce the computational effort (in total, there are 39 titratable residues in trypsin and 81 in thrombin). This results in the following titratable residues for trypsin: Tyr39, His40, His57, Tyr59, Lys60, Tyr94, Asp102, Lys145, Tyr151, Tyr172, Asp189, Asp194, Lys224, Tyr228. In the case of thrombin, the following titratable residues lie within the sphere: Glu18, Glu39, Asp56E, His57, Tyr60A, Lys60F, Tyr94, Glu97A, Asp100, Asp102, Lys145, Glu146, Lys169, Tyr184A, Asp189, Glu192, Asp194, Glu217, Asp221, Asp222, Lys224, Tyr225, Tyr228.

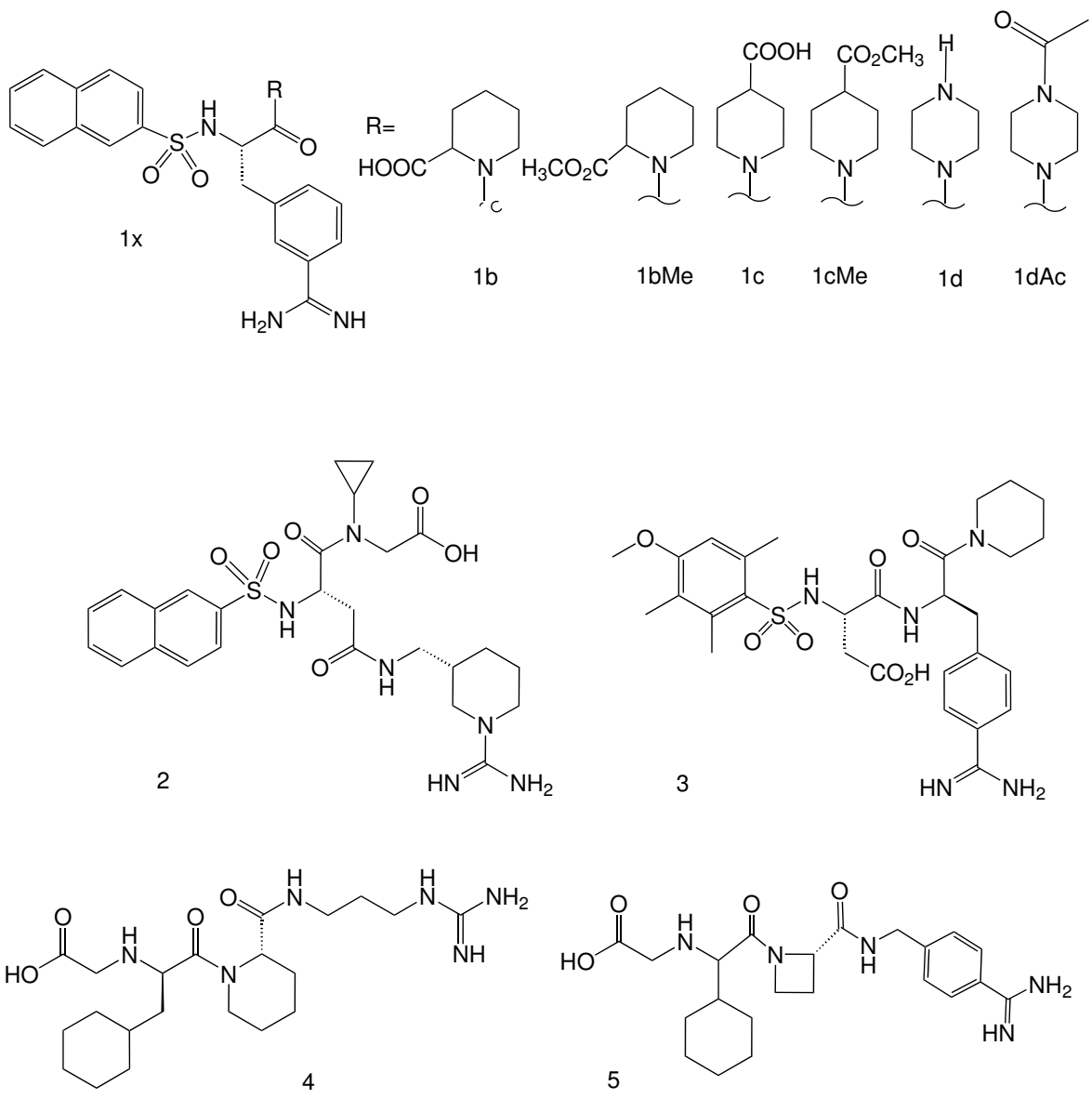

Figure 3.3: Chemical formulae of the ligands studied by Dullweber et al. [22] and used in this investigation; for clarity the original numbering used from the previous contribution is maintained without changes.

4 ATYPICAL PROTONATION STATES IN THE ACTIVE SITE OF HIV-1 PROTEASE: A COMPUTATIONAL STUDY

4.1 INTRODUCTION

The **A**cquired **I**mmuno**D**eficiency **S**yndrome (AIDS) is a disease affecting the human immune system first characterized about 25 years ago. The causative agent of this disease is the retrovirus named **H**uman **I**mmunodeficiency **V**irus (HIV) [102], which appears in two forms, HIV-1 and HIV-2. The best-studied drug target of the virus is the HIV-1 protease (HIVP). Currently, there are eight approved drugs on the market that aim at the inhibition of HIVP (http://www.hiv.net). Besides its significant pharmaceutical relevance, HIVP is a prominent example for atypical protonation states in the active site [25, 24, 103, 26]. In this context, atypical protonation means a non-standard protonation, i.e. a protonated acid or a deprotonated base. We therefore selected this target to study the predictive power of Poisson-Boltzmann-based pKa calculations using our newly developed PEOE_PB partial charge model [85].

HIVP forms a symmetric homo-dimer, with each chain containing 99 residues. The central core of the dimer forms the active site (see Figure 4.1). The reaction catalyzed by HIVP is a peptide-bond cleavage, usually between a Tyr and a Pro or between a Phe and a Pro. The binding pocket is composed of the catalytic triad Asp-Thr-Gly. The two aspartates (one from each domain of the homo-dimer) are of main relevance, and are therefore named the *catalytic dyad*. Besides the central cavity which is formed by the active site, three specificity pockets are accessible (S_1, S_2, and S_3; and due to symmetry, S_1', S_2' and S_3'). These specificity pockets preferably accommodate the hydrophobic parts of the ligand, which are named accordingly P_1, P_2, and P_3 (and P_1', P_2', and P_3', respectively).

Two design strategies have been followed in the development process of HIVP inhibitors: in the first generation, it was attempted to mimic the transition state by peptidomimetics mostly with one central hydroxylic group donating a hydrogen bond to the catalytic dyad. Inhibitors of the second generation contain a carbonyl group mostly embedded into a central ring moiety to displace the "flap water". The latter is a conserved water molecule hydrogen-bonding to the flap region via Ile50; the flap region of HIVP is the highly flexible part of the protein. This flap water is a special structural element unique for this viral protease. We

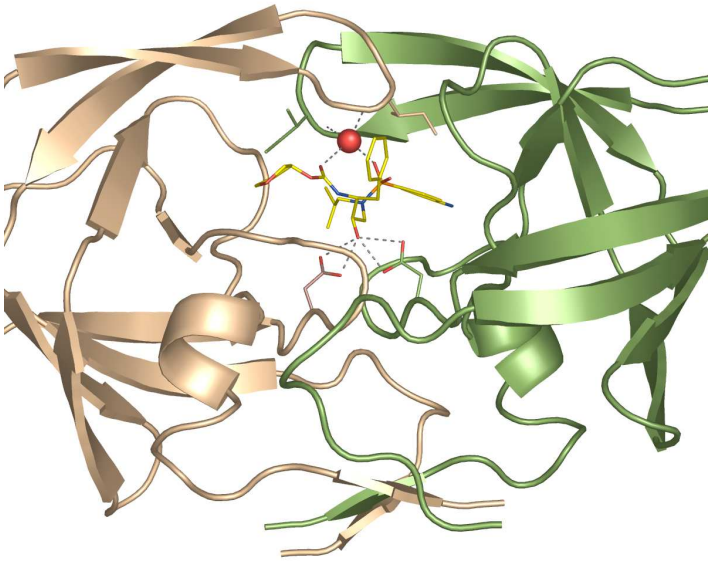

Figure 4.1: The complex formed by HIVP with the inhibitor **VX-478** (**Amprenavir**, see Figure 4.2): the ligand, the catalytic dyad and Ile50 of the flap region are shown in stick representation. The flap water, hydrogen-bonding to the backbone of Ile50, is given as a red sphere.

will perform systematic studies on the inclusion or exclusion of this water molecule to study the relevance of explicitly considered water molecules for future pKa calculations.

The five inhibitors that were considered in our study are shown in Figure 4.2. **KNI-272**, **VX-478** and **S6** represent members of the first generation of HIVP inhibitors, while **DMP-323** and **S7** belong to the second generation. The ligand **DMP-323** induces a protonation change of the catalytic dyad: from the apo state with one of the two aspartates protonated [25, 24] to a double-protonated state with both aspartates uncharged [103]. Thus, obviously HIVP displays already as apo enzyme an atypical protonation state for the catalytic dyad, which can be modulated upon ligand binding. The correct prediction of such effects displays the main focus of our computational study. We start with calculations for the uncomplexed and complexed HIVP where experimental data on the adopted protonation states are available. This will give insight into the power of our method to correctly predict protonation states and will serve as a benchmark study. Furthermore, we will face our results to a former computational pKa study on HIVP complexes [52]. Finally, we will predict the protonation states for HIVP complexes where up to now no experimental data on the protonation states of the catalytic dyad have been recorded.

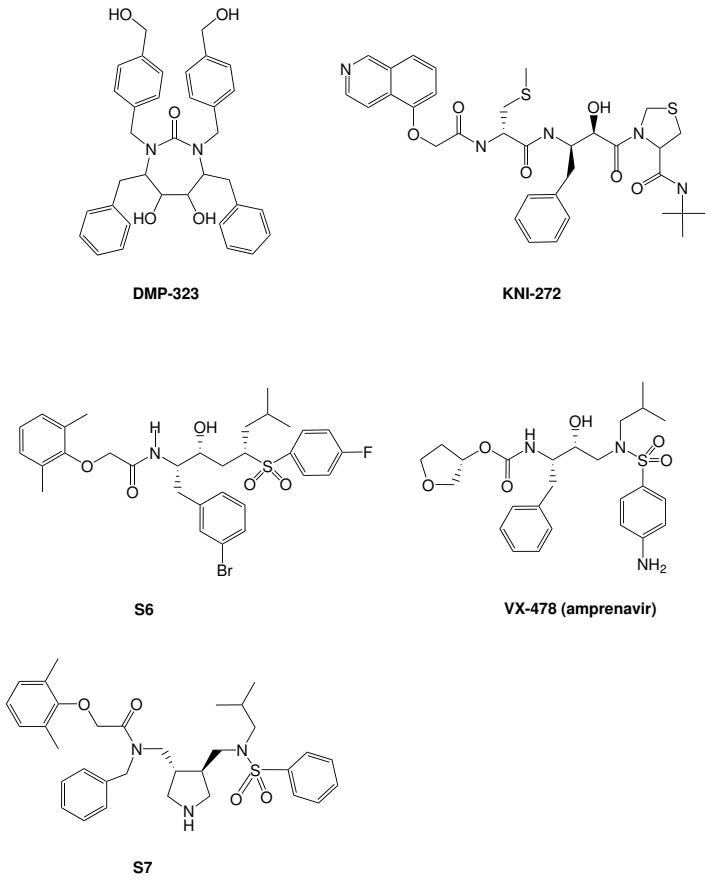

Figure 4.2: Inhibitors of HIVP used for the pKa study: **DMP-323**, **KNI-272**, **S6**, **S7** and **VX-478**.

4.1.1 Setting of the dielectric constant and handling of coupled systems

Prior to the discussion of possibly limiting factors, we will start with a brief introduction to PB-based pKa calculations. Three different energy contributions to the ΔG of a protonation reaction are determinant for the final pKa value and provoke the shift from the model pKa value, pK_a^{mod}, (corresponding to the pKa value of the isolated residue in aqueous solution) to the pKa value in the actual protein environment: the solvation energy, the background interaction energy, and the site-site interaction energy. The solvation energy considers the penalty from the desolvation; the interaction with the permanently present dipoles and charges is represented by the background interaction; the interaction of the titratable residues with each other is covered by the site-site interaction energy term. The first two quantities represent the *intrinsic pKa value* (pK_a^{int}): although they are artificial (all titratable residues besides the one of interest are in the neutral state) and cannot be experimentally detected, they possibly indicate the predominance of one contribution. The consideration of the site-site interactions reveals the final pKa value in the protein environment.

One central point of discussion with respect to pKa calculations is the choice of the protein dielectric constant ($\epsilon_{Protein}$). The interested reader is referred to Warshel's excellent review on this topic [34]. In previous studies, it was shown that the best overall agreement (with the experimentally determined pKa values) can be achieved setting $\epsilon_{Protein}$ to 20 [18, 85]. However, the analyses primarily involved solvent-exposed residues which show only minor pKa shifts. For active-site residues experiencing large pKa shifts, the assignment of $\epsilon_{Protein} = 4$ generates better results. This could be shown for lysozyme (Glu35: pKa^{exp}=6.2) [18, 85] and xylanase (Glu172: pKa^{exp}=6.7) [85]. Competitive studies illustrate a broad range of applied dielectric constants: Demchuk and Wade used in their PB-based pKa calculations $\epsilon_{Protein}$ of 80 for solvent-exposed residues and 15 for desolvated sites [19]. Nielsen et al. used a single value of 8 for all residues [21, 80]. King et al. found in microscopic MD simulation studies that the (macroscopic) dielectric constant could be as high as 10 in sites of catalytic importance [104]. This shows that there is no *universal* value for $\epsilon_{Protein}$. In order to validate our pKa calculation methodology, and having experimental data for some of the studied complexes as a reference, we performed our pKa calculations at three different values of $\epsilon_{Protein}$: 4, 10 and 20. These values cover a large range for the dielectric constant of the protein, and we hope in light of the experimental reference points to draw some conclusion concerning the preferred setting of $\epsilon_{Protein}$.

A second issue of pKa calculations on HIVP structures concerns electrostatically strongly coupled systems. The two aspartates of the catalytic dyad undergo a strong mutual interaction on short distance resulting in such a system. This was previously reported in the pKa study on HIVP of Trylska et al. [52] and can also be observed in our calculations. The two aspartates approach each other in space by approximately 3 Å. This proximity leads to the formation of a coupled system and results in a significant non-sigmoidal behaviour (see Figure 4.3, thin lines). With respect to the computational setup, in the case of strongly coupled systems, it is necessary to perform a post-processing step after the actual pKa calculation. This was done in analogy to Trylska et al. [52], who superimposed the single titration curves of Asp25 and Asp25'. This gives a titration curve similar to that of a diprotic acid, e.g., maleic acid. The pKa values for the two aspartates correspond then to the situation where $\Sigma H^+ = 1.5$ and 0.5, corresponding to 25 or 75 % protonation of the two coupled aspartates (see Figure 4.3, thick line and horizontal lines).

4.2 MATERIALS AND METHODS

MEAD was employed as PB solver [105]. The settings of the pKa calculations and computational details to our developed PEOE_PB partial charges are given elsewhere [85]. PDB2PQR

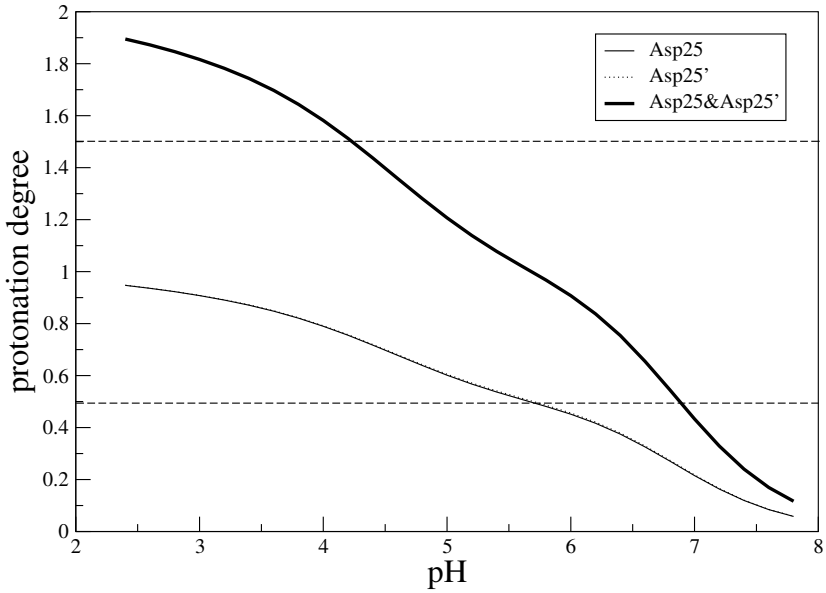

Figure 4.3: Titration curves of apo HIVP (crystal structure 1HHP, $\epsilon_{Protein} = 10$). The individual titration curves of the virtually independent aspartates are shown as thin (solid and dashed) lines, both curves are almost indistinguishable. The superimposed titration curve of the coupled diprotic acid system is given as thick line. The dashed straight lines at 1.5 and 0.5 correspond to situations of the coupled system where either 25 % or 75 % of both aspartates are protonated (corresponding pKa values are here 4.22 and 6.89).

was used to generate hydrogens for the protein residues [27]. A direct comparison to our former model [85] for the simulation of the uncharged form of aspartate and glutamate in which implicit hydrogens were placed to model the neutral state showed an improvement when using explicit hydrogens instead. Therefore, we placed explicit hydrogens on the OD2 (Asp) and OE2 (Glu) atom. SYBYL was applied to add hydrogens to the ligands [106]. Manual adjustment of the considered local conformation was done for the central hydroxylic groups of the ligand being in close contact to the catalytic dyad (named "rot-scan"). Where available, experimental knowledge about the formed hydrogen bonds was used: this is the case for the symmetric cyclic urea compound **DMP-323** [103] and the asymmetric peptidomimetic **KNI-272** [26]. In case pKa calculations were performed considering explicit water molecules, the orientation of the hydrogen atoms was manually adjusted.

To study the protonation effects upon ligand binding, two pKa calculations are required: one without the bound ligand and a second with the bound ligand. The net protonation effect is the resulting difference between the degrees of protonation between the two forms. The protonation states are calculated at a pH of 5.0, as HIVP optimally catalyzes the peptide bond cleavage within a pH range of 4.0 and 6.0 [24]. The notation H^+ is used for the protonation

degree of one particular residue, i.e. $H^+ = 1.0$ corresponds to full protonation and $H^+ = 0.0$ to full deprotonation. The notation ΣH^+ describes the sum of the protonation degrees for the complexed and uncomplexed form of the protein-ligand system. The following crystal structures (given by their PDB codes) for complexed and uncomplexed HIVP structures were used: 1HHP (apo), 3HVP (apo), 1HPX (**KNI-272**), 1QBS (**DMP-323**), 1XL5 (**S6**), 1XL2 (**S7**), 1HPV (**VX-478**). The ligands are given in Figure 4.2.

The structural differences between the two apo structures are rather marginal: the C_α fit gives an RMSD value of 0.40 Å, and both crystal structure show the same space group ($P4_12_12$). Their resolution is 2.7 Å (1HHP) and 2.8 Å (3HVP). A comparison of the apo structures and the complexes reveals C_α-RMSD values between 0.96 and 1.35 Å. The complexes are structurally more similar among each other: the RMSD lies betwen 0.47 and 0.77 Å. The resolution of the complexes is higher compared to the apo structures and ranges between 1.5 and 2.0 Å. The complexes appear in different space groups: $P2_12_12_1$ (1XL2), $P6_1$ (1QBS, 1HPV), $P2_12_12$ (1XL5).

We calculated pKa values for the following titratable residues (predefined model pKa values are given in parenthesis): aspartate (4.0), glutamate (4.4), histidine (6.3), cysteine (8.7), tyrosine (9.6), lysine (10.4). The pKa values of the C- and N-termini were not considered, since they are 18 Å apart from the catalytic dyad and assumed to remain in their standard protonation state upon complexation. With respect to the ligands, only **S7** contains a titratable group, which is a pyrrolidine ring.

In Table 4.1, the slight differences in the primary sequences of the evaluated HIV-1 proteases are listed. In the case of the 3HVP structure, all mutations to natural amino acids

Table 4.1: Comparison of HIVP sequences (Aba: α-amino-N-butyric acid; Cso: S-hydroxycysteine).

Position	1HHP/1HPX	3HVP	1QBS	1XL2/1XL5	1HPV
3	Ile	Ile	Val	Ile	Ile
14	Lys	Arg	Lys	Lys	Lys
37	Ser	Asn	Ser	Ser	Ser
41	Arg	Lys	Arg	Arg	Arg
63	Leu	Pro	Leu	Leu	Leu
64	Ile	Val	Ile	Ile	Ile
67	Cys	Aba	Cso	Cys	Cys
95	Cys	Aba	Ala	Cys	Cys

regard solvent-exposed residues and occur at least 16 Å from the catalytic dyad. Mutations to non-natural amino acids such as Aba and Cso (3HVP, 1QBS) are observed at least 10

Å apart from the catalytic center and thus, the influence on the catalytic dyad should be marginal. These residues (Aba, Cso) were mutated to cysteine for the pKa calculations.

4.3 RESULTS

4.3.1 Calculations on two apo structures

In Table 4.2, the experimentally determined pKa values (determined via kinetic measurements) for apo HIVP are listed, which suggest mono-protonation for the catalytic dyad at a pH of 5.0. Whereas the experimental range for the first titration step of the coupled aspar-

Table 4.2: Experimentally determined pKa values for the two catalytic aspartates in apo HIVP. The two ranges from reference [25] result from measurements on different substrates.

	pK_{a1} *(aspartate$_1$)*	pK_{a2} *(aspartate$_2$)*
reference [25]	3.4-3.7	5.5-6.5
reference [25]	3.1-3.3	4.9-5.3
reference [24]	3.3	6.8

tates falls into a range between 3.1 and 3.7 the second step has been determined to occur between 4.9 and 6.8. The spread of nearly two log units makes a comparison to our computed results difficult. However, the protonation state at pH of 5 appears well-characerized and suggests mono-protonation of the catalytic dyad. The results from NMR experiments [107] are not listed, since they only provide a range (< 5.9) for both pKa values.

The results of our pKa calculations taking the two different apo protein structures (1HHP, 3HVP) as starting geometries are given in Table 4.3 and 4.4. They show that the computed

Table 4.3: pKa calculations on apo HIVP (1HHP).

						(Trylska et al. [52])				
$\epsilon_{Protein}$	pK_{a1} *(aspartate$_1$)*	$H^{+\,a}$	pK_{a2} *(aspartate$_2$)*	$H^{+\,a}$	$\Sigma H^{+\,b}$	pK_{a1} *(aspartate$_1$)*	$H^{+\,a}$	pK_{a2} *(aspartate$_2$)*	$H^{+\,a}$	$\Sigma H^{+\,b}$
4	6.06	0.92	9.90	1.00	1.92	3.4	0.02	6.9	0.99	1.01
10	4.37	0.19	6.97	0.98	1.17					
20	3.58	0.04	5.93	0.89	0.93	3.1	0.01	5.3	0.67	0.68

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5

Table 4.4: pKa calculations on apo HIVP (3HVP).

						(Trylska et al. [52])				
$\epsilon_{Protein}$	pK_{a1}	$H^{+\ a}$	pK_{a2}	$H^{+\ a}$	$\Sigma H^{+\ b}$	pK_{a1}	$H^{+\ a}$	pK_{a2}	$H^{+\ a}$	$\Sigma H^{+\ b}$
	$(aspartate_1)$		$(aspartate_2)$			$(aspartate_1)$		$(aspartate_2)$		
4	4.83	0.40	9.69	1.00	1.40	2.1	0.00	6.7	0.98	0.98
10	3.17	0.01	6.62	0.98	0.99					
20	2.65	0.00	5.44	0.73	0.73	2.9	0.01	5.3	0.67	0.68

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5

results are in better agreement with the experimental values once the higher $\epsilon_{Protein}$ (20) is chosen. Decreasing the value of $\epsilon_{Protein}$ results in increasing pKa differences between the two acids; in all cases, the difference is at least two log units. Nonetheless, the net protonation state of the two catalytic aspartates can be correctly predicted by five of the six pKa calculations shown in Table 4.3 and 4.4. Thus, both starting geometries seem appropriate over a wide range of $\epsilon_{Protein}$ to correctly predict the mono-protonated state of the catalytic dyad at a pH of 5.0.

From a structural point of view, both apo structures are very similar: a C_α-fit reveals a root mean square deviation (RMSD) of 0.4 Å. Similarily, the catalytic triad is structurally conserved: the RMSD values amount to 0.5 Å (Asp25), 0.06 Å (Thr26), 0.59 Å (Gly27) after the C_α-fit for all C_α atoms. The distances between the oxygens of the catalytic dyad of the different apo structures vary by about 0.2 Å.

Similar PB-based pKa-calculations were performed for several complexed and uncomplexed HIVP structures by Trylska et al. [52]. For comparison, the results from this study are shown in the last two columns of Table 4.3 and 4.4. They used different values for $\epsilon_{Protein}$ in combination with two different charge models (*full-charge*: $\epsilon_{Protein} = 4$; *single-site*: $\epsilon_{Protein} = 20$). The *single-site* model handles ionization by adding ± 1 point charge to one central atom [18] (using CHARMM22 charges and OPLS radii), whereas in the *full-charge* model the formal charges are distributed over several atoms [108] (using PARSE charges and radii). The latter treatment is more similar to our charge model. Using a dielectric constant of 20, the differences between our results and the study of Trylska et al. are smaller than one log unit. For the calculations with $\epsilon_{Protein} = 4$, larger deviations are observed.

4.3.2 Calculations for the complexes

Uncharged symmetric cyclic urea inhibitor DMP-323

DMP-323 is a symmetric cyclic urea derivative which binds with a K_i-value of 340 pM [109]. Its benzyl moieties are accommodated in the S_1,S_1'-pockets, whereas the benzyl alcohol portions address the S_2,S_2'-pockets. The structural differences with respect to the apo HIVP structure (1HHP) are relatively large: the RMSD of the C_α-fit amounts to 1.35 Å. However, this difference is mostly provoked by the flap region, which is known to be rather adaptive [110]. The carbonyl group of the cyclic urea ring displaces the flap water, which was one of the initial goals in the development of this type of inhibitor.

The experimentally determined pKa values of the two catalytic aspartates in the **DMP-323** complex are > 7.2 [103] suggesting that both residues are protonated at pH = 5. The pKa values of all other aspartates and glutamates have also been determined by NMR experiments. Of these, only Asp29 deviates by more than one log unit from its model pKa, exhibiting a pKa value between 1.97 and 2.06. This lowered pKa value is induced by the formation of salt bridges to Arg8 and Arg87'.

The results for the **DMP-323**/HIVP complex are given in Table 4.5. We focus mainly on the correct determination of the proton uptake upon ligand binding. To correctly predict this effect, $\Delta\Sigma H^+$ must be +1, since the catalytic dyad changes from mono-protonation to double-protonation upon ligand binding. This quantity is comparable to Δn (molar net exchange of protons), which can be measured using ITC. In a previous study, we successfully predicted such Δn for a series of trypsin and thrombin complexes by means of similar pKa calculations [111]. For the **DMP-323** complex, the uptake of additional protons is qualitatively correctly predicted, but quantitatively the full amount of one proton on molar ratio is not accurately estimated: we compute an uptake of +0.63 protons ($\epsilon_{Protein} = 10$). However, we consider this as a correct prediction of the trend. Furthermore, the pKa shifts of the catalytic dyad for this $\epsilon_{Protein}$ amount to +1.6 and +3.5. Such changes can be considered as significant enough to assume protonation of both aspartates. Applying a dielectric constant of 4 or 20 the prediction is less satisfactory, though the correct direction of the pKa shifts is suggested.

A comparison of our results for the **DMP-323** complex with the study of Trylska et al. is given in Table 4.6. It can be seen that (using $\epsilon_{protein} = 20$) the difference is less than one log unit. The aspartate with the lower pKa value seems to be overestimated in our calculations at $\epsilon = 4$, while in the case of Trylska et al., the use of $\epsilon = 4$ gives better agreement

Table 4.5: pKa calculations on HIVP complexed with **DMP-323** (1QBS).

	apo (ligand-deleted)					*complexed with **DMP-323***					
ϵ	pK_{a1} (*"aspartate$_1$"*)	H^+ [a]	pK_{a2} (*"aspartate$_2$"*)	H^+ [a]	ΣH^+ [b]	pK_{a1} (*"aspartate$_1$"*)	H^+ [a]	pK_{a2} (*"aspartate$_1$"*)	H^+ [a]	ΣH^+ [b]	$\Delta\Sigma H^+$ [c] $\approx \Delta n$
4	5.79	0.86	10.19	1.00	1.86	9.96	1.00	20.34	1.00	2.00	+0.14
10	3.73	0.05	7.20	0.99	1.04	5.30	0.67	10.71	1.00	1.67	+0.63
20	2.97	0.01	5.95	0.89	0.90	3.62	0.04	7.06	0.99	1.03	+0.13

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5

[b] net degree of protonation of the dyad at pH=5

[c] net change of protonation upon ligand binding

Table 4.6: Comparison with results from Trylska et al. for the HIVP/**DMP-323** complexes. PEOE_PB stands for our charge model used throughout the pKa calculations.

*HIVP complexed with **DMP-323***										
	PEOE_PB				Trylska [52]				NMR [103]	
	pK_{a1}	H^{+} [a]	pK_{a2}	H^{+} [a]	pK_{a1}	H^{+} [a]	pK_{a2}	H^{+} [a]	pK_{a1}	pK_{a2}
$\epsilon = 4$	9.96	[1.00]	20.34	[1.00]	7.9	[1.00]	20.5	[1.00]	> 7.2	> 7.2
$\epsilon = 20$	3.62	[0.04]	7.06	[0.99]	3.6	[0.04]	6.5	[0.97]	> 7.2	> 7.2

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5

with experiment, at least for the aspartate with the lower pKa. Concerning the extreme pKa values (>20), the authors argue that such values are unreasonable; nevertheless, also these calculations indicate that both residues definitely remain protonated.

Uncharged asymmetric inhibitor KNI-272

The asymmetric inhibitor **KNI-272** differs significantly from the cyclic urea compound studied in the previous section (see Figure 4.2). The inhibitor binds with a K_i of 5.5 pM [112]. Furthermore, only one hydroxylic group is placed between the two catalytic aspartates, qualifying **KNI-272** as member of the first generation of HIVP inhibitors. Two water molecules can be observed in the crystal structure: W301 and W607. W301 is the "flap water", and W607 is in close vicinity to one of the catalytic aspartates (with a distance of 3 Å). It forms a hydrogen bond to the ligand's hydroxylic group. We performed pKa calculations both with and without W607 to probe its influence on the protonation states. The water molecule W301 was not included explicitly due to its large distance from the catalytic dyad.

We explored the hydrogen-bond formation capability of the ligand's hydroxylic group to both aspartates by means of rotating the OH group in steps of 30 degrees ("rot-scan"). For each step, a pKa calculation was performed. We observed dependencies on the calculated pKa values: using $\epsilon_{Protein} = 20$, the pKa value of one aspartate fluctuates within two log units, and that of the second by one log unit. In the crystal structure, the distance of the hydroxy oxygen amounts to 2.6 Å with respect to the carboxylate oxygen of the first aspartate, whereas the distance to the second displays a value of 3.1 Å. For the pKa calculations presented in Table 4.7 and 4.8, an orientation of the ligand OH group has been considered which reveals the lowest pKa for the proximate aspartate (Asp25 labeled in the crystal structure): in this orientation, a hydrogen bond can be formed to this aspartate. This evidence is in agreement with NMR experiment [26] where this aspartate forms two hydrogen bonds,

one to W607, and a second to the ligand. The NMR experiment also revealed the pKa values for the catalytic dyad; they were found to be < 2.5 and > 6.2 for the coupled system of the two aspartates.

This is in agreement with the predictions based on semi-empirical quantum mechanical calculations that the system of aspartates is in mono-protonated state once complexed with **KNI-272** [113]. In summary, all calculations suggest that the protease system remains in the same mono-protonated state upon complexation of the ligand **KNI-272**.

A comparison of the results in Table 4.7 and 4.8 shows that the influence of the presence or absence of the water molecule W607 can modulate the pKa value of the proximal aspartate by up to 0.5 log units. However, this has only minor impact on the resulting protonation states and inventory. The influence of different orientations of the water molecule was not probed, but it is assumed that the influence is of minor importance, because an "ideal" hydrogen bond between the W607 and Asp25 has been modelled for our calculations.

In Table 4.9, we compare our results with those of Trylska et al.; the experimentally determined pKa values are also listed. The results are similar for Asp25', whereas large deviations are observed for Asp25. Suprisingly, opposing trends are suggested for both studies while decreasing $\epsilon_{Protein}$ from 20 to 4: in our case, the pKa value of Asp25 is increased, whereas in the Trylska et al. study a decrease is suggested. Possibly in the latter study (which uses different charge models for the different dielectric constants) the PARSE charges show better performance for $\epsilon_{Protein} = 4$. Best agreement to experiment is achieved with our PEOE_PB charges using $\epsilon_{Protein} = 20$. Most importantly, all calculations suggest the overall trend correctly: the catalytic dyad remains in the mono-protonated state.

S6

The hydroxyethylene sulfone **S6** is another asymmetric inhibitor. The binding mode of its hydroxylic group resembles the binding mode of the central hydroxylic group of **KNI-272**. Inhibitor **S6** has been developed in our group and binds with a K_i of 45 nM [114, 115]. Similar to the **KNI-272** complex, **S6** provokes different local environments next to the two aspartates of the catalytic dyad. Interestingly, a water molecule close to one of the aspartates is missing, although the **S6** and **KNI-272** structures are similar. No experimental data regarding the protonation states of the catalytic aspartates are available.

As the crystal structure determination cannot eludicate the orientation of hydrogens, the hydrogen-bonding partner of the ligand's hydroxy group is not fully characterized (both

Table 4.7: pKa calculations on HIVP complexed with **KNI-272** (1HPX).

	apo (ligand-deleted from the complex)					*complexed with **KNI-272***					
ϵ	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	ΣH^+ [b]	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	ΣH^+ [b]	$\Delta\Sigma H^+$ [c]
	("aspartate$_1$")		*("aspartate$_2$")*			*("aspartate$_1$")*		*("aspartate$_1$")*			$\approx \Delta n$
4	5.98	0.89	11.10	1.00	1.89	5.51	0.76	28.71	1.00	1.76	-0.13
10	3.78	0.06	7.24	0.99	1.05	3.84	0.06	13.43	1.00	1.06	+0.01
20	2.93	0.01	5.88	0.88	0.89	3.1	0.01	7.99	1.00	1.01	+0.12

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5

[b] net degree of protonation of the dyad at pH=5

[c] net change of protonation upon ligand binding

Table 4.8: pKa calculations on HIVP complexed with **KNI-272** (1HPX), but considering one explicit water molecule (W607) in the complexed form.

	*complexed with **KNI-272***					
ϵ	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	ΣH^+ [b]	$\Delta\Sigma H^+$ [c]
	(*"aspartate$_1$"*)		(*"aspartate$_2$"*)			$\approx \Delta n$
4	5.20	0.61	29.63	1.00	1.61	-0.28
10	3.37	0.02	13.60	1.00	1.02	-0.03
20	2.71	0.01	7.98	1.00	1.01	+0.12

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5
[c] net change of protonation upon ligand binding

Table 4.9: Comparison with results from Trylska et al. [52] and experimentally determined pKa values.

	*HIVP complexed with **KNI-272***									
	PEOE_PB				Trylska et al.				NMR [26]	
	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	pK_{a1}	pK_{a2}
	(*"aspartate$_1$"*)		(*"aspartate$_2$"*)		(*"aspartate$_2$"*)		(*"aspartate$_2$"*)			
$\epsilon = 4$	5.20	0.61	29.63	1.00	1.9	0.00	27.0	1.00	< 2.5	> 6.2
$\epsilon = 20$	2.71	0.01	7.98	1.00	4.5	0.26	7.0	0.99	< 2.5	> 6.2

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5

aspartates are close enough to form a hydrogen bond); the distance of the carboxylic oxygen amounts to 2.52 Å to the first aspartate and 2.61 Å to the second. Accordingly, we also performed a "rot-scan" of this ligand OH-group in steps of 30 degrees to detect the dependence of the pKa values on the orientation of the OH group. For the results shown in Table 4.10, we selected an orientation in which the ligand's hydroxylic group forms a hydrogen bond to the proximal aspartate.

A comparison with the results from **KNI-272** shows that the pKa value of the aspartate which is hydrogen-bonded to the ligand's hydroxylic group is much lower for the **S6**/HIVP complex. Most likely this is due to the shorter distance between the oxygen atoms of the hydrogen-bond in the **S6** complex (2.52 Å) compared to the **KNI-272** complex (2.96 Å).

The pKa values resulting from our calculations must be seen in the light of the relative shifts. Hence, we can assume that a conserved mono-protonated state of the catalytic dyad seems most likely for the complexation with **S6**.

Table 4.10: pKa calculations on HIVP complexed with **S6** (1XL5).

	apo (ligand-deleted)					*complexed with **S6***					
ϵ	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	ΣH^+ [b]	pK_{a1}	H^+ [a]	pK_{a2}	H^+ [a]	ΣH^+ [b]	$\Delta\Sigma H^+$ [c]
	("aspartate$_1$")		*("aspartate$_2$")*			*("aspartate$_1$")*		*("aspartate$_2$")*			$\approx \Delta n$
4	4.05	0.10	11.95	1.00	1.10	-0.52	0.00	31.10	1.00	1.00	-0.10
10	2.13	0.00	7.24	0.99	0.99	0.53	0.00	15.20	1.00	1.00	+0.01
20	1.93	0.00	5.77	0.85	0.85	1.08	0.00	8.45	1.00	1.00	+0.15

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5
[c] net change of protonation upon ligand binding

VX-478

In the design of **S6**, **VX-478** (now on the market and named amprenavir), was taken as a parent scaffold [114, 115] bearing a similar central hydroxyl group. It binds with a K_i of 0.60 nM [116]. **VX-478** addresses the S_1 and S_2 (and S_1' and S_2', respectively) pockets in a similar way as **S6**.

Analogous to the study of **KNI-272** and **S6**, we performed a "rot-scan" of the central ligand hydroxylic group to detect its optimal orientation. Regardless of the orientation, the catalytic dyad is in the mono-protonated state. We chose the orientation which revealed the lowest pKa value for the deprotonated aspartate. Similar to the complexes with **KNI-272** and **S6**, we can assume that the catalytic dyad remains mono-protonated upon ligand binding.

S7

The ligand **S7** is another HIVP inhibitor that was developed in our group. It bears a pyrrolidine ring as new central core [114, 115]. The (R,R)-enantiomer of **S7** binds with a K_i of 1.5 μM [114, 115], which makes it a weaker inhibitor than the other compounds studied in this contribution. One interesting structural feature is a water molecule, which is located in close vicinity of one of the two aspartates. The position of this water molecule is almost identical with the position of the water molecule W607 in the **KNI-272** complex.

The experimentally determined pKa value for unsubstituted pyrrolidine amounts to 11.3 [117]; we assigned a pKa value of 11 for our calculations. We studied the charge setting for the neutral form of the pyrrolidine: either the charge of one of the bonded hydrogens was set to zero, or both hydrogens were "equally neutralized" (the charge settings are shown in Figure 4.4). The second model considers to some degree a possible dynamic fluctuation of the remaining hydrogen between the two orientations to the neighboring aspartates. However, both charge models produce similar pKa predictions, and the results in Table 4.12 and 4.13 correspond to the "equally neutralized"-model. The results suggest very low pKa values for both aspartates. The large pKa difference between the two aspartates can be explained by the asymmetric position adapted by the pyrrolidine ring. It approaches one of the aspartates 0.3 Å closer than the other. This aspartate reveals the lower pKa value. In contrast, as expected for a strong salt bridge, the pKa of the ligand nitrogen increases compared to the unbound state.

The influence of the explicit water molecule (named WAT98 in the PDB file) was also

Table 4.11: pKa calculations on HIVP complexed with **VX-478** (1HPV).

	apo (ligand-deleted)					*complexed with **VX-478***					
ϵ	pK_{a1} (*"aspartate$_1$"*)	H^+ [a]	pK_{a2} (*"aspartate$_2$"*)	H^+ [a]	ΣH^+ [b]	pK_{a1} (*"aspartate$_1$"*)	H^+ [a]	pK_{a2} (*"aspartate$_1$"*)	H^+ [a]	ΣH^+ [b]	$\Delta\Sigma H^+$ [c] $\approx \Delta n$
4	5.07	0.54	11.36	1.00	1.54	3.84	0.06	26.89	1.00	1.06	-0.48
10	3.45	0.03	7.29	0.99	1.02	2.71	0.01	12.52	1.00	1.01	-0.01
20	2.88	0.01	5.94	0.90	0.91	2.49	0.00	7.52	1.00	1.00	+0.09

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5
[c] net change of protonation upon ligand binding

Table 4.12: pKa calculations on HIVP complexed with **S7** (1XL2). The model pKa of the unbound **S7** is 11.0.

	apo (ligand-deleted)					*complexed with **S7***						
ϵ	pK_{a1} (*"aspartate$_1$"*)	H^+ [a]	pK_{a2} (*"aspartate$_2$"*)	H^+ [a]	ΣH^+ [b]	pK_{a1} (*"aspartate$_1$"*)	H^+ [a]	pK_{a2} (*"aspartate$_1$"*)	H^+ [a]	ΣH^+ [b]	$\Delta\Sigma H^+$ [c] $\approx \Delta n$	
4	5.40	0.72	10.48	1.00	1.72	27.38	-21.0	0.00	-4.55	0.00	0.00	-1.72
10	3.54	0.03	7.23	0.99	1.02	19.59	-9.60	0.00	-1.00	0.00	0.00	-1.02
20	2.92	0.01	5.97	0.90	0.91	16.66	-4.18	0.00	0.42	0.00	0.00	-0.91

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5
[c] net change of protonation upon ligand binding

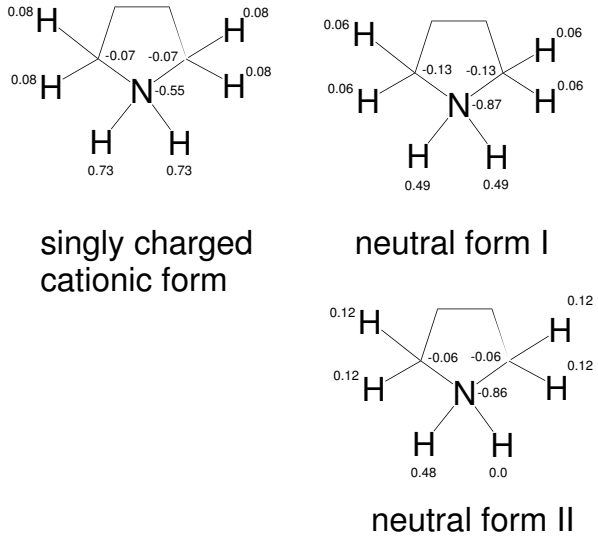

Figure 4.4: Charge settings for the neutral and charged states of the ligand **S7**. The model "equally neutralized" is displayed on the top right, whereas the charges for explicit deletion of one hydrogen are given in the bottom right corner. The difference in the net charge for both neutral charge models is caused by the differing number of atoms. "q" corresponds to the sum of the labeled charges.

Table 4.13: pKa calculations on HIVP complexed with **S7**, where one explicit water molecule next to the more acidic aspartate has been considered (in contrast to the results in Table 4.12).

	complexed with **7**					
ϵ	pK_{a1} (*"aspartate$_1$"*)	$H^{+\ a}$	pK_{a2} (*"aspartate$_2$"*)	$H^{+\ a}$	$\Sigma H^{+\ b}$	$\Delta\Sigma H^{+\ c}$ $\approx \Delta n$
4	-21.00	0.00	-4.93	0.00	0.00	-1.72
10	-11.13	0.00	-1.17	0.00	0.00	-1.02
20	-4.85	0.00	0.32	0.00	0.00	-0.91

[a] net degree of protonation of aspartate$_1$ or aspartate$_2$ at pH=5
[b] net degree of protonation of the dyad at pH=5
[c] net change of protonation upon ligand binding

studied, and the results show a decreased pKa value for the neighboring aspartate (labeled Asp25, see Table 4.13), which forms a hydrogen bond to this water molecule. However, this shift has no impact on the predicted protonation states.

Overall, the pKa values of the two aspartates are shifted to a pKa range where (at pH = 5) the catalytic dyad is supposedly fully deprotonated and, thus, present with two negatively charged carboxylate groups.

4.4 DISCUSSION

The two investigated apo crystal structures exhibit some deviations in the geometry, in consequence the computed pKa values differ slightly. With increasing $\epsilon_{Protein}$ these differences decrease, an observation which correlates with a growing damping of the electrostatic interactions. For this reason larger $\epsilon_{Protein}$ values have been recommended as they better consider possible relaxations or local conformational multiplicity of the protein [18]. This might serve as an explanation why in general the values obtained with $\epsilon_{Protein} = 4$ deviate rather strongly from experiment. Furthermore, site-site interactions are responsible for the increased pKa value of one of the catalytic aspartates in the apo structure. Possibly such site-site interactions are overestimated with $\epsilon_{Protein} = 4$, as too extreme pKa values are predicted. Likely, again the electrostatic interactions on short distance are exaggerated.

The initial goal for this contribution was to benchmark our PEOE_PB charges, particularly to estimate changes of protonation states upon ligand binding. Experiments reveal the catalytic dyad to be mono-protonated for the uncomplexed enzyme. In case of the study of Trylska et al. [52], this is only predicted using a $\epsilon_{Protein} = 4$. With our methodology, reasonable values are obtained with a setting of $\epsilon_{Protein} = 10$ or 20 for both apo structures. The use of different charge models provides a probable explanation for these differences.

As mentioned, the geometrical differences of the two apo structures result in somewhat deviating assignments of pKa values. Nevertheless, the relative differences are similarly predicted. This observation points to a crucial aspect: What is the most reasonable approximation of the uncomplexed geometry, or, vice versa, by how much does the uncertainty in the reference geometry effect the accuracy of the computed pKa calculations? It can be argued that a *ligand-deleted* structure of the protease is only a very approximate represensation of the actual apo geometry, however, it is straightforward to produce. We want to compare

Table 4.14: Comparison of mean pKa values for the *real* apo structures and the *ligand-deleted* HIVP structures

	pKa^{exp}	mean pKa^{apo}			mean $\text{pKa}^{ligand-deleted}$		
		$\epsilon = 4$	$\epsilon = 10$	$\epsilon = 20$	$\epsilon = 4$	$\epsilon = 10$	$\epsilon = 20$
pK_{a1}	3.1-3.7	5.45	3.77	3.12	5.26	3.33	2.73
pK_{a2}	4.9-6.5	9.79	6.79	5.69	11.02	7.24	5.90

the pKa values obtained using different structures as input. In Table 4.14, the mean pKa values averaged across the five different *ligand-deleted* structures are summarized. Facing them with the pKa values obtained for the two experimental protein structures reveals a maximum deviation of 1.3 log units for $\epsilon_{Protein} = 4$. Considering the more reasonable settings

of $\epsilon_{Protein} = 10$ or 20, the mean deviations reduce to 0.4 log units. As these deviations are smaller than the consequences of modifying the $\epsilon_{Protein}$ settings, the use of a *ligand-deleted* geometry appears a justifiable and very practical approximation for the intended calculations.

To predict the properties of the **DMP-323** complex, $\epsilon_{Protein} = 10$ seems to be optimal. The twofold protonation of the catalytic dyad is predicted correctly. Also for the **KNI-272** complex, $\epsilon_{Protein} = 10$ produces the best results. Here the protonation state of the catalytic dyad is estimated mono-protonated in agreement with experiment. A zero net uptake of protons is correctly predicted. The placement of an explicit water close to the catalytic dyad has only minor impact on the computed results.

The binding modes of **S6** and **VX-478** are rather similar, particularly with respect to the central hydroxylic group. Nonetheless, the resulting pKa values of both aspartates reveal different pKa values for the two complexes, particularly for $\epsilon_{Protein} = 4$. Independent of the actual values for $\epsilon_{Protein}$, however, the catalytic dyad is predicted to adopt mono-protonation, and zero uptake of protons is predicted upon complex formation.

The inhibitor **S7** induces a protonation pattern of the catalytic dyad which deviates from other complexes: both carboxy functions are predicted to be deprotonated. Although rather negative (most likely exaggerated) pKa values are computed for both catalytic aspartates the release of one proton upon complex formation is predicted, particularly for a setting of $\epsilon_{Protein} = 10$ or 20.

A central hydroxy group is present in the three ligands **KNI-272**, **S6** and **VX-478**. Experimental evidence (by means of NMR measurements [26]) about the orientation of this hydroxy group is only available for the **KNI-272** complex. We explored different orientations of the hydroxy group by systematically scanning all rotamers. The hydrogen bond formation takes influence on the pKa values of the catalytic dyad, as fluctuations of one to two pKa units are observed. We selected the orientation that achieves the lowest pKa value of all rotamers in agreement with experimental evidence.

4.5 CONCLUSIONS

Inhibitor binding to HIVP can induce a change of protonation of the two catalytic aspartates, as confirmed experimentally. We performed pKa calculations on five different HIVP complexes and two apo structures. Depending on the type of ligand, different protonation states for the catalytic dyad are predicted. In cases where experimentally determined pKa

values are available our results have been compared to these reference values. Furthermore, a similar theoretical study [52] has been consulted for comparison. For the apo form of HIVP, the mono-protonated state of the catalytic dyad is evidenced by kinetic measurements [25, 24]. Our calculations suggest very similar results either based on the geometry of two apo structures or of five complex structures from which the ligand has been removed. This fact underlines that the straightforward strategy of deleting the ligand from the complex structure is sufficiently accurate to produce relevant pKa calculations (in case of HIVP).

The binding of **KNI-272** to HIVP involves no change of protonation of the catalytic dyad. The pKa calculation correctly reproduces this experimental result. In contrast, upon binding of **DMP-323** one proton is picked-up by the catalytic dyad. Also, this experimentally observed change is correctly reproduced by our calculations.

In the second part of the study, we have predicted protonation states of the catalytic dyad for three HIVP complexes where no experimental data on the protonation states are available. For the complexes with one central ligand hydroxyl group facing the dyad (**S6**, **VX-478**), the mono-protonated state is suggested. Most interestingly, for the **S7** complex with a formal positive charge on the central ligand's pyrrolidine moiety interacting with the dyad, both aspartates are predicted to occur in the deprotonated state. To the best of our knowledge, this protonation pattern has not yet been suggested or experimentally observed for any HIVP complex.

Knowledge about the most likely protonation state of ligands and protein residues is an important prerequisite to correctly setup computational drug design techniques such as docking and virtual screening. Possibly deviating settings can retrieve alternative hits since the assigned physicochemical properties (e.g. with respect to hydrogen-bond donor or acceptor functionalities) of the catalytic aspartates determine what type of ligand functional groups will be accepted as potential binder of the protease. Furthermore, with respect to ITC experiments the present calculations can indicate whether superimposed changes of protonation might affect the measured heat signal. In case of overlaid protonation steps elaborate corrections have to be performed to obtain the net heat of binding which allows together with the recorded binding constant to factorize the binding affinity into entalpic and entropic contributions.

5 PROTONATION EFFECTS IN HUMAN ALDOSE REDUCTASE

5.1 INTRODUCTION

Aldose reductase (AR) is a NADPH-dependent oxidoreductase [118] catalyzing the reduction of a wide variety of carbonyl compounds to the corresponding alcohols. Human aldose reductase (hAR) consists of a single polypeptide chain of 315 residues with a molecular mass of 35.8 kDA [119] (see Figure 5.1). The active site of hAR is segmented into two parts: the anion pocket composed by the residues Asp43, Tyr48, Lys77 and His110 and the hydrophobic specificity pocket flanked by the residues Trp111, Thr113, Leu300 and Cys303. The enzymatic reaction catalyzed by hAR involves, subsequently to the reductive hydride transition, a proton transfer. It is postulated that either Tyr48 [120] or His110 [121] are engaged in this process. This unresolved question makes hAR an attractive test case for our pKa calculation methodology.

Although the physiological role of hAR in healthy individuals is still unclear, the enzyme is believed to be of primary importance in the development of severe degenerative complications of diabetes mellitus [122]. Therefore, hAR represents a putative target for the treatment of this disease. In general, two inhibitor classes exist for the inhibition of hAR: compounds with a carboxylic head group and spiro-hydatoins. Both classes bind similarily to the anion binding pocket, but differ substantially with respect to the specificity pocket: if the ligand bears appropriate hydrophobic parts, the specificity pocket opens up, otherwise it remains in closed state (see Figure 5.2).

In an extensive experimental study [23], we performed a full thermodynamic and crystallographic characterization of the hAR/ligand binding process. In the context of this study, we were able to detect protonation changes occuring in due course of the complexation process. The ligands used for our investigation are shown in Figure 5.3. In this contribution we try to collect evidence for the structural interpretation of the detected protonation effects by means of pKa calculations. In two previous studies on protein-ligand complexes where protonation transfer steps have been recorded experimentally [111, 123], we were able to compute atypical protonation states which explained the change in the protonation inventory. In this context, atypical protonation means a non-standard protonation, i.e. a protonated acid or a deprotonated base.

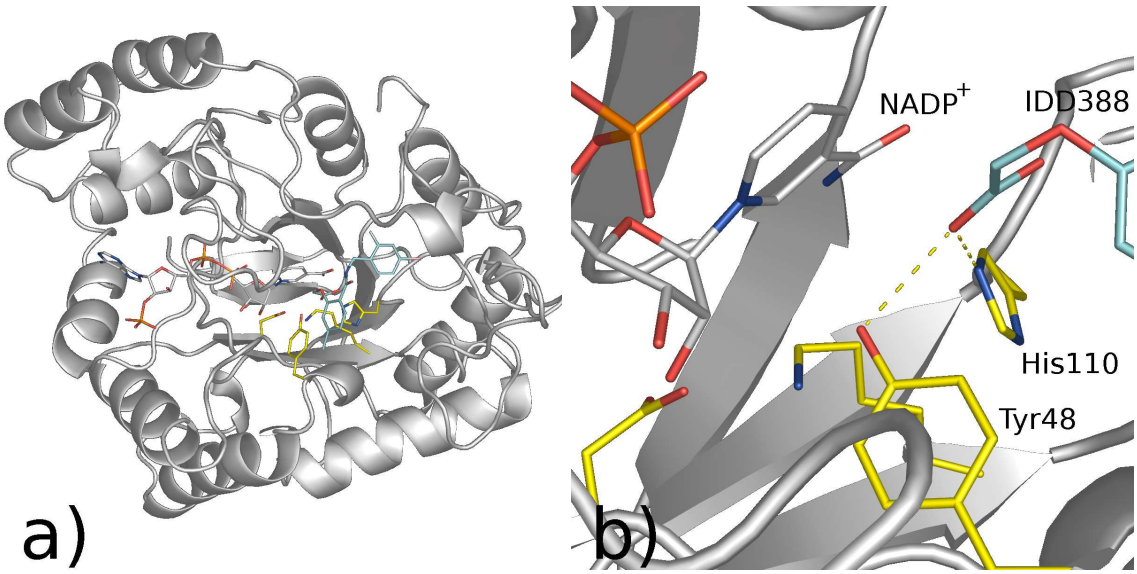

Figure 5.1: Crystal structure of hAR adapts a TIM-barrel fold (eight-stranded α-β fold). The cofactor, the ligand **IDD388** and the anion pocket are given in stick representation. (a) Overall view of the structural elements. (b) Hydrogen bonds formed by the carboxylic head group of the ligand towards Tyr48 and His110.

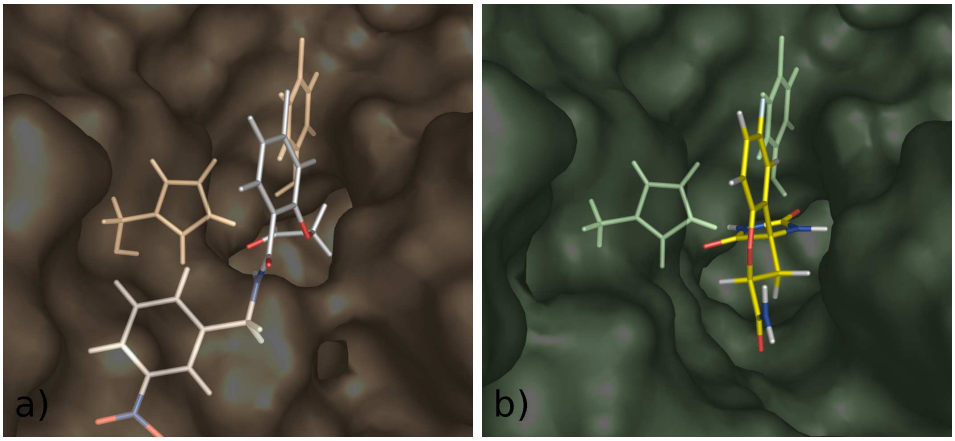

Figure 5.2: Complexes of human aldose reductase with the inhibitors (a) **IDD388** (b) **Fidarestat**. Tyr48 and His110 are shown in stick representation (see Figure 5.3). The specificity pocket (lower part) is closed in the complex with **Fidarestat**, whereas it opens up when **IDD388** is bound.

5.2 RESULTS

5.2.1 Calculations for five holo structures

We used five *ligand-deleted* structures of hAR (from the studied complexes) to investigate the conformational influence on the resulting pKa values. The analysis of ligand-deleted structures is a very pragmatic approach, as a geometry for the uncomplexed protein is generated

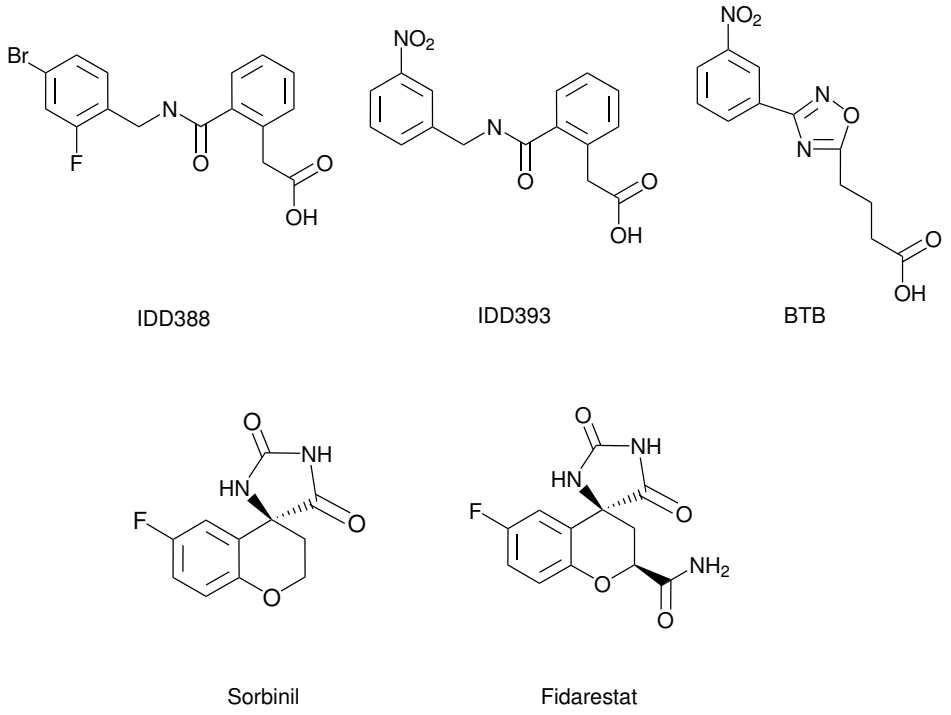

Figure 5.3: Chemical formulae of the aldose reductase inhibitors used in this investigation.

by simply removing the ligand from the complex. One apo structure has recently become available [124], but as the bound cofactor is missing in the active site, this structure is not of relevance for our studies. We explored the following different settings in our calculations on the holo enzyme:

- Usage of different values for the dielectric constant (10 or 20).
- Impact of the orientation of the Tyr48 OH group on its pKa value.
- Oxidation state of the cofactor: oxidized ($NADP^+$) or reduced (NADPH).

The findings with respect to the different applied settings will help to accomplish the most appropiate parameter adjustment. All results are listed in Tables 5.1 and 5.2. The two considered orientations of the Tyr48 OH group are shown in Figure 5.4. It appeared reasonable to us to place the OH-group in different orientations, since such a group is known to adopt different conformational states.

In case of His110, N_δ has been defined as sole titratable group. In the complexes to be simulated in the following, N_ϵ will be permanently involved in a hydrogen bond to the ligands. For consistency in our calculations, we only considered N_δ as titratable group also

[1]To differentiate the different starting geometries the entires are named by their originally bound ligand.

Table 5.1: Variation of the **Tyr48** pKa depending on the orientation of the Tyr48 OH group: oriented either towards *Asp43* or *His110* (see Figure 5.4). The cofactor is given in the oxidized and reduced state, as well.

		ligand-deleted AR [1]					
ϵ	cofactor	"**IDD388**"	"**IDD393**"	"**BTB**"	"**Fidarestat**"	"**Sorbinil**"	mean (std dev)
10_{Asp43}	$NADP^+$	5.82	6.27	6.79	7.14	7.10	6.62 (0.51)
10_{Asp43}	NADPH	11.45	11.32	11.53	11.60	11.32	11.44 (0.11)
10_{His110}	$NADP^+$	8.45	8.52	9.09	8.90	9.02	8.80 (0.26)
10_{His110}	NADPH	13.94	13.80	14.40	13.44	13.29	13.77 (0.39)
20_{Asp43}	$NADP^+$	7.70	7.84	8.15	8.19	8.23	8.02 (0.21)
20_{Asp43}	NADPH	10.98	11.05	11.15	11.60	11.32	11.22 (0.22)
20_{His110}	$NADP^+$	8.72	8.78	9.13	8.97	9.08	8.94 (0.16)
20_{His110}	NADPH	12.68	12.99	13.25	12.34	12.66	12.78 (0.31)

Table 5.2: Variation of the **His110** pKa depending on the orientation of the Tyr48 OH-group: oriented either towards *Asp43* or *His110* (see Figure 5.4). The cofactor is given in the oxidized and reduced state, as well.

		ligand-deleted AR					
ϵ	cofactor	"**IDD388**"	"**IDD393**"	"**BTB**"	"**Fidarestat**"	"**Sorbinil**"	mean (std dev)
10_{Asp43}	$NADP^+$	-1.40	-1.70	-1.47	-0.91	-1.03	-1.30 (0.29)
10_{Asp43}	NADPH	0.09	-1.70	-1.47	-0.91	-1.03	0.27 (0.31)
10_{His110}	$NADP^+$	-1.75	-2.02	-1.80	-1.19	-1.36	-1.62 (0.30)
10_{His110}	NADPH	-0.59	-2.02	-1.80	-1.19	-1.36	-0.50 (0.51)
20_{Asp43}	$NADP^+$	2.34	2.07	2.36	2.44	2.46	2.33 (0.14)
20_{Asp43}	NADPH	3.67	2.07	2.36	2.44	2.46	3.71 (0.21)
20_{His110}	$NADP^+$	1.97	1.78	2.02	2.17	2.15	2.02 (0.14)
20_{His110}	NADPH	3.18	1.78	2.02	2.17	2.15	3.25 (0.19)

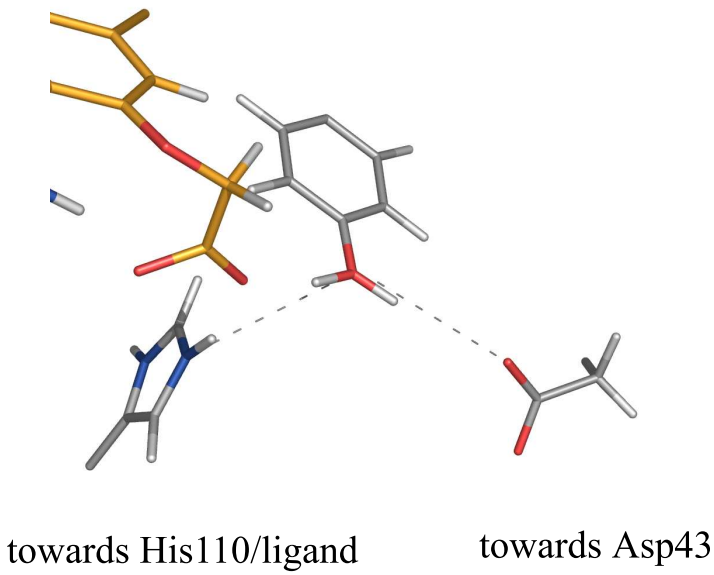

Figure 5.4: The two studied orientations of the Tyr48 OH group: it is either oriented towards Asp43 or His110/ligand. For better orientation, the carboxylic head group of **IDD388** is shown on the left hand side.

in the uncomplexes situation. For the two remaining titratable residues of the active site (Asp43, Lys77), no pKa values indicating an atypical protonation state are observed and, hence, they are not listed in the Table.

For **Tyr48**, it is observed that the oxidation state of the cofactor has an eminent influence on the pKa value; in general, the consideration of $NADP^+$ instead of NADPH results in mean pKa values lowered by 3-5 log units. All calculations with the reduced cofactor suggest a fully protonated tyrosine, which is not the case considering the oxidized cofactor. This behaviour may serve as a first indicator that Tyr48 is deprotonated in the holo enzyme. The influence of $\epsilon_{Protein}$ is opposing for the different forms of the cofactor: the mean Tyr48 pKa value is reduced when $NADP^+$ is bound and a lower $\epsilon_{Protein}$ is applied, whereas it increases when NADPH is present. The adopted orientation of the OH-group takes a remarkable influence (a further indicator for the perturbed protonation state of Tyr48). For xylanase, we were able to show the influence of the Tyr-OH orientation on the pKa value of an active site glutamate [85]. Choosing an orientation towards Asp43 ($\epsilon_{Protein} = 10$, $NADP^+$) gives a mean pKa value of 6.62 which corresponds to a protonation degree of 4 %. For **His110**, very low pKa values are calculated in all cases (negative values for a dielectric constant of 10). The oxidation state of the cofactor shows no influence on an altered protonation state of the histidine residue. Hence, all shifts fall into a range suggesting that this residue remains in neutral state and is not involved in any changes.

5.2.2 Calculations for compounds with carboxylic head groups

The structures of hAR complexed with **IDD388**, **IDD393** and **BTB** have been crystallographically determined in-house. ITC revealed a proton uptake of the protein-ligand complexes from the buffer of $+0.8 \pm 0.2$ (**IDD388**, **BTB**) or $+0.9 \pm 0.2$ (**IDD393**) moles of protons. The results for the different cofactor oxidation states are given in Table 5.3. They suggest that the carboxylic group of the ligand should remain deprotonated (neglecting the NADPH/$\epsilon_{Protein} = 10$ calcualtion). Furthermore, the calculations propose (definite protonation states of the two crucial residues: Tyr48 appears in the protonated form. For the calculations with NADPH, a partially protonated His110 is suggested for the **BTB** complex with $\epsilon_{Protein} = 20$, whereas it is fully deprotonated for the $NADP^+$ calculations.

Our calculations on the holo structures showed (see previous section) that the protonation degree of Tyr48 before ligand binding amounts to 4 % (Tyr48 OH group orientation towards Asp43, $\epsilon_{Protein} = 10$, $NADP^+$). Using these settings, a Δn_{calc} of +0.96 is obtained for all three complexes. This in excellent agreement with experiment.

ITC experiments with IDD388 and NADPH instead of $NADP^+$ gave a Δn of $+0.2 \pm 0.04$. In the NADPH calculations for the holo structure, the results suggest a fully protonated Tyr48. The calculations for the **IDD388** complex reveal that the Tyr48 pKa is increased. Taken together with the partially protonated His110 ($\epsilon_{Protein} = 20$), this gives a Δn_{calc} of 0.09 which matches remarkably with the experiment. Aslo the setting to $\epsilon_{Protein} = 10$ suggest neglectable change in protonation which correctly predicts the trends seen in the experiments. For the sake of completeness, the results for the two other complexes are also given in Table 5.3.

5.2.3 Calculations for compounds containing spiro-hydatoins

Spiro-hydatoins belong to the second class of inhibitors studied in this contribution. They bind with a cyclic hydantoin moiety in a similar region of the binding pocket as the carboxylic head group of the carboxylic acid type inhibitors. The experimentally determined pKa values (in aqueous solution) for **Fidarestat** and **Sorbinil** are 7.9 [125] and 8.7 [126]: these pKa values fall very close to the applied pH conditions and make these inhitors sensitive to pka shifts resulting in protonation changes.

ITC experiments indeed revealed Δn to be $+0.3 \pm 0.1$ for **Fidarestat** and $+0.1 \pm 0.02$ for **Sorbinil** (hAR and $NADP^+$). The results (for the different cofactor oxidation states)

Table 5.3: Computed pKa values for Tyr48 and His110 for complexes of hAR with three inhibitors (**IDD388**, **IDD393**, **BTB**). The cofactor is given in the oxidized and reduced state, as well. Δn_{calc}: change in protonation upon ligand binding with respect to the holo form. The ligand pKa values are not considered for any protonation change (cf. NADPH/$\epsilon_{Protein} = 10$.

		IDD388			***IDD393***			***BTB***			
ϵ	cofactor	pKa (COO)	pKa (Tyr48)	pKa (His110)	pKa (COO)	pKa (Tyr48)	pKa (His110)	pKa (COO)	pKa (Tyr48)	pKa (His110)	Δn_{calc}
10	$NADP^+$	-5.54	16.15	1.92	-2.79	17.16	1.61	-3.78	31.10	2.54	+0.96
10	NADPH	0.36	31.10	5.24	6.28	31.10	-3.03	8.48	31.10	-1.85	±0.0
20	$NADP^+$	-2.32	13.67	5.48	-1.04	14.37	5.14	-1.90	16.33	5.60	+0.51
20	NADPH	-0.55	22.25	7.02	0.64	22.11	6.90	1.79	23.72	7.29	+0.09/+0.07/+0.16

from our calculations on the respective complexes are summarized in Table 5.4. The trend observed for the pKa's of Tyr48 and His110 is similar to the trend seen in the previous section, whereas the pKa values of the ligands are only modulated by less than one log unit. Since this pKa shift is too small to be regarded as significant, we do not give any Δn_{calc} values for the respective cases.

ITC experiments with NADPH instead of $NADP^+$ gave a Δn of -0.3 ± 0.05 (**Fidarestat**) and Δn of -0.5 ± 0.2 (**Sorbinil**). Our calculations estimate a protonated tyrosine residue in the holo state with the reduced cofactor. Therefore, the experimentally detected change in the protonation inventory must result from the titratable group of the ligand. Our calculations can not predict this. Finally, we explored the influence of the formal charge distribution on the proton inventory. The computational details are given in the **Materials and Methods** section. The changed settings result in lower pKa values of the ligand (see Table 5.5), but these values are still not sufficient to predict ligand binding in the deprotonated state.

5.3 DISCUSSION

5.3.1 Holo structures

The ITC experiments together with results from mutagenesis give clear evidence that Tyr48 adopts a deprotonated state in the complex with the oxidized cofactor prior to ligand binding - this was confirmed by measurements for the Tyr48Phe mutant which revealed a Δn of 0.0 for the **IDD388** complex [23]. We obtain a (median) pKa value of 6.62 for Tyr48 (OH oriented towards Asp43, $\epsilon_{Protein} = 10$, $NADP^+$) which corresponds to an almost fully deprotonated tyrosine residue.

The orientation of the OH-group of Tyr48 shows remarkable influence on its pKa value (see Table 5.1). We postulate that two different orientations for the Tyr48 OH-group might exist: before ligand binding, it points towards Asp43, and after complexation it orients towards the ligand and His110 (see Figure 5.4). A similar case was reported in a recent PB-based pKa study on an alcohol dehydrogenase, for which the binding pocket exhibits (amongst others) a lysine, tyrosine and NAD^+. Multiple orientations of the Tyr OH-group were extensively investigated and a significant influence on its pKa value has been registered [127]. One other issue concerns the structural situation prior to ligand binding: since no holo structure is available, we can only speculate about the preferred OH orientation in the active site prior to ligand binding.

Table 5.4: Computed pKa values for Tyr48, His110 and the ligand for two complexes (**Sorbinil**, **Fidarestat**.) The cofactor is given in the oxidized and reduced state, as well.

		Fidarestat			***Sorbinil***		
ϵ	cofactor	pKa (Ligand)	pKa (Tyr48)	pKa (His110)	pKa (Ligand)	pKa (Tyr48)	(His110)
10	$NADP^+$	8.56	28.71	-6.24	8.37	29.87	-5.68
10	NADPH	12.22	31.10	-4.27	10.87	31.10	-3.85
20	$NADP^+$	7.70	14.52	0.59	8.45	16.61	0.96
20	NADPH	9.77	21.96	1.67	23.03	18.54	2.00

Table 5.5: Computed pKa values for **Fidarestat**/hAR using different formal charge models. For further explanations on the different models see Figure 5.6. The cofactor is in the oxidized form.

model [see Figure 5.6]	ϵ	pKa (**Fidarestat**)
(a)	10	8.37
	20	8.45
(b)	10	7.34
	20	7.51
(c)	10	7.46
	20	7.60

According to kinetic studies, the pKa value of Tyr48 is 8.25 in the binary $NADP^+$/aldose reductase complex [120]. In a theoretical study by Várnai and Warshel [128], who used the protein-dipole Langevin-dipole method, the pKa of Tyr48 was estimated to be ~ 8.5 in the holo structure with the cofactor in the oxidized form. For the NADPH/aldose reductase complex, the experimentally determined pKa value is 8.7 [120]. We obtain a mean value of about 11 ($\epsilon_{Protein} = 10$) or 13 ($\epsilon_{Protein} = 20$), see Table 5.1. Possibly our calculations overestimate the upward pKa shift of Tyr48, but the correct trend is predicted.

For the *ligand-deleted* **IDD388** structure, variation of the initial settings were examined: (a) explicit consideration of one water molecule which donates a hydrogen-bond from beyond the catalytic site to the N_δ atom of His110, (b) titration of the N_ϵ atom of His110 instead of the N_δ atom (see above). The effect of the explicit consideration of an additional water molecule is rather marginal - the resulting impact is less than 0.1 pKa units on the pKa of Tyr48. On the other hand, variation of the titratable nitrogen of His110 results in pKa values for Tyr48 which deviate by 0.5 log units (predicting a higher pKa value for Tyr48 compared to the N_δ titration).

Analyzing the different effects that determine the calculated pKa value of Lys77 which is suggested to to exhibit a pKa beyond > 14 in most calculations shows the following: the intrinsic pKa value neglecting the influence of the local neighborhood reveals a pKa of 5 ($\epsilon = 10$) or 8 ($\epsilon = 20$), however, the site-site interactions with Asp43 and Tyr48 induce a strongly shifted pKa value of this residue. In contrast, the intrinsic pKa value of Tyr48 is about 10, but the site-site interactions with Lys77 and His110 reduce its pKa value. This also supports the assumption of a deprotonated state of Tyr48 and illustrate the complex nature of the binding pocket.

Former studies have speculated that His110 serves as proton acceptor in the catalytic mech-

anism of aldose reductase [121]. In none of our calculations we collect evidence for such a doubly protonated state of the imidazole moiety (see results for the complexes) and a recent neutron diffraction study of complex with a bound carbonic acid-type inhibitor clearly shows His110 in the neutral state [129].

5.3.2 Carboxylic head groups

Our calculations support the assumption that Tyr48 is protonated and His110 is deprotonated upon ligand binding, the large absolute values and extreme shifts should only by interpreted as an indicator for the induced trends.

- For an $\epsilon_{Protein} = 10$, the calculations suggest an unreasonably high Tyr48 pKa value. Some aspects have to be considered as possible explanation for these extreme and exaggerated values. In the previously studied systems, we maximally encountered three titratable groups in close vicinity, and a straightforward explanation was possible [111, 123]. Here in the present case we have to consider five titratable sites and thus the locally crowded charged groups make the handling of such a system rather complex. For example, Tyr48 and Lys77 form a coupled system in the studied hAR complexes: both possess close intrinsic pKa values and they interact with the titratable groups of the ligand through short distance.

- The agreement to experiment for the **IDD393** complex with the reduced cofactor seems only fairly convincing ($\Delta n = +0.09$). However, we regard the experimentally recorced protonation effect of $\Delta n = +0.2 \pm 0.04$ as rather small to be quantitatively correctly predictable by our pKa calculations methodology. But it is notable that the correct trend is predicted.

5.3.3 Spiro-hydatoins

The ITC measurements discussed so far strongly suggest that Tyr48 is deprotonated in the oxidized holo enzyme [23] and gets protonated upon ligand binding (**IDD388**, **IDD393**, **BTB**). Complexing hAR with ligands possessing a spiro-hydatoin moiety transfers a basic function into the active site. Due to their pKa of about 8-9 in solution they will be partially protonated and most likely they release their proton upon complexation to Tyr48. Indeed, ITC supports such a scenario: For **Fidarestat**, the experiment detected an uptake of $\Delta n = +0.3 \pm 0.1$ moles of protons for wild type, whereas the Tyr48Phe mutant shows a release of $\Delta n = -0.5 \pm 0.1$. The latter experiment shows that, besides Tyr48, the ligand's

cyclic hydantoin functionality is involved in the change of the protonation inventory. For **Sorbinil**, Δn was measured to be $+0.1 \pm 0.02$. This difference compared to **Fidarestat** is explained by the higher pKa value of the latter ligand in aqueous solution (8.7 [126] instead of 7.9 [125]). For both spiro-hydatoins, the results from ITC can be explained by assuming a fully deprotonated ligand once bound to hAR.

Our calculations are not in line with this experimental finding, because two issues become apparent:

- Only marginal pKa shifts are observed for the ligand upon binding, however, as they fall very close to the pH of the buffer, small changes would provoke strong changes in the protonation inventory. The modulation of the formal charge distribution according to the different models does not take significant impact. Furthermore, the solvation contribution of the ligands leads to intrinsic pKa values higher than the model pKa, and the site-site interactions further shift the pKa to higher values. Obviously, the computational approach does not favor the deprotonated state of the hydantoin.
- The formation of doubly (ligand/Tyr48) or triply (ligand/Tyr48/Lys77) coupled systems with close pKa values depending on the predefined $\epsilon_{Protein}$ results in a very complex system and complicates an accurate assignment of pKa values.

What can we learn from this disagreement? Conclusive answers are difficult to find and we therefore want to consult some structural evidences (see Figure 5.5): An inspection of the binding pocket in the bound state suggests that the deprotonated ligand is favorably stabilized by hydrogen bonds to His110 and Tyr48. These two residues supposedly interact strongly with the ligand, which is also reflected by the strong contributions from the site-site interactions. Possibly, this mutually polarizing system is further stabilized and these effects are not captured by our calculations.

5.4 CONCLUSIONS

In this contribution, we attempted to identify residues involved in a proton transfer observed experimentally by ITC in aldose reductase [23]. We postulate that the active site residue Tyr48 is deprotonated in the holo structure with oxidized cofactor but neutral in the situation with reduced NADPH. Examples have been reported in literature, where a deprotonated tyrosine occurs in the active site:

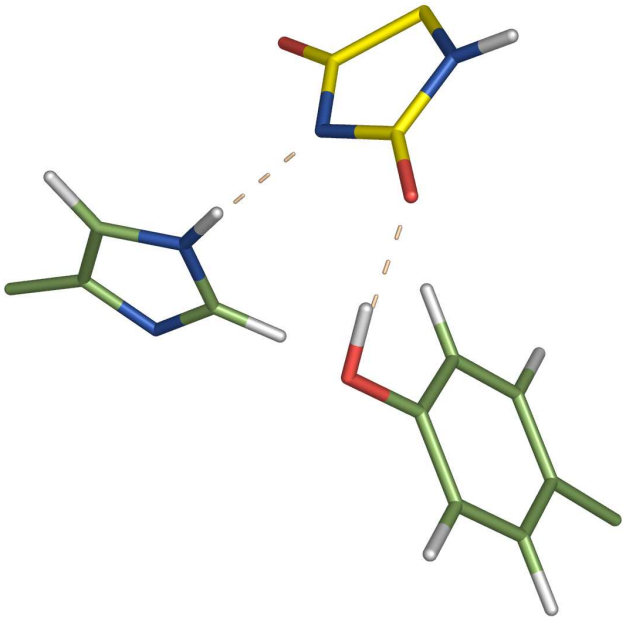

Figure 5.5: Network of H-bonds form by the spiro-hydatoin moiety of **Sorbinil** and **Fidarestat**.

- For two members of the short-chain dehydrogenase/reductase family, namely dehydratase and epimerase, kinetic measurements for an active site tyrosine residue reveal pKa values of 6.41 (dehydratase) [5] and 6.08 (epimerase) [6]. In both enzymes, the crucial tyrosine is neighbored by a lysine residue and the cofactor NAD^+. Both are made responsible for stabilizing the deprotonated tyrosine. In case of the epimerase, it was even possible to experimentally determine the contributions fostering the more acidic pKa value: approximately 3 units come from the positive electrostatic field contribution generated by NAD^+ and the neighbored lysine residue, whereas the hydrogen-bonding to a serine residue contributes about 0.6 units [6].

The structural similiarity of our hAR case with the latter epimerase/dehydratase examples underlines that a deprotonated tyrosine is likely given in holo hAR.

ITC revealed that protons are picked-up, most likely by Tyr48, upon binding of ligands with carboxylic head groups (**IDD388**, **IDD393**, **BTB**). We were able to provide evidence for this interpretation as a large pKa shift for Tyr48 is observed. In the case of **Fidarestat** and **Sorbinil**, the ITC experiment proved that the ligands bind in the deprotonated state. Our calculations propose pKa values for the bound ligand close to the pH of the buffer thus at least partial protonation would have to be assumed. Possibly, this insufficient accuracy results from short-comings of our computational approach which is not capable to correctly

cope with strong polarization effects and the complexity of a coupled protonation system of three adjacent groups with close pKa values.

5.5 MATERIALS AND METHODS

We used the following (publically available) crystal structure of complexed hAR: 1PWM (**Fidarestat**). In-house structures were evaluated for the hAR complexes of **Sorbinil**, **IDD388**, **IDD393** and **BTB**. The 2D structures of the ligands are shown in Figure 5.3. For **Fidarestat** and **Sorbinil**, the experimentally determined pKa values are 7.9 [125] and 8.7 [126]. The experimentally determined pKa value of IDD594 is 2.9 [130]. The structural similarity to the ligands **IDD388** and **IDD393** tempted us to perform the pKa calculations with a pKa value of 3.0 for the carboxylic head group. For the sake of consistency, a pKa value of 3.0 was also assumed for the ligand **BTB**. However, changing the pKa value to 2.0 or 4.0 does not suggest an altered proton inventory.

We used different models to distribute the formal charge in **Fidarestat**, because the negative formal charge of the deprotonated state is, at least partially, reallocated over the hydatoin ring: with our PEOE_PB algorithm, it is only possible to deal with one such possible reallocation (this is shown in Fiure 5.6a). All details on the settings, including the resulting atomic partial charges, are given in Figure 5.6.

For the pKa calculations, we applied the same methodology as recently described elsewhere [85]. MEAD was employed as PB solver [64], whereas we performed calculations with a dielectric constant of 20 and 10, in parallel, to demonstrate its influence. REDUCE was used to generate hydrogens for the residues of the protein [75]. For the calculations of the complexes, the orientation of the Tyr48 OH group was chosen to point towards the ligand. SYBYL was applied to add hydrogens to the cofactor and the ligands [101]. To study the protonation effects upon ligand binding, two pKa calculations are required: One without ligand and a second with the bound ligand. Only the titratable residues within a 12 Å sphere around the active site were selected for evaluation of the site-site interactions in order to reduce the computational effort (in total, there are 91 titratable residues in hAR). This results in the following 26 titratable residues: Lys21, Asp43, Cys44, His46, Tyr48, Glu51, Lys77, Cys80, Tyr82, His83, His110, Lys116, Lys119, Glu120, Glu185, Cys186, Tyr209, Asp216, Lys262, Cys298, Cys303, His306, Lys307, Asp308, Tyr309, His312. We used the higher populated conformer for amino acids in case a split conformation has been assigned to in the crystal structure.

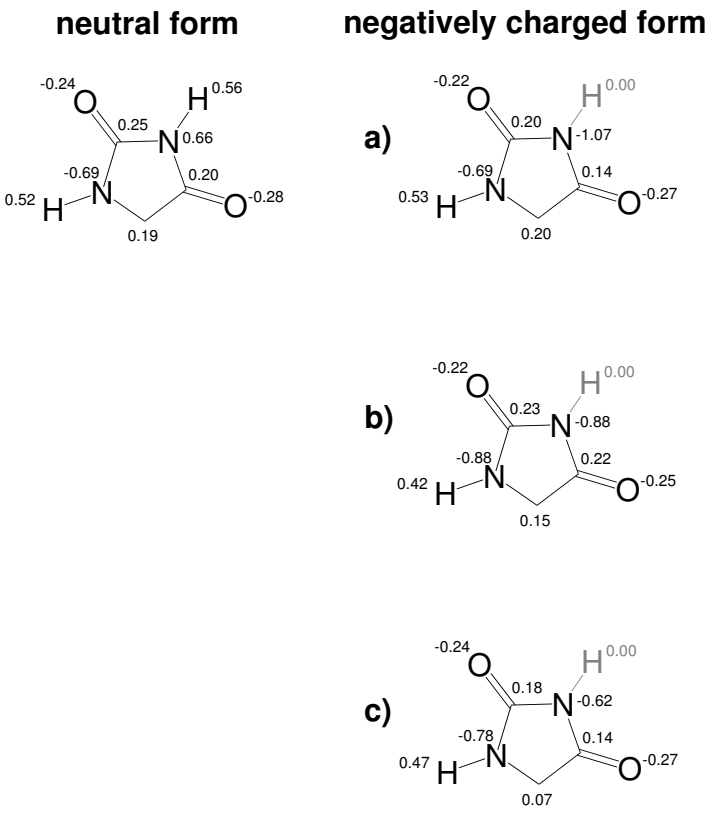

Figure 5.6: Atomic partial charges for the neutral and charged form of **Fidarestat**. Three different variants have been applied for the deprotonated state: (a) The acidic nitrogen atom is assigned to an initial formal charge of -1.0. (b) Both ring nitrogens are assigned initially to a formal charge of -0.5. (c) Each ring atom bears an initial formal charge of -0.2.

In contrast to our standard settings for the titration of histidine residues, the N_δ atom of His110 was titrated (see **Results** section).

For (doubly or triply) coupled systems, we superimposed the titration curves and derived the pKa values from the intersections at, e.g. $pKa_{2.5}/pKa_{1.5}/pKa_{0.5}$.

6 PDB2PQR AS A NEW TOOL FOR THE SETUP OF PKA CALCULATIONS ON PROTEIN-LIGAND COMPLEXES

6.1 INTRODUCTION

The setup of pKa calculations in a fairly straight-forward manner of protein-ligand complexes has become feasible on the basis of our PEOE_PB charges [85]. Their applications shows convincing agreement to experimental data for a series of complexes [111, 123]. However, the current setup still requires an elaborate procedure especially for the handling of the titratable groups of a ligand. Furthermore, all previous calculations have been performed using the PBE (Poisson-Boltzmann equation) solver MEAD [64], which is no longer maintained and supported: this fact puts further development in question. To circumvent these issues, we decided to combine our algorithm for the PEOE_PB charge assignment into within the framework of PDB2PQR [27] and we extend it to the automated handling of ligands. PDB2PQR is a convenient tool for the generation of PQR files, which serve as input for the PBE solver APBS [131]; the PQR file contains charges and radii for all protein atoms.

The workflow for setting up a pKa calculation with PDB2PQR is shown schematically in Figure 6.1. Generally, it is sufficient to submit a PDB file for the protein and a MOL2 file for the ligand, and all necessary steps are conveniently performed by PDB2PQR without interference of the user. The organization of this technically oriented report is as follows: in section 6.2, the general workflow will be presented. This description will be instructive for the interested non-expert to gain insights into the concept of PDB2PQR. In the subsequent section 6.3, a detailed characterization of the involved scripts is given, which is of main relevance for developers.

6.2 PDB2PQR PRINCIPLES

PDB2PQR is a Python (version 2.3.5, www.python.org) based command line tool, which is controlled via various flags. The standard invocation (addition of optimized hydrogen positions to a protein structure and a debumping of missing atoms) is as follows:

```
pdb2pqr.py --ff=charmm --apbs-input 2LZT.pdb 2LZT_H.pdb
```

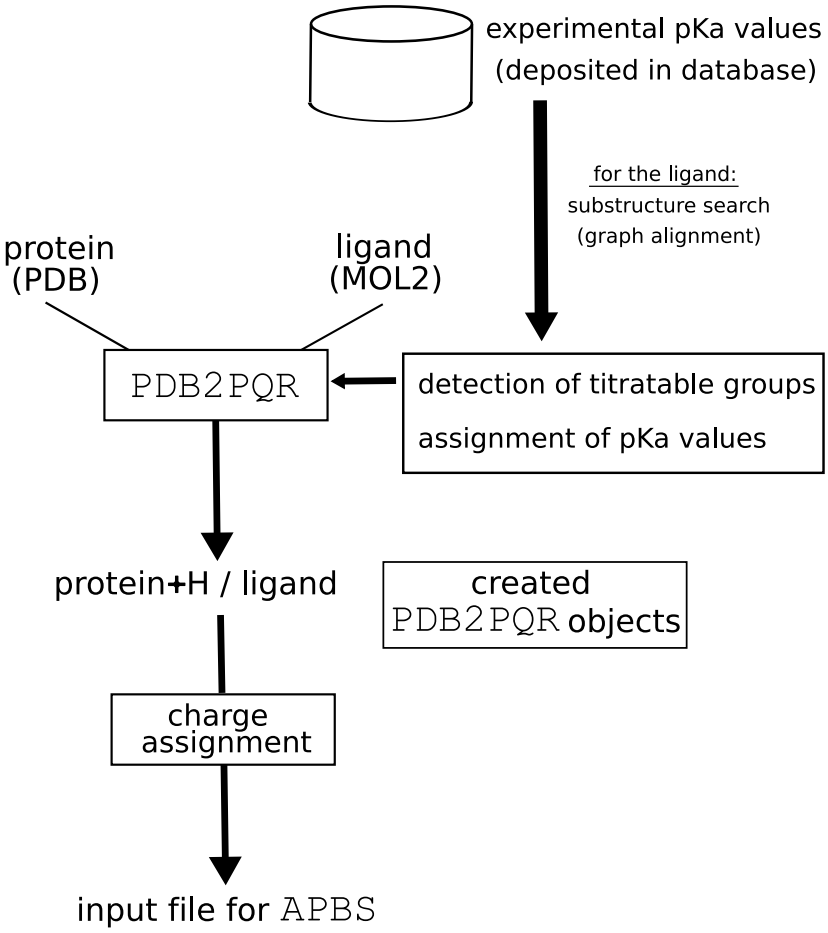

Figure 6.1: Workflow for the setup of pKa calculations for a protein-ligand complex with PDB2PQR.

The `--ff=charmm` flag switches to the CHARMM22 [77] parameters for the protein, other available parameter sets are AMBER99 [132] and PARSE [65]. The option `--apbs-input` generates an input file for a calculation with APBS. Other optional switches are `--nodebump` and `--noopt`, which suppress the debumping and hydrogen position optimization, respectively.

For the preparation of an input suitable for pKa calculations, the Python script `pka.py` has been written by Jens Erik Nielsen, and we included the extensive ligand processing. The script `pka.py` is invoked as follows:

```
pka.py --ff=peoepb --lig test.mol2 test.pdb
```

Basically, `pka.py` creates all PDB2PQR objects for the ligand (the details are given in section 6.3) and makes a full flexible treatment of the ligand possible within `pka.py`. The PEOE_PB charges are assigned in the beginning, and a substructure search for the detection of titratable groups follows. The substructure matching is necessary, since no general rule exists of how to detect the titratable groups of a ligand and to derive the corresponding model pKa values. For this purpose, a database (http://polymerase.ucd.ie/pka_lig_tool) was created [133] in which structural data for molecules with experimentally determined pKa values (in aqueous solution) are stored. Two possibilities exist for the deposition of data:

- via the graphical online front end: it can be accessed by a web browser, the data is entered to the system by an appropiate interface on a web page, and the required 3D structure is uploaded as MOL2 file.
  This procedure represents a convenient way for populating the database, although it can be not done in a *high-throughput* manner.

- via a Python script (`molpka.py`): it uses 2D MDL MOL files as input, the program CORINA [71] converts these structures into 3D MOL2 files, and the necessary entries can be entered interactively.
  At first glance, this procedure seems rather tedious, but displays its strength when processing large quantities of structural data.

The basis for the experimental data is built by a pKa compilation (http://research.chem.pse.edu/brpgroup/pKa_compilation.pdf). In the current version, there are ca. 350 entries in the database covering a large variety of different chemical structures.

After a successful matching, the query molecule from the database is converted into a

PDB2PQR object. Then, `pka.py` assembles all settings necessary for the titration of the ligand (model pKa value, atom(s) being protonated/deprotonated, type of titratable group) together with the protein settings. In a final step, the input file for APBS is generated.

6.3 SCRIPTS

All routines called from `pka.py` for the ligand processing are given in a schematic workflow shown in Figure 6.2. In the top part of this figure, all scripts added and modified in the course

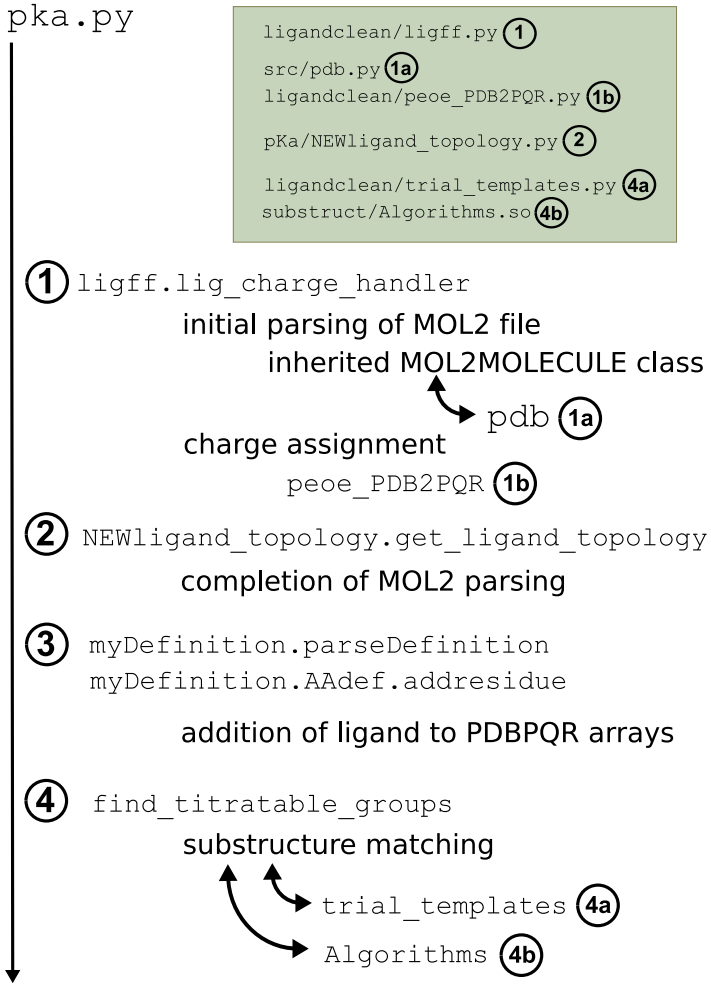

Figure 6.2: Routines called by `pka.py`. All scripts added and modified for the ligand extension are given in the grey box in the top right corner. Only protein-specific definitions are set in step (3), no modifications have been done here and, thus, it is not annotated in the top of this figure.

of this project are listed. The invoked routines are commented in the following enumeration (the numbering corresponds to the indices in Figure 6.2):

1. In the first step, the parsing of the ligand MOL2 file is initiated. The `MOL2MOLECULE`

class has been added to the `pdb` module to implement MOL2 parsing. Furthermore, the `HETATM` class has been modified in the `pdb` module. The different objects created for the ligand by PDB2PQR are shown in Figure 6.3.
Furthermore, the ligand parsing process includes the charge assignment procedure (`peoe_PDB2PQR`).

2. The MOL2 parsing is completed after the ligand object has inherited all attributes from the `get_ligand_topology` class, which belongs to the `NEWligand_topology` module.

3. Amino acid and rotamer definions are set by `myDefintion` which calls the `definitions` module. However, no such information is available for the ligand and, thus, the corresponding classes are set empty. Accordingly, no modification were done in this module.

4. Finally, the substructure matching is called within `pka.py` by the routine `find_titratable_groups`. This function itself uses a shared object for the clique detection algorithm (`Algorithms.so`), see Section 6.3.1 for further details.
The library of molecules with experimentally determined pKa values is deposited in the module `trial_templates`. An examplary entry of this module is given in Figure 6.4.

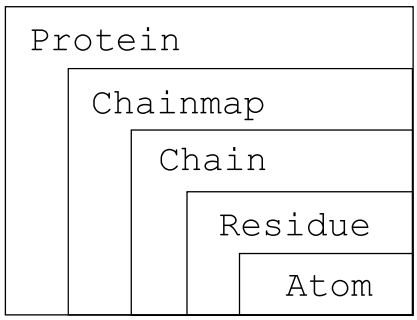

Figure 6.3: Hierarchy of the PDB2PQR classes.

```
templates['COO'] =
{'O10':{'sybylType':'O.co2','neighbours':['C3'],'alreadyvisited':False},
 'O20':{'sybylType':'O.co2','neighbours':['C3'],'alreadyvisited':False},
 'C3': {'sybylType':'C.2',  'neighbours':['O10','O20'],'alreadyvisited':False}}
```

Figure 6.4: Examplary entry of the module `trial_templates` containing the query molecules for substructure matching..

6.3.1 Substructure matching

The substructure search is based on a graph alignment algorithm implemented in our group [134], it is a routine written in *C++* and can be called as shared object from Python. The

Bron-Kerbosch-Algorithm forms the methodological basis of the substructure matching: it detects the maximum common subgraph by finding the maximum clique of a graph [135] (see Figure 6.5). Prior to the substructure search, the ligand and query molecules are converted

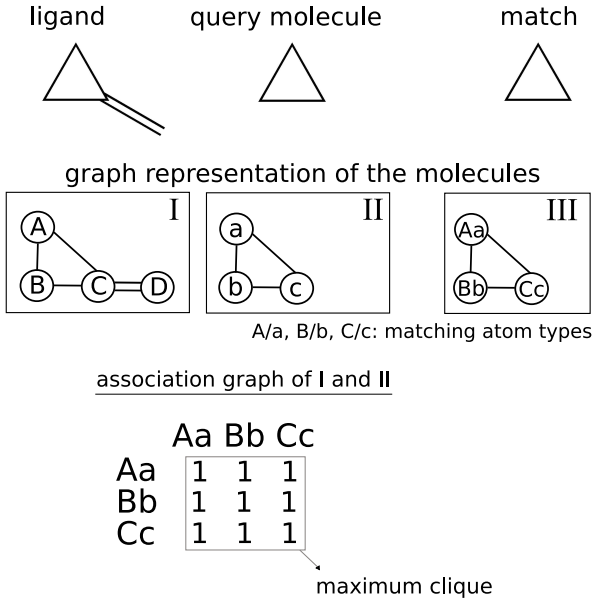

Figure 6.5: Simple example of the graph alignment algorithm detecting the maximum common subgraph. The ligand (I) and the the query molecule (II) are given as molecules and the corresponding graphs. The association graph of the two graphs is given in the lower part of this Figure. In the present (trivial) case, the maximum clique is formed by the complete association graph. The graph (III) represents the maximum common substructure, which is reported to PDB2PQR.

into an association graph, which forms a suitable basis for the clique detection. When the substructure matching is performed, we look for the maximum common substructure/subgraph: this means that a match is only reported if the whole query fragment is found in the ligand.

6.3.2 pka_lig_tool

The database `pka_lig_tool` containing structural data for molecules with experimentally determined pKa values is written in MySQL (version 4.1.16, www.mysql.com). It was initated by the cooperation with Jens-Erik Nielsen and maintained by Chresten Søndergaard. The organization is given in Figure 6.6. Four different tables have been created in which the data is organized:

- the `ligands` table: it covers the structural data of the ligand such as the MOL2 file, xyz file (for the JMOL visualisation on the web page), an optional smiles code, ID and name for the ligand

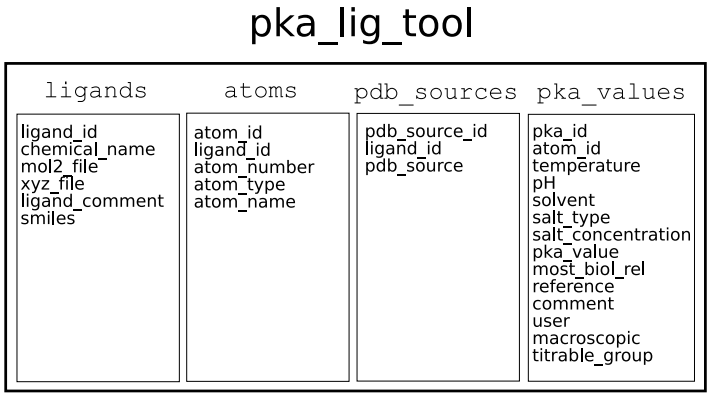

Figure 6.6: Organization of the database pka_lig_tool containing experimentally determined pKa values (in aqueous solution) of small organic molecules.

- the `atoms` table: it contains atomic entries such as name, ID and MOL2 type and the ID of the ligand
- the `pdb_source` table: if ligand data originates from a protein-ligand complex, the corresponding PDB code is deposited here
- the `pka_values` table: the experimental values are stored here plus annotations such as literature references, solvent, temperature etc.

The content of the database can be converted to a format suitable for the substructure matching (see Figure 6.4 for an exemplare entry) by a Python script (`convert.py`).

7 SUMMARY,ZUSAMMENFASSUNG

7.1 Summary

Recent hardware development increase the computing power, in consequence many biological and chemical processes can now be successfully modelled in a way which was not to imagine 20 years ago. Examples of such processes are molecular dynamics studies of large biomolecules, the prediction of free energy of binding for protein-ligand complexes, investigations of reaction paths in enzymes, to mention only a few. One issue which is still unresolved concerns the accurate estimation of protonation states in protein-ligand complexes.

In this thesis, we present the development of a novel charge assignment procedure named PEOE_PB (Partial Equalisation of Orbital Electronegativities - optimized for Poisson-Boltzmann calculations), which represents a method for the assignment of atomic partial charges. It works reliably with both proteins and small organic molecules using a consistent approach. Such charges are a key parameter in Poisson-Boltzmann (PB) calculations, which are a well-established method for pKa calculations of proteins. The development of the PEOE_PB charges is necessary, because no generic procedure exists to perform PB-based pKa calculations of protein-ligand complexes. The development of the PEOE_PB charges is described in section 2.

We adapted the PEOE formalism to optimally predict solvation free energies of small organic molecules. Modifications were performed in a heuristic manner, i.e. purely result-oriented. The changes considered exclusively parameter a of the PEOE polynomial (see section 2.2.1). In our optimization protocol, we first attempted to reproduce best solvation free energies of the polar amino acids (r^2 is 0.94, RMSD is 0.84) and continued with a dataset of 80 small organic molecules (r^2 is 0.78, RMSD is 1.57). The latter step underlines the generic character of our PEOE_PB charges. Subsequently, we performed calculations on a dataset of nine (apo) proteins with 132 experimentally determined pKa values and obtained an overall RMSD of 0.88. The dielectric constant of the protein is set to 20. Active site residues with highly shifted pKa values are given for two enzymes of the dataset. For these, the following issues were indicated:

- The value of the dielectric constant has to be lowered from 20 to 4, which can be partially explained by the degree of burial of the binding pocket.
- The orientation of the tyrosine OH group was observed to have a remarkable influence

on the pKa value of an active site residue (a glutamate with a highly increased pKa value). This fact emphasizes that the position of polar hydrogens can be crucial.

In the final step of our PEOE_PB validation study, we performed pKa calculations for three protein-ligand complexes, where experiment revealed a proton transfer. All our calculations agree with the experimental findings.

In section 3, we describe pKa calculations for a series of ligands binding to the serine proteases **trypsin** and **thrombin**. For the studied complexes, we previously performed an extensive ITC and crystallographic study and were able to identify protonation changes for four complexes [22]. However, since ITC measures only the overall proton exchange, it does not provide structural insights into the functional groups involved in the proton transfer. By using Poisson-Boltzmann calculations based on our PEOE_PB charges, we compute pKa values for all complexes from our previous work in order to reveal the residues with altered protonation states. The results indicate that His57, a member of the catalytic triad, is responsible for the most relevant pKa shifts resulting in the experimentally detected protonation changes. This finding is in contrast to our previous assumption that the observed protonation changes occur at the carboxylic group of the ligands. The newly detected proton acceptor is used for a revised factorization of the ITC data, which is necessary in cases where the protonation inventory changes upon complexation. pKa values of complexes showing no change of protonation in the ITC experiment are reliably predicted in most cases, whereas predictions of strongly coupled systems remain problematic. Such coupled systems appear if two (or more) strongly interacting titratable groups are close to each other.

The **HIV protease** (HIVP) is a prominent example of successful structure-based drug design, and it is a well-studied system for which, as experimental evidence shows, protonation changes in the active site occur upon ligand binding (section 4). In the apo enzyme, the catalytic dyad consisting of two aspartates is in the mono-protonated state. This protonation state can be altered by ligands bearing a cyclic urea moiety. Our PEOE_PB charge model reliably suggests the experimentally determined protonation states in the active site of HIVP. Furthermore, we perform pKa calculations for two HIVP complexes with novel types of inhibitors developed and synthesized in our group [115, 114]. For these complexes, no experimental knowledge on the protonation states is given. For one of the compounds, containing a central pyrrolidine ring, the calculations predict that both catalytic aspartates should be deprotonated upon ligand binding (Table 4.12 and 4.13). Such a protonation pattern has yet not been observed in any HIVP complex.

Similarly to the experimental trypsin/thrombin study, a combined crystallographic and

thermodynamic investigation of the ligand complexation process of **human aldose reductase** (hAR) was performed in our group [23]. The ITC measurements detected a proton transfer induced by the ligand binding process. Our calculations suggest an active site tyrosine residue (Tyr48) as proton acceptor, which is equivalent to a deprotonated tyrosine in the holo enzyme. This is in agreement with the ITC results for the Tyr48Phe mutant. Good quantitative agreement with the ITC experiment is obtained for the hAR complexes with inhibitors bearing a carboxylic head group. In contrast, the calculations cannot reliably predict the properties of inhibitors with a cyclic hydantoin moiety. A possible explanation for the deficiency is the fact that the ligand shows a strong electrostatic interaction with the active site tyrosine and lysine residues. Furthermore, its pKa value in aqueous solutions falls next tothe physiological pH range which makes the system very sensitive to the actual pKa shifts.

One limiting factor for the large-scale application of the pKa calculation methodology presented here concerns the ligand processing. For this purpose, we included the PEOE_PB algorithm in the PDB2PQR program. This tool serves as the input file generator for the Poisson-Boltzmann solver APBS. Besides the fully flexible ligand consideration within the PDB2PQR framework, substructure matching has been enabled for the ligand. With this technique, it is possible to automatically detect titratable groups of the ligand and assign pKa values. These pKa values originate from a database which has designed built and currently contains 348 molecules with experimentally determined pKa values.

7.2 Zusammenfassung

Die ständige Weiterentwicklung der Computer-Hardware und die daraus resultierende Steigerung der Rechenleistung ermöglicht heutzutage eine erfolgreiche Modellierung von chemischen und biologischen Prozessen, die vor 20 Jahren noch undenkbar war. Als Beispiele sind Molekulardynamik-Simulationen grosser Biomoleküle, die Berechnung freier Bindungsenergien von Protein-Ligand-Komplexen oder auch Untersuchungen von Reaktionswegen in Enzymen zu nennen. In einem Bereich mangelt es jedoch weiterhin an akkuraten Methoden: die Abschätzung von Protonierungszuständen in Protein-Ligand-Komplexen.

In der vorliegenden Arbeit zeigen wir die Entwicklung einer neuen Methode der Ladungszuweisung, genannt PEOE_PB (*Partial Equalisation of Orbital Electronegativities* - optimiert für *Poisson-Boltzmann* Rechnungen). Diese Methode stellt eine konsistente Ladungszuweisung sowohl für Proteine als auch für kleine organische Moleküle dar. Die Ladungen sind entscheidende Parameter bei Poisson-Boltzmann (PB)- Rechnungen. PB-Rechnungen stellen eine

etablierte Methode bei der Bestimmung von pKa-Werten in Proteinen dar. Die Entwicklung von PEOE_PB-Ladungen ist notwendig geworden, da es keine generische Methode gibt, PB-basierte pKa-Berechnungen in Protein-Ligand-Komplexen durchzuführen.

Der PEOE-Ansatz wurde gewählt, um zunächst die freien Solvatationsenergien kleiner organischer Moleküle bestmöglich vorherzusagen. Modifikationen wurden heuristisch, d.h. ergebnisorientiert vorgenommen, wobei Änderungen lediglich den Parameter a des PEOE-Polynoms betreffen. Bei unserer Optimierung versuchten wir zunächst, die experimentell bestimmten freien Solvatationsenergien polarer Aminosäuren ($r^2 = 0.94$, RMSD = 0.84) und anschließend eines Datensatzes von 80 kleinen organischen Molekülen ($r^2 = 0.78$, RMSD = 1.57) zu reproduzieren. Die Verwendung des letztgenannten Datensatzes zeigt den generischen Charakter unserer PEOE_PB-Ladungen. Abschliessend führten wir Rechnungen an einem Datensatz von neun (apo-)Proteinen mit 132 experimentell bestimmten pKa-Werten durch und erzielten einen RMSD von 0.88. Die Dielektrizitätskonstante im Protein war hierbei auf 20 gesetzt. Aminosäurereste in Bindetaschen mit stark verschobenen pKa-Werten lagen bei zwei Enzymen des Datensatzes vor, in diesen Fällen wurden folgende Beobachtungen gemacht:

- Die Dielektrizitätskonstante musste von 20 auf 4 gesenkt werden, was teilweise durch die Vergrabenheit der Bindetasche erklärt werden kann.
- Die Orientierung der Hydroxylgruppe des Tyrosins hatte einen beachtlichen Einfluss auf den pKa-Wert eines Aminosäurerestes in der Bindetasche (ein Glutamat mit stark erhöhtem pKa-Wert). Diese Tatsache unterstreicht den entscheidenden Einfluss der Orientierung polarer Wasserstoffatome.

Im letzten Schritt unserer PEOE_PB-Validierung führten wir pKa-Berechnungen für drei Protein-Ligand-Komplexe (die im Experiment einen Protonentransfer zeigten) durch: in allen Fällen stimmten unsere Berechnungen mit dem Experiment überein.

In einer folgenden reinen Anwendungsstudie führten wir pKa-Rechnungen für eine Serie von Liganden, die an die Serin-Proteasen **Trypsin** und **Thrombin** binden, durch. Für diese Komplexe waren bereits ausführliche ITC- und Kristallographie-Studien gemacht worden, und für vier dieser Komplexe konnten Änderungen in den Protonierungszuständen detektiert werden [22]. Da ITC-Experimente jedoch nur gesamtheitliche Änderungen in der Protonierung messen, konnten diese Experimente keinen Aufschluss darüber geben, welche funktionellen Gruppen tatsächlich am Protonentransfer beteiligt sind. Um diese Gruppen identifizieren zu können, führten wir PB-Rechnungen, basierend auf unseren PEOE_PB-Ladungen, durch. Die resultierenden pKa-Werte zeigen, dass His57 (einer der drei katalytisch

aktiven Reste) für die wichtigsten pKa-Änderungen, die sich im Experiment als Änderungen im Protonierungszustand zeigen, verantwortlich ist. Dies steht im Widerspruch zu unserer früheren Annahme, dass die Änderungen im Protonierugszustand an der Carboxylgruppe der Liganden stattfinden. Der neuentdeckte Protonenakzeptor wurde für die Refaktorisierung der ITC-Daten eingesetzt; dies ist wichtig für Fälle, in denen sich die Protonierung während der Komplexbildung ändert. Die pKa-Werte von Komplexen, die im ITC-Experiment keine Änderung im Protonierungszustand zeigen, werden in den meisten Fällen verlässlich vorhergesagt, während dies in Fällen stark koppelnder Systeme schwierig bleibt. Solche Fälle treten auf, wenn zwei (oder mehr) interagierende titrierbare Gruppen räumlich nahe beiander liegen.

Die **HIV-Protease** (HIVP) ist ein bekanntes Beispiel für erfolgreiches strukturbasiertes Wirkstoffdesign und stellt ein gut untersuchtes System dar, bei dem Änderungen des Protonierungszustandes während der Ligandenbindung auftreten, wie im Experiment gezeigt wurde. Das System der HIVP stellt einen Ausgangspunkt für eine weitere Anwendungsstudie unserer PEOE_PB-Ladungen dar. Bei dem Apo-Enzym befindet sich die zwei katalytisch aktiven Reste (Aspartate) im monoprotonierten Zustand. Dieser kann sich ändern, wenn Liganden binden, die eine zyklische Harnstoff-Gruppe enthalten. Unser PEOE_PB-Modell reproduziert den experimentell bestimmten Protonierungszustand. Ferner führten wir pKa-Berechnungen für zwei HIVP-Komplexe mit neuartigen Inhibitoren, die unserer Gruppe entwickelt und synthetisiert wurden [115, 114], durch. In diesen Fällen gibt es keinerlei experimentelle Daten für die Protonierungszustände. Einer der Inhibitoren enthält ein Pyrrolidin-Ring: hier sagten die Berechnungen voraus, dass beide katalytisch aktiven Aspartate nach Ligandenbindung deprotoniert vorliegen. Solch ein Protonierungsmuster wurde bisher in keinem HIVP-Komplex beobachtet, weder experimentell noch mittels einer Berechnungsmethode.

Neben den experimentellen Trypsin/Thrombin-Studien wurden auch kombinierte kristallographische und thermodynamische Untersuchungen der Ligandenbindung an **humaner Aldose-Reduktase** (hAR) in unserer Gruppe vorgenommen [23]. Die ITC-Messungen zeigten einen durch die Ligandenbindung induzierten Protonentransfer. Unsere pKa-Rechnungen lassen darauf schließen, dass ein Tyrosin-Rest der Bindetasche (Tyr48) als Protonenakzeptor fungiert, was bedeutet, dass das Tyrosin im Holo-Enzym deprotoniert vorliegt. Dies stimmt mit den Ergebnissen von ITC-Messungen an Tyr48Phe-Mutanten überein. Während bei hAR-Komplexen von Inhibitoren mit einer Carboxyl-Kopfgruppe die Rechnungen gut im Einklang mit den ITC-Experimenten standen, zeigten sich Ungenauigkeiten bei der Vorhersage von Inhibitoren mit einer zyklischen Hydantoin-Gruppe. Eine mögliche Erklärung hierfür ist die starke elektrostatische Wechselwirkung zwischen Ligand und den Tyrosin bzw. Lysin-Resten der Bindetasche. Ferner liegen die pKa-Werte in wässriger Lösung

nahe dem physiologischen pH-Bereich, was das System sehr anfällig für kleine Änderungen des pKa-Wertes macht.

Ein limitierender Faktor für die breite Anwendung unserer PEOE_PB Ladungsmethode bei pKa-Rechnungen stellt die vorangehende Prozessierung der Liganden dar. Zu diesem Zweck implementierten wir den PEOE_PB-Algorithmus in das PDB2PQR-Programm (dieses erzeugt Input-Dateien für das PB-Programm APBS). Liganden wurden in der PDB2PQR-Umgebung als voll flexibel betrachtet und es wurde eine Suchprozedur für gemeinsame Substrukturen eingeführt. Mit dieser Technik ist es möglich, titrierbare Gruppen des Liganden automatisch zu erkennen und ihnen pKa-Werte zuzuweisen. Diese pKa-Werte stammen aus einer Datenbank, die momentan 348 Moleküle mit experimentell bestimmten pKa-Werten enthält.

Bibliography

[1] D. Xie, S. Gulnik, L. Collins, E. Gustchina, L. Suvorov, and J.W. Erickson. Dissection of the ph dependence of inhibitor binding energetics for an aspartic protease: Direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry*, 36:16166–16172, 1997.

[2] S. Kuramitsu and K. Hamaguchi. Analysis of the acid-base titration curve of hen lysozyme. *J Biochem (Tokyo)*, 87(4):1215–9, 1980.

[3] M.F. Jeng and H.J. Dyson. Direct measurement of the aspartic acid 26 pKa for reduced Escherichia coli thioredoxin by 13C NMR. *Biochemistry*, 35(1):1–6, 1996.

[4] P.A. Tishmack, D. Bashford, E. Harms, and R.L. Van Etten. Use of 1H NMR spectroscopy and computer simulations to analyze histidine pKa changes in a protein tyrosine phosphatase: experimental and theoretical determination of electrostatic properties in a small protein. *Biochemistry*, 36:11984–11994, 1997.

[5] B. Gerratana, W.W. Cleland, and P.A. Frey. Mechanistic roles of Thr134, Tyr160, and Lys 164 in the reaction catalyzed by dTDP-glucose 4,6-dehydratase. *Biochemistry*, 40(31):9187–95, 2001.

[6] Y. Liu, J.B. Thoden, J. Kim, E. Berger, A.M. Gulick, F.J. Ruzicka, H.M. Holden, and P.A. Frey. Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli. *Biochemistry*, 36(35):10675–84, 1997.

[7] L.A. Highbarger, J.A. Gerlt, and G.L. Kenyon. Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115. *Biochemistry*, 35(1):41–6, 1996.

[8] C. Tanford. Protein denaturation. *Adv Protein Chem*, 25:1–95, 1970.

[9] M.F. Perutz. Electrostatic effects in proteins. *Science*, 201:1187–1191, 1978.

[10] A. Warshel. Calculations of Enzymic Reactions: Calculations of pKa, Proton Transfer Reactions, and General Acid Catalysis Reactions in Enzymes. *Biochemistry*, 20:3167–3177, 1981.

[11] A. Warshel and S.T. Russell. Calculations of Electrostatic Interactions in Biological Systems and in Solutions. *Quart. Rev. Biophys.*, 17:283–422, 1984.

[12] J.B. Matthew. Electrostatic effetcs in proteins. *Annu Rev Biophys Biomol Struct*, 14:387–417, 1985.

[13] K. Sharp, R. Fine, K. Schulten, and B. Honig. Brownian dynamics simulation of diffusion to irregular bodies. *J Phys Chem*, 91:3624–3631, 1987.

[14] M.K. Gilson and B. Honig. Calculation of the electrostatic energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. *Proteins*, 4:7–18, 1988.

[15] A. Warshel and J. Åqvist. Electrostatic Energy and Macromolecular Function. *Ann. Rev. Biophys. Biophys. Chem.*, 20:267–298, 1991.

[16] A. Fersht. *The pH dependance of enzyme catalysis. Enzyme structure and mechanism.* 1985.

[17] D. Bashford and M. Karplus. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, 29(44):10219–25, 1990.

[18] J. Antosiewicz, J.A. McCammon, and M.K. Gilson. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.*, 238:415–436, 1994.

[19] E. Demchuk and R.C. Wade. Improving the continuum dielectric approach to calculating pKas of ionizable groups in proteins. *J. Phys. Chem.*, 100:17373–17387, 1996.

[20] J.E. Nielsen and G. Vriend. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. *Proteins*, 43(4):403–12, 2001.

[21] J.E. Nielsen and J.A. McCammon. Calculating pKa values in enzyme active sites. *Protein Sci*, 12:1894–1901, 2003.

[22] F. Dullweber, M.T. Stubbs, D. Musil, J. Stürzebecher, and G. Klebe. Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J Mol Biol*, 313(3):593–614, 2001.

[23] H.M. Steuber. *Structural and thermodynamic characterization of inhibitor binding to Aldose Reductase by X-ray crystallography and isothermal titration calorimetry in order to understand driving forces and selectivity determinants.* PhD Thesis, Philipps-University Marburg Germany, 2006.

[24] E. Ido, H. Han, F.J. Kezdy, and J. Tang. Kinetic studies of Human Immunodeficiency Virus type 1 protease and its active-site hydrogen bond mutant A28S. *J. Biol. Chem.*, 266:24349–24366, 1991.

[25] L.J. Hyland, T.A. Tomaszek, and T.D. Meek. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, 30:8454–8463, 1991.

[26] Y.X. Wang, D.I. Freedberg, T. Yamazaki, P.T. Wingfield, S.J. Stahl, D. Kaufman, Y. Kiso, and D.A. Torchia. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry*, 35:9945–9950, 1996.

[27] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, and N.A. Baker. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*, 32:665–667, 2004.

[28] K. Linderstrom-Lang. Om Proteinstoffernes Ionisation. *CR Trav Lab Carlsberg*, 15:1–29, 1924.

[29] C. Tanford and J.G. Kirkwood. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J Am Chem Soc*, 79:5333, 1957.

[30] C. Tanford and R. Roxford. Interpretation of protein titration curves. Application to lysozyme. Biochemistry. *Biochemistry*, 11:2192–2198, 1972.

[31] A. Warshel and M. Levitt. Theoretical Studies of Enzymatic Reactions: Dielectric Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.*, 249:103:227, 1976.

[32] Y.Y. Sham, Z.T. Chu, and A. Warshel. Consistent Calculations of pKa's of Ionizable Residues in Proteins: Semi-Microscopic and Macroscopic Approaches. 101:4458–4472, 1997.

[33] A. Warshel, F. Sussman, and G. King. Free Energy of Charges in Solvated Proteins: Microscopic Calculations Using a Reversible Charging Process. *Biochemistry*, 25(8368-8372), 1986.

[34] C.N. Schutz and A. Warshel. What are the Dielectric "Constants" of Proteins and How To Validate Electrostatic Models? *Proteins*, 44:400–417, 2001.

[35] M.S. Wisz and H.W. Hellinga. An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. *Proteins*, 51(3):360–77, 2003.

[36] H. Li, A.D. Robertson, and J.H. Jensen. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61:704–721, 2005.

[37] E.L. Mehler and F. Guarnieri. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys J*, 77(1):3–22, 1999.

[38] U. Börjesson and P.H. Hünenberger. Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to simple amines. *J. Chem. Phys.*, 114:9706–9719, 2001.

[39] A.M. Baptista, P.J. Martel, and S.B. Petersen. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins*, 27:523–544, 1997.

[40] A.R. Klingen and G.M. Ullman. Negatively-charged Aminoacids and Hydrogen Bond Pattern Tune the pKa Values of the Rieske-type Iron-Sulfur Proteins. *Biochemistry*, 43,:12383 –12389, 2004.

[41] E.J.W. Verwey and J.T.G. Overbeek. *Theory of the Stability of Lyophobic Colloids. The Interaction of Sol Particles Having an Electric Double Layer*. Elsevier, New York, 1948.

[42] M.K. Gilson. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins*, 15(3):266–82, 1993.

[43] A.S. Yang, M.R. Gunner, R. Sampogna, K. Sharp, and B. Honig. On the calculation of pKas in proteins. *Proteins*, 15(3):252–65, 1993.

[44] B. Brooks, R.E. Bruccoleri, B.DOlafson, D.J. States, S. Swaminathan, and M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187 ff, 1983.

[45] W. D. Cornell, P. Cieplak, and C.I. Bayly. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc*, 117:5179–5197, 1995.

[46] W.L. Jorgensen and J. Tirado-Rives. The OPLS potential function for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657 ff, 1988.

[47] D. Bashford, D.A. Case, C. Dalvit, L. Tennant, and P.E. Wright. Electrostatic calculations of side chain pKa values in myoglobin and comparison with nmr data for histidines. *Biochemistry*, 32(31):8045–8056, 1993.

[48] S.T. Russell and A. Warshel. Calculations of Electrostatic Energies in Proteins; The Energetics of Ionized Groups in Bovine Pancreatic Trypsin Inhibitor. *J. Mol. Biol.*, 185:389–404, 1985.

[49] A. Warshel and Z.T. Chu. *Calculations of Solvation Free Energies in Chemistry and Biology.* 1994.

[50] D. Sitkoff, N. Ben-Tal, and B.Honig. Calculation of alkane to water solvation free energies using continuum solvent models. *J. Phys. Chem.*, 100:2744–2752, 1996.

[51] W.R. Cannon, B.J. Garrison, and S.J. Benkovic. Consideration of the pH-dependent inhibition of dihydrofolate reductase by methotrexate. *J Mol Biol*, 271(4):656–68, 1997.

[52] J. Trylska, J. Antosiewicz, M. Geller, C.N. Hodge, R.M. Klabe, M.S. Head, and M.K. Gilson. Thermodynamic linkage between the binding of protons and inhibitors to HIV-1 protease. *Protein Sci*, 8:180–95, 1999.

[53] T. Hansson and J. Åqvist. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Engineering*, 8:1137–1144, 1995.

[54] J. Marelius, M. Graffner-Nordberg, and Hansson T. Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *Journal of Computer-Aided Molecular Design*, 12:119 – 131, 1998.

[55] J.W. Storer, D.J. Giesen, C.J. Cramer, and D.G. Truhlar. Class IV Charge Models: A New Semiempirical Approach in Quantum Chemistry. *J. Comput.-Aid. Mol. Des.*, 9:87–110, 1995.

[56] C. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, 97:10269–10280, 1993.

[57] J. Gasteiger and M. Marsili. Iterative Partial Equalization of Orbital Electronegativity - A rapid Access to Atomic Charges. *Tetrahedron*, 36:3219–3228, 1980.

[58] G. Del Re. A simple MO-LCAO method for the calculation of charge distributions in saturated organic molecules. *J. Chem. Soc.*, 2:4031–4040, 1958.

[59] R.S. Mulliken. A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *J. Chem. Phys*, 2:782–793, 1934.

[60] J. Hinze and H.H. Jaffe. Electronegativity. I. Orbital Electronegativity of neutral atoms. *J. Am. Chem. Soc.*, 84:540–546, 1962.

[61] J. Hinze and H.H. Jaffe. Electronegativity. IV. Orbital electronegativities of the neutral atoms of the periods three a and four a and of positive ions of periods one and two. *J Phys. Chem.*, 67:1501–1506, 1963.

[62] J. Gasteiger. *Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds.* Springer, 1988.

[63] R. Wolfenden. Interaction of the peptide bond with solvent water: a vapor phase analysis. *Biochemistry*, 17:201–204, 1978.

[64] D. Bashford. *An object-oriented programming suite for electrostatic effects in biological molecules.* ISCOPE97, 1997.

[65] D. Sitkoff, N. Ben-Tal, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem*, 98:1978–1988, 1994.

[66] A. Pedretti, L. Villa, and G. Vistoli. VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script-programming. *J. Comput.-Aid. Mol. Des.*, 18:167–173, 2004.

[67] V.N. Viswanadhan, A.K. Ghose, U.C. Singh, and J.J. Wendoloski. Prediction of solvation free energies of small organic molecules: additive-constitutive models based on molecular fingerprints and atomic constants. *J. Chem. Inf. Comput. Sci.*, 39:405–412, 1999.

[68] Galaxy 2.3 Molecular Modeling Software Package. *Galaxy 2.3 Molecular Modeling Software Package*. San Antonio, Texas, USA, 1997.

[69] B.J. Smith. Solvation parameters for amino acids. *J. Comp. Chem.*, 20:428–442, 1999.

[70] D. Bashford and M. Karplus. Multiple-Site Titration Curves of Proteins: An analysis of exact and approximate methods for their calculation. *J. Phys. Chem*, 95:9556–9561, 1991.

[71] J. Gasteiger, C. Rudolph, and J. Sadowski. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comp Method*, 3:537–547, 1990.

[72] Suite 1600. Montreal Chemical Computing Group I, 1255 University St. Molecular Operating Environment (MOE 1999.05), 1999.

[73] M.F. Sanner, J.C. Spehner, and A.J. Olson. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320, 1996.

[74] A. Bondi. *Journal of Physical Chemistry*, 68:441–451, 1964.

[75] J.M. Word, S.C. Lovell, J.S. Richardson, and D.C. Richardson. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol*, 285(4):1735–47, 1999.

[76] J. Fokkens. *Microcalorimetric Studies to Correlate Thermodynamic and Structural Properties of Inhibitors of the Blood Clotting Cascade*. PhD thesis, Philipps University Marburg, 2005.

[77] J.A.D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B.Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, D. Wirkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B*, 102:3586–3616, 1998.

[78] G. King, F.S. Lee, and A. Warshel. Microscopic Simulations of Macroscopic Dielectric Constants of Solvated Proteins. *J Chem Phys*, 95:4366–4377, 1992.

[79] A. Warshel, S.T. Russell, and A.K. Churg. Macroscopic Models for Studies of Electrostatic Interactions in Proteins: Limitations and Applicability. *Proc. Natl. Acad. Sci. USA*, 81:4785–4789, 1984.

[80] M.D. Joshi, G. Sidhu, J.E. Nielsen, G.D. Brayer, S.G. Withers, and L.P. McIntosh. Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. *Biochemistry*, 40:10115–10139, 2001.

[81] L. Cocco, J.P. Groff, C. Temple Jr., J.A. Montgomery, N.A. London, and R.L. Blakely. Carbon-13 nuclear magnetic resonance study of protonation of methotrexate and aminopterin bound to dihydrofolate reductase. *Biochemistry*, 20:3972–3978, 1981.

[82] S.R. Stone and J.F. Morrison. The pH-dependence of the binding of dihydrofolate and substrate analogues to dihydrofolate reductase from Escherichia coli. *Biophys. Acta*, 745:247–258, 1983.

[83] L. Cocco, B. Roth, C. Temple Jr., J. A. Montgomery, and N.A. London. Protonated state of methotrexate, trimethoprim, and pyrimthamine bound to dihydrofolate reductase. *Arch. Biochem. Biophys. Acta*, 226:567–577, 1983.

[84] R.E. London, E.E. Howell, M.S. Warren, J. Kraut, and R.L. Blakely. Nuclear magnetic resonance study of the state of protonation of inhibitors bound to mutant dihydrofolate reductase lacking the active-site carboxyl. *Biochemistry*, 25:7229–7235, 1986.

[85] P. Czodrowski, I. Dramburg, C.A. Sotriffer, and G. Klebe. Development, validation, and application of adapted PEOE charges to estimate pKa values of functional groups in protein-ligand complexes. *Proteins*, 65:424–437, 2006.

[86] *The PyMOL Molecular Graphics System on http://www.pymol.org*, 2002.

[87] Kesvatera, T., Jonsson, B., Thulin, E., and Linse, S. Ionization behavior of acidic residues in calbindin. *Proteins*, 37:106–115, 1999.

[88] Onufriev, A., Case, D.A., and Ullmann, G.M. A novel view of pH titration in biomolecules. *Biochemistry*, 40:3413–3419, 2001.

[89] D. Bashford and K. Gerwert. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.*, 224:473–486, 1992.

[90] Carlson, H.A., Briggs, J.M., and McCammon, J.A. Calculation of the pKa values for the ligands and side chains of Escherichia coli D-alanine:D-alanine ligase. *J. Med. Chem.*, 42:109–117, 1999.

[91] A.R. Klingen, E. Bombarda, and M. Ullmann. Theoretical investigation of the behavior of titratable groups in proteins. *Photochemical and Photobiological Sciences,*, 6:588–596, 2006.

[92] O. Guvench, D.J. Price, and C.L. Brooks. Receptor rigidity and ligand mobility in trypsin-ligand complexes. *Proteins*, 58:407–417, 2005.

[93] Edgecomb S. and Murphy K. Variability in the pka of histidine side-chains correlates with burial within proteins. *Proteins*, 49:1–6, 2002.

[94] Plesniak, L.A., Connelly, G.P., Wakarchuk, W.W., and McIntosh, L.P. Characterization of a buried neutral histidine residue in bacillus circulans xylanase: Nmr assignments, ph titration, and hydrogen exchange. *Protein. Sci.*, 5:2319–2328, 1996.

[95] Becktel W.J. Anderson D.E. and Dahlquist F.W. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry*, 29:2403–2408, 1990.

[96] Toseland, C.P., McSparron, H., Davies, M.N., and Flower, D.R. PPD v1.0–an integrated, web-accessible database of experimentally determined protein pKa values. *Nucleic Acids Res.*, 34:199–203, 2006.

[97] Kaslik, G., Westler, W.M., Graf, L., and Markley, J.L. Properties of the His57-Asp102 dyad of rat trypsin D189S in the zymogen, activated enzyme, and alpha1-proteinase inhibitor complexed forms. *Arch. Biochem. Biophys.*, 362:254–264, 1999.

[98] M. Tanokura. 1H-NMR study on the tautomerism of the imidazole ring of histidine residues: I. microscopic pk values and molar ratios of tautomers in histidine-containing peptides. *Biochimica et Biophysica Acta*, 742:576–585, 1983.

[99] P. Gerber and K. Müller. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comp. Aid. Mol. Des.*, 9:251–268, 1996.

[100] Jones, G., Willett, P., and Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 245:43–52, 1995.

[101] Tripos Inc. *Sybyl.* 699 South Hanley Rd., St. Louis, Missouri, 63144, USA 6.6 Auflage, 2000.

[102] J. Coffin, A. Haase, J.A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, L. Toyoshima, H. Varmus, P. Vorgt, and R. Weiss. Human Immunodeficieny Virus. *Science*, 232:697–700, 1986.

[103] T. Yamazaki, K. Nicholson, D.A. Torchia, P. Wingfield, S.J. Stahl, J.D. Kaufman, C.J. Eyerman, C.N. Hodge, P.Y.S. Lam, Y. Ru, P.K. Jadhav, C. Chang, and P.C. Weber. NMR and X-ray evidence that the HIV-1 protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.*, 116:10791–10792, 1994.

[104] G. King, F.S. Lee, and A. Warhsel. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.*, 95:4366–4377, 1991.

[105] D. Bashford. An object-oriented programming suite for electrostatic effects in biological molecules. *Lecture Notes in Computer Science*, 1343:233–240, 1997.

[106] *SYBYL Molecular Modeling Software, 7.0 ed.; Tripos Inc.: St. Louis, MO*, 2002.

[107] R. Smith, I.M. Breton, R.Y. Chai, and S.B.H. Kent. Ionization states of the catalytic residues in HIV-1 protease. *Nature Structural Biology*, 3:946–950, 1997.

[108] J. Antosiewicz, J.M. Briggs, A.H. Elcock, M.K. Gilson, and J.A. McCammon. Computing the ionization states of proteins with a detailed charge model. *J Comput Chem*, 17:1633–1644, 1996.

[109] P.Y. Lam, Y. Ru, P.K. Jadhav, P.E. Aldrich, G.V. DeLucca, C.J. Eyermann, C.H. Chang, G. Emmett, E.R. Holler, W.F. Daneker, L. Li, P.N. Confalone, R.J. McHugh, Q. Han, R. Li, J.A. Markwalder, S.P. Seitz, T.R. Sharpe, L.T. Bacheler, M.M. Rayner, R.M. Klabe, L. Shum, D.L. Winslow, D.M. Kornhauser, and C.N.Hodge. Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2' structure-activity relationship, and molecular recognition of cyclic ureas. *J. Med. Chem.*, 39:3514–25, 1996.

[110] R. Ishima, D.I. Freedberg, Y.X. Wang, J.M. Louis, D.A. Torchia, R.D. Tung, and M.A. Navia. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure*, 7:1047–1055, 1999.

[111] P. Czodrowski, C.A. Sotriffer, and G. Klebe. Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments. 2006. submitted.

[112] T. Mimoto and Y. Kiso. Kynostatin (KNI)-227 and 272, highly potent anti-HIV agents: conformationally-constrained tripeptide inhibitors of HIV protease containing allophenylnorstatine. *Chem. Pharm. Bull.*, 40:2251–2253, 1992.

[113] E.T. Baldwin, N. Bhat, S. Gulnik, B. Liu, I.A. Topol, Y. Kiso, T. Mimoto, H. Mitsuya, and J.W. Erickson. Sructure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure*, 3:581–590, 1995.

[114] E. Specker, J. Böttcher, S. Brass, A. Heine, H. Lilie, A. Schoop, G. Müller, N. Griebenow, and G. Klebe. Unexpected novel binding mode of pyrrolidine-based aspartyl protease inhibitors: design, synthesis and crystal structure of HIV protease. *ChemMedChem*, 1:106–117, 2006.

[115] E. Specker, J. Böttcher, A. Heine, C.A. Sotriffer, H. Lilie, A. Schoop, G. Müller, N. Griebenow, and G. Klebe. Hydroxyethylene sulfones as new scaffold to address aspartic proteases: desgin, synthesis and structural classification. *J. Med. Chem.*, 48:6607–6619, 2005.

[116] E.E. Kim, C .T. Baker, M.D. Dwyer, M.A. Murcko, B.G. Rao, R.D. Tung, and M.A. Navia. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.*, 117:1181–1182, 1995.

[117] S. Cabani, G. Conti, and L. Lepori. Thermodynamic study on aqueous dilute solutions of organic compounds. *Transactions of the Faraday Society*, 67:1933–1942, 1971.

[118] K.M. Bohren, B. Bullock, B. Wermuth, and K.H. Gabbay. The aldo-keto reductase superfamily. cDNAs and deduded amino acid sequences of human aldehyde and aldose reductase. *J. Biol. Chem.*, 264:9547–9551, 1989.

[119] D.K. Wilson, K.M. Bohren, K.H. Gabbay, and F.A. Quiocho. An unlikely sugar substrate in the 1.65 å structure of human aldose reductase holoenzyme implicated in diabetic complications. *Science*, 257:81–84, 1992.

[120] C.E. Grimshaw, K.M. Bohren, C.J. Lai, and K.H. Gabbay. Human aldose reductase: pK of tyrosine 48 reveals the preferred ionization state for catalysis and inhibition. *Biochemistry*, 34(44):14374–84, 1995.

[121] Y.S. Lee, M. Hodoscek, B.R. Brooks, and P.F. Kador. Catalytic mechanism of aldose reductase studied by the combined potentials of quantum mechanics and molecular mechanics. *Biophys Chem*, 70(3):203–16, 1998.

[122] E.R. Larson, C.A. Lipinski, and R. Sarges. Medicinal chemistry of aldose reductase inhibitors. *Med. Res. Rev.*, 8:159–186, 1998.

[123] P. Czodrowski, C.A. Sotriffer, and G. Klebe. Atypical protonation states in the active site of HIV-1 protease: A computational study. 2006. to be submitted.

[124] K.M. Bohren, J.M. Brownlee, A.C. Milne, K.H. Gabbay, and D.H.T. Harrison. The structure of apo R268A human aldose reductase: hinges and latches that control the kinetic mechanism. *Biochimica et Biophysica Acta*, 1748:201–212, 2005.

[125] M. Oka. personal communication.

[126] R. Sarges, R.C. Schnur, John L. Belletire, and M.J. Peterson. Spiro hydantoin aldose reductase inhibitors. *J. Med. Chem.*, 31:230–243, 1988.

[127] A. Koumanov, J. Bernach, S. Atrian, R. Gonzalez-Duarte, A. Karshikoff, and R. Ladenstein. The catalytic mechanism of *Drosophila* alcohol dehydrogenase: Evidence for a proton relay modulated by the coupled ionization of the active site lystine/tyrosine pair and NAD^+ ribose oh switch. *PROTEINS*, 51:289–298, 2003.

[128] P. Varnái and A. Warshel. Computer simulation studies of the catalytic mechanism of human aldose reductase. *J. Am. Chem. Soc.*, 122:3849–3860, 2000.

[129] I. Hazemann, M. T. Dauvergne, M. P. Blakeley, F. Meilleur, M. Haertlein, A. Van Dorsselaer, A. Mitschler, D. A. A. Myles, and A. Podjarny. High-resolution neutron protein crystallography with radically small crystal volumes: application of perdeuteration to human aldose reductase. *Acta Crystallographica Section D*, 61(10):1413–1417, Oct 2005.

[130] E.I. Howard, R. Sanishvili, R.E. Cachau, A. Mitschler, B. Chevrier, P. Barth, V. Lamour, M. van Zandt, E. Sibley, C. Bon, D. Moras, T.R. Schneider, A. Joachimiak, and A. Podjarny. Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. *Proteins*, 55:792–804, 2004.

[131] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci., USA*, 2001.

[132] J.M. Wang and P.A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21:1049–1074, 2000.

[133] C. Søondergaard, P. Czodrowski, G. Klebe, and J.E. Nielsen. pka_lig_tool: A freely accessible database of small molecule pKa values. 2006. in preparation.

[134] N. Weskamp. *Discrete algorithms module - access from python to basic algorithms on sequences and graphs.* 2004.

[135] C. Bron and J. Kerbosch. Algorithm 457. Finding all cliques of an undirected graph. *Commun. ACM*, 16:575–577, 1973.

Mein herzlicher Dank gilt:

- Herrn Prof. Dr. Gerhard Klebe für seine Unterstützung und Gewährung vieler akademischer Freiheiten.
- Dr. Christoph A. Sotriffer für seine fortwährende Hilfe und Rat im Rahmen der gesamten Doktorarbeit.
- Dr. Jens E. Nielsen für einen sehr produktiven Aufenthalt in seiner Gruppe in Dublin und die Teilnahme am PDB2PQR-Projekt.
- Prof. Dr. Annick Dejaegere für einen Aufenthalt in Ihrer Gruppe in Strasbourg, um einen fundierten Umgang mit CHARMM und UHBD zu erlernen.
- Christof Gerlach für eine feine Zeit in Marburg und seine Seelenruhe.
- Holger Steuber für die unzähligen Diskussionen zur Aldose Reduktase und seine hochwertigen und zuverlässigen Daten.
- Torsten Luksch und Naomi Tidten für viele nette Stunden neben der Promotion.
- Sebastian Dohm für seine Hilfe bei der pKa-Datenbank.
- Chresten Søondergaard für die Pflege und Administration von pka_lig_tool.
- Fergal O'Meara, Martin Sippel, Helen Webb für das Korrekturlesen der Arbeit.
- Dr. Andreas Evers und Dr. Daniel Kuhn für die Überlassung der LaTeX-Vorlage.
- meinen Eltern für Ihre Unterstützung jeglicher Art.

Aus der vorliegenden Arbeit sind folgende Posterbeiträge, Vorträge und Publikationen hervorgegangen:

Posterbeiträge:

- Czodrowski, P., Dramburg, I., Sotriffer, C.A., Klebe, G., *Protonation states of residues in the binding site of aldose reductase studied by calculations of the electrostatic.* Poster präsentiert bei der Modelling 2004-Tagung, Erlangen (2004)
- Czodrowski, P., Dramburg, I., Sotriffer, C.A., Klebe, G., *Protonation states of residues in the binding site of aldose reductase studied by calculations of the electrostatic.* Poster präsentiert bei der High Resolution Drug Design Meeting, Strasbourg (2004)
- Czodrowski, P., Dramburg, I., Sotriffer, C.A., Klebe, G., *To charge or not to charge? Protonation states in proteins and their complexes.* Poster präsentiert bei der MGMS Internation Conference, Dublin (2005)

Vorträge:

- Czodrowski, P., Klebe, G., *Continuum electrostatics calculations on the human aldose reductase.* Vortrag präsentiert beim bilateralen CERC3-Treffen, Marburg (2003).
- Czodrowski, P., Klebe, G., *Protonation states in protein/ligand complexes.* Vortrag präsentiert an École Polytechnique Féderal, Lausanne (2004).
- Czodrowski, P., Klebe, G., *Protonation states in protein/ligand complexes.* Vortrag präsentiert an Université Louis Pasteur, Strasbourg (2004).
- Czodrowski, P., Klebe, G., *To charge or not to charge? Protonation states in proteins and their complexes.* Vortrag präsentiert bei der Modelling 2005-Tagung, Erlangen (2005). (Der Vortrag wurde mit einem Vortragspreis ausgezeichnet.)
- Czodrowski, P., Klebe, G., *To charge or not to charge? Protonation states in proteins and their complexes.* Vortrag präsentiert beim Roche Symposium for Leading Chemists of the Next Deacde, Basel (2005). (Der Vortrag wurde mit einem Vortragspreis ausgezeichnet.)
- Czodrowski, P., Klebe, G., *To charge or not to charge? Protonation states in proteins and their complexes.* Vortrag präsentiert beim Young Modellers' Forum, London (2005). (Der Vortrag wurde mit einem Vortragspreis ausgezeichnet.)

Aufsätze:

- Czodrowski, P., Dramburg, I., Sotriffer, C.A., Klebe, G., *Development, Validation, and Application of Adapted PEOE Charges to Estimate pKa Values of Functional Groups in Protein-Ligand Complexes Proteins*, 65 (2006), 424-437.
- Czodrowski, P., Sotriffer, C.A., Klebe, G., *Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments.* eingereicht
- Czodrowski, P., Sotriffer, C.A., Klebe, G., *Atypical protonation states in the active site of HIV-1 protease: A computational study* eingereicht
- Steuber, H., Czodrowski, P., Sotriffer, C.A., Klebe, G., *Where does the proton go? Proton linkage and thermodynamic factorization of inhibitor binding to Aldose Reductase* Manuskript in Vorbereitung

Erklärung

Ich versichere, dass ich meine Dissertation

Prediction of protonation states in ligand-protein complexes upon ligand binding

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

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch in keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 13. Oktober 2006

(Paul Czodrowski)

Lebenslauf

Paul Czodrowski

Geburtstag:	19. April 1977
Geburtsort:	Olsztyn (Polen)
Staatsangehörigkeit:	deutsch
1983 – 1987	*Grundschule* in Langenselbold
1987 – 1988	*Gesamtschule* in Freigericht
1988 – 1996	*Cusanus-Gymnasium* Wittlich
Jun. 1996	Abitur
Jul. 1996 – Aug. 1997	Zivildienst bei *Essen auf Rädern*, Caritas, Wittlich
1997 – 2002	Chemiestudium an der *Technischen Universtität M"unchen* *Vordiplomprüfungen* Okt. 1999 *Diplomprüfungen* Apr. 2002 Schwerpunkt: *Theoretische Chemie*
Mai. 1999 – Nov. 1999	*Diplomarbeit* im Department Molecular Pharmacology, *Brown University, Rhode Island, USA*, im Arbeitskreis von Prof. D. Mierke *Spezifische Inhibitoren von SAP90: Strukturbasiertes Design*
Feb. 2003	Beginn der Arbeiten zur vorliegenden Dissertation unter Betreuung von Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie der *Philipps-Universität* Marburg
Feb. 2003 – Sep. 2006	Wissenschaftlicher Angestellter am Institut für Pharmazeutische Chemie der Philipps-Universität Marburg; Betreuung des Studentenpraktikums im ersten Semester