Computational Modelling of Protonation Equilibria and Reaction Mechanism of HIV-1 Protease

Joanna Trylska

Praca doktorska
Wydział Fizyki Uniwersytetu Warszawskiego
Warszawa 2001

Promotor: dr hab. Maciej Geller

Recenzenci: prof. dr hab. Lucjan Piela
dr hab. Wiesław Nowak

Podziękowania

Składam serdeczne podziękowania wszystkim, którzy okazali mi pomoc w mojej pracy oraz udzielali mi cennych rad i wskazówek, w szczególności: dr. hab. Janowi Antosiewiczowi, dr. Pawłowi Grochowskiemu, dr. hab. Maciejowi Gellerowi, dr. hab. Piotrowi Bale oraz prof. dr. hab. Bogdanowi Lesyngowi.

Wszystkie obliczenia, których wyniki są przedstawione w tej pracy, wykonane zostały w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego. Praca była także finansowana przez Komitet Badań Naukowych (projekty badawcze nr 8 T11F 001 13 oraz 8 T11F 016 16).

Contents

Preface vii

I Protonation Equilibria in HIV-1 Protease 1
- 1.1 HIV-1 Protease its Function and Structure . . . 1
  - 1.1.1 Enzymatic System . . . 1
  - 1.1.2 Reaction Mechanism . . . 2
- 1.2 Computer Modelling of Protonation Equilibria in HIV-1 Protease . . . 4
  - 1.2.1 Introduction . . . 4
  - 1.2.2 Modelled Systems . . . 4
  - 1.2.3 Methods . . . 5
    - Theoretical Method . . . 6
    - Poisson-Boltzmann Model . . . 9
    - Modelling the Ionization Processes . . . 11
    - Electrostatic Calculations . . . 12
    - Treatment of Multiple Ionization State Problem . . . 13
    - Summary of the Algorithm . . . 14
    - Comparison with Measured pK_as . . . 14
    - Restrictions of the Method . . . 15
  - 1.2.4 Coordinates and Parameters . . . 16
  - 1.2.5 Modifications of the Computational Methodology . . . 18
- 1.3 Results - The Protonation States of Titratable Amino Acids . . . 20
  - 1.3.1 Free HIV-1 PR . . . 20
  - 1.3.2 HIV-1 PR Complexed with Neutral, Symmetric Cyclic Urea Inhibitors 22
  - 1.3.3 HIV-1 PR Complexed with the Neutral, Asymmetric Inhibitor KNI-272 24
  - 1.3.4 HIV-1 PR Complexed with the Cationic, Asymmetric Inhibitor MVT-101 26
  - 1.3.5 Apoenzymes of HIV-1 PR : MVT-101 and HIV-1 PR : XK-263 . . . 27
  - 1.3.6 HIV-1 PR Complexed with a Model Substrate . . . 27
- 1.4 Conclusions . . . 28
  - 1.4.1 Analysis of pK_a Calculations . . . 28
  - 1.4.2 Comparison of Single-site/20 and Full-charge/4 Models . . . 29
  - 1.4.3 Possible Improvements . . . 30

II Parameterization of the Approximate Valence Bond Method to Describe Potential Energy Surface in the Reaction Catalysed by HIV-1 Protease 31
- 2.1 Introduction . . . 31
- 2.2 Approximate Valence Bond Method . . . 31
  - 2.2.1 Basics of Valence Bond Methods . . . 31
  - 2.2.2 The AVB Structures . . . 33

2.2.3 AVB Description of an Isolated Molecule . . . 35
2.2.4 Description of a Molecular System . . . 39
2.3 Parameters Characterizing Intramolecular Interactions in the AVB Region . . 40
2.4 Parameters Characterizing the Nonbonding Interactions in the AVB Region 50
2.4.1 Proton Transfer in the Water – Ionized Carboxylic Acid Complex . . 50
2.4.2 The Nucleophilic Attack . . . 51
2.4.3 Proton Transfer onto O^- in $CH_3CO^-OHNHCH_3$ (I1) – Neutral Carboxylic Acid Complex . . . 53
2.4.4 Proton Transfer from Neutral Carboxylic Acid onto the Nitrogen of the $CH_3C(OH)_2NHCH_3$ (I2) Molecule . . . 54
2.4.5 The Breakage of $CH_3C(OH)_2NH_2^+CH_3$ (I3) Molecule . . . 54
2.4.6 Proton Transfer between Doubly Protonated and Ionized Carboxylic Acids . . . 56
2.4.7 Nonbonding Interactions between Oxygens of Neutral and Ionized Carboxylic Acids . . . 58
2.4.8 Proton transfer from Neutral Carboxylic Acid onto the Nitrogen of the $CH_3CO^-OHNHCH_3$ (I1) Molecule . . . 58
2.4.9 Summary . . . 60

III Molecular Dynamics Simulations of the Reaction Mechanism Catalysed by HIV-1 Protease **63**
3.1 Introduction . . . 63
3.2 Equilibration and Thermalization of the HIV-1 PR : Substrate Complex . . . 63
3.2.1 Energy Minimization . . . 63
3.2.2 Thermalization Procedure . . . 64
3.3 MD/AVB Simulations . . . 67
3.4 Steered MD/AVB Simulations of the First Step of the Reaction . . . 71
3.5 Energy Checkpoint Between the AVB Parameterization and DFT Calculations 76
3.6 MD/AVB Simulations Following the Nucleophilic Attack . . . 78
3.7 MD/AVB Simulations of Further Steps of the Reaction . . . 78
3.7.1 Proton Transfer from Asp125 onto Peptide Nitrogen . . . 78
3.7.2 Inclusion of the C–N Bond Breakage . . . 80
3.8 Concluding Remarks . . . 80

Summary **83**

Preface

The aim of this thesis was to study the protonation equilibria and the reaction mechanism of the human immunodeficiency virus type 1 protease (HIV-1 PR).

Properties of large biomolecular systems such as enzymes are often described theoretically using the methods of molecular mechanics (MM) or molecular dynamics (MD). In these methods atomic nuclei are represented as classical point-like particles that move on the Born-Oppenheimer potential energy surface corresponding to the ground electronic state. In MM methods such a surface is approximated by a sum of simple analytical functions that describe effective interactions between nuclei. This approximation is, however, not suitable for studying chemical reactions because it does not allow for rearrangement of chemical bonds on-line with a dynamical simulation.

The approach often used while studying time-dependent chemical processes in large biological structures is to divide the system into separate regions and characterize them at various approximation levels. Generally, the most important part of an enzymatic system is a fragment of a substrate and residues of an active site directly involved in catalysis. The surrounding atoms which influence the active site through steric constraints and electrostatic interactions can be treated by computationally less expensive MM methods using a classical force field. The active site atoms[1] however, must be treated more rigorously by quantum-mechanical (QM) methods such as *ab initio* or density functional theories (DFT) to describe atom-atom interactions more precisely.

All proteins contain residues which are capable of exchanging protons with their environment. This process depends on pH and determines the electrostatic properties of enzymes. The charge distribution plays a crucial role in functioning of an enzyme. It contributes to steering the substrate near the active site and stabilizing possible transition states. It influences the binding of ligands and catalysis. Thus, while performing theoretical calculations starting from an X-ray structure, as this was the case, one has to model properly the number and positions of hydrogens attached to some heavy atoms. This may not be determined from crystallographic studies due to small electron density of the hydrogen protons which location is unclear at the resolution of a typical X-ray structure. As the electrostatic behaviour of enzymes is influenced by the pH of the medium it is primarily important in the calculations to determine the number and precise location of covalently bound hydrogens (more loosely protons). The number of hydrogens attached to aliphatic carbons and peptide nitrogens results from their hybridization. Therefore, in the modelled protein one has to determine mainly the protonation states of the ionizable (titratable) residues. Even though experimental ionization constants (pK_as) for amino acids in the water environment are determined, the apparent pK_as of titratable groups in proteins may not necessarily correspond to that measured for model compounds. Hence,

[1]The word atoms does not apply to an isolated atom but to a set or collection of nuclei and electrons in a molecular environment.

if electrostatic characteristics of the protein are to be considered, precise placement of hydrogens is essential for further computational studies of the system.

As the molecular model of the protein build up in this study was based on crystallographic data, additional information about protonation, not provided by X-ray structures, was required. Therefore, the first part of this thesis concentrates on predicting ionization states of titratable residues for the modelled enzyme. The object of the study is the human immunodeficiency virus type 1 protease (HIV-1 PR). This enzyme most probably promotes an acid-base catalysis and its activity requires the catalytic residues to acquire appropriate protonation states [1, 2]. The cases studied included free HIV-1 PR, HIV-1 PR complexed with a model substrate, and HIV-1 PR complexed with four competitive inhibitors. The approach used was based on the Poisson-Boltzmann (PB) model of the protein-solvent system and the finite-difference method of solving the PB equation.

In the second Chapter the aim was to parameterize, within the approximate valence bond (AVB) formalism [3], the potential energy surface to characterize possible paths of the enzymatic reaction catalysed by HIV-1 PR. An empirical version of the valence bond (VB) method was first proposed by Warshel [4] to describe energy surfaces for enzymatic reactions. The similar AVB approach uses *ab initio* type of calculations and applies a different parameterization strategy [3]. This allows calculating of the Born-Oppenheimer energy surface for the nuclei in an enzyme active site. The method was coupled with a classical force field description of the rest of the enzyme. It may be used in microscopic classical or quantum-classical molecular dynamics (MD/AVB or QCMD/AVB) simulations of enzymatic processes. The QCMD/AVB was successfully applied for simulations of the whole enzymatic reaction catalysed by phospholipase A_2 [5, 6].

The third Chapter describes the MD/AVB calculations used to investigate the reaction mechanism of HIV-1 PR since there are many questions that had not yet been solved regarding this process. First, typical energy minimization and thermalization of the HIV-1 PR complexed with a model substrate was performed. The potential energy surface included the AVB parameters in the active centre and standard Gromos'96 [7] force field description for the rest of the enzyme. Second, the classical MD/AVB calculations of the system were carried out. They did not lead to the transition of the substrate due to the fact that most enzymatic reactions occur on a time scale much larger than 1ns (typical time scale for MD simulations). To check if the AVB description properly characterizes the interactions in the active centre and whether they allow for the reaction, the force driven molecular dynamics calculations were performed. The system was forced through a barrier on a shorter time scale than in reality and the first step of the reaction occurred. Next, the following proton transfer was modelled in the MD/AVB simulations. Various molecular conformations as transition states were tested.

I am the coauthor of four publications which have been already published in four journals (see [8, 9, 10, 11]). The papers include the results of the first and second Chapter. I also presented those results as 11 communications at domestic and international conferences. The AVB parameterization was the subject of a talk that I gave at XIIIth International Biophysics Congress held in New Delhi in 1999. The results presented in the last Chapter are also being prepared for publication.

Chapter I

Protonation Equilibria in HIV-1 Protease

1.1 HIV-1 Protease its Function and Structure

1.1.1 Enzymatic System

The human immunodeficiency virus type 1 protease (HIV-1 PR) is one of the three enzymes encoded by the viral genome. It exhibits its main role during the course of viral maturation. HIV-1 protease processes the *gag* and *pol* polyproteins, which further generate important viral enzymes and structural proteins, including HIV-1 PR itself. Human immunodeficiency virus (HIV) is a causative agent of acquired immunodeficiency syndrome (AIDS) disease. It was observed that virions that lack HIV-1 protease are non-infectious [12]. Hence agents that inhibit this enzyme are used to treat patients infected with the virus and may be crucial in chemotherapy of AIDS.

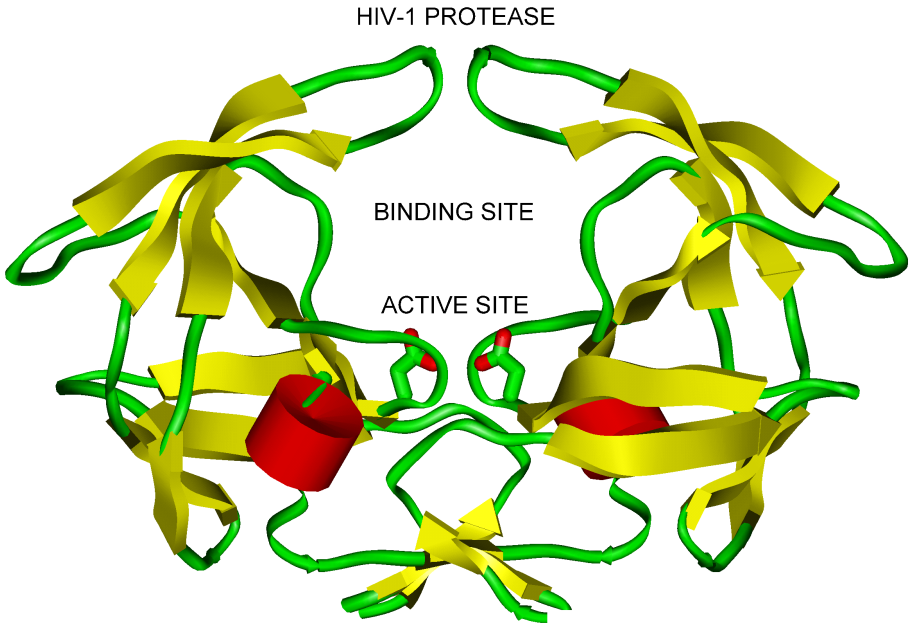

Figure 1.1: Secondary structure of the two subunits of HIV-1 protease.

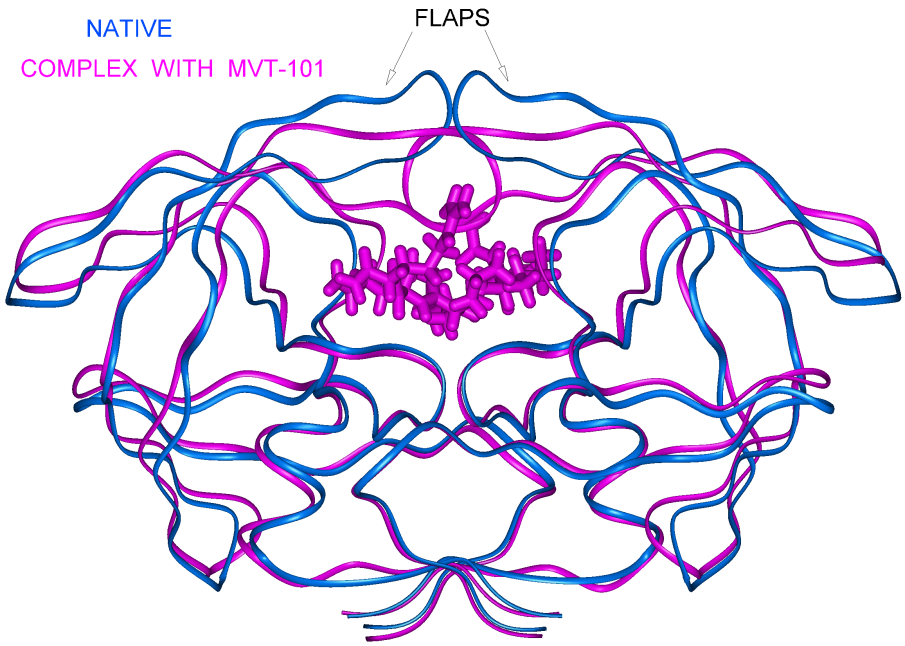

Figure 1.2: C_α traces of the native HIV-1 protease and in the complex with the MVT-101 inhibitor [14].

HIV-1 protease acts and crystallises as a homo-dimer, with 99 residues in each monomer, one related to the other by a crystallographic twofold axis (see Figure 1.1). β-strands dominate the secondary structure of the enzyme. There is only one short α-helix in each monomer. The substrate cleavage site is situated at the interface between identical, adjacent subunits. It involves a pair of a catalytic triad, Asp-Thr-Gly, which identifies HIV-1 PR as an aspartyl protease. The two aspartyl side chains, one from each monomer, are closely apposed and form a symmetric dyad [13]. The binding site has a shape of an extended ravine formed by both monomers and perpendicular to the twofold C_{2v} symmetry axis of the dimer. It can accommodate six to eight amino acids, which bind in an extended β conformation. The specifity is broad but hydrophobic residues are preferred. The entry to the ravine is controlled by two flexible flaps which form a gate for an approaching ligand (see Figure 1.2). Each flap has a β-hairpin form maintained by several hydrogen bonds. There are four hydrophobic pockets inside the side walls of the ravine. Binding of a substrate or inhibitor induces substantial conformational changes in the flaps which tips move during this process as far as 7 Å. The area of the monomer-monomer interface constitutes nearly one third of the solvent accessible surface of each monomer [8].

1.1.2 Reaction Mechanism

HIV-1 PR catalyses the cleavage of peptide bonds. There have been several proposals for the reaction mechanism [15, 1, 16, 2, 17]. However, most studies e.g. [1, 2] indicate an acid-base mechanism in which the peptide bond hybridization is changed from sp^2 to sp^3 through a nucleophilic attack on its carbonyl group by a lytic water molecule. The activity of this enzyme varies with pH to yield a "bell-shaped" graph of the catalytic constant *versus* pH with a maximum at pH $\sim 5-6$ [1, 18]. The customary explanation of this fact

is that the active form is monoprotonated and exists only at intermediate pH (the dyad is fully protonated at low pH and fully deprotonated at high pH). The proposed mechanism is presented in Figure 1.3. First, a hydrogen transfer occurs (Fig. 1.3.1) between the lytic water molecule and Asp125[1] followed by the nucleophilic attack of the resulting hydroxy anion on the peptide bond carbon. Secondly, proton transfer between Asp25 and the carbonyl oxygen of the peptide bond occurs followed by or together with a conformational transition of the C–N bond allowing the nitrogen lone pair to accept the hydrogen of Asp125 (Fig. 1.3.3). Next, the C–N bond breaks and the hydrogen is transferred between one of the hydroxyl groups of the tetrahedral intermediate onto Asp25 (Fig. 1.3.4).

[1]Residues of one monomer are usually numbered 1–99 and of the other 101–199.

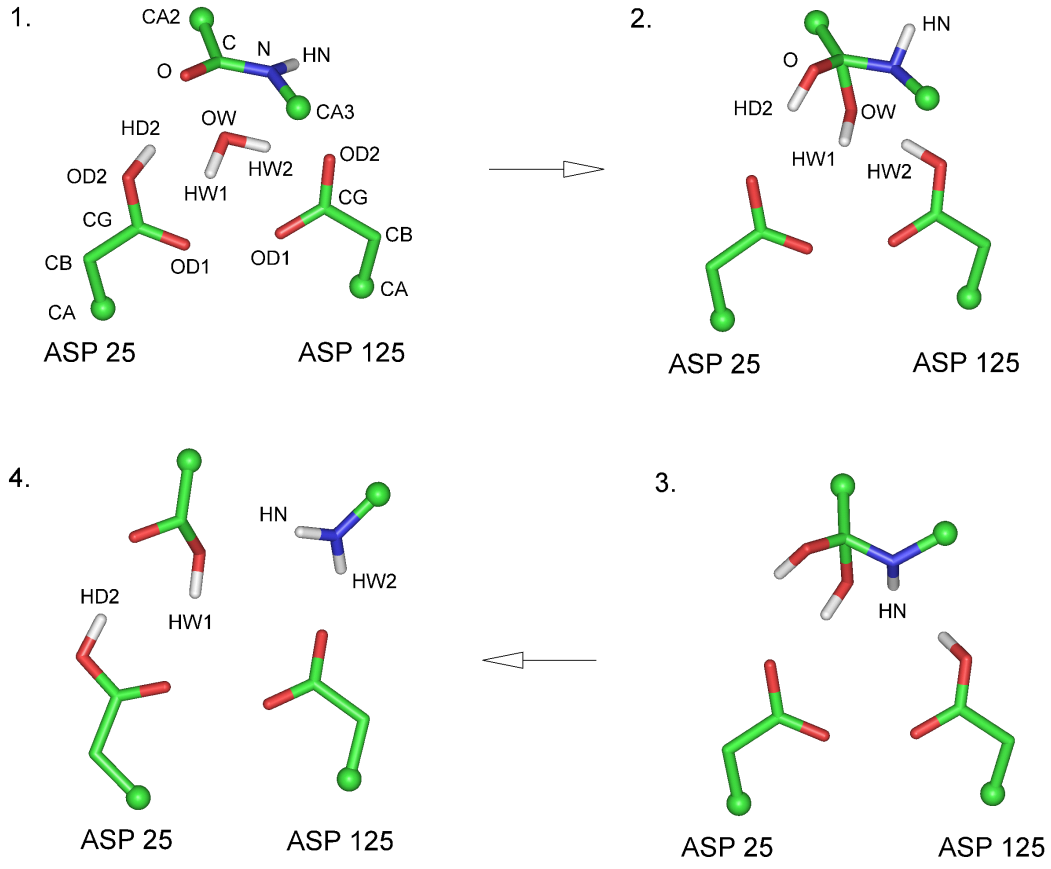

Figure 1.3: One of the proposed reaction mechanisms. 1. Enzyme reaction centre: catalytic aspartates (25 and 125), peptide bond and lytic water molecule. 2. Tetrahedral intermediate following HW2 transfer between water molecule and OD2 of Asp125, nucleophilic attack of the resulting hydroxy anion (OW–HW1) on the peptide carbon (C), and HD2 transfer between Asp25 and the carbonyl oxygen of the peptide bond. 3. An *anti-gauche* to *gauche* transition of the C–N bond allowing the nitrogen lone pair to accept HW2 proton of Asp125. 4. C–N bond rupture and hydrogen transfer between one of the hydroxyl groups of the tetrahedral intermediate onto Asp25. Balls represent carbon atoms bonded to other fragments of the protein or the substrate.

1.2 Computer Modelling of Protonation Equilibria in HIV-1 Protease

1.2.1 Introduction

The ability of protein residues to exchange protons with their environment and the dependence of this process on pH influences the electrostatic properties of enzymes. At extreme values of pH, for instance, proteins unfold and enzymes cease to function. The rate of enzymatic catalysis and also the binding of inhibitors, as has been observed experimentally, depends upon pH. These processes are therefore thermodynamically linked to protonation states of the system. Protonation states of ionizable residues determine the overall charge distribution of enzymes. This distribution may play a crucial role in the reaction path e.g. drawing the substrate near the active site or stabilizing possible transition states. Also the appropriate protonation state of active site residues is often required to activate an enzyme because most enzymes contain residues that promote acid-base catalysis. Moreover, the protonation state of ionizable residues surrounding the active site may also be important since their electrostatic field can stabilize transition states of the reaction path. The properties of HIV-1 PR are also affected by pH. The activity of this enzyme varies with pH and is peaked at pH $\sim 5-6$ [1, 18]. Studies of the pH dependence of the kinetics of HIV-1 PR have been used to deduce the titration behaviour of the active site aspartyl dyad. These studies yield two distinct pK_as for the dyad in the free enzyme [1, 18, 19]. Competitive inhibitors of HIV-1 PR bind in the active site, displacing water and forming contacts with the carboxyl groups of the dyad. As a consequence, they also can alter the pK_as of the dyad. So studies of protonation states of the residues of HIV-1 PR are essential for further modelling of its properties e.g. molecular dynamics. The computations directed at elucidating the changes in protonation of the dyad when inhibitors or substrates bind are presented in this Chapter. The Poisson-Boltzmann model of electrostatic interactions is used to compute the apparent pK_as of the dyad and other titratable amino acids; in the free enzyme, in complex with a model substrate, and in complexes with four different inhibitors. The calculations are based upon X-ray structures of the free enzyme and the enzyme-inhibitor complexes, and the results are compared with available experimental data.

1.2.2 Modelled Systems

The present study uses the PB method to compute the protonation states of the aspartyl dyad of free HIV-1 PR, HIV-1 PR complexed with a model substrate and of HIV-1 PR complexed with four inhibitors that interact intimately with the aspartyl dyad. One cationic inhibitor, MVT-101 [14] (Figure 1.4), contains a basic secondary amine that lies within $3-4$ Å of the carboxyl oxygens of the aspartyl dyad [14, 20]. The second inhibitor is the asymmetric inhibitor KNI-272 [21] (Figure 1.5), a peptidomimetic transition state mimic. The other two inhibitors, DMP-323 [22], and XK-263 [23] (Figure 1.6), are electrically neutral cyclic ureas with two hydroxyl groups that hydrogen bond to the carboxyls of the dyad [23]. These four inhibitors leave no room for bulk solvent at the dyad.

The enzyme-substrate complex was prepared using the crystallographic structure of HIV-1 PR complexed with MVT-101 inhibitor [N-acetyl-Thr-Ile-Nleu-Ψ[CH_2NH_2]-Nleu-Gln-Arg-amide] [20]. This inhibitor was constructed on the basis of the hydrolyzed X/p9 junction [Thr-Ile-Met-/-Met-Gln-Arg] of the HIV-1 polyprotein. Hence, to mimic a sub-

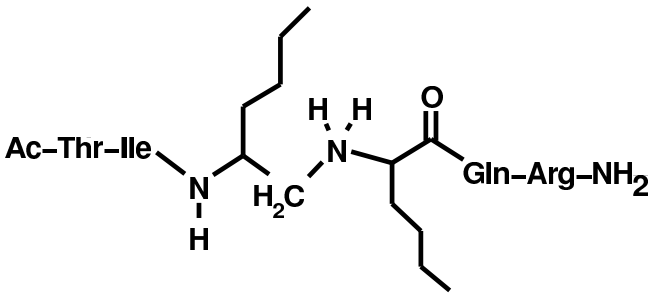

Figure 1.4: The cationic, asymmetric inhibitor MVT-101 (N-acetyl-Thr-Ile-NLeu-Ψ[CH2-NH2]-NLeu-Gln-Arg-amide).

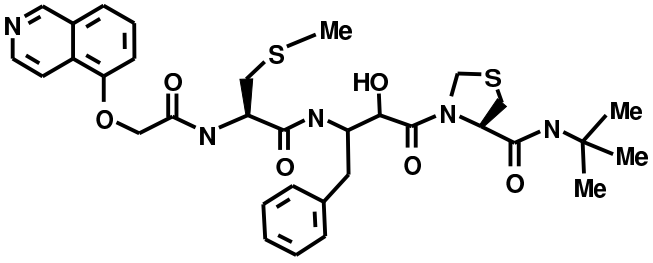

Figure 1.5: The neutral, asymmetric inhibitor KNI-272.

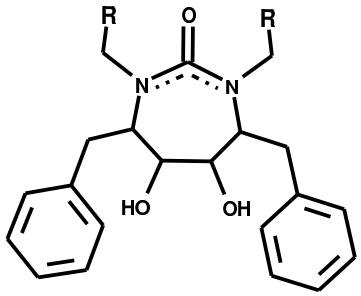

Figure 1.6: Three symmetric cyclic urea inhibitors of HIV-1 PR. DMP-323: R=*p*-(hydroxymethyl)benzyl. XL-075: R=benzyl. XK-263: R=2-naphthylmethyl.

strate, each norleucine of the MVT-101 inhibitor was replaced by a methionine, and the reduced $\Psi[CH_2NH_2]$ link was modified into a regular peptide bond. The (cata)lytic water was also added to the structure as it was not observed in the HIV-1 PR : MVT-101 complex. This is probably due to the fact that transition state analogues have no space for water molecule next to the catalytic aspartic acid residues.

1.2.3 Methods

The bases of the method for predicting the protonation states of titratable residues applied herein have been described previously [24, 25, 26, 27]. This Section reports briefly on the method and shows the modifications and improvements which had to be applied to treat properly the protonation equilibria in HIV-1 protease.

Theoretical Method

Absolute ionization constants (pK_as) of titratable groups in proteins cannot be predicted reliably therefore the method predicts only the pK_a shifts from experimental values determined for model compounds in water. The problem of a protonation state of a group in the protein is considered referring to the protonation of that group in water. The influence of the protein medium on the pK_a of a residue is accounted for and calculated as a shift from experimental value.

Generally, protonation is described by the acidic dissociation constant, K_a. Such dissociation constant is related to the equilibrium concentrations of the species and the standard free energy change ΔG°_{w} by the equation

$$\exp(\frac{-\Delta G^{\circ}_{w}}{RT}) = \frac{[A^-][H^+]}{[AH]} = \mathrm{K}_a \tag{1.1}$$

where index w refers to water environment; R is the gas constant and T is the absolute temperature. Introducing the definition of pK_a

$$\mathrm{pK}_a^{(w)} = -\log_{10} \mathrm{K}_a \tag{1.2}$$

and transforming equation 1.1 one gets

$$\mathrm{pK}_a^{(w)} = \frac{1}{2.303RT}\Delta G^{\circ}_{w}. \tag{1.3}$$

To find a formula for the pK_a of a group in a protein environment, $\mathrm{pK}_a^{(\mathrm{p})}$, it is convenient to refer to a thermodynamic cycle shown in Figure 1.7.

$$\begin{array}{ccc} AH(w) & \xrightarrow{\Delta G^{\circ}_{w}} & A^-(w) \;+\; H^+(w) \\ \Big\downarrow{\scriptstyle \Delta G^{AH}_{transfer}} & & \Big\downarrow{\scriptstyle \Delta G^{A^-}_{transfer}} \\ AH(p) & \xrightarrow{\Delta G^{\circ}_{p}} & A^-(p) \;+\; H^+(w) \end{array}$$

Figure 1.7: Thermodynamic cycle showing the free energy contributions to the pK_as of molecule AH in water (w) and in the protein (p).

From the cycle the following relation is obtained:

$$0 = \Delta G^{0}_{p} - \Delta G^{A^-}_{transfer} - \Delta G^{AH}_{transfer} + \Delta G^{0}_{w}. \tag{1.4}$$

Transforming the above using equations analogous to 1.1 – 1.3, $\mathrm{pK}_a^{(p)}$ can be calculated as

$$\mathrm{pK}_a^{(p)} = \mathrm{pK}_a^{(w)} + \frac{\Delta G^{A^-}_{transfer} - \Delta G^{AH}_{transfer}}{2.303RT} = \mathrm{pK}_a^{(w)} + \frac{\Delta G^{AH,A^-}_{protein} - \Delta G^{AH,A^-}_{water}}{2.303RT} \tag{1.5}$$

where $\mathrm{pK}_a^{(w)}$ is the pK_a in an aqueous state obtained from experiment, $\Delta G^{AH,A^-}_{protein}$ and $\Delta G^{AH,A^-}_{water}$ are the free energy differences between protonated (AH) and deprotonated (A^-)

states in the protein and in water, respectively. As it is simpler to determine the free energy difference between structures that are more similar and in the same environment, the terms appearing in the second equality of the above equation, instead of transfer energies, are used in the calculations.

However, the above equation refers only to one ionizable group and it does not allow to calculate the ionization constant of a group in protein, $\mathrm{pK}_a^{(p)}$, since proteins posses many titratable residues and the multiple ionization problem appears. If a group has two possible ionization states (either protonated or deprotonated), one has to deal with 2^M possible states (M being the number of ionizable groups). According to first principles the evaluation of the average pH-dependent properties of a protein would require the knowledge of its energies in all possible ionization states. The subsequent evaluation of the average property would then be based on the Boltzmann's distribution. But for most systems of interest the total number of ionization states is extremely large and some other more efficient methods must be used instead of enumerating all possible states. This task can be treated, e.g. with a Monte Carlo algorithm [28]. Assuming one has a formula for calculating the Gibbs free energy (ΔG) of the protein dependent on pH and its protonation state. The most probable protonation states are then evaluated by changing (in each Monte Carlo step) the protonation states instead of coordinates. This ΔG may be introduced in the following way. Using the definition of pH

$$\mathrm{pH} = -\log_{10}[H^+] \tag{1.6}$$

the second equality of equation 1.1 can be converted to the form

$$\log_{10}\frac{[A^-]}{[AH]} = \mathrm{pH} - \mathrm{pK}_a^{(w)}. \tag{1.7}$$

Assuming that the concentrations are proportional to the probabilities of being in a given state, and with the use of the Boltzmann's distribution one can write,

$$\frac{[A^-]}{[AH]} = \exp(\frac{-\Delta G}{RT}). \tag{1.8}$$

Hence, the free energy change of an ionization of a single state is obtained as

$$\Delta G = 2.303RT(\mathrm{pH} - \mathrm{pK}_a^{(w)}) \tag{1.9}$$

where ΔG can be considered as the free energy level of the protonated state (AH) relative to the deprotonated one (A^-) [29]. Similar expression may be used for a titratable site in the protein.

The method for calculating apparent pK_as in the protein environment is based on these energy levels with the knowledge of $\mathrm{pK}_a^{(w)}$. The method assumes that the response of the protein to protonation or deprotonation of each residue is linear. Considering a protein with M ionizable sites, its free energy in a given ionization state may be expressed as a sum of the above energy terms. If the neutral state of all ionizable residues is chosen as the reference state, the free energy of the ionization state $(x_1, ..., x_M)$ of the whole molecule is

$$\Delta G(\mathrm{pH}, x_1, ..., x_M) = 2.303RT\sum_{i=1}^{M} x_i\gamma_i(\mathrm{pH} - \mathrm{pK}_i^{(p)}) \tag{1.10}$$

where x_i is 1 when the group is ionized, and 0 when it is neutral; γ_i is $+1$ for bases, and -1 for acids.

The last equation can be expressed as a sum of the standard free energy change of ionizing a single molecule in water taking into account the influence of the protein environment in terms of interaction free energy matrix $\|G_{ij}\|$

$$\Delta G(\mathrm{pH}, x_1, ..., x_M) = 2.303RT \sum_{i=1}^{M} x_i \gamma_i (\mathrm{pH} - \mathrm{pK}_i^{(w)}) + \sum_{i=1}^{M} x_i G_{ii} + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} x_i x_j G_{ij}. \tag{1.11}$$

A symmetric $\|G_{ij}\|$ matrix of interaction free energies among the ionizable groups may be constructed based on another assumption that the difference in protonation behaviour of a group in solution and in the protein is purely electrostatic in origin. Each diagonal electrostatic free energy interaction matrix element is a self-energy term; the work of ionizing group i in the protein with all other groups neutral, relative to the work of ionizing that group in solution, i.e.

$$G_{ii} = \Delta G_{protein,i}^{electrostatic} - \Delta G_{model,i}^{electrostatic} \tag{1.12}$$

where

$$\Delta G_{model,i}^{electrostatic} = W_{model,i}^{(ionized)} - W_{model,i}^{(neutral)} \tag{1.13}$$

and

$$\Delta G_{protein,i}^{electrostatic} = W_{protein,i}^{(ionized)} - W_{protein,i}^{(neutral)}. \tag{1.14}$$

Index $model$ refers to a model compound in water. W is the energy necessary to assemble the whole set of n point charges in an arbitrary dielectric body, with dielectric constant of ϵ_2, immersed in an infinite medium with another dielectric constant of ϵ_1, assuming no field-dependence of the dielectric permittivity [30].

$$W = \frac{1}{2} \sum_{i=1}^{n} q_i \phi_i \tag{1.15}$$

where ϕ_i is the electrostatic potential at the location of the charge q_i, and is considered to be the sum of the Coulombic potentials due to all other charges except the charge on i, plus the total reaction field at the location of charge i. W is calculated for appropriate set of point charges representing the neutral and the ionized form of a residue in its isolated state and as a titratable group in protein.

The non-diagonal matrix element G_{ij}, i.e. the interaction energy, between sites i and j is expressed as:

$$G_{ij} = \Delta G_{protein,i,j}^{electrostatic} - (\Delta G_{protein,i}^{electrostatic} + \Delta G_{protein,j}^{electrostatic}) \tag{1.16}$$

where

$$\Delta G_{protein,i,j}^{electrostatic} = W_{protein,i,j}^{(ionized)} - W_{protein,i,j}^{(neutral)}. \tag{1.17}$$

$\Delta G_{protein,i,j}^{electrostatic}$ has a similar meaning as $\Delta G_{protein,i}^{electrostatic}$ in equation 1.14 but both sites, i and j, are ionized. G_{ij} is the change in the work of ionizing group i that results when group j is in its ionized form.

Poisson-Boltzmann Model

Calculations of electrostatic potentials required for evaluating the ionization and interaction energies are based on the Poisson-Boltzmann model of electrostatic interactions. The foundation of the PB model constitutes the fundamental equation of electrostatics, differential Poisson equation that describes the electrostatic potential $\Phi(\mathbf{r})$ in a medium with a dielectric scalar field $\epsilon(r)$ and with a charge density $\rho(r)$

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}). \tag{1.18}$$

When one deals with a macromolecule (in this case the protein) immersed in an aqueous medium with mobile ions, the charge density $\rho(r)$ can be separated into two components:

$$\rho(\mathbf{r}) = \rho^f_{int} + \rho^m_{ext} \tag{1.19}$$

where ρ^f_{int} describes the interior charge distribution of the fixed (^{f}) positions of all the charges in the protein and ρ^m_{ext} is a mobile (^{m}) exterior charge density modelled by a Boltzmann distribution. Equation 1.18 takes the form

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) = -4\pi \left[\rho^f_{int}(\mathbf{r}) + \lambda(\mathbf{r}) \sum_i q_i n_i \exp(-q_i \Phi(\mathbf{r})/kT) \right] \tag{1.20}$$

where n_i is the bulk number density of type-i ions, q_i is the ionic charge, k is the Boltzmann constant, and λ equals 1 for ion-accessible regions and 0 elsewhere.

Linearizing the exponential terms in the mobile charge distribution and assuming equal number of positively and negatively charged ions one obtains

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) = -4\pi \left[\rho^f_{int}(\mathbf{r}) + \lambda(\mathbf{r}) \sum_i q_i n_i q_i \Phi(\mathbf{r})/kT) \right]. \tag{1.21}$$

Next, the modified Debye-Hückel parameter is introduced

$$\bar{\kappa}^2 = \frac{4\pi \sum_i n_i q_i^2}{kT} = \frac{8\pi e^2 N_A I}{1000kT} \tag{1.22}$$

where N_A is the Avogadro constant, e the charge of an electron, and I the ionic strength. The ionic strength is represented as

$$I = 1/2 \sum_i c_i z_i^2 \tag{1.23}$$

where c_i and z_i are the molar concentration and the valency of the mobile ions, respectively.

Introducing $\bar{\kappa}^2$ in 1.21 one obtains the linearized Poisson-Boltzmann equation where the protein is represented as a low dielectric medium containing fixed charges and the solvent is represented as a medium of dielectric constant of 80 which contains mobile ions that screen the fixed charges according to the Debye-Hückel model

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho_{int}(\mathbf{r}) + \lambda(\mathbf{r}) \bar{\kappa}^2 \phi(\mathbf{r}). \tag{1.24}$$

The charges are usually located at nucleic positions determined by X-ray crystallographic methods or high-resolution Nuclear Magnetic Resonance (NMR) spectroscopy.

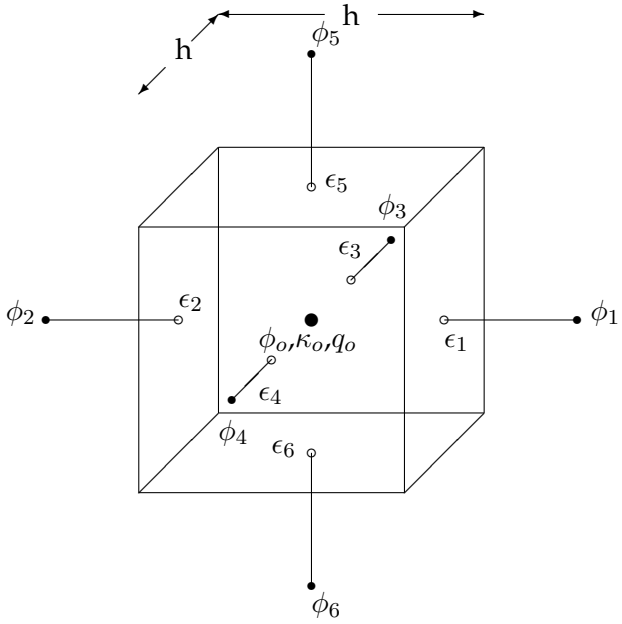

Figure 1.8: Main elements of the finite-difference method used for the solution of the Poisson-Boltzmann equation. Associated with each grid point (with a spacing of h Å) is a charge (q), Debye-Hückel inverse length (κ), and the electrostatic potential (ϕ) solved for each grid point. A value for the dielectric constant (ϵ) is associated with the midpoints of the lines between the grid points.

For more details about this equation one may see [31, 32, 33, 34]. The linearized PB equation usually must be solved numerically. A finite-difference method as implemented in the University of Houston Brownian Dynamics (UHBD) program [35, 36] was used to solve the equation and to calculate the self- and interaction-energies of the ionizable groups in protein.

The finite-difference approximation involves mapping of the molecule onto a three-dimensional cubic grid (see Figure 1.8). The atomic charges are partitioned to the eight grid points that form the cube around the actual charge location according to the tri-linear weighting function. The fractional charge assigned to that grid point is

$$q_0 = q(1-a)(1-b)(1-c) \tag{1.25}$$

where a, b, and c are the fractional distances along the grid axes giving the actual location of charge q from the grid point.

Continuous functions are approximated by distinct values at regular grid points which represent an appropriate average of the continuous function over the volume surrounding the given point. Equation 1.24 is integrated over the volume occupied by the cube shown in Figure 1.8.

The Gauss' theorem is applied ($\int_V (\nabla \cdot \mathbf{A}) \mathrm{d}v = \oint_\Sigma \mathbf{A} \cdot \mathbf{n} \mathrm{d}\sigma$), derivatives are approximated by the ratio of the differences, continuous functions (ϵ, Φ, ρ) are approximated by distinct functions at certain points of the cubic grid (see Figure 1.8). The following equation relating

the potentials, charges and dielectric constants is obtained

$$h^2 \sum_{i=1}^{6} \frac{\epsilon_i(\phi_i - \phi_o)}{h} = -4\pi q_o + \lambda(\mathbf{r}_o)\bar{\kappa}^2 \phi_o h^3. \tag{1.26}$$

Equation (1.26) represents the basis of the finite-difference method [37, 38, 39, 40]. It shows that the electrostatic potential at each point is linearly related to the potentials at the neighbouring points and must be found by an iteration method.

Modelling the Ionization Processes

Ionization refers to a process of loosing or accepting a proton. Two different Poisson-Boltzmann titration models for neutral and ionized forms of titratable groups are used in this study. The two models differ in the treatment of the charge changes that occur upon ionization, and in the dielectric constant assigned to the molecular interior. The single-site/20 titration model [24] includes all the partial atomic charges of the neutral form of each ionizable group, but treats ionization as an addition of a single $\pm$ 1 point-charge to one central atom. This single-site/20 model works well when is used with an internal dielectric constant of 20.

The full-charge/4 model [25] provides a more realistic description of the charge changes associated with ionization. The force field charges assigned to the neutral form of each ionizable group are replaced by the full set of partial charges describing the ionized form. Since more details are included than in the single-site/20 method this model is used with a solute dielectric constant of 4.

Examples of both models are shown in Table 1.1 for aspartic acid. The polar-hydrogen-only atomic charges are taken from the `CHARMm22` [41, 42] parameter set. The atomic radii are taken from the OPLS parameter set [43] for the single-site/20 model, and from the `PARSE` parameter set developed by Honig and coworkers [44] for the full-charge/4 model.

Table 1.1: Neutral and charged parameter set used for aspartic acid residue in the single-site/20 and full-charge/4 titration models.

Asp	single-site/20				full-charge/20			
Atom name	*neutral*		*charged*		*neutral*		*charged*	
	q	r	q	r	q	r	q	r
N	-0.350	1.625	-0.350	1.625	-0.400	1.500	-0.400	1.500
H	0.250	1.200	0.250	1.200	0.400	1.000	0.400	1.000
CA	0.100	1.900	0.100	1.900	0.000	2.000	0.000	2.000
CB	0.000	1.952	0.000	1.952	0.000	2.000	0.000	2.000
CG	0.700	1.875	0.700	1.875	0.000	2.000	0.100	1.700
OD1	-0.550	1.480	-0.550	1.480	-0.495	1.400	-0.550	1.400
OD2	-0.400	1.480	-1.400	1.480	-0.490	1.400	-0.550	1.400
HD	0.250	1.200	0.250	1.200	0.435	1.000	0.000	0.000
C	0.550	1.875	0.550	1.875	0.550	1.700	0.550	1.700
O	-0.550	1.480	-0.550	1.480	-0.550	1.400	-0.550	1.400

When the residue is considered neutral, the atomic charges and radii have values as in the *neutral* parameter set, when it is considered ionized, the atomic charges and radii have values taken from a *charged* parameter set. For the full-charge/4 method in the charged form of the aspartic acid residue, both the charge and the radius of the HD atom are set to zero.

Electrostatic Calculations

The method described herein can be applied to proteins with coordinates of heavy atoms established by solving the three-dimensional structures crystallographically or by multidimensional NMR spectroscopy, or generated via homology modelling. For the purpose of the Poisson-Boltzmann approach, the coordinates of hydrogens need to be established. It is questionable which method to use to initially place them and if all of them or, e.g. only polar ones. In these simulations an intermediate approach was used, i.e. all polar and all aromatic hydrogens were added. Their positions were generated with the HBUILD [45] command of `CHARMm22` [42]. Then their positions were optimized with 500 steps of the steepest descents energy minimization method. The number of added hydrogens corresponds to the fully protonated structures at low pH conditions.

Next, the electrostatic calculations with the UHBD program were performed. The temperature was set to 293K, ionic strength to 150mmol/litre and solvent dielectric constant to 80. The protein dielectric boundary was taken to be a Richards' probe-accessible surface [46] computed with a spherical probe of radius 1.4 Å and an initial density of 500 dots per atom [47].

The full-charge/4 titration method presented herein requires a minimum of four UHBD calculations for each titratable group in the protein; two for the protein and two for the model compound environment. The latter describes an amino acid in the water environment extracted from the protein, without changing its conformation and the N- and C-termini. So the overall charge on these termini sums to zero. The following values of $pK^{(w)}_{i,model}$ in equation 1.11 are used: ARG 12.0; LYS 10.4; ASP 4.0; GLU 4.4; TYR 9.6; HIS 6.3; CYS 8.3; N-terminus 7.5; C-terminus 3.8 [24, 48, 49]. They cannot be referred directly to experimental values of pK_as due to a different N- and C-termini. Each UHBD calculation is carried out with either neutral or charged set of parameters with all other charges set to zero.

In the single-site/20 method to obtain all the necessary energy terms only two UHBD calculations for each titratable group are required. Due to the application of "focusing" algorithm [47, 50, 24] more calculations are actually performed for each titratable group. However, the focusing method saves computer time by permitting the use of less extensive finite-difference grids in solving the Poisson-Boltzmann equation.

In the end such sets of calculations generate a file containing all the energies ($\|G_{ij}\|$ matrix) required to solve for the pH-dependent properties of interest. Such an output file is organized into a form shown in Figure 1.9. The first line contains the number of ionizable sites, M. Subsequent lines are organized into blocks describing each titratable group. The first line of each block contains information about the group itself, i.e. the model compound pK_a in water ($pK_{i,model}$), the basic or acidic type of the group, (γ_i), the electrostatic free energy difference for ionization of the group in the protein with all other sites in their reference state, relative to the same change in the model compound, (G_{ii}), and the index of the group. The remaining $M-i$ lines of each block contain the absolute values of the effective interaction potentials, (Ψ_{ij}), with the remaining $M-i$ sites ($j = i+1, ..., M$).

$$
\begin{array}{llll}
M & & & \\
\mathrm{pK}_{1,model} & \gamma_1 & G_{11} & 1 \\
\Psi_{1,2} & & & \\
\vdots & & & \\
\Psi_{1,M} & & & \\
\mathrm{pK}_{2,model} & \gamma_2 & G_{22} & 2 \\
\Psi_{2,3} & & & \\
\vdots & & & \\
\Psi_{2,M} & & & \\
\vdots & & & \\
\mathrm{pK}_{M,model} & \gamma_M & G_{MM} & M
\end{array}
$$

Figure 1.9: Format of the output file containing all the information necessary to solve the pH properties of proteins.

Treatment of Multiple Ionization State Problem

The above data from electrostatic calculations may be used to compute the electrostatic free energy of the protein (Eq. 1.11) in any of its 2^M protonation states. As mentioned before, according to first principles, the evaluation of pK_as and mean charges would require the knowledge of ΔG in all possible ionization states of a protein. There are, however, more efficient methods which allow for not enumerating all possible states. One of them, based on the Metropolis Monte Carlo algorithm [28], is implemented in the the DOPS (Distribution of Protonation States) program [51, 52, 53]. It extracts mean charges, pK_as and the most probable ionization states of the protein. Another one, based on grouping the titratable groups in clusters, is implemented in HYBRID [54] program and evaluates mean charges, pK_as, and electrostatic free energies of proteins.

The method based on Monte Carlo algorithm in the DOPS program [51, 52, 53] works in the following way. The initial state is taken with all residues in the system protonated. The Gibbs free energy for this starting state, $G1$, is computed based on the equation analogous to Eq. 1.10 but the reference state for the system is with all sites deprotonated. In this case the free energy of the molecule with the protonation state $(x_1^{'}, ..., x_M^{'})$ is

$$
\begin{aligned}
\Delta G(x_1^{'}, ..., x_M^{'}) = 2.303RT \sum_{i=1}^{M} x_i^{'}(\mathrm{pH} - \mathrm{pK}_i^{(w)}) + \sum_{i=1}^{M} x_i^{'}\gamma_i G_{ii} + \\
\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} [(q_i^o + x_i^{'})(q_j^o + x_j^{'}) - q_i^o q_j^o]\Psi_{ij}
\end{aligned}
\tag{1.27}
$$

where $x_i^{'}$ is 1 when the group is protonated and 0 when it is deprotonated, q_i^o is the formal charge of the deprotonated site, and Ψ_{ij} is absolute value of G_{ij}.

Next, a number is randomly chosen for each titratable group in the protein in the range $0 - 1$. Its ionization state is switched for those groups for which the randomly chosen number is greater than some value s. At each step an average of $(1 - s) \cdot 100\%$ of the ionizable groups are changed. The value of s should be adjusted so that approximately 50% of new states are accepted. The free energy, $G2$, is again calculated based on equation 1.27 and the second ionization state is either accepted or rejected. If the free energy is lower

or equal than the one in the previous step, the new state is accepted; if it is greater, the new state is accepted according to the probability based on the value of $\exp(-(G2 - G1)/RT)$. Again, a number is randomly chosen in the range $0 - 1$ and if it is smaller than $\exp(-(G2 - G1)/RT)$, the new state is accepted. New states are sampled for a large value of configurations depending on the size of the system. Due to the fact that the starting state is always far from equilibrium, the initial Monte Carlo steps are discarded in computing thermodynamic averages. The DOPS presents the lowest energy states and one is able to see if any degenerate low energy states exist and which residue should be treated as ionized in further computational studies, e.g. in molecular dynamics simulations.

The approach used in the HYBRID program [54] is based on separation of ionizable groups into clusters. There are strong correlations between ionization states of individual groups inside the cluster and inter-cluster correlations are small. A full partition function ($\Sigma_i e^{-G_i/\mathrm{RT}}$) is evaluated for each cluster and fractional ionizations of titratable groups belonging to the cluster are evaluated. Contributions to energies of titratable groups in cluster-cluster interactions are treated by the mean field approach [55]. In the mean field approximation, the ionization equilibrium of group i is influenced by group j according to $\zeta_j G_{ij}$, where ζ_j is the fractional ionization of group j. The starting state is with all groups ionized so that there are strong interactions between clusters. The partition function and fractional ionizations of the groups of the first cluster are evaluated with the usage of initially guessed charges. Using fractional ionizations of the first cluster and the initial, guessed charges for groups in other clusters, the partition function and fractional ionizations for the second cluster are obtained. Such procedure is repeated until the last cluster is reached and then it starts again with the first cluster. All clusters are updated in every cycle and the procedure is self-consistent. The process is iterated until it has converged, i.e. the changes in the fractional charges fall below a certain small value.

Both methods allow for calculations of ionization constants of titratable groups, net average charges, and electrostatic free energies as functions of pH.

Summary of the Algorithm

To compute the ionization constants of titratable residues in peptides or proteins a file containing the x, y, z coordinates of heavy atoms with their names is required. Having such, the following steps are performed. First, the hydrogens are added and their positions are optimized. Secondly, the finite-difference Poisson-Boltzmann method is used to calculate the self- and interaction-energies of titratable groups in the protein and in water. Next, based on the energy formula dependent on $(\mathrm{pH}, x_1^{'}, ..., x_M^{'})$, the Monte Carlo algorithm or any hybrid approach is used to determine the ionization constants, net average charges and electrostatic free energies.

Comparison with Measured pK$_a$s

This thesis compares computed pK$_a$s with measured pK$_a$s reported in previous publications. Two sources of experimental data are referred to, one are the kinetic studies and the other one is NMR spectroscopy. The use of kinetic studies to determine pK$_a$s is unfortunately associated with certain pitfalls [56, 57]. Due to the fact that the titration curve of a protein is the result of contributions of many titratable groups, of particular concern is the assignment of apparent pK$_a$s, determined from fitting to this curve, to specific ionizable groups. However, when structural information is available literature contains examples of

kinetic complete assignments of pK_as. Such is the case of hen egg white lysozyme [58]. In the case of HIV-1 PR researchers investigating the pH dependence of enzyme kinetics agree that the aspartyl dyad accounts for the apparent pK_as observed kinetically [1, 18]. NMR spectroscopy determines the ionization constants in proteins based on the variation in chemical shift which is sensitive to electronic environment around the nucleus of an atom [59]. In proximity to an ionizable group, changes in protonation state lead to changes in chemical shifts on the nuclei. The observed chemical shift is an average over protonated and deprotonated species. Plots of chemical shifts *versus* pH allow the pK_a values to be determined for individual titratable groups [60]. NMR spectroscopy can therefore provide unambiguous assignments of apparent pK_as to certain amino acid residues (e.g. hen and turkey egg white lysozyme [61]).

In order to compare the computational results with experimental kinetic studies one needs to compute the pH-dependence of K_I^{app} and k_{cat}^{app}/K_S^{app} [15, 18]. This means computing the fractional occupancy of the monoprotonated form of the aspartyl dyad, with and without bound inhibitor. Binding polynomial theory [62, 63, 54] provides a basis for computing these quantities. The apparent pK_as of the free enzyme are then determined by fitting the computed graph of $f(pH)$ to a two-pK_a model. This is the same approach that is used to extract the pK_as from experimental plots of k_{cat}^{app}/K_S^{app} *versus* pH [15, 18]. Therefore, the computed pK_as may be compared directly with experiment. The binding polynomial theory is also used to compute the pH at which the dyad possesses 0.5 and then 1.5 protons. These apparent pK_as are compared with the NMR studies.

Restrictions of the Method

Neither of the titration models used here accounts in detail for the flexibility of the protein or inhibitors. It is evident that the assumption of a single and rigid enzyme conformation is a possible source of error. Changes in conformation may be coupled with changes in ionization. For instance, solvent-exposed groups may reorient freely, so a single conformation may not necessarily be a good representative in the model. On the other hand, it seems probable that the use of an elevated dielectric constant helps to correct for the lack of an explicit treatment of flexibility [64, 65, 66]. However, in principle, it would be preferable to treat flexibility in detail. Such calculations would be computationally much more intensive and would require further developments. Some progress in this direction has been made, and modest improvements in accuracy, where limited flexibility is included [67] especially for key protons in or near ionizable groups [66], have been obtained. Neglecting the tautomerism in the calculations is also a probable source of error. Asp, Glu, Arg and His residues can exist in two neutral forms and in most computations a single neutral tautomeric form is assumed (e.g. OD2 atom of Asp and OE2 atom of Glu residue are protonated by default). For some residues their pK_as may be different depending on the choice of the neutral tautomeric form. In order to capture some of these gains in accuracy, the present study allows for all possible protonation states of the aspartyl dyad, explicitly considering tautomeric forms for this groups.
Another important source of problems may be the choice of partial charges, atomic radii and dielectric constant. Neither of the models presented herein accounts for the Van der Waals interactions between atoms in determining the pK_as. It is assumed that the response of the system upon ionization of a residue is purely electrostatic and the PB calculations include only electrostatic interactions. This is probably the mainly contributing element, however, one must be aware of the restrictions of the model while analysing the results.

The overall efficiency of the method, i.e. the agreement between predicted and experimental pK_as was tested for the single-site/20 [24, 26, 68] and full-charge/4 [26, 25] methods and correlation is sufficient.

1.2.4 Coordinates and Parameters

The present calculations are based upon six different crystal structures of HIV-1 PR. Five of them are drawn from the Protein Data Bank [69]. Two structures of the free enzyme are used: 3HVP [13] and 1HHP [70]. The former is a synthetic enzyme with cysteines 67 and 95 replaced by α-amino-N-butyric acid. The latter is obtained from the BRU isolate and expressed in *E. coli*. Four enzyme-inhibitor structures are used. One is form B of a new refinement of the HIV-1 PR/MVT-101 complex [20] that had previously been solved to 2.3 Å resolution [14], (Figure 1.4). The sequence of the enzyme in this complex is identical to that of in PDB file 3HVP. The second enzyme-inhibitor structure is 1HVR [23], HIV-1 PR complexed with the cyclic urea inhibitor XK-263 (Figure 1.6). The third structure is 1QBS [71], HIV-1 PR complexed with the cyclic urea inhibitor DMP-323 [22], (Figure 1.6). The fourth enzyme-inhibitor structure is 1HPX [72], HIV-1 PR complexed with the thioproline inhibitor KNI-272, (Figure 1.5). The sequence of the enzyme for this complex is identical to that of the free structure 1HHP. Table 1.2 lists the differences among the HIV-1 PR sequences in the various structures. All of the variations conserve electrical charge and are remote from the active site. Table 1.3 gives the resolution of the crystal structures and the pH of the mother liquors.

The crystal structure with MVT-101 includes a crystallographic water roughly 7 Å from the aspartyl dyad. This water molecule (no 301) links the inhibitor with the active-site flaps via two pairs of hydrogen bonds (see Figure 1.10). It is well localized and is not in contact with bulk solvent. Molecular dynamics simulations [73] showed that all four hydrogen bonds formed by this water molecule are stable over time, with a mean and root mean square deviation interaction energy of -61.1 ± 3.3 kJ/mol. In the present calculations this water is treated explicitly as part of the solute and assigned dielectric constant of 4 for the full-charge/4 model and 20 for the single-site/20 model; i.e. it is not replaced by continuum solvent. However, the rest of the solvent is modelled as continuum.

Table 1.2: Comparison of HIV-1 protease sequences.

Position	1HHP/1HPX	3HVP	1HVR	1QBS
3	ILE	ILE	VAL	VAL
14	LYS	ARG	LYS	LYS
37	SER	ASN	SER	SER
41	ARG	LYS	ARG	ARG
63	LEU	PRO	LEU	LEU
64	ILE	VAL	ILE	ILE
67	CYS	ABA[a]	CYS	CSO[b]
95	CYS	ABA	ALA	ALA

[a] α-amino-N-butyric acid
[b] S-hydroxycysteine

Table 1.3: Crystal structures used in calculations.

	3HVP[a]	1HHP[b]	1HVR[c]	HIV-1 PR/MVT-101[d]	1QBS[e]	1HPX[f]
pH of liquor	7.0	7.0[g]	5.4	5.4	5.4	5.6-5.9
Resolution (Å)	2.8	2.7	1.8	2.0	1.8	2.0

[a][13]; [b][70]; [c][23]; [d][20]; [e][71]; [f][72]; [g]Dr R. Poljak, personal communication.

Two crystallographic water molecules are located in the active site of the KNI-272 complex: waters 301 and 607 [72]. Water 301 is analogous to the crystallographic water molecule in the MVT-101 structure, and is again included explicitly as part of the low dielectric solute in the present pK_a calculations. Water 607 is of particular interest because it contacts OD2 of Asp125. It has been proposed that this water molecule is important in determining the pK_a value of Asp125 [72, 74]. The pK_as reported herein are computed with water 607 treated explicitly as part of the low-dielectric solute and oriented to donate a hydrogen bond to Asp125, which is believed to be preferentially ionized [74]. Calculations were also done with this water in other orientations and treated as bulk solvent.

The structure of an enzyme-substrate complex was achieved by modifications of the crystal structure of HIV-1 PR : MVT-101 as described earlier. The catalytic water (207) was added explicitly.

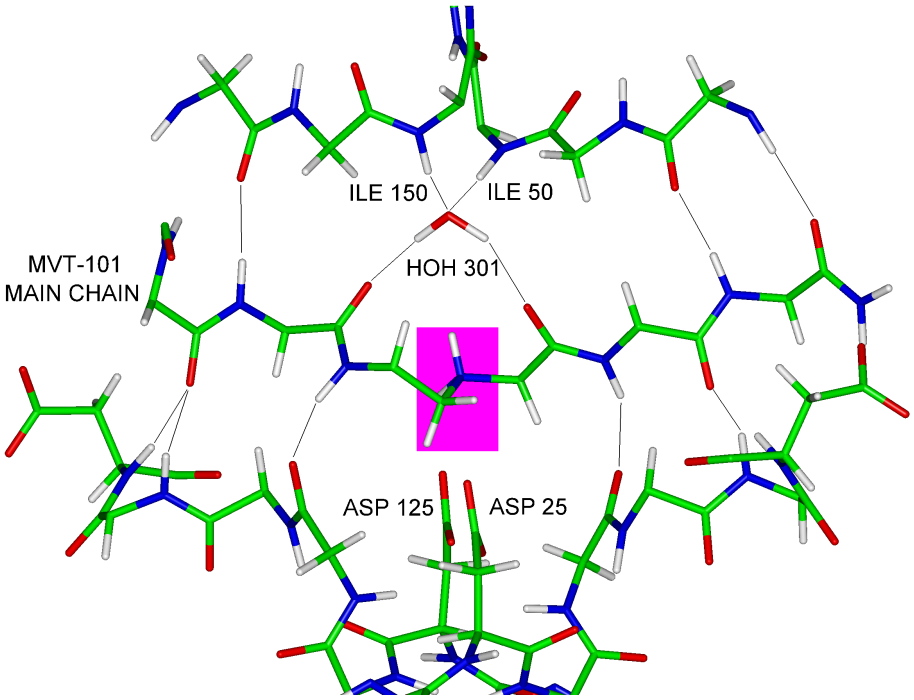

Figure 1.10: Hydrogen bonds linking the MVT-101 inhibitor with HIV-1 protease. The box denotes the reduced link of the inhibitor.

1.2.5 Modifications of the Computational Methodology

Several modifications of the computational methodology are introduced and described in this Section. First, the present calculations treat the protonation equilibria of the aspartyl dyad in more detail allowing for both tautomeric forms. Secondly, the new methodology accounts for the symmetry of HIV-1 PR. Next, the protons of the aspartic acid side chains are positioned "by hand", rather than with the HBUILD [45] procedure of `CHARMm` [41].

Previously, the calculations used a reference state in which no groups were ionized. One oxygen of each carboxylic group regarded as protonated in this neutral reference state had to be selected [24, 25]. In contrast, the present procedure allows the carboxyl groups of the aspartyl dyad to adopt either of two neutral forms, thus establishing an equilibrium between three forms of each group: deprotonated, protonated at OD1, and protonated at OD2. Therefore, four different neutral forms are considered for free HIV-1 PR and for HIV-1 PR complexed with a non-protonatable inhibitor. The inhibitor MVT-101 possesses a basic secondary amine with two different positions for its proton in the neutral form. Calculations for the complex with this inhibitor allow for both positions of the proton in the neutral form, as well as for the two neutral forms of each active site carboxyl, to yield eight different neutral protonation states.

In order to allow for these mutually exclusive neutral states of the active site aspartyl groups (and protonatable MVT-101 inhibitor), the present calculations use a fully ionized reference state instead of the fully neutral one described in Section 1.2.3. The change of the reference state is accomplished as follows.

First, a full set of PB energy calculations is carried out for each of the four neutral states of the system. In the case of the HIV-1 PR: MVT-101 complex there are eight neutral states because of the additional protonatable site of the reduced link of this inhibitor. These calculations yield four (eight in the case of HIV-1 PR : MVT-101 complex) complete interaction matrices with self- and interaction energies [75], G_{ij}, $i,j \in [1,M]$, for the titratable groups. These matrices are computed for the neutral reference state and are next altered so that they refer to a reference state in which all titratable groups are ionized. The energy of the fully ionized reference state relative to the neutral reference state is

$$G_{ion} = \sum_{j \geq i}^{M} G_{ij}\gamma_i\gamma_j \tag{1.28}$$

where γ_i is $+1$ for bases, and -1 for acids (see equations 1.11, 1.12, 1.16). Taking group K and assuming that only this group is neutral, the self energy term G_{KK} is now the work of neutralizing group K in the protein with all other groups ionized, relative to the work of neutralizing that group in solution. With respect to the fully ionized reference state, G_{KK} may be written as the difference between the energy with only group K neutral ($G_{ion} - \sum_{i=1}^{M} G_{ik}\gamma_i\gamma_k$) and the reference energy G_{ion}. Therefore, the self-energy term G_{KK} is

$$G_{KK} = -\sum_{i=1}^{M} G_{ik}\gamma_i\gamma_k. \tag{1.29}$$

The non-diagonal matrix element, the interaction energy between sites K and L with both groups neutral, G^{int}_{KL}, relative to the new reference state, is the energy with both groups neutral, G_{KL}, minus energies with only K and L neutral, G_{KK} and G_{LL},

$$G^{int}_{KL} = G_{KL} - G_{KK} - G_{LL} - G_{ion} \tag{1.30}$$

where G_{LL} is given by the formula analogous to that for G_{KK} above. The energy with both groups neutral, G_{KL}, relative to the new reference state is

$$G_{KL} = G_{ion} - \sum_{i=1}^{M} G_{ik}\gamma_i\gamma_k - \sum_{i=1}^{M} G_{il}\gamma_i\gamma_l + G_{kl}\gamma_k\gamma_l. \tag{1.31}$$

Substituting all the terms into equation 1.30, the interaction energy of groups K and L in the all-ionized reference state is

$$G_{KL}^{int} = G_{kl}\gamma_k\gamma_l. \tag{1.32}$$

The calculations yield separate energy matrices for each set of neutral states of the two or three groups with multiple neutral sites. Many of the entries, i.e. interaction energies in these matrices are redundant because HIV-1 PR possesses rotational symmetry. For example, the interactions among groups outside the active site (the place where multiple protonation is accepted) should be the same among all the matrices. However, values that should be equivalent do vary somewhat. This results partially from numerical noise and partially from slight changes in hydrogen placement that result from the different neutral states. Here, ambiguity is avoided by averaging of redundant interaction energies. To merge the separate energy matrices into one an averaging procedure has been introduced.

Assuming one takes the case of a protein with M ionizable groups indexed by I. In general, the number of neutral states possible for the group I is $N(I)$. $N(I)$ equals 1 for all the groups apart from a maximum of three groups, for which $N(I)$ equals 2, i.e. the active site aspartates and the amine of MVT-101. If $N(I)$ equals 2 for these three groups, the titration calculations yield eight energy matrices: $G(1,1,1,i,j)$, $G(1,2,1,i,j)$, $G(1,2,2,i,j)$, $G(2,1,1,i,j)$, etc., where the first three integers denote the assumed neutral state of the three groups, e.g. for aspartic acid at the OD1 or OD2 atom. The eight matrices are merged into a single one $g(i,j)$ where i and j range over all titration sites ($i,j \in 1, M+3$). If one takes as an example a molecule with 5 titratable groups, where groups 2, 3 and 4 (J, K and L) have two accounted for neutral sites $n(i)$, one may write:

i	1	2	3	4	5	6	7	8
group number $I(i)$	1	2	2	3	3	4	4	5
neutral site $n(i)$	1	1	2	1	2	1	2	1

In the final interaction matrix $g(i,j)$ a given element involves 0, 1 or 2 of the special sites J, K, L. The averaging is performed over all unspecified neutral forms of these groups. Therefore, when neither $I(i)$ nor $I(j)$ are the special sites we have

$$g(i,j) = 1/8 \sum_{n1=1}^{2} \sum_{n2=1}^{2} \sum_{n3=1}^{2} G(n1, n2, n3, I(i), I(j)). \tag{1.33}$$

If $I(i)$ equals J then

$$g(i,j) = 1/4 \sum_{n2=1}^{2} \sum_{n3=1}^{2} G(n(i), n2, n3, I(i), I(j)) \tag{1.34}$$

and the averaging involves two groups. When $I(i) = J$ and $I(j) = K$, i.e. two of the groups are involved, the average involves only the third group

$$g(i,j) = 1/2 \sum_{n3=1}^{2} G(n(i), n(j), n3, I(i), I(j)). \tag{1.35}$$

If $I(i) = I(j)$ and $n(i) = n(j)$ the self-energy of site $n(i)$ of group $I(i)$ is obtained. To prohibit simultaneous protonation of both neutral forms a large value of 210 kJ/mol is assigned to $g(i,j)$ if $I(i) = I(j) = J, K$ or L and $n(i) \neq n(j)$. It serves as a large repulsion that prevents simultaneous protonation of both oxygens of a single carboxyl group between the protonation sites at OD1 and OD2 of each aspartyl side-chain. In order to correct for the statistical effect of adding a second protonation site, which by itself favours the neutral form of the side-chain, the model compound pK_a of each aspartyl group must be changed from its normal value of 4.0 to $4.0 - \log 2 = 3.7$. Such correction for statistical effects is analogous in the MVT-101 inhibitor case.

Last modification is considered with the noticed fact that the HBUILD command of `CHARMm` occasionally generates unlikely proton positions for the neutral aspartyl dyad. For example, Asp25 HD2 might be placed *anti* to Asp25 OD1 without an evident energetic basis. Such locations lead to erratic computed pK_as, especially for calculations with a low protein dielectric constant. Therefore, the coordinates of the carboxylic protons for the neutral forms of the aspartyl dyad are constructed in their *syn* positions, and the energy is then minimized with only the aspartyl hydrogens free to move, using a dielectric constant of 1 and in the absence of the protein environment. Similarly, the orientation of water 607 and of the hydroxyl of KNI-272 are adjusted "by hand" to optimize their interactions with the nearest hydrogen-bond acceptor; *viz.*, Asp125 OD2. Finally, the ring hydroxyl hydrogens of XK-263 and DMP-323 were rotated to match the arrangement diagrammed previously [76]. All other hydrogen coordinates are generated with HBUILD [45], as previously described [24].

1.3 Results - The Protonation States of Titratable Amino Acids

This Section presents the calculated pK_as for free and inhibited HIV-1 PR. The results for the complex with the model substrate are also shown. The computations include the interactions and energetics of all ionizable groups in the system, and yield titration curves for all of them. However, for brevity, results are presented mainly for those groups for which experimental data are available.

1.3.1 Free HIV-1 PR

The calculations for both the single-site/20 and full-charge/4 models are shown in Table 1.4. The Table also compares the computed pK_as with those extracted from experimental curves. They agree quite well with experiment.

The calculations yield apparent pK_as of $2.1 - 3.4$ and $5.3 - 6.9$, while the experimental pK_as are $3.1 - 3.7$ and $4.9 - 6.8$; and the results are similar for both crystal structures. The full-charge/4 calculations yield a somewhat larger gap between pK1 and pK2 than the single-site/20 calculations. The large range of experimental values makes it difficult to determine which titration model is more accurate in this case. Table 1.4 also includes the result of an NMR study of the free enzyme, which suggested that that aspartyl pK_as are less than 5.9. If correct, this would imply that the single-site/20 calculations are more accurate here than the full-charge/4 calculations. It is worth re-emphasizing that for the free enzyme, pK1 and pK2 cannot each be assigned to one of the catalytic aspartyl groups. The two aspartyl groups are chemically equivalent because of the symmetry of the enzyme, and the two pK_as belong to the dyad as a whole. This is illustrated in Figure 1.11, which

Table 1.4: Apparent pK_as of the aspartyl dyad of free HIV-1 PR, measured experimentally and computed with free-enzyme crystal structures 1HHP and 3HVP.

	Expt 1[a]	Expt 2[b]	Expt 3[c]	Expt 4[d]
pK1	3.4-3.7	3.3	3.1-3.3	<5.9
pK2	5.5-6.5	6.8	4.9-5.3	<5.9

	Calculated			
	1HHP		3HVP	
	single-site/20	full-charge/4	single-site/20	full-charge/4
pK1	3.1	3.4	2.9	2.1
pK2	5.3	6.9	5.3	6.7

[a] From apparent catalytic constant [1]
[b] From apparent catalytic constant [18]
[c] From apparent inhibition constant [1]
[d] From chemical shift of aspartyl dyad in the free enzyme [19]

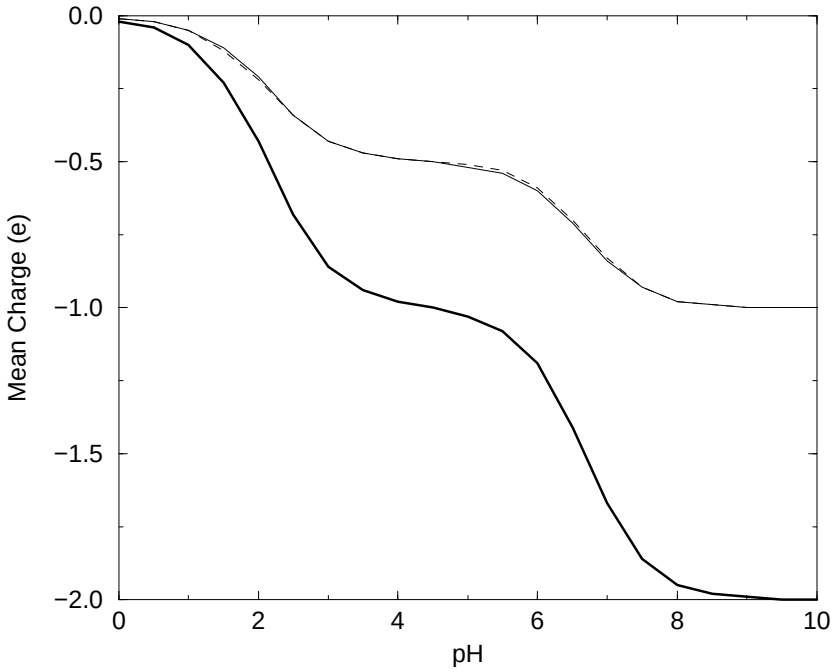

Figure 1.11: Computed mean charge of aspartyl dyad of HIV-1 PR *versus* pH. Thin solid and dashed lines – separate aspartyl groups; heavy solid line – total charge of aspartyl dyad. Results are for 3HVP, full-charge/4 model.

shows the computed mean charge of the two separate aspartates and of the complete dyad as a function of pH. The results shown are for 3HVP with full-charge/4 parameters.

Table 1.5 shows the pK_a results for other ionizable groups in the native HIV-1 PR. Results are presented for the single-site/20 method for 3HVP [13] crystal structure for one out of four possible protonation sites of aspartyl groups (OD2 of Asp25 and OD1 of Asp125). The model values of the ionization constant $pK_{a,model}$ and the mean charges at pH 7 are also shown. The group designates the atom at which a charge of ± 1 is added while modelling the ionization of the residue. The mean charge of the whole protein at pH 7 is 4.53 [e]. On the average all aspartic and glutamic acids are negatively charged and all lysines and arginines have a positive charge. Histidine pK_as are shifted up by 0.3 units. For the single-site/20 model the protein mean charge at pH 7 for the 1HHP crystal structure is 4.78 [e]. In the full-charge/4 calculations the mean charges of the protein at pH 7 are 5.15 and 5.34 [e] for the 3HVP and 1HHP structures, respectively.

1.3.2 HIV-1 PR Complexed with Neutral, Symmetric Cyclic Urea Inhibitors

Table 1.6 presents the computed and measured pK_as of the aspartyl dyad and of other carboxylic acid side chains for the complexes of HIV-1 PR with the cyclic urea inhibitors: DMP-323 and XK-263. The computed pK_as of the dyad are defined as the pH values for which its net charge is -0.5 and -1.5. The pK_as of the other groups are the pH values for which the net charge is -0.5. The pK_as measured by NMR [76] for the complex with DMP-323 are included for comparison. The two inhibitors DMP-323 and XK-263 are chemically similar (Figure 1.6) and bind to HIV-1 PR in extremely similar conformations (see 1HVR [23] and 1QBS [71]). Not surprisingly, then, the pK_as computed for the two different complexes are quite similar to each other.

For the catalytic dyad the calculations using the full-charge/4 titration model, for both inhibitors, agree rather well with the NMR measurements. They indicate that the dyad is doubly protonated at least to pH 7. The values of 20.5 and 22.0 are obviously implausible because the protein would not be stable at such a high pH. These theoretical values clearly show that for the given structure it is highly unlikely to remove both protons from the dyad, i.e. the dyad retains one proton to the maximal pH at which the protein is stable. The single-site/20 model does not agree with the NMR results for the dyad.

The difference between the two models results largely from the difference in the solute dielectric constants they use. The full-charge/4 model yields high pK_as for the dyad because the low-dielectric inhibitor displaces high-dielectric solvent from the dyad without adding back new stabilizing interactions that would compensate for the loss of solvation. This makes ionization of the dyad energetically expensive. When the single-site/20 model is used, the high internal dielectric constant leads to a smaller desolvation penalty, making it easier to ionize the dyad, and yielding pK_as that are rather similar to the values for the free enzyme.

Table 1.6 reports separate values for the pK_as of the two monomers of HIV-1 PR because the crystal structures (1HVR and 1QBS) break the symmetry of the dimers. However, because the real complex is symmetric, NMR experiments do not distinguish between equivalent residues in the two monomers, and only one experimental value is given. For most residues, the computed pK_as are similar for both crystal structures, and the computed pK_as also correlate well with the experiment. Only one computed value for the complex with DMP-323 inhibitor and three computed values for the complex with XK-263 inhibitor deviate more than 1 pH unit from the experimental value. All are full-charge/4 results for

Table 1.5: pK_as and mean charges at pH 7 for ionizable residues in the free HIV-1 PR (3HVP).

Residue no	Group	Residue name	$pK_{a,model}$	pK_a	q at pH 7 [e]
1	N	PRON	7.5	8.6	0.960
8	CZ	ARG	12.0	13.6	1.000
14	CZ	ARG	12.0	12.3	1.000
20	NZ	LYS	10.4	10.5	1.000
21	CD	GLU	4.4	4.2	-0.997
25	CG	ASP	4.0	3.5	-0.992
29	CG	ASP	4.0	1.7	-1.000
30	CG	ASP	4.0	3.8	-0.998
34	CD	GLU	4.4	3.7	-0.999
35	CD	GLU	4.4	3.6	-0.999
41	NZ	LYS	10.4	10.3	0.999
43	NZ	LYS	10.4	10.3	0.999
45	NZ	LYS	10.4	10.8	1.000
55	NZ	LYS	10.4	10.3	0.999
57	CZ	ARG	12.0	15.1	1.000
59	OH	TYR	9.6	10.5	-0.001
60	CG	ASP	4.0	3.5	-1.000
65	CD	GLU	4.4	3.3	-1.000
69	NE2	HIS	6.3	6.6	0.286
70	NZ	LYS	10.4	10.9	1.000
87	CZ	ARG	12.0	13.1	1.000
99	C	PHEC	3.8	2.1	-1.000
101	N	PRON	7.5	8.6	0.962
108	CZ	ARG	12.0	13.5	1.000
114	CZ	ARG	12.0	12.3	1.000
120	NZ	LYS	10.4	10.5	1.000
121	CD	GLU	4.4	4.2	-0.997
125	CG	ASP	4.0	4.8	-0.982
129	CG	ASP	4.0	1.5	-1.000
130	CG	ASP	4.0	3.7	-0.998
134	CD	GLU	4.4	3.7	-0.999
135	CD	GLU	4.4	3.6	-0.999
141	NZ	LYS	10.4	10.3	0.999
143	NZ	LYS	10.4	10.3	0.999
145	NZ	LYS	10.4	10.8	1.000
155	NZ	LYS	10.4	10.3	0.999
157	CZ	ARG	12.0	15.1	1.000
159	OH	TYR	9.6	10.5	-0.001
160	CG	ASP	4.0	3.4	-1.000
165	CD	GLU	4.4	3.3	-1.000
169	NE2	HIS	6.3	6.6	0.288
170	NZ	LYS	10.4	10.9	1.000
187	CZ	ARG	12.0	13.0	1.000
199	C	PHEC	3.8	2.1	-1.000

the "B" monomer. In the case of XK-263, the discrepancies evidently result from several unlikely conformations in the "B" monomer of the structures. For example, a carboxyl oxygen of Asp B60 in 1HVR structure lies only 2.96 Å from the side-chain amide oxygen of Gln B58. This accounts for the high computed pK_as of this group in the "B" monomer of 1HVR structure. In contrast, the corresponding residue in the "A" monomer of the 1HVR structure, Asp A60, has a more reasonable position, with one oxygen 2.19 Å from an amide hydrogen of Gln B58. The full-charge/4 calculations are sensitive to these conformational details, as is evident from Table 1.6. It is not so clear why the pK_a of Glu B65 in the DMP-323 complex is in error by more than 1.

1.3.3 HIV-1 PR Complexed with the Neutral, Asymmetric Inhibitor KNI-272

Table 1.7 presents experimental [77] and computed pK_as for HIV-1 PR complexed with the asymmetric inhibitor KNI-272. Results are presented for the aspartyl dyad and for six other aspartyl groups for which experimental data are available. The two monomers are not equivalent when the asymmetric inhibitor is bound. In particular, the two aspartyl residues of the active site of HIV-1 PR, Asp25 and Asp125, are distinguishable in both the NMR experiment and the calculations. Therefore, the various groups are identified by their numbers in the PDB file.

For the dyad the full-charge/4 calculations again agree relatively well with experiment. In particular, they are consistent with the experimental result that Asp25 is protonated and Asp125 deprotonated from pH 2.5 to 6.5. As above, the extreme pK_as for the dyad indicate a charge state that does not change with pH until the protein itself becomes unstable. The

Table 1.6: Computed pK_as for complex of HIV-1 PR with cyclic urea inhibitors DMP-323 and XK-263 (crystal structures 1QBS and 1HVR, respectively), compared with previous measurements for complex with the cyclic urea inhibitor DMP-323. The two computed results in each column, A and B, are for the two subunits of the enzyme; the crystal structures are not symmetric.

		DMP-323		XK-263	
	Expt[a]	single-site/20	full-charge/4	single-site/20	full-charge/4
Dyad pK1	>7.2	3.6	7.9	3.6	7.1
Dyad pK2	>7.2	6.5	20.5	6.4	22.0
subunits –>		A B	A B	A B	A B
Asp29	2.0	1.7, 2.5	2.1, 2.2	1.1, 1.4	1.6, -0.6
Asp30	4.0	3.3, 3.5	3.2, 3.7	3.0, 3.3	3.6, 3.3
Asp60	3.1	2.4, 3.3	2.9, 3.9	2.1, 3.9	2.6, 5.6
Glu21	4.5	4.4, 4.6	4.7, 5.1	4.2, 4.2	4.8, 4.8
Glu34	4.9	4.9, 4.6	5.8, 5.4	4.3, 4.4	4.9, 4.1
Glu35	3.7	3.8, 3.6	3.6, 4.4	3.6, 3.9	3.8, 5.5
Glu65	3.7	3.6, 4.2	3.9, 5.0	3.7, 3.6	3.9, 3.7

[a] [76]

Table 1.7: Computed pK_as for complex of HIV-1 PR with asymmetric inhibitor KNI-272 (crystal structure 1HPX), compared with measurements for this complex [77]. Two buried crystallographically resolved water molecules (301 and 607 [72]) were included as the solute structure in calculations. The crystal structure is not symmetric and NMR spectra show distinguishable chemical shifts for some of the residues.

	Expt[a]	single-site/20	full-charge/4	
		waters 301 and 607	waters 301 and 607	water 301 only
Asp25	>6.2	7.0	27.	26.
Asp125	<2.5	4.5	1.9	4.6
Asp29	<2.5	2.4	–0.1	–0.3
Asp30	3.88	2.2	3.3	2.9
Asp60	2.99	3.0	1.0	1.0
Asp129	<2.5	2.5	–1.6	–1.6
Asp130	3.78	3.3	3.8	3.5
Asp160	2.99	2.5	1.4	1.3

[a] [77]

difference between the pK_as of the two aspartates correlates with structure in an intuitively reasonable way: Asp125 can receive three hydrogen bonds: one from the hydroxyl of KNI-272, one from Gly127, and one from water 607. Asp25, in contrast, receives only one hydrogen bond, from Gly27. Furthermore, one carboxyl oxygen of Asp25 is in close contact (2.55 Å in the crystal structure) with a carbonyl oxygen of the inhibitor.

Additional calculations with the full-charge/4 model indicate that water 607 is important in setting the pK_a of Asp125, as mentioned previously in Section 1.2.4 [72]. Thus, when water 607 is not represented as part of the solute, but only as a droplet of high-dielectric bulk solvent, the computed pK_a of Asp125 rises from 1.9 to 4.6. This change in the model causes changes of only a few tenths in the other computed pK_as. The calculations presented in Table 1.7 have water 607 forming a geometrically optimal hydrogen bond with the carboxyl group of Asp125, the one believed to be preferentially ionized, based upon the NMR measurements. Alternative orientations of water 607 raise the computed pK_a of Asp125 to values that disagree with the NMR results.

For the dyad the single-site/20 calculations do not agree as well with the NMR results as the full-charge/4 calculations. In particular, they fail to yield the lowered pK_a of Asp125, although they do yield good results for the other pK_as. The single-site/20 calculations are not sensitive to the presence of either water 607 or 301 in the active site. Table 1.7 shows that the calculated pK_as for aspartyl groups outside the active site are reasonably accurate for both titration models. The NMR study suggests that Asp29 and Asp129 do not titrate in the pH range $2.5 - 6.2$, in at least qualitative agreement with the calculations. The experimentally observed variation in their chemical shift over this pH range seem to result from interactions with Asp30 and Asp130, which do titrate in this range [77]. The low pK_as of Asp29 and Asp129 result from their strong interactions with two arginine residues each. The computed pK_as of Asp29 and Asp129 vary significantly among the different structures examined here (Tables 1.6 and 1.7), when the full-charge/4 model is used. These changes

Table 1.8: Computed pK_as for complex of HIV-1 PR with amine inhibitor MVT-101 [73]. Water corresponding to water 301 in 1HPX is included explicitly. Crystal structure is not symmetric.

	single-site/20	full-charge/4
Asp25	4.7	10.8
Asp125	0.2	-6.0
Amine of MVT-101	15.	23.

reflect differences in the local environment of these groups which border the active site. The sensitivity of the full-charge/4 calculations to conformation is a significant source of noise in the results. Optimally, this would be dealt with by appropriate averaging over a number of energetically accessible conformations.

1.3.4 HIV-1 PR Complexed with the Cationic, Asymmetric Inhibitor MVT-101

The computed pK_as of the aspartyl dyad in the presence of the asymmetric, basic inhibitor MVT-101 are presented in Table 1.8. The pK_a of the reduced amide nitrogen of the inhibitor itself is also included. The initial pK_a of the secondary amine in MVT-101 was set to 10 based upon the pK_as of related compounds [78]. The calculated mean charge of the protein at pH 7 is shifted up, compared to the native enzyme, and is 4.57 and 6.14 [e] for the single-site/20 and full-charge/4 model, respectively. No experimental pK_as are available for this complex. The full-charge/4 parameters predict the aspartyl dyad to be monoprotonated across a range of pH that spans the experimental range of stability of the enzyme itself. This is similar to the case of KNI-272, except that both pK_as are shifted lower for MVT-101. This shift results from the stabilizing effect of the aspartyl groups with the cationic inhibitor, which is predicted to be strongly basic.

Although the pK_as of this complex have not been measured experimentally, the present results may be compared with previous calculations and measurements for cationic inhibitors. A previous molecular dynamics study of HIV-1 PR complexed with MVT-101 suggested that the pK_as of the aspartyl dyad are lower than 5.4 [79]. This disagrees with the full-charge/4 calculations, and agrees with the single-site/20 calculations. However, a similar molecular dynamics study [73] based upon a more highly refined version of the crystal structure [20] suggests that the dyad is, in fact, monoprotonated at this pH, in agreement with the full-charge/4 calculations. Kinetic studies [15] of a proline-containing cationic HIV-1 PR inhibitor [80] are also of interest. The measurements suggest that the prolyl inhibitor binds preferentially to the doubly deprotonated form of the aspartyl dyad in the pH range ~ 3.5 to ~ 8 [15], implying that the dyad is doubly ionized in this range when the inhibitor is bound. This result agrees better with single-site/20 calculations, but it should be borne in mind, that the inhibitors are not the same. In particular, the prolyl inhibitor has a much bulkier cationic group, which may assume a significantly different conformation in the active site.

Examination of the crystal structure of this complex shows why the calculations yield a very low pK_a for Asp125. This group is close (2.96 Å) to the cationic amino group of the inhibitor, and it also receives a weak hydrogen-bond from Gly127: the N–O distance

Table 1.9: Computed pK_as for the apoenzyme of the HIV-1 PR : MVT-101 and HIV-1 PR : XK-263 complexes. Coordinates of the inhibitors were not included in the calculations.

	Apoenzyme MVT-101		Apoenzyme XK-263	
	single-site/20	full-charge/4	single-site/20	full-charge/4
Dyad pK1	3.3	2.9	2.6	2.5
Dyad pK2	5.9	9.3	5.3	7.2

is 3.07 Å and the N-H-O angle is 146°. The crystal structure of the complex also shows why the calculations predict Asp25 to have a high pK_a of 10.8. The closest approach of Asp25 to the positively charged amino group is 4.46 Å, but Asp25 approaches negatively charged Asp125 to within 2.68 Å. In addition, Asp25 receives only one weak hydrogen bond, donated by Gly27: the N–O distance is 3.05 Å and the N-H-O angle is 134°. It should be noted that neither of the aspartates contacts a crystallographic solvent molecule or is accessible to bulk solvent. In summary, the electrostatic calculations are at the least consistent with intuition, based upon graphical examination of the structure.

1.3.5 Apoenzymes of HIV-1 PR : MVT-101 and HIV-1 PR : XK-263

The calculations for the apoenzymes of HIV-1 proteases were performed in order to check the influence of the inhibitors' charges on the pK_as of the active site apartyl side chains. The results were compared with those obtained for appropriate complexes with inhibitors (see Tables 1.8 and 1.6). The single-site/20 titration model predicts moderate changes. Like in complexes with inhibitors both aspartic acids titrate below pH 7. Full-charge/4 model gives more significant changes which are observed for both apoenzyme structures. In the complex with XK-263 inhibitor the calculated pK_as show that it is much harder to extract the proton from either of the aspartates when inhibitor is present, what is in agreement with intuition as the dyad is inaccessible to bulk solvent (see Tables 1.6 and 1.9). This effect is smaller for the complex with MVT-101 inhibitor (Tables 1.8 and 1.9). At pH 7 one of the aspartates is protonated and the other deprotonated, both in the apoenzyme structure and in the complex with inhibitor. However, the MVT-101 inhibitor imposes a large gap in the pK_as as one of them is shifted up markedly. The negative and positive charge of aspartyl groups in the complex with MVT-101 inhibitor is therefore more stable. The effects which result only from changes in the structure of the protein caused by binding of the ligand may be noticed. The direct effect of the presence of the inhibitor influences the pK_as of the dyad. The interactions with ligands, e.g. hydrogen bonds with strongly basic MVT-101 inhibitor, impose a larger gap between the ionization constants in both exemplar calculations.

1.3.6 HIV-1 PR Complexed with a Model Substrate

The computed pK_as of the aspartyl dyad in the HIV-1 PR complexed with a model substrate are presented in Table 1.10. Catalytic water 207 in the active site and the water corresponding to water 301 in 1HPX structure are either included explicitly or as a test case these waters are removed and treated as a part of bulk solvent.

Table 1.10: Computed pK_as for the complex of HIV-1 PR with a model substrate. Results are presented for structures including all crystal waters, water molecules 207 and 301 only, and without crystal waters. The structure is not symmetric.

	All crystal waters	
	single-site/20	full-charge/4
Asp25	6.7	21.
Asp125	4.0	2.3

	Waters 207 and 301 only	
	single-site/20	full-charge/4
Asp25	6.6	23.
Asp125	4.0	3.4

	No crystal waters	
	single-site/20	full-charge/4
Asp25	6.9	24.
Asp125	3.5	3.2

The full-charge/4 calculations show that in the presence of a model substrate Asp25 possesses a hydrogen. The pK_as shift in comparison with the same data for the HIV-1 PR : MVT-101 complex suggesting that the active site environment needs more positive charge than with the inhibitor. The single-site/20 calculations also shift the pK_as but only a few units. However, they still indicate that both aspartates are deprotonated at pH 7. This difference may result from the fact that the single-site/20 model is not sensitive to small changes of conformation and leads to underestimates of charge-charge interactions. This may be of great importance for the dyad situated inside the protein and being inaccessible to bulk solvent. The full-charge/4 calculations are consistent with the acid-base model of the reaction mechanism presented in Section 1.1.2 which shows that one of the aspartyl groups has to be protonated.

1.4 Conclusions

1.4.1 Analysis of pK_a Calculations

In this Chapter, computed pK_as were compared with experiment for three forms of HIV-1 protease: the free enzyme, the complexes with inhibitors DMP-323 and KNI-272, and for the two titration models. The single-site/20 model yields good agreement with experiment for the free enzyme, but it fails to reproduce the large pK_a shifts found experimentally for the inhibited enzyme. The full-charge/4 model reproduces these shifts much better, and also performs quite well for the free enzyme. However, some pK_as computed with the full-charge/4 model show exaggerated shifts that are clearly unrealistic. For example, the

pK_a of Asp25 is computed to be 27 for the KNI-272 complex (Table 1.7). Such values are interpreted as indications that the group remains fixed in a single ionization state at any pH for which the complex is actually stable in solution. Calculations for HIV-1 PR with the cyclic urea inhibitor, XK-263, yield results very similar to those for DMP-323.

Computed pK_as were also reported for the complex of the cationic inhibitor MVT-101 with HIV-1 PR. The full-charge/4 calculations suggest that the aspartyl dyad will be monoprotonated over a wide pH range, and certainly to the lowest pH at which the complex is stable. Given the uncertainty in the calculations, it is conceivable that the second aspartyl group might deprotonate at about neutral pH rather than near 11 (Table 1.8). Measured pK_as are not available in this case.

Calculations for the complex of the model substrate with HIV-1 PR were also performed. The pK_as for aspartic dyad were reported. There are no experimental data to compare to, nevertheless, the reported reaction mechanisms [1, 2] suggest protonation of one of the aspartates in the first stage of the reaction what is in agreement with the full-charge/4 calculations.

1.4.2 Comparison of Single-site/20 and Full-charge/4 Models

While comparing the two titration models from the previous studies the single-site/20 model has been found to be more accurate statistically than the full-charge/4 model [26]. However, the cases used to generate these statistics are biased toward ionized groups that are highly accessible to solvent. For desolvated groups, more accurate results were sometimes obtained with the full-charge/4 model; a good example is Glu35 of hen egg white lysozyme [26, 25]. The HIV-1 PR system provided an interesting and important test of the relative validity of the two titration models, because its active site is complex and the aspartyl dyad is highly desolvated when inhibitors are bound.

It is significant from a theoretical standpoint that, for the dyad, the full-charge/4 calculations apparently agree better with experiment than the single-site/20 calculations. For the ionizable groups on the surface, however, the high dielectric single-site/20 model is somewhat more accurate, in agreement with other previous study [26]. There are several likely reasons for this difference.

First, most of the ionizable groups considered in [26] are at the protein surface, reflecting the usual distribution of ionizable groups in proteins. Such groups are relatively mobile and their electrostatic interactions tend to be weak. For such groups, the use of an unrealistically [64, 81, 82] high solute dielectric constant of 20 probably helps correct for the use of a rigid model of the protein in the calculations [24]. Using a more realistic solute dielectric constant of 4 for such groups tends to overestimate pK_a shifts, because the model neglects side-chain motions. Also, because the interactions are weak, a simplified charge model is as good as a more detailed one. It is consistent with this view that the pK_as of ionizable groups other than the aspartyl dyad in HIV-1 PR are reproduced somewhat more accurately by the single-site/20 model than the full-charge/4 model (see Tables 1.6 and 1.7).

In contrast, the aspartyl dyad is inaccessible to bulk solvent, depending upon the state of inhibition. The dyad is therefore conformationally restrained, so modelling the protein as rigid is more appropriate for the dyad than it is for a surface group. In addition, the active site is highly desolvated when inhibitors are bound. As a consequence, the carboxyl groups are expected to interact strongly with each other and with other nearby groups. Using a solute dielectric constant of 20 is expected to lead to underestimates of the desolvation

penalty and of the charge-charge interactions. Finally, the use of realistic atomic charges is more important in this case where interactions are strong and desolvation is marked. Similar cases have been observed previously. A good example is Glu35 of hen egg white lysozyme, whose experimental pK_a is shifted up to about 6; this shift is reproduced by the full-charge/4 model but not by the single-site/20 model [26]. It is worth noting, that a method has been proposed for predicting, from the crystal structure, what protein dielectric constant will yield the most accurate pK_a for a given ionizable group [83].

It should also be noted, that the titration calculations presented herein are somewhat more sophisticated than those used previously. This is because the various possible protonations state of the aspartyl dyad are explicitly accounted for. This in effect provides a degree of "conformational relaxation" that is otherwise absent from the model.

1.4.3 Possible Improvements

Generally, modelling the protein as rigid causes the titration calculations to generate "microscopic" pK_as that are formally appropriate only for the selected conformation. For example, a pK_a of 27 for Asp25 with KNI-272 bound (Table 1.7) is obviously unrealistic because the protein is expected to denature at a much lower pH than 27. Accurate pK_as may nonetheless be obtainable in cases where the conformation is indeed insensitive to the charge states of the ionizable groups, or where one can predict a modest number of alternate conformations that will be important [66]. As noted above, accurate pK_as can also be obtained in some cases by using an artificially high protein dielectric constant to correct for conformational relaxation. However, this approach is not likely to be helpful when the conformation of the protein is, in fact, restricted, as for the aspartyl dyad when an inhibitor is bound.

These considerations suggest that pK_a calculations can be made more accurate by improved methods of accounting for the coupling between changes in charge and changes in conformation. Such methods are associated with significant pitfalls, as pointed out in [84], but fast approximations [85, 86] to the classical electrostatics models used here may facilitate this effort. One preliminary work in this direction indicates that at least one implementation of the generalized Born model [87] works quite well in calculations of the pK_as of small difunctional molecules when conformational flexibility is accounted for [88]. However, this approximation may need additional adjustment to work reliably for larger molecules, such as proteins [88].

Of course, accounting for flexibility in pK_a calculations will make the calculations more time-consuming. In many applications, though, only certain groups are of interest; for example, those in a binding site. The computational demands could thus be moderated by restricting the treatment of conformational flexibility to these key areas.

It is also worth mentioning, that a detailed treatment of conformational flexibility would account explicitly for the reorientation of protein dipoles. In this case, the only implicit contribution to the dielectric constant of the protein would be electronic polarization, as discussed in [64]. As a consequence, it would become appropriate to use a protein dielectric constant less than 4; in principle, a value of $1.5 - 2$ would be most realistic [64]. On the other hand, the empirical force fields that are typically used to simulate proteins are parameterized for use with a dielectric constant of 1. This value might well cause the strength of charge-charge interactions within a protein to be overestimated. Incorporating electronic polarizability explicitly in the force field might be the optimal solution to this problem.

Chapter II

Parameterization of the Approximate Valence Bond Method to Describe Potential Energy Surface in the Reaction Catalysed by HIV-1 Protease

2.1 Introduction

The aim of this Chapter was to describe the performed parameterization of the AVB method for the reaction catalysed by HIV-1 PR and to present the modifications of this method which have been applied for modelling this enzyme.

The molecular system for the modelled reaction was based initially on the crystallographic structure of HIV-1 PR complexed with the MVT-101 inhibitor which was modified into a substrate as previously described in Section 1.2.2. The modelled HIV-1 PR structure was divided into two regions. One, described with the AVB method, comprises molecular fragments involved directly in the enzymatic reaction, i.e. catalytic aspartates (25 and 125), the lytic water, and hydrolyzable peptide bond between both methionines (Met203 and Met204). The second (classical) region consists of the surrounding protein and solvent environment and will be described with a classical force field in further molecular dynamics simulations.

To parameterize the AVB method, one needs to perform QM calculations in vacuum for various molecular fragments involved in the reaction. The fragments chosen were: (a) the side-chains of catalytic aspartates with C_β carbon modified into the CH_3 group, (b) a catalytic water molecule, and (c) a peptide bond with C_α carbons modified into the CH_3 group. The QM calculations were performed using the density functional theory (DFT) implemented in the Gaussian'94 program [89].

2.2 Approximate Valence Bond Method

2.2.1 Basics of Valence Bond Methods

In the valence bond (VB) methods one constructs a matrix of the electronic Hamiltonian expressed in the basis of valence structures. The structures represent different arrangements of covalent or ionic bonds in a molecular system. The ground many-electron state

is approximated by a linear combination of electronic wave functions Ψ_I corresponding to various valence structures

$$\Phi = \sum_{I=1}^{N} c_I \Psi_I. \tag{2.1}$$

The ground-state energy, E, and the corresponding combination coefficients, c_I, are calculated from a set of equations resulting from the variational principle

$$\sum_{J=1}^{N} (H_{IJ} - ES_{IJ}) c_J = 0, \qquad I = 1, \ldots, N \tag{2.2}$$

where $H_{IJ} = \langle \Psi_I | \hat{H} | \Psi_J \rangle$ and $S_{IJ} = \langle \Psi_I | \Psi_J \rangle$ are the elements of the Hamiltonian and overlap matrices, respectively.

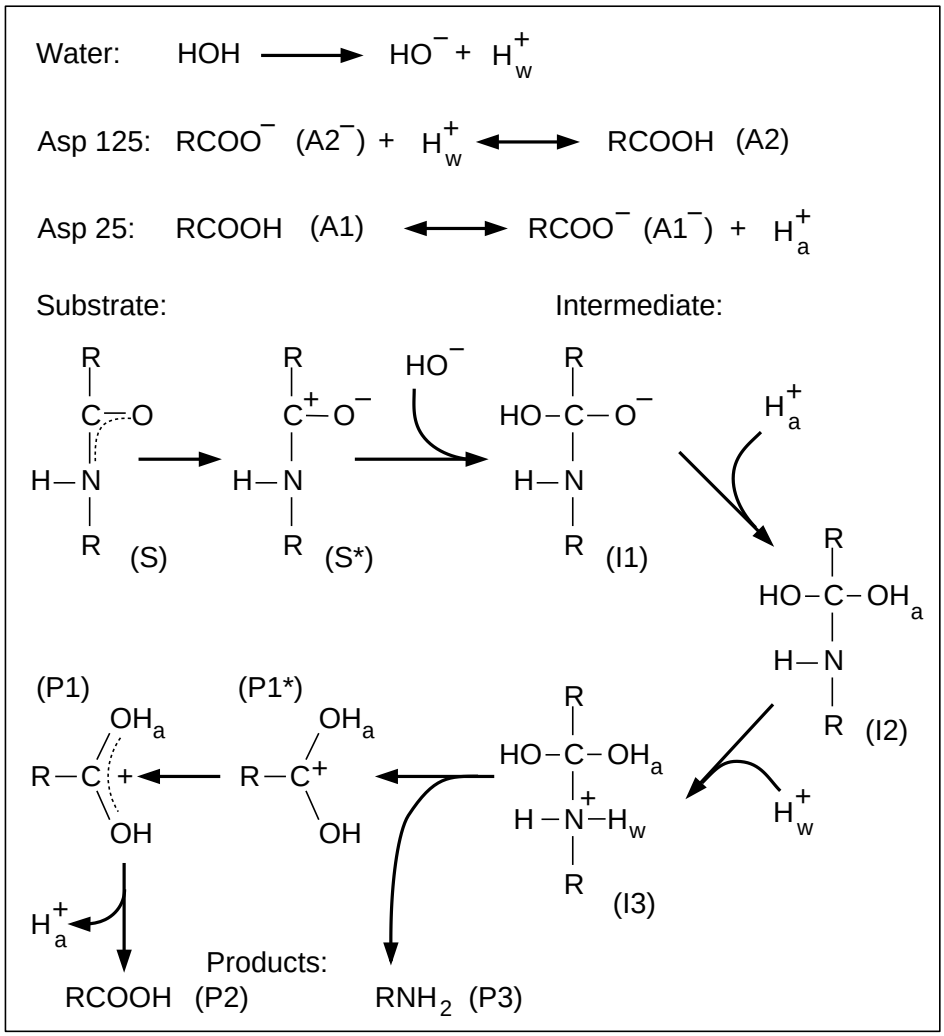

Figure 2.1: A minimal set of molecular structures in the active site allowing one to reproduce the mechanism of the enzymatic reaction. The key protons are labelled with letters, w (for water) and a (for acid). Arrows denote the direction of the enzymatic reaction. (R=CH_3)

2.2.2 The AVB Structures

In the AVB method the chemical reactions of interest are decomposed into elementary steps. A single step represents a dissociation or an association of one chemical bond. This means either a dissociation of a single bond leading to separation of ionic molecular fragments, or a partial dissociation of a double bond leading to polarized molecular structures.

Molecules and ionic molecular fragments, which appear at different steps of the reaction, represent a minimal set of molecular valence structures with which one constructs the valence structures of the whole molecular system. The molecules and their fragments appearing in one of the proposed models of the enzymatic reaction of HIV-1 PR are shown and labelled in Fig. 2.1.

Molecules such as OH^-, $A1^-$, $A2^-$, P3, and P2 are treated as indivisible in the considered chemical process. They are described by a single valence structure representing the ground state of the molecule. Other molecules are either dissociable ones or result from association of molecular fragments. Their ground state is described by a superposition of different valence structures. For example, the water molecule is described by one covalent structure (pure covalent HO–H_w bond) and one ionic structure (HO^- + H_w^+) which, combined together, represent the ground state of the molecule with one dissociable O–H_w bond. An analogous two-structure description applies to molecules A1, A2, S, I1 and I2, with corresponding ionic structures, $A1^-$ + H_a^+, $A2^-$ + H_w^+, S^*, S^* + OH^-, and I1 + H_a^+, respectively. In turn, the ground state of the I3 molecule is described by three structures: the covalent one (with pure covalent N^+–H_w and C–N bonds) and two ionic structures, I2 + H_w^+ and $P1^*$ + P3. A three-structure description applies also to the P1 molecule having one covalent and two ionic structures, $P1^*$ and P2 + H_a^+.

Valence structures for the whole molecular process are generated by setting together different combinations of molecular structures representing either indivisible molecular fragments in their ground states or covalent structures of dissociable molecules. For each structure another one, differing by dissociation or formation of one chemical bond, may be found. In case of the HIV-1 PR active site, the molecular fragments presented in Figure 2.1 there are twenty three valence structures (subscripts "g" and "c" denote molecular fragments in their ground and covalent state, respectively):

1. $A1_c + H_2O_c + A2_g^- + S_c$
2. $A1_c + OH_g^- + H_w^+ + A2_g^- + S_c$
3. $A1_c + OH_g^- + A2_c + S_c$
4. $A1_c + H_2O_c + A2_g^- + S_c^\star$
5. $A1_c + OH_g^- + H_w^+ + A2_g^- + S_c^\star$
6. $A1_c + OH_g^- + A2_c + S_c^\star$
7. $H_2O_c + A2_g^- + S_c + H_a^+ + A1_g^-$
8. $A2_g^- + H_w^+ + OH_g^- + S_c + H_a^+ + A1_g^-$
9. $OH_g^- + A2_c + S_c + H_a^+ + A1_g^-$
10. $H_2O_c + A2_g^- + S_c^\star + H_a^+ + A1_g^-$
11. $A2_g^- + H_w^+ + OH_g^- + S_c^\star + H_a^+ + A1_g^-$
12. $OH_g^- + A2_c + S_c^\star + H_a^+ + A1_g^-$
13. $A1_c + A2_g^- + H_w^+ + I1_c$

14. $A1_c + A2_c + I1_c$
15. $A2_g^- + H_w^+ + H_a^+ + A1_g^- + I1_c$
16. $A2_c + H_a^+ + A1_g^- + I1_c$
17. $A2_g^- + H_w^+ + A1_g^- + I2_c$
18. $A2_c + A1_g^- + I2_c$
19. $A2_g^- + A1_g^- + I3_c$
20. $A2_g^- + A1_g^- + P1_c^\star + P3_g$
21. $A2_g^- + A1_g^- + P1_c + P3_g$
22. $A1_c + A2_g^- + P3_g + P2_g$
23. $A2_g^- + H_a^+ + A1_g^- + P3_g + P2_g$

If an acid-base catalysis with a lytic water donating the hydroxy anion for the nucleophilic attack is assumed, there still are various possibilities of the formation of the reaction intermediate and its breakage. The first model (Figure 2.1) proposed that after the I1 molecule is formed, H_a^+ proton is transferred onto one of the oxygens what results in I2 molecule. This is followed by the H_w^+ proton transfer onto nitrogen where I3 is formed. However, after the formation of I1 some other paths may be possible what is shown in Figure 2.2. The I1 molecule may first accept H_w^+ proton on its nitrogen. The I2′ molecule is formed and it can either dissociate into products through a polarized molecule (P2*) or accept the H_a^+ proton on the peptide oxygen O^- and form the I3 molecular fragment. The I3 molecule dissociates into products in the same manner as presented in Figure 2.1. A three-structure description applies to I2′ molecule: the covalent one and two ionic structures I1 + H_w^+ and P2* + P3. P2 molecule has a corresponding ionic structure P2*.

Considering the path I1 → I2′ → I3 and then dissociation into products also twenty three structures describe the whole reaction (as shown above) but structures 17^{th} and 18^{th} are different:

17′. $A2_g^- + H_a^+ + A1_g^- + I2'_c$
18′. $A1_c + A2_g^- + I2'_c$

Considering the path I1 → I2′ → P2* + P3 total of twenty two structures model the system with the first sixteen similar to those presented above. The further structures are shown below:

17′. $A2_g^- + H_a^+ + A1_g^- + I2'_c$
18′. $A1_c + A2_g^- + I2'_c$
19. $A1_c + A2_g^- + P2_c^\star + P3_g$
20. $A2_g^- + H_a^+ + A1_g^- + P2_c^\star + P3_g$
21. $A1_c + A2_g^- + P2_g + P3_g$
22. $A2_g^- + H_a^+ + A1_g^- + P2_g + P3_g$

Therefore, depending on the choice of the set of structures one may model various mechanisms of the reaction. Of course, all the models may be accounted for and the whole set of molecular fragments in Figures 2.1 and 2.2 may be used to construct the set of AVB structures.

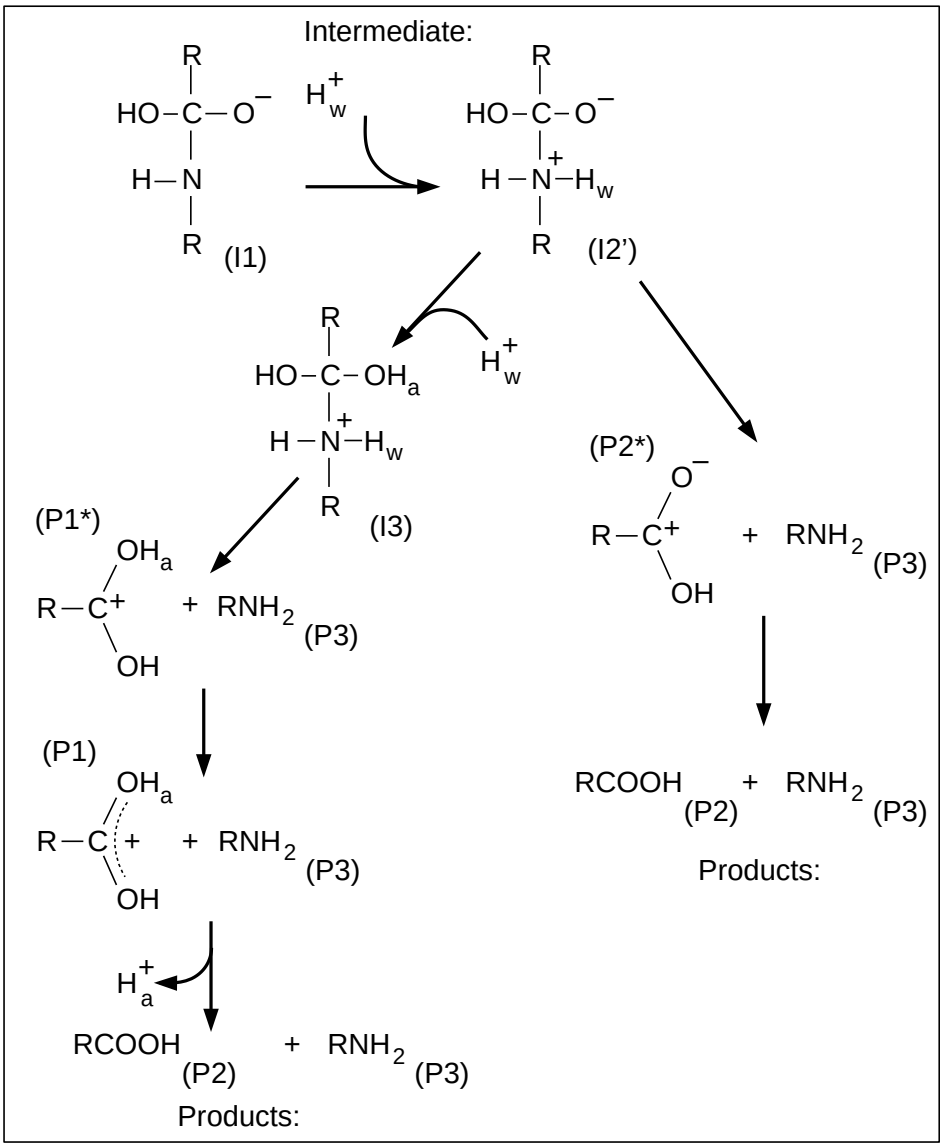

Figure 2.2: A set of molecular fragments used to reproduce other possibilities of the reaction path following the formation of the I1 molecule. The key protons are labelled with letters, w (for water) and a (for acid). Arrows denote the direction of the reaction. (R=CH_3)

2.2.3 AVB Description of an Isolated Molecule

The evaluation of the AVB Hamiltonian matrix applied in this study differs somewhat from the original method presented in [3], and was developed in order to simplify and systematize the parameterization procedure.

Let us denote E_g^A as the Born-Oppenheimer energy surface of an isolated molecule A in the ground electronic state. In the AVB method E_g^A is approximated by a sum of bonding and nonbonding type of elementary potentials, similar to those applied in the MM. The bonding potentials depend on bond lengths (r), planar angles (α), torsional angles (θ), and

improper dihedral angles (δ) according to the following formulae:

$$V_{\rm bond} = \begin{cases} D_0 \left[1 - e^{-a(r-r_0)}\right]^2 \\ \frac{1}{2}B(r-r_0)^2 \end{cases}, \tag{2.3}$$

$$V^0_{\rm ang} = \frac{1}{2}A_0(\cos\alpha - \cos\alpha_0)^2, \tag{2.4}$$

$$V^0_{\rm tors} = \sum_{k=1}^{3}[A_k + |A_k|\cos(k\psi)], \tag{2.5}$$

$$V^0_{\rm impr} = \frac{1}{2}A(\delta - \delta_0)^2. \tag{2.6}$$

In case of $V_{\rm bond}$ the Morse function is used for the dissociable bonds and the harmonic function for all the others.

For the dissociating molecules the above (planar, torsional and improper dihedral) angular deformation potentials, that have been denoted with index 0, are applied only for the dissociable bond lengths of $r \leq r_0$, i.e. close to equilibrium or shorter. If we allow molecules to dissociate and increase the bond lengths (r) unlimitedly, the angular deformation potentials are modified in a manner presented further in this Section.

The nonbonding potentials depending on interatomic distances (r_{ab}) are applied for the pairs of atoms (a and b) linked by three or more chemical bonds. The electrostatic and Lennard-Jones terms are used:

$$V^{\rm ES}_{\rm (a,b)} = k\frac{q_{ga}q_{gb}}{r_{ab}}, \tag{2.7}$$

$$V^{\rm LJ}_{\rm (a,b)} = \frac{A}{r^{12}_{ab}} - \frac{B}{r^{6}_{ab}}. \tag{2.8}$$

where q_{ga} and q_{gb} denote charges of atoms a and b of an isolated molecule in its ground state. These charges are determined for the optimized geometry of the molecule by fitting to the molecular electrostatic potential resulting from QM calculations (ESP charges) [89]. The parameters of the Lennard-Jones potential, as well as of bonding potentials, are fitted to reproduce the QM energy of the molecule in various deformed geometries, as described in Sections 2.3 and 2.4. Some less important parameters, i.e. for atoms that do not participate directly in the reaction, may also be taken from the standard MM libraries. In the AVB method it is important to ensure that E^A_g has a minimum equal to zero for the optimal geometry. For this reason an additional constant energy term in E^A_g is applied that compensates possible tensions and nonbonding intramolecular energy in the optimal geometry.

If a molecule A is a dissociable one, then the energies of its covalent and ionic forms, as well as the corresponding non-diagonal elements of the Hamiltonian matrix, are to be determined. For this purpose a general algorithm was developed to describe a molecule with many ionic dissociation channels in the form of $A \leftrightarrow B_i + C_i$ or $A \leftrightarrow A^*_i$, where $i = 1, \ldots, n$ numbers of channels; B and C denote molecular ionic fragments resulting from ionic dissociation of a chemical bond; A^*_i denotes a polarised molecule resulting from reduction of a double bond into a single one. The ground state of molecule A is represented by a combination of a single covalent structure and a set of ionic structures that correspond to products of various ionic dissociations. Valence structures are assumed orthogonal, i.e. $S_{ij} = \delta_{ij}$. The non-diagonal elements of the Hamiltonian matrix corresponding to pairs of

different ionic structures are neglected. So the AVB eigensystem for the isolated molecule A is

$$\begin{pmatrix} H^A_{cc} - E^A_g & H^A_{c1} & \cdots & H^A_{cn} \\ H^A_{c1} & H^A_{11} - E^A_g & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H^A_{cn} & 0 & \cdots & H^A_{nn} - E^A_g \end{pmatrix} \begin{pmatrix} c_c \\ c_1 \\ \vdots \\ c_n \end{pmatrix} = 0. \tag{2.9}$$

This system can be used to determine the Hamiltonian matrix element H^A_{cc} corresponding to the energy of the molecule A in its covalent form, and the non-diagonal elements $H^A_{ci}, i = 1, \ldots, n$, corresponding to the coupling between covalent and ionic structures. First, the diagonal elements corresponding to ionic structures are calculated as:

$$H^A_{ii} = E^{\text{diss}}_i + E^{B_i}_g + E^{C_i}_g + E^{B_i C_i}. \tag{2.10}$$

Here E^{diss}_i is the difference between the energy of molecule A and the energy of separated ionic fragments B_i and C_i in their ground states and in their optimal geometries calculated with QM methods; E^B_g and E^C_g denote the ground state energy surfaces of separated ionic fragments, calculated as a sum of elementary potentials (Eqs 2.3-2.8) and parameterized with respect to QM calculations of separate ionic fragments as in case of molecule A. The last term in (2.10) denotes the nonbonding interaction between each pair of atoms of the two ionic fragments. This term is represented as a sum of electrostatic and Lennard-Jones potentials. The electrostatic interaction between atoms a and b is

$$V^{\text{ES}}_{(a,b)} = k \frac{q_{ia} q_{ib}}{r_{ab}} \tag{2.11}$$

where q_{ia} and q_{ib} denote charges of atoms a and b in i-th ionic structure and are equivalent to the ESP charges of separated ionic fragments. For atoms that form an ionic bond between fragments B and C, resulting from a dissociation of a corresponding covalent bond in molecule A, a more flexible potential function instead of the Lennard-Jones one was found and is applied

$$V^{\text{IB}}_{\text{a,b}} = \left(\frac{A}{r_{ab}} - B \right) \exp(-\alpha r_{ab}). \tag{2.12}$$

The parameters of this potential are determined by fitting to the QM energy surface for the bimolecular complexes, as described in the next Section.

When a double bond "dissociates" into a single one, E^{diss}_i is interpreted as the energy required to diminish the bond order, and the following equation replaces 2.10:

$$H^A_{ii} = E^{\text{diss}}_i + E^{A^*} \tag{2.13}$$

where E^{A^*} is the energy surface of a polarized molecule A calculated based on Eqs 2.3–2.8 with parameters and charges on atoms modified with respect to the ground state of A, to model the decrease of the bond order due to polarization. E^{diss}_i is subject to fit in the parameterization, as described in Section 2.4 concerning the parameterization of nonbonding interactions.

Having described ionic structures of a dissociating molecule one may now use the $E^{B_i}_g$ and $E^{C_i}_g$ energies of Eq. 2.10 to improve the angular deformation potentials presented in Eqs 2.4, 2.5, 2.6 and constituting E^A_g. The angular deformation potential of the dissociating molecule A in the ground state E^A_g now takes the form

$$V_{\text{def}} = w V^0_{\text{def}} + (1 - w) V^{\text{ion}}_{\text{def}} \tag{2.14}$$

where $V^0_{\rm def}$ is a sum of angular deformation potentials shown in Eqs 2.4, 2.5, 2.6:

$$V^0_{\rm def} = V^0_{\rm ang} + V^0_{\rm tors} + V^0_{\rm impr}. \tag{2.15}$$

The $V^{\rm ion}_{\rm def}$ element is the angular deformation potential in molecules B_i and C_i. The w weighing function is

$$w = \begin{cases} 2e^{-a_w(r-r_0)} - e^{-2a_w(r-r_0)} & r > r_0 \\ 1 & r < r_0 \end{cases} \tag{2.16}$$

where r and r_0 are, respectively, the actual and equilibrium length of the dissociating bond in the ith structure, and a_w is the parameter. The form of $V_{\rm def}$ may be generalized for more than one dissociation in molecule.

The second set of quantities calculated in the system (2.9) are the combination coefficients c_c and c_i. According to the properties of the AVB method derived in [3], the actual combination coefficients are related to the effective atomic charges of the molecule

$$q^{\rm eff}_a = c_c^2 q_{ca} + \sum_i c_i^2 q_{ia} \tag{2.17}$$

where q_{ca} and q_{ia} denote charges of atoms in the covalent and ionic structures, respectively. The coefficients c should be therefore parameterized so that the AVB effective atomic charges reproduce the ESP charges derived from the QM calculations. According to QM ground-state calculations, all considered dissociation processes of hydrogens in isolated molecules lead to radical products. Generally, this is not true for dissociation of the C–O or the C–N bond. However, one may assume that the products of radical dissociation of such bonds have, in vacuum, lower energy than the products of ionic dissociation where one of the products is described with a polarized structure. In the AVB model the radical dissociation should correspond to vanishing coefficients of ionic structures. This is achieved by describing c_i of an isolated molecule with an exponentially decreasing function of the dissociable bond lengths. On the other hand, one has to ensure that $\sum_i c_i^2 \leq 1$. Therefore, the following formulae are applied:

$$c_i = \frac{S_i f_i}{\sqrt{1 + \sum_i S_i^2 (f_i^2 - 1)}}, \tag{2.18}$$

$$c_c = \sqrt{1 - \sum_i c_i^2}, \tag{2.19}$$

$$f_i = \exp[-a_f(r_i - r_{i0})] \tag{2.20}$$

where r_i are the actual bond lengths, r_{i0} are the bond lengths in the optimal geometry, and a_f and S_i are the parameters. For the molecule with optimal geometry, $r_i = r_{i0}$, the equation (2.17) takes the form

$$q_{ga} = \left(1 - \sum_i S_i^2\right) q_{ca} + \sum_i S_i^2 q_{ia} \tag{2.21}$$

and is used to determine S_i from the atomic charges. As the number of equations in Eq. 2.21 is equal to the number of atoms in molecule A, and is usually greater than the number of considered dissociations in molecule A, the following procedure for solving (2.21) is applied. The charges q_{ca} are first assigned to atoms of a dissociating bond, ensuring

that the covalent structure can be divided into radical fragments with proper integer net charges. The parameters S_i are then derived from a subset of equations from Eq.2.21. Finally, other covalent charges q_{ca} are determined from the rest of the equations using the known values of S_i.

From the Hamiltonian matrix for an isolated molecule (Eq. 2.9) H^A_{cc} and H^A_{ci}, $i = 1, \ldots, n$ may be now calculated as

$$H^A_{ci} = \frac{c_i}{c_c}(E^A_g - H^A_{ii}), \tag{2.22}$$

$$H^A_{cc} = E^A_g - \sum_i \frac{c_i^2}{c_c^2}(E^A_g - H^A_{ii}). \tag{2.23}$$

2.2.4 Description of a Molecular System

A diagonal element of the Hamiltonian matrix corresponding to the I^{th} valence structure of the whole molecular system is expressed as

$$H_{II} = \sum_{A(I)} E^A + \sum_{A<B(I)} E^{AB} \tag{2.24}$$

where A and B are the indices numbering molecules or molecular fragments appearing in the I^{th} valence structure of the system. E^A represents the energy of an isolated molecule in its ground or covalent state. The first case, $E^A = E^A_g$, applies if A is an indivisible molecule. The second case, $E^A = H^A_{cc}$ applies if A is a dissociable molecule. E^{AB} represents intermolecular interactions, including electrostatics and nonbonding interatomic potentials. The latter are the Lennard-Jones type of potentials, V^{IB} (2.12) potentials, or a modified potential form applied for atoms forming a hydrogen bond

$$V^{\mathrm{HB}}_{\mathrm{a,b}} = A\exp(-\alpha r) - \frac{B}{r^6}. \tag{2.25}$$

The full form of this potential is used for the heavy atoms of the hydrogen bond. For the interactions of a heavy atom and a hydrogen $B = 0$ is set. The parameters of nonbonding potentials between the molecules are determined by fitting to the QM energy surface of bimolecular complexes.

The non-diagonal elements of the AVB Hamiltonian are assumed nonzero only for those pairs of structures that differ by dissociation of one chemical bond. It occurs if an I^{th} structure involves molecule A in the covalent form and a J^{th} structure involves product(s) of i^{th} ionic dissociation of molecule A. In such a case we approximate the element H_{IJ} with the term H^A_{ci}.

The coupling between the AVB and the classical regions is also accounted for in the AVB method by adding to the diagonal elements of the full Hamiltonian matrix a potential term

$$\Delta H_{II} = k\sum_{a,b} q_{a,I}q_b/r_{ab} \tag{2.26}$$

where a, b designate the atoms in the AVB and classical regions, respectively, $q_{a,I}$ are the atomic charges assigned to the I^{th} valence structure, and r_{ab} are the interatomic distances. The coupling of the non-diagonal terms of the Hamiltonian with the classical region is neglected.

Given the AVB Hamiltonian of the molecular system and assuming $S_{IJ} = \delta_{IJ}$, the eigensystem (2.2) is solved for the ground-state energy E and the corresponding normalized combination coefficients c_I. The forces acting on the atoms of the AVB region are determined from the equation

$$\mathbf{F}_a = -\frac{\partial E}{\partial \mathbf{r}_a} = -\sum_{IJ} \frac{\partial H_{IJ}}{\partial \mathbf{r}_a} c_I c_J \tag{2.27}$$

where the gradient of the Hamiltonian is calculated analytically. The Coulomb forces resulting from charges in the AVB region and acting on atoms in the classical region are also calculated with an analytical formula [3]

$$\mathbf{F}_b = -k \sum_a q_a^{\text{eff}} q_b \mathbf{r}_{ab} / r_{ab}^3, \quad q_a^{\text{eff}} = \sum_{I=1}^{N} c_I^2 q_{a,I}. \tag{2.28}$$

2.3 Parameters Characterizing Intramolecular Interactions in the AVB Region

This Section presents a set of calculated AVB parameters used to describe separate molecules and their ionic fragments (Fig. 2.1). To determine intramolecular interaction parameters, the DFT calculations using the Gaussian'94 program [89], applying the nonlocal hybrid density functional B3LYP [90, 91] and 6-31G(d,p) basis set, were performed.

This method was chosen as being comparable to the second-order Møller-Plesset (MP2) and other *ab initio* methods and giving reasonable results for the systems of interest at a much lower computational cost. For example B3LYP reproduces in a satisfactory way the reaction paths and barriers of the proton transfers (e.g. in carboxylic acid dimers [92]), optimized geometries (e.g. of N-methylacetamide [93]), hydrogen bond distances and energies (e.g. [94]), and proton affinities (e.g. [95]).

For each molecule its optimal geometry, the ESP atomic charges, the energy surface for deformations from optimal geometry, and the asymptotic energies of ionic dissociations were determined. The dissociation energies, which are most sensitive to atomic bases, were calculated with the 6-31G+(d,p) basis set.

The optimal geometries of molecules are presented in Tables 2.1 and 2.2. Each atom of a molecule was assigned a valence type, the ground-state charge q_g (corresponding to the ESP charge), and the covalent state charge q_c (see Tables 2.3–2.8). For non-polar hydrogens the united atom model was used where the ESP charge of a carbon is presented as a sum of charges of a carbon and its hydrogens. The ionic dissociation energies, E_i^{diss}, and the parameters S_i are presented in Table 2.9. The parameter a_f appearing in (2.20) was determined based on the changes of ESP charges in a dissociating water molecule ($a_f = 0.695$ Å^{-1}) and this value was used elsewhere for dissociable O–H and N–H bonds. For dissociable C–O and C–N bonds a roughly optimized value of $a_f = 1$ Å^{-1} was used. In case of partial dissociation of a double bond the dependence of charges on the bond length in an isolated molecule is neglected($a_f = 0$). For the parameter in equation 2.16 a roughly optimized value $a_w = 3a_f$ is used.

Tables 2.10–2.12 show the parameters for the bonding interactions (Eqs 2.3–2.6) fitted to the DFT energy surface for molecular deformations. For all crucial bonds that are changed during the reaction the Morse potential was applied (Table 2.10). Fit of the Morse potential

Table 2.1: Optimized distances [Å], planar angles and torsions for the RCONHR (S), $RCO^{-}OHNHR$ (I1), $RC(OH)_2NHR$ (I2), $RC(OH)_2NH_2^+R$ (I3) and $RC(OH)_2^+R$ + RNH_2 (P1 + P3) molecules, (R=CH_3).

	S	I1	I2	I3	P1 + P3
bond lengths					
C-O	1.225	1.274	1.411	1.385	1.288
C-CA2	1.523	1.564	1.526	1.517	1.475
C-N	1.366	1.573	1.463	1.581	-
N-HN	1.008	1.023	1.017	1.026	1.017
N-CA3	1.453	1.463	1.464	1.500	1.464
HD2-O	-	-	0.969	0.969	0.977
HW1-OW	-	0.968	0.970	0.971	0.977
C-OW	-	1.516	1.410	1.371	1.288
N-HW2	-	-	-	1.025	1.017
bond angles					
CA2-C-O	123.1	115.0	106.5	115.7	117.9
CA2-C-N	115.2	104.6	117.1	109.8	-
O-C-N	121.8	116.6	106.1	99.4	-
C-N-HN	119.7	102.8	107.9	107.7	-
C-N-CA3	121.1	108.6	116.9	116.4	-
CA3-N-HN	119.3	105.7	110.4	111.2	109.7
CA2-C-OW	-	105.1	106.6	110.3	117.9
OW-C-N	-	100.8	108.0	107.2	-
OW-C-O	-	113.1	112.7	113.6	124.2
HD2-O-C	-	102.7	106.0	110.8	117.7
HW1-OW-C	-	-	106.2	109.3	117.7
C-N-HW2	-	-	-	105.2	-
CA3-N-HW2	-	-	-	110.2	109.7
HN-N-HW2	-	-	-	105.6	105.9
dihedral angles					
CA2-C-N-CA3	-180.0	-172.7	-58.3	-67.6	-
O-C-N-HN	180.0	67.4	-51.9	-63.8	-
O-C-OW-HW1	-	161.9	-71.7	-17.9	-0.6
OW-C-N-HN	-	-169.9	-173.0	177.8	-
OW-C-O-HD2	-	-	74.7	-90.9	0.6
HW2-N-C-CA2	-	-	-	170.2	-

Table 2.2: Optimized distances (l), planar angles (ϕ) and torsions (Θ), derived from the DFT calculations for the RCOOH (A1, A2) and $RCOO^-$ ($A1^-$, $A2^-$) molecules, (R=CH_3).

	l(A-B) [Å]		ϕ(A-B-C)[°]		Θ(A-B-C-D)[°]
ABCD sequence	RCOOH	$RCOO^-$	RCOOH	$RCOO^-$	RCOOH
OD2HD2	0.972	-	-	-	-
CGOD2HD2	1.358	1.257	105.9	-	-
CBCGOD1	-	-	126.1	115.0	-
CBCGOD2HD2	1.507	1.575	111.4	115.0	180.0
OD1CGOD2HD2	1.210	1.257	122.5	129.8	0.0

Table 2.3: Valence types, ESP charges of the ground state, q_g, and atomic covalent charges, q_c, of the H_2O and OH^- molecule.

		H_2O		OH^- + H^+	
Atom	Type	q_g	q_c	Type	q_g
O	ow	-0.730	-0.494	ow^-	-1.141
HW1	hw	0.365	0.494	hw	0.141
HW2	hw	0.365	0.000	h+	1.000

was performed only for such a range of bond lengths where closed-shell calculations are valid (see Figure 2.3).

In this and all other figures of this Section the solid line represents the molecular ground state AVB energy (E_g) while circles and triangles represent the DFT energy. For other non-dissociable bonds (not listed in Table 2.10) the harmonic potential was used and the parameters were taken from the Gromos'96 force field [7]. Potentials for planar, dihedral, and improper angles adjacent to dissociable bonds were parameterized based on the DFT calculations (Tables 2.11–2.12). For all other angles the Gromos'96 force field [7] parameters were used as for non-dissociable bonds. Figures 2.4–2.6 show fits for the CH3-C-N-CH3 dihedral angle in the $CH_3CO^-OHNHCH_3$ (I1), $CH_3C(OH)_2NHCH_3$ (I2), and $CH_3C(OH)_2NH_2^+CH_3$ (I3) molecules, respectively.

A conformational transition of the C–N bond can be achieved in the model by CH3-C-N-CH3 rotation and/or inversion of the nitrogen atom. Possible rotations of the hydroxyl groups in the transition state are also accounted for. Figure 2.7 shows the fit for the O-C-OW-HW1 angle in the $CH_3C(OH)_2NHCH_3$ (I2) molecule.

Table 2.4: Valence types, ESP charges of the ground state, q_g, and atomic covalent charges, q_c, of the RCOOH (A1, A2) and $RCOO^-$ ($A1^-$, $A2^-$) molecules, (R=CH_3).

	A1, A2			$A1^-$, $A2^-$	
Atom	q_g	q_c	Type	q_g	Type
CB	0.000	0.154	ch3	-0.214	ch3
CG	0.698	0.664	c	0.748	c
OD1	-0.529	-0.371	ok	-0.760	ok^-
OD2	-0.574	-0.447	ok	-0.761	ok^-
HD2	0.405	0.000	h	–	–

Table 2.5: Valence types, ESP charges in the ground state, q_g, and atomic covalent charges, q_c, of the RCONHR (S) and RC^+O^-NHR (S*) molecules, (R=CH_3).

	S			S^*	
Atom	q_g	q_c	Type	q_g	Type
CA2	-0.077	-0.077	ch3	-0.077	ch3
C	0.641	0.116	cn+	1.116	cn
O	-0.525	0.000	o	-1.000	o
N	-0.505	-0.505	n	-0.505	n
HN	0.303	0.303	hn	0.303	hn
CA3	0.163	0.163	ch3	0.163	ch3

Table 2.6: Valence types, ESP charges in the ground state, q_g, and atomic covalent charges, q_c, of the RCO^-OHNHR (I1), $RC(OH)_2NHR$ (I2) and $RCO^-OHNH_2^+R$ (I2′) molecules, (R=CH_3).

	I1			I2			I2′	
Atom	Type	q_g	q_c	Type	q_g	q_c	Type	q_g	q_c
CA2	ch3	-0.275	-0.396	ch3	0.007	0.234	ch3	-0.068	-0.128
C	cn	0.938	0.829	cn	0.707	0.638	c	0.662	0.208
O	om	-0.847	-0.754	oa	-0.615	-0.537	om	-0.806	-1.090
HD2	–	–	–	ha	0.360	0.000	–	–	–
OW	oa	-0.673	-0.387	oa	-0.609	-0.734	oa	-0.628	-0.370
HW1	ha	0.294	0.387	ha	0.394	0.398	ha	0.389	0.380
N	n	-0.785	-0.956	n	-0.810	-0.728	n	-0.120	0.957
HN	hn	0.266	0.243	hn	0.356	0.419	hn	0.219	0.064
HW2	–	–	–	–	–	–	hn	0.229	0.000
CA3	ch3	0.082	0.033	ch3	0.210	0.310	ch3	0.123	-0.020

Table 2.7: Valence types, ESP charges, q_g and q_c, of the $RC(OH)_2N^+H_2R$ (I3), $RC^+(OH)_2$ (P1*) and RNH_2 (P3) molecules, (R=CH_3).

	I3		P1* and P3	
Atom	Type	q_g	q_c	Type	q_g
CA2	ch3	0.071	-0.221	ch3	0.191
C	cn	0.517	-0.108	c	1.000
O	oa	-0.624	-0.710	ok	-0.568
HD2	ha	0.473	1.040	h	0.472
OW	oa	-0.534	-0.395	ok	-0.568
HW1	ha	0.442	0.393	h	0.472
N	n	-0.250	0.469	na	-0.920
HN	hn	0.335	0.295	hn	0.339
HW2	hn	0.317	0.000	hn	0.339
CA3	ch3	0.253	0.236	ch3	0.242

Table 2.8: Valence types, ESP charges in the ground state, q_g, and atomic covalent charges, q_c, of the $RC(OH)_2^+$ (P1) and RCOOH (P2) molecules, (R=CH_3).

	P1			P2	
Atom	Type	q_g	q_c	Type	q_g
CA2	ch3	0.191	0.362	ch3	0.000
C	c	0.727	0.000	c	0.698
O	ok	-0.431	0.073	ok	-0.574
HD2	h	0.472	0.000	–	–
OW	ok	-0.431	0.033	ok	-0.529
HW1	h	0.472	0.532	h	0.405

Table 2.9: Ionic dissociation energies (E_i^{diss}, Eqs 2.10 and 2.13), and parameters of the ionic structure coefficients (Eqs 2.18 and 2.20).

Dissociation Type	E_i^{diss} [kJ/mol]	S_i^2	r_0 [Å]
$H_2O \rightarrow H^+ + OH^-$	1656	0.365	0.965
RCOOH (A1,A2) $\rightarrow RCOO^- + H^+$	1477	0.405	0.972
RCONHR (S) $\rightarrow RC^+O^-NHR$ (S*)	581	0.525	–
I1 $\rightarrow RC^+O^-NHR$ (S*) + OH^-	626	0.379	1.505
I2 $\rightarrow$ I1 + H^+	1537	0.360	0.969
I2'$\rightarrow$ I1 + H^+	1430	0.050	1.030
I3 $\rightarrow$ I2 + H^+	961	0.200	1.030
I3 $\rightarrow RNH_2$ (P3) + $RC^+(OH)_2$ (P1*)	224	0.345	1.581
$RC(OH)_2^+$ (P1) $\rightarrow$ RCOOH (P2) + H^+	796	0.472	0.977
$RC(OH)_2^+$ (P1) $\rightarrow RC^+(OH)_2$ (P1*)	102	0.313	–
$RC^+(OH)_2$ (P1*) $\rightarrow$ RCOOH (P2) + H^+	694	0.472	0.977
I2'$\rightarrow RNH_2$ (P3) + RC^+O^-OH (P2*)	720	0.527	1.581
RCOOH (P2) $\rightarrow RC^+O^-OH$ (P2*)	581	0.698	–

Table 2.10: Fitted Parameters of the Morse potentials, (R=CH_3), (Eq. 2.3).

	Bond	D_0 [kJ/mol]	a[$Å^{-1}$]	r_0[Å]
Water	OW-HW1,HW2	561.5	2.157	0.965
RCOOH	OD2-HD2	488.7	2.232	0.978
RCOHNHR	O-HD2	499.9	2.213	0.977
RCO^-OHNHR (I1)	OW-C	123.1	2.122	1.601
	O-C	529.2	2.087	1.313
	C-N	92.3	2.180	1.666
	OW-HW1	453.	2.280	0.979
$RC(OH)_2NHR$ (I2)	O-C,OW-C	293.0	2.040	1.460
	O-HD2,OW-HW1	453.0	2.280	0.979
	C-N	296.4	1.937	1.512
$RC(OH)_2NH_2^+R$ (I3)	C-OW	407.5	2.000	1.401
	C-O	407.5	2.000	1.413
	OW-HW1,O-HD2	531.6	2.133	0.971
	C-N	129.9	2.054	1.646
	N-HN,HW2	487.6	2.007	1.035
$RC(OH)_2^+$ (P1)	O-HD2,OW-HW1	459.2	2.265	0.980

Table 2.11: Fitted parameters of the torsional angle potentials (R=CH_3), (Eq. 2.5).

	Angle	A_1 [kJ/mol]	A_2 [kJ/mol]	A_3 [kJ/mol]
RCO^-OHNHR (I1)	O-C-N-CA3	8.8	17.8	25.2
	OW-C-N-CA3	14.3	8.5	-27.1
	HD2-O-C-OW	1.0	-10.9	7.7
	HD2-O-C-N	8.5	-1.7	-5.7
$RC(OH)_2NHR$ (I2)	HD2-O-C-OW	-0.1	6.7	6.4
	HD2-O-C-N	3.0	2.7	-4.6
	HW1-OW-C-O	-0.1	6.7	6.4
	HW1-OW-C-N	3.0	2.7	-4.6
	O-C-N-CA3	16.3	-4.6	-32.8
	OW-C-N-CA3	11.7	2.1	36.4
$RC(OH)_2NH_2^+R$ (I3)	O-C-N-CA3	1.0	1.5	4.1
	OW-C-N-CA3	2.6	3.3	0.2
	HD2-O-C-OW	-0.1	6.7	6.4
	HD2-O-C-N	3.0	2.7	-4.6
	HW1-OW-C-O	-0.1	6.7	6.4
	HW1-OW-C-N	3.0	2.7	-4.6

Table 2.12: Fitted parameters of the planar angle potentials (R=CH_3), (Eq. 2.4).

Angle	A_0 [kJ/mol]	α_0 [°]		
		RCO^-OHNHR (I1)	$RC(OH)_2NHR$ (I2)	$RC(OH)_2NH_2^+R$ (I3)
CA2-C-N	1140.	104.	109.	104.
CA2-C-OW	1140.	103.	100.	104.
CA2-C-O	830.	107.	-	-
CA2-C-O	1140.	-	100.	109.
N-C-OW	1100.	108.	108.	103.
N-C-O	1260.	116.	-	-
N-C-O	1100.	-	108.	100.
O-C-OW	1700.	112.	111.	112.

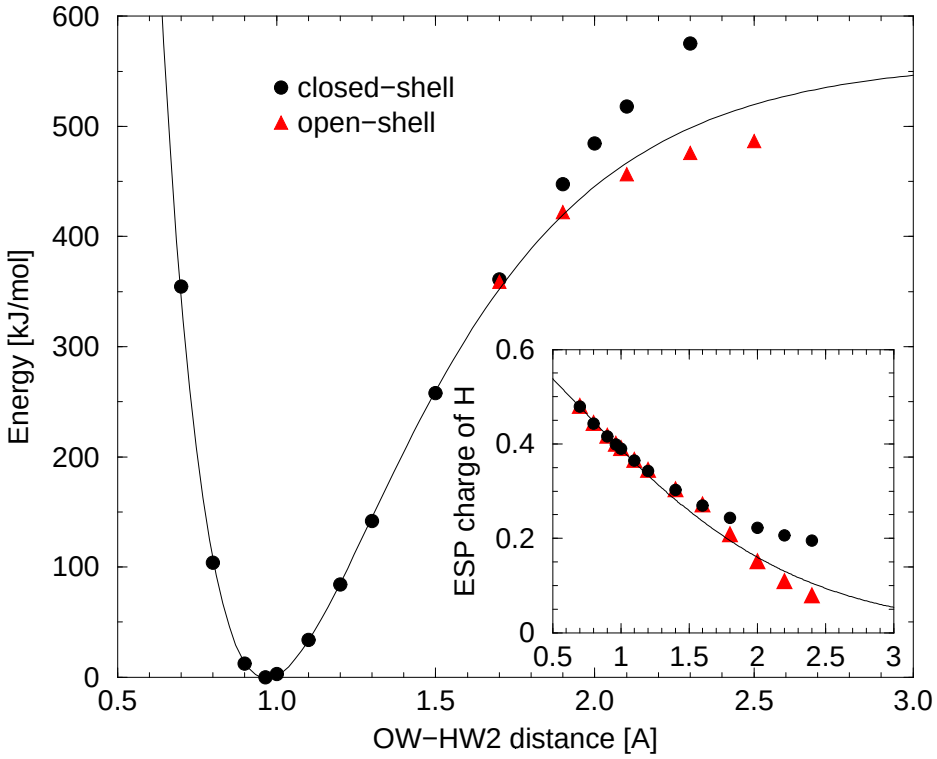

Figure 2.3: Fitted energy profile for the stretching of the OW–HW2 bond of a water molecule. Circles and triangles present the DFT results for restricted closed-shell and open-shell calculations, respectively. Small plot shows how AVB charges reproduce the ESP charges. Fitted parameters are presented in Table 2.10.

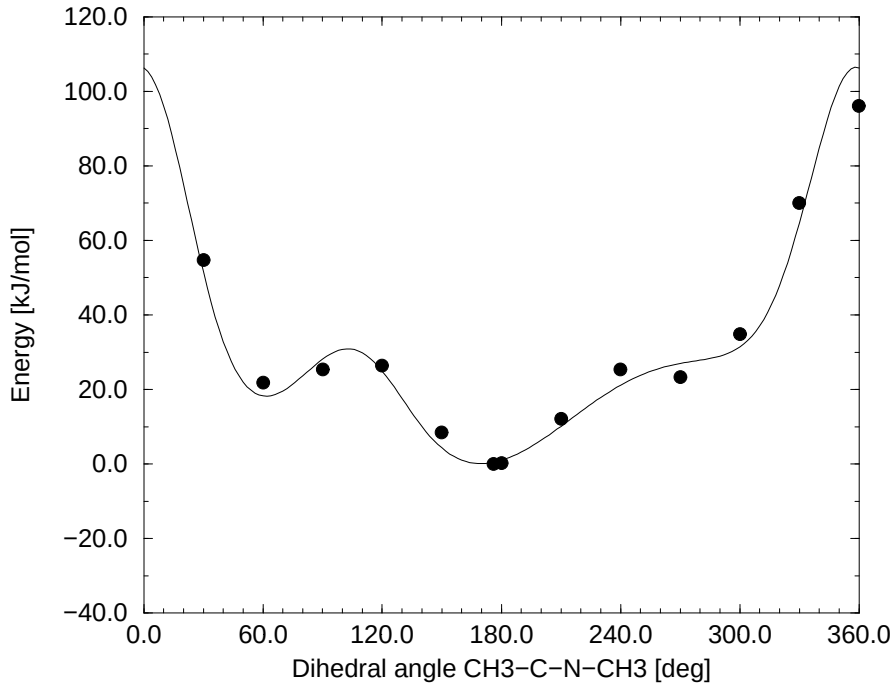

Figure 2.4: Fitted energy profile for the CH3-C-N-CH3 torsional angle in the $CH_3CO^-OHNHCH_3$ (I1) molecule. Fitted parameters are presented in Table 2.11.

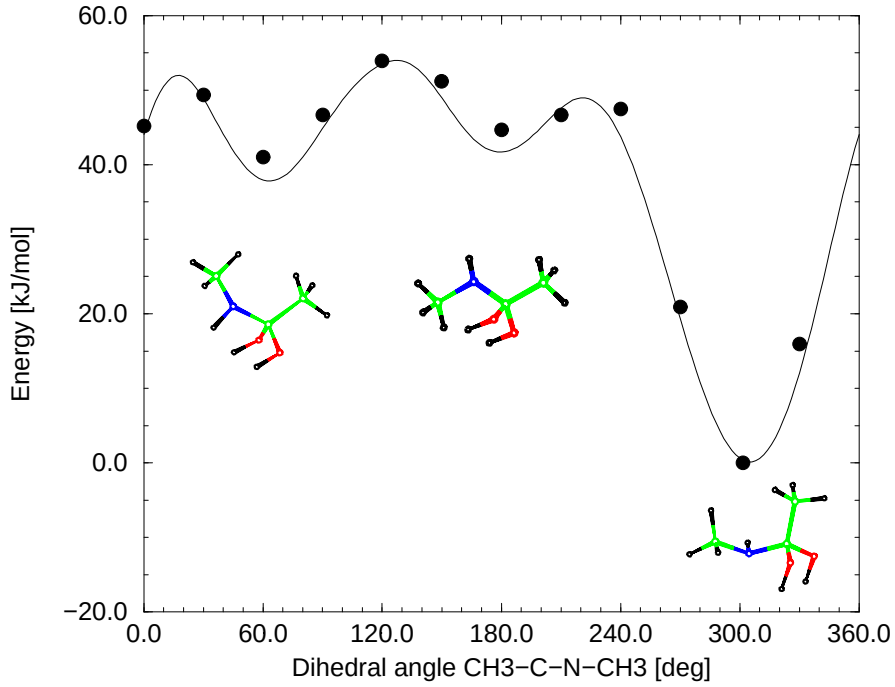

Figure 2.5: Fitted energy profile for the CH3-C-N-CH3 torsional angle in the $CH_3C(OH)_2NHCH_3$ (I2) molecule. Fitted parameters are presented in Table 2.11.

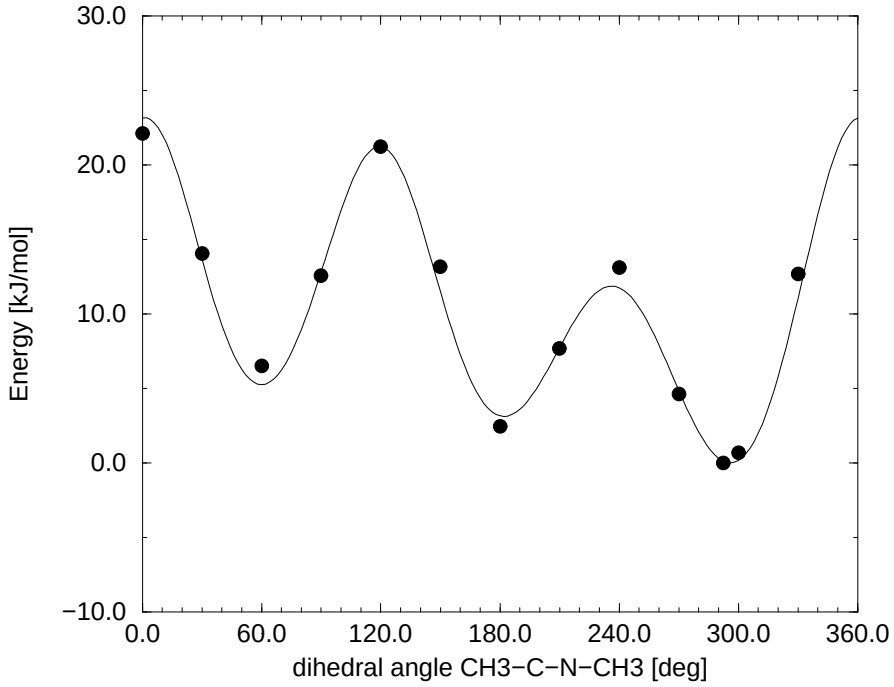

Figure 2.6: Fitted energy profile for the CH3-C-N-CH3 torsional angle in the $CH_3C(OH)_2NH_2^+CH_3$ (I3) molecule. Fitted parameters are presented in Table 2.11.

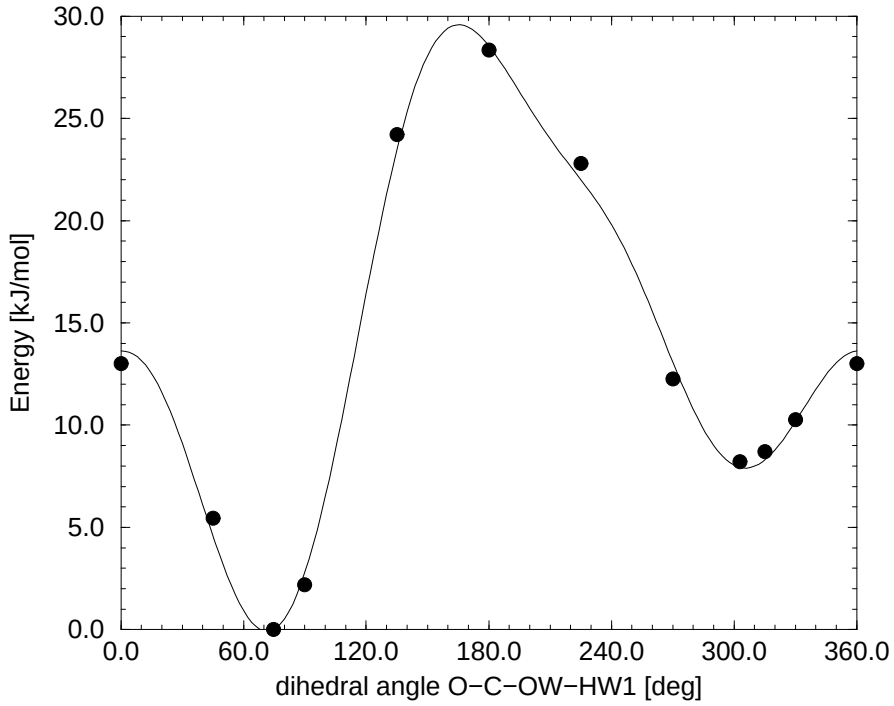

Figure 2.7: Fitted energy profile for the O-C-OW-HW1 torsional angle in the $CH_3C(OH)_2NHCH_3$ (I2) molecule. Fitted parameters are presented in Table 2.11.

2.4 Parameters Characterizing the Nonbonding Interactions in the AVB Region

To parameterize the nonbonding interactions (Eqs 2.8, 2.12, 2.25) several model complexes, consisting of two molecular fragments that take part in the model reaction path, were formed. The molecules were taken in their optimal geometries. A complete set of nonbonding valence interactions' parameters is presented in Tables 2.13 and 2.14. The parameters were fitted to the results of the DFT calculations using the B3LYP method and the 6-31G+(d,p) basis set. The names of molecules used in this Section are shown in Figures 2.1 and 2.2

2.4.1 Proton Transfer in the Water – Ionized Carboxylic Acid Complex

A model complex was formed with a water molecule and an ionized carboxylic acid ($A2^-$) placed in the same plane so that atoms OW, HW2 and OD2 formed a linear hydrogen bond. The DFT energy of such a complex was computed for various OW–OD2 distances with the molecules fixed in their optimal geometries. The binding energy has its lowest value, -64.0 kJ/mol, for an OW–OD2 distance of 2.7 Å where the zero energy level corresponds to the separated molecules H_2O and $A2^-$. Next, the transfer of the HW2 hydrogen was studied for selected OW–OD2 distances with other internal degrees of freedom also fixed. The energy profiles are presented in Figures 2.8 and 2.9.

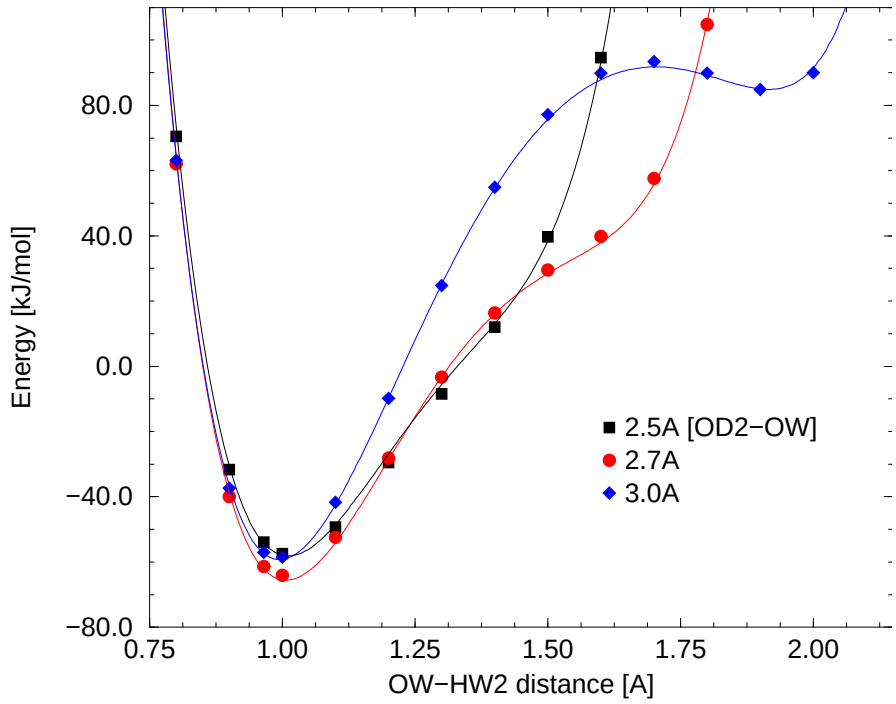

Figure 2.8: The energy of a hydrogen bonded complex of a water molecule and an ionized carboxylic acid ($A2^-$) plotted with respect to the position of the proton between the water oxygen (OW) and acid oxygen (OD2). Zero energy level corresponds to separated molecules H_2O and $A2^-$. Solid lines present the AVB energy profiles. Squares, circles and diamonds present the DFT results for OW–OD2 distances of 2.5, 2.7 and 3.0 Å, respectively.

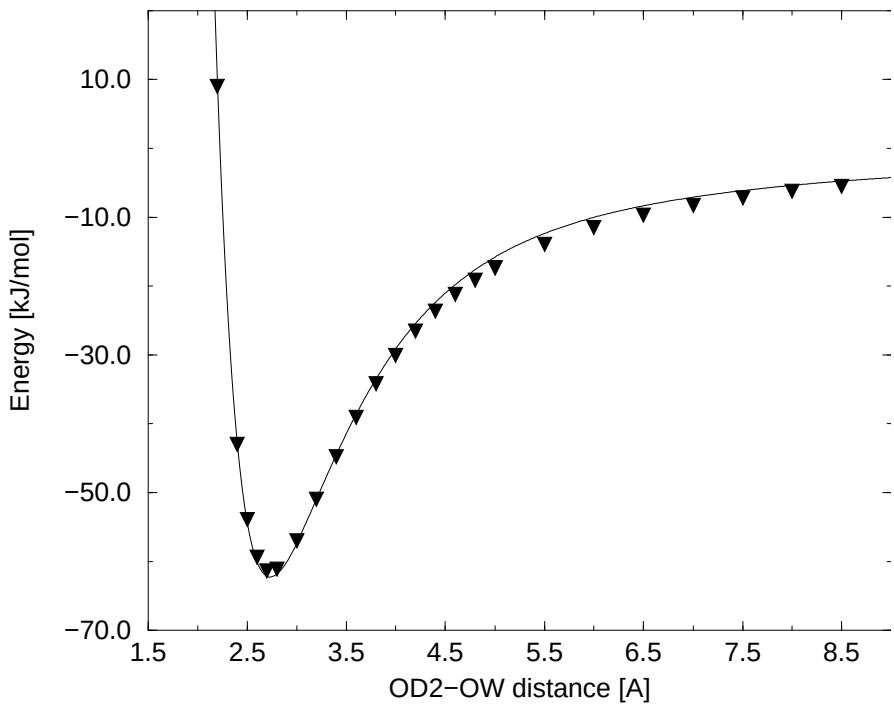

Figure 2.9: The energy profile for the OW of a water molecule approaching the OD2 atom of an ionized carboxylic acid ($A2^-$). Zero energy level corresponds to separated molecules H_2O and $A2^-$. The solid line presents the AVB energy, and triangles the DFT results, for various OD2–OW distances.

The solid line represents the ground energy eigenvalue of the three-state AVB Hamiltonian (H_2O + $A2^-$, HO^- + H^+ + $A2^-$, HO^- + A2) with intermolecular potentials V^{HB} and V^{IB} (Eqs 2.12 and 2.25) parameterized to fit the ground-state energy of the three-state AVB Hamiltonian to the DFT results. The V^{HB} potential was applied to ow*$\cdots$ ok*, ow$^-\cdots$ h, ok$^-\cdots$ hw atomic types. For atomic pairs of type ow$^-\cdots$ h$^+$ and ok$^-\cdots$ h$^+$, the V^{IB} potential was applied. The results are presented in Tables 2.13–2.14.

2.4.2 The Nucleophilic Attack

In the AVB method the nucleophilic attack of the hydroxy ion on the peptide bond carbon is modelled with three valence structures: $CH_3CONHCH_3$ (S) + OH^-, $CH_3C^+O^-NHCH_3$ (S^*) + OH^-, and $CH_3CO^-OHNHCH_3$ (I1). The parameters of intramolecular interactions in molecules S, OH^-, and I1 were determined in the previous section. The parameters of S^* were taken the same as in S except for the equilibrium length of the C–N and C–O bonds which were taken from the molecule I1. The C–O bond in molecule S^* was additionally polarized by shifting the partial charge from the O to the C atom to yield a -1 charge on oxygen. In this Section the polarization energy (E^{diss}) for molecule S^* and the parameters of nonbonding interactions between OH^- and S or S^* molecules were determined. These quantities were optimized simultaneously by fitting the ground-state energy of the three-state AVB Hamiltonian to the DFT data.

The DFT calculations were performed for the system in two configurations. The first corresponds to the state before the nucleophilic attack, with the peptide bond planar (sp^2

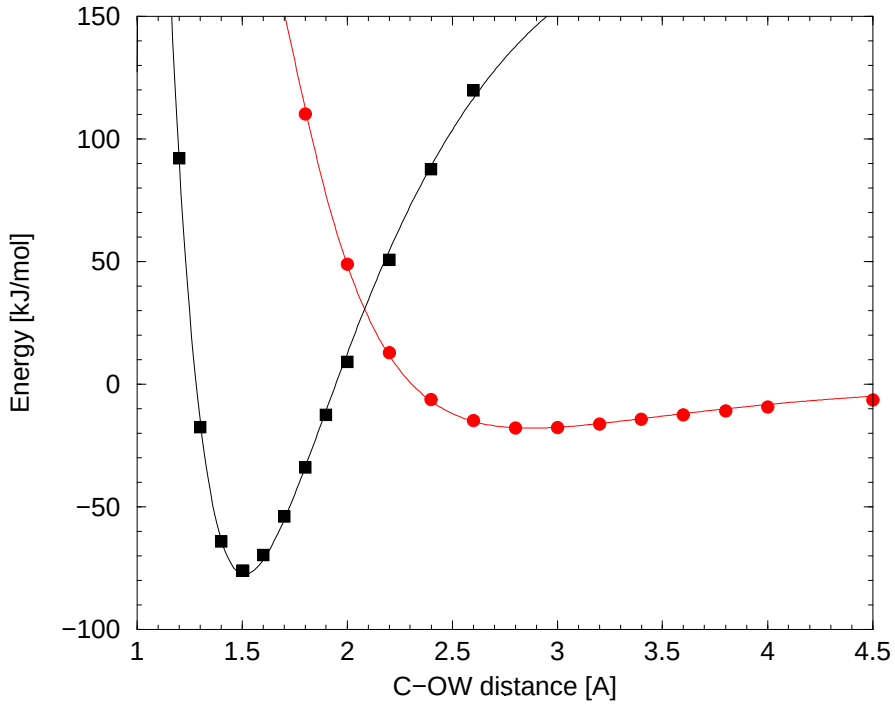

Figure 2.10: The energy profiles for nucleophilic attack of the hydroxy anion on the peptide bond, plotted with respect to the distance between nucleophile oxygen and carbonyl carbon. Zero energy level corresponds to the separated molecules. Circles and squares represent the DFT results obtained for planar and tetrahedral configurations of the peptide carbon, respectively. The solid lines represent the fitted ground-state energy eigenvalue of the three-state AVB Hamiltonian (see Tables 2.13–2.14).

hybridization of the peptide carbon) and the hydroxy ion approaching the peptide carbon from a direction perpendicular to the plane of the peptide bond. The fragments were located in such a way that the OW–HW1 bond was coplanar to the C=O of the peptide bond, the C-OW-HW1 planar angle was set to 120° and the O-C-OW-HW1 dihedral angle to 0°.

In the second case the carbon was in a tetrahedral configuration (sp^3 hybridization) as in the I1 molecule. In both configurations the DFT energy was calculated for various C–OW distances. The minimum energy was achieved for a C–OW distance of 2.8 Å and 1.6 Å for the planar and tetrahedral configurations, respectively, (see Fig. 2.10). These data were used to fit the polarization energy and parameters of the ionic potential V^{IB} between the C and OW atoms.

In addition the DFT energies for the complex with a planar peptide bond and the hydroxy ion approaching the neighbour atoms of the peptide carbon were computed. These data were used to optimize the parameters of nonbonding interactions, described with the V^{HB} type potential, between OW and CH3, N or O atoms in the peptide fragment.

2.4.3 Proton Transfer onto O^- in $CH_3CO^-OHNHCH_3$ (I1) – Neutral Carboxylic Acid Complex

This proton transfer was modelled with three valence structures: I1 + A1, I1 + H^+ + $A1^-$, and I2 + $A1^-$. The complex was formed using optimal geometries of molecules I1 and A1. The molecules were positioned so that the O-HD2-OD2 atoms were colinear and the angles CG-OD2-HD2 and HD2-O-C set to 120°. The binding energy was calculated for various OD2–O and OD2–HD2 distances and minima of -109.65 kJ/mol for OD2–O = 2.6 Å and OD2–HD2 =1.1 Å were obtained (zero energy level corresponds to the separated molecules, A1 and I1). These data were used to fit the parameters of nonbonding interactions V^{IB} between oxygens and H^+, and V^{HB} between oxygens of I1 (types om and oa) and A1 (types ok and ok-) as well as oxygens and appropriate hydrogens (types h and ha) of these molecules (see Tables 2.13–2.14). The energy profiles are shown in Figures 2.11 and 2.12.

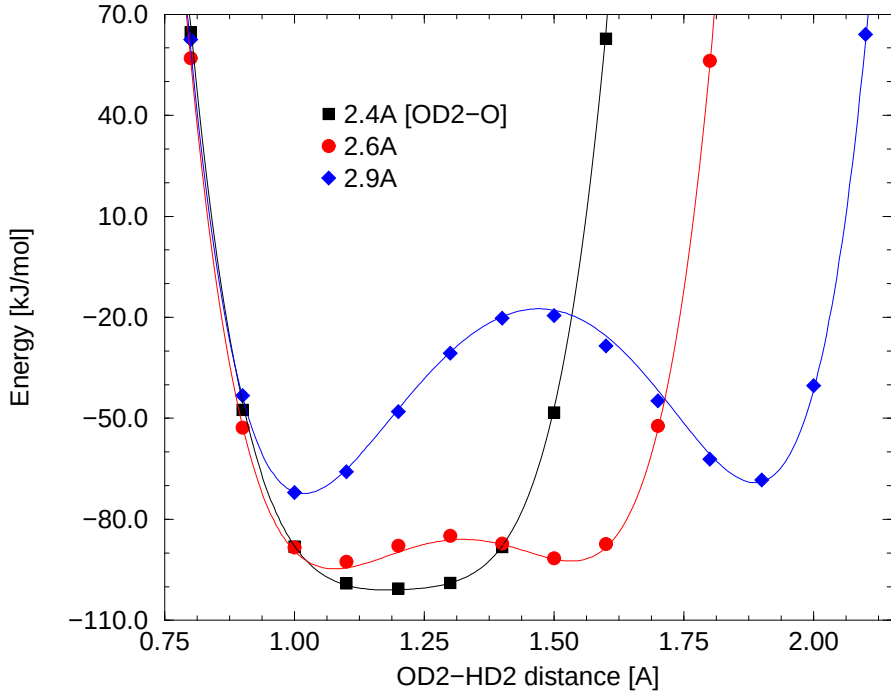

Figure 2.11: Plot showing selected energy profiles for proton transfer from neutral aspartic acid onto the oxygen of the I1 molecule for selected OD2–O distances. Zero energy level corresponds to separated molecules A1 and I1. Squares, circles and diamonds present the DFT results for OD2–O distances of 2.4, 2.6 and 2.9 Å, respectively. The solid lines represent the ground energy eigenvalues of the three-state AVB Hamiltonian (see Tables 2.13–2.14).

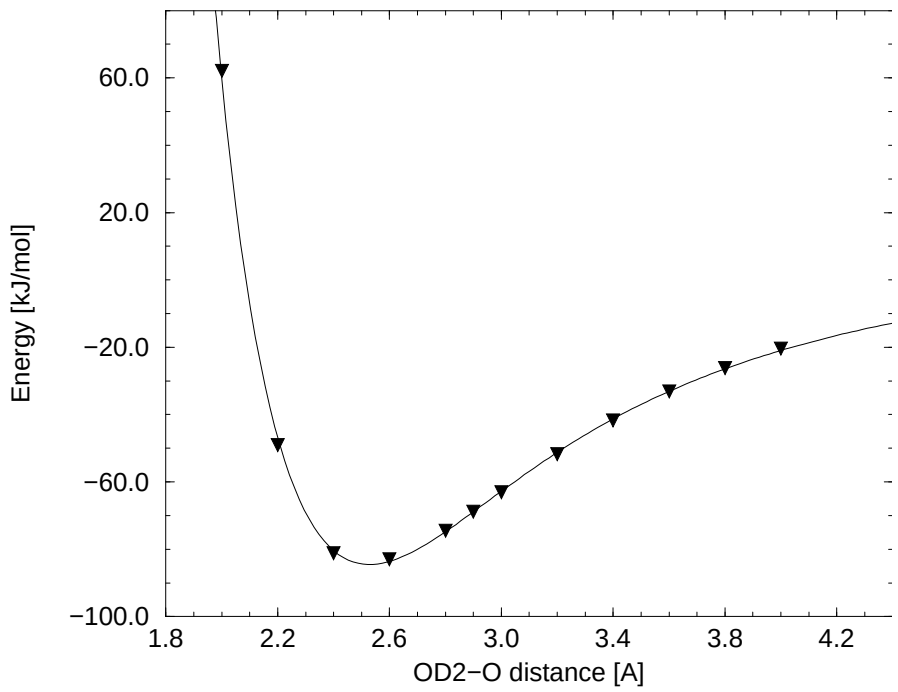

Figure 2.12: Plot showing the selected energy profile for OD2 of neutral aspartic acid approaching the oxygen atom of the I1 molecule. Zero energy level corresponds to separated molecules A1 and I1.

2.4.4 Proton Transfer from Neutral Carboxylic Acid onto the Nitrogen of the $CH_3C(OH)_2NHCH_3$ (I2) Molecule

The complex was modelled with a three-state AVB Hamiltonian with the following structures: I3 + $A2^-$, I2 + H^+ + $A2^-$ and I2 + A2. The $CH_3C(OH)_2NH_2^+CH_3$ (I3) and $RCOO^-$ ($A2^-$) molecules were taken and a linear hydrogen bond formed between the OD2–HW2 and N atoms. The HW2-OD2-CG angle was set to 120° and atoms N, HW2, OD2, and CG were located coplanar. The energy of the system was calculated for various OD2–N and N–HW2 distances. Energy profiles are shown in Figures 2.13 and 2.14. The intermolecular potentials were parameterized to fit the DFT results in a similar way as in the previous cases (see Tables 2.13–2.14).

2.4.5 The Breakage of $CH_3C(OH)_2NH_2^+CH_3$ (I3) Molecule

The breakage of the tetrahedral intermediate I3 is a process analogous to nucleophilic attack, but occurring in the opposite direction. In the AVB method the breakdown of the intermediate is described with three valence structures: $CH_3C(OH)_2NH_2^+CH_3$ (I3), $CH_3C(OH)_2^+$ (P1) + CH_3NH_2 (P3), and $CH_3C^{+1}(OH)_2$ (P1*) + P3. The DFT calculations were performed for two molecular configurations. The first one was built up of P1 and P3 molecules taken in their optimal geometries. The N-C-CA2 and CA3-N-C angles were set to 88° and 121°, respectively. The CA3-N-C-CA2 dihedral angle was set to –160°. The second configuration corresponds to the optimal geometry of the I3 molecule. In both configurations the DFT energy was calculated at various distances of C–N (see Fig. 2.15). These data were used to fit the ground state energy of the three-state AVB Hamiltonian.

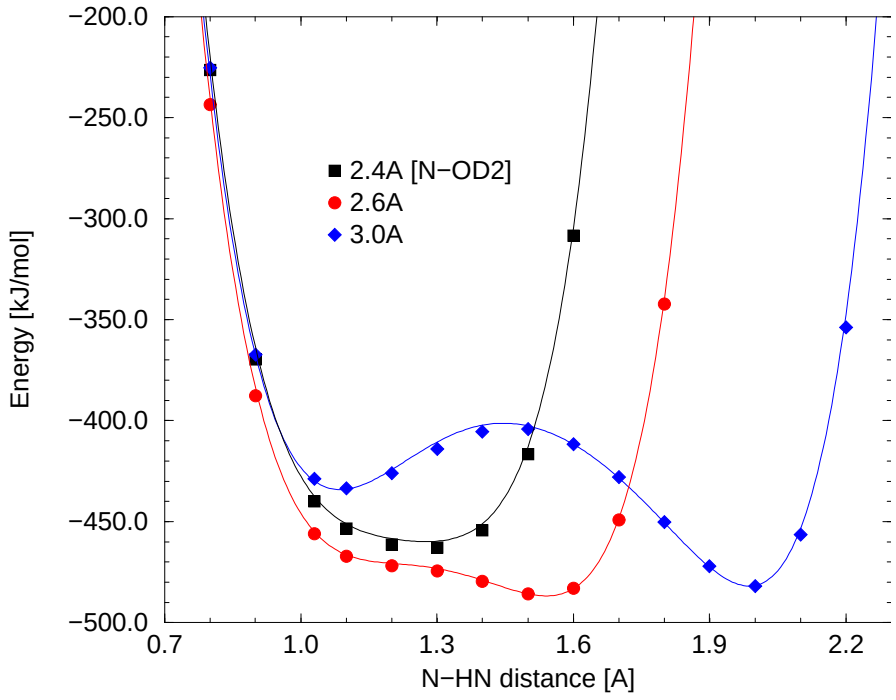

Figure 2.13: Plot showing selected energy profiles for proton transfer from neutral aspartic acid onto the nitrogen of the I2 molecule for selected N–OD2 distances (zero energy level corresponds to the separated molecules $A2^-$ and I3). Solid lines represent the ground energy eigenvalue of the three-state AVB Hamiltonian (see Tables 2.13–2.14).

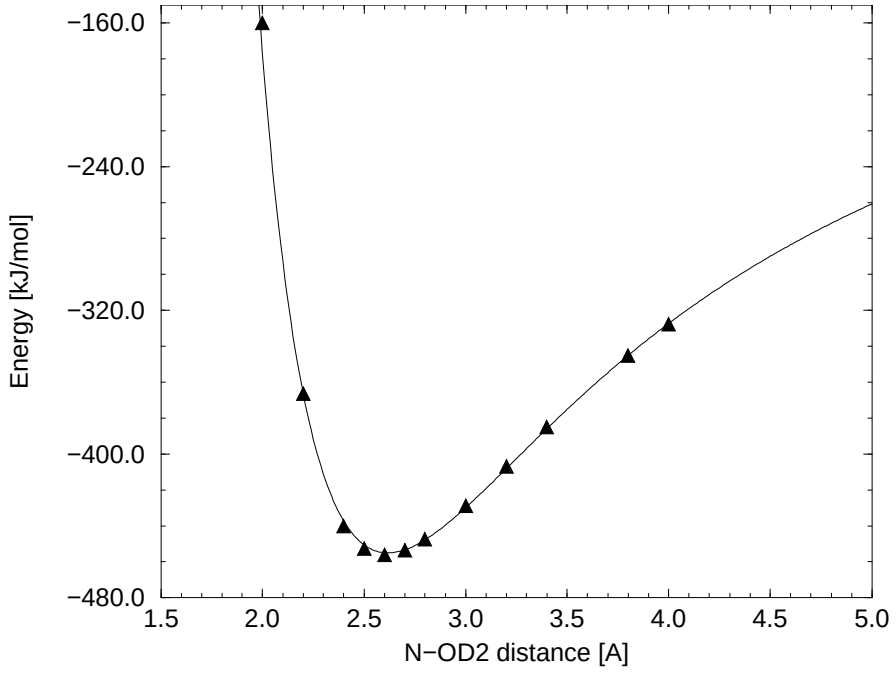

Figure 2.14: Plot showing the selected energy profile for the approach of the OD2 atom of aspartic acid to the nitrogen atom of the I2 molecule (zero energy level corresponds to the separated molecules $A2^-$ and I3). Fitted parameters are presented in Tables 2.13–2.14.

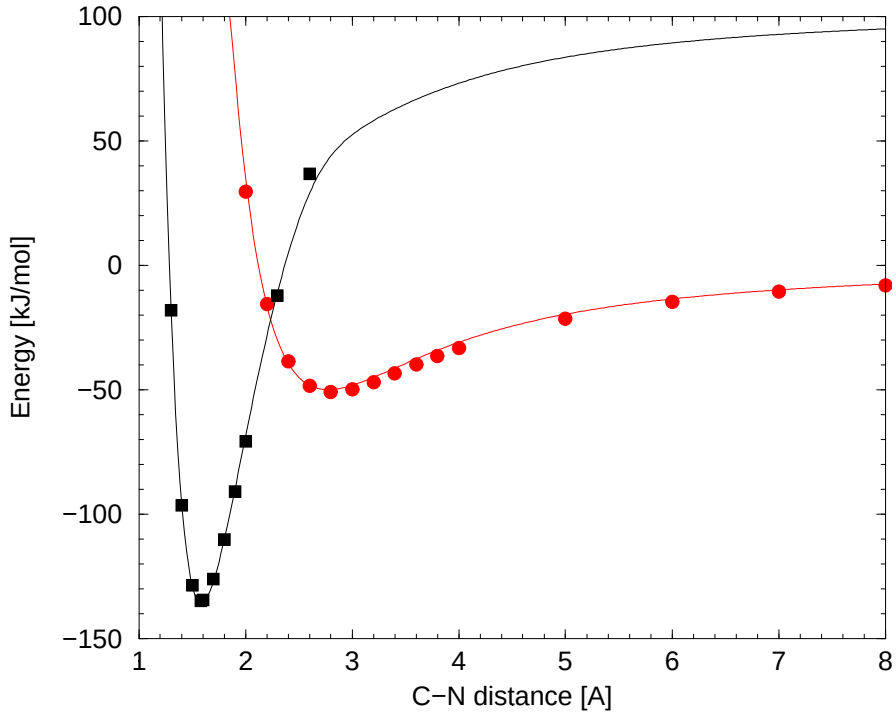

Figure 2.15: Energy profiles describing the cleavage of the reaction intermediate (I3). The profiles are plotted with respect to the distance between peptide nitrogen and peptide carbon. Zero energy level corresponds to the separated P1 and P3 molecules. Circles and squares represent the DFT results obtained for planar and tetrahedral carbon configurations, respectively. The solid lines represent the ground energy eigenvalue of the three-state AVB Hamiltonian (I3, P1* + P3, P1 + P3) with intermolecular potentials parameterized to fit the DFT results (see Tables 2.13-2.14).

The model complex of $CH_3C(OH)_2^+$ (P1) and CH_3NH_2 (P3) molecules was used to model the intermolecular interactions in the breakdown of the tetrahedral intermediate. In the fit the polarization energy of P1 and interaction parameters (V^{IB}) for C and N atoms were optimized.

2.4.6 Proton Transfer between Doubly Protonated and Ionized Carboxylic Acids

To obtain the intermolecular interaction parameters for this transfer a complex using $CH_3C(OH)_2^+$ (P1) and CH_3COO^- ($A1^-$) molecules was formed. A linear hydrogen bond was formed between O, HD2, and OD2 atoms. The C-O-OD2-CG dihedral angle was set to 180° and the HD2-OD2-CG angle to 120°. The O–OD2 distance was modified, the DFT energy calculations performed, and the minimum attained for 2.4 Å. The AVB parameters were fit in the same way as previously described for other proton transfer processes. The energy profiles are shown in Figures 2.16 and 2.17.

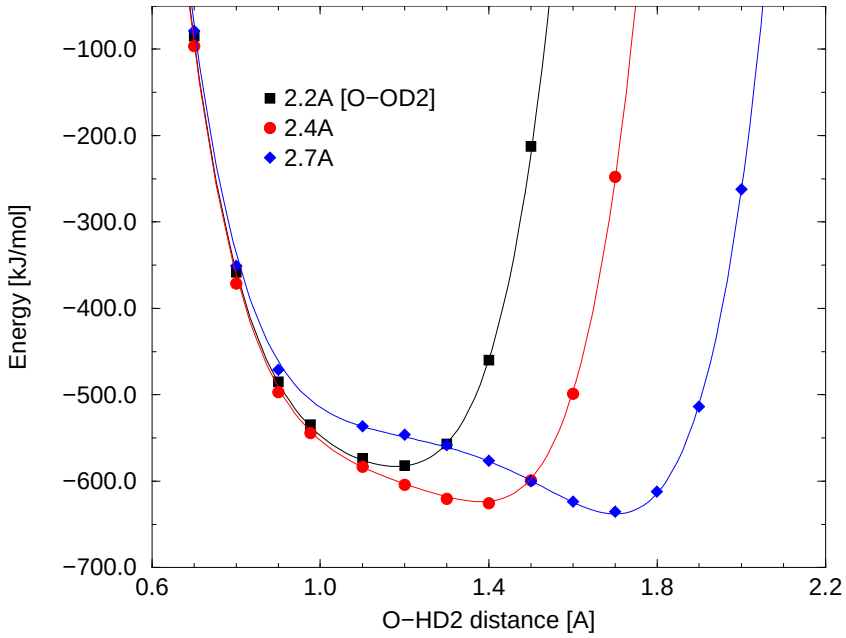

Figure 2.16: Plot showing selected energy profiles for proton transfer from a doubly protonated carboxylic acid onto an oxygen of an ionized carboxylic acid for selected oxygen-oxygen distances (zero energy level corresponds to the separated molecules, P1 and $A1^-$). Solid lines represent the ground energy eigenvalue of the three-state AVB Hamiltonian (P1 + $A1^-$, P2 + H^+ + $A1^-$, P2 + A1). Fitted parameters are presented in Tables 2.13–2.14.

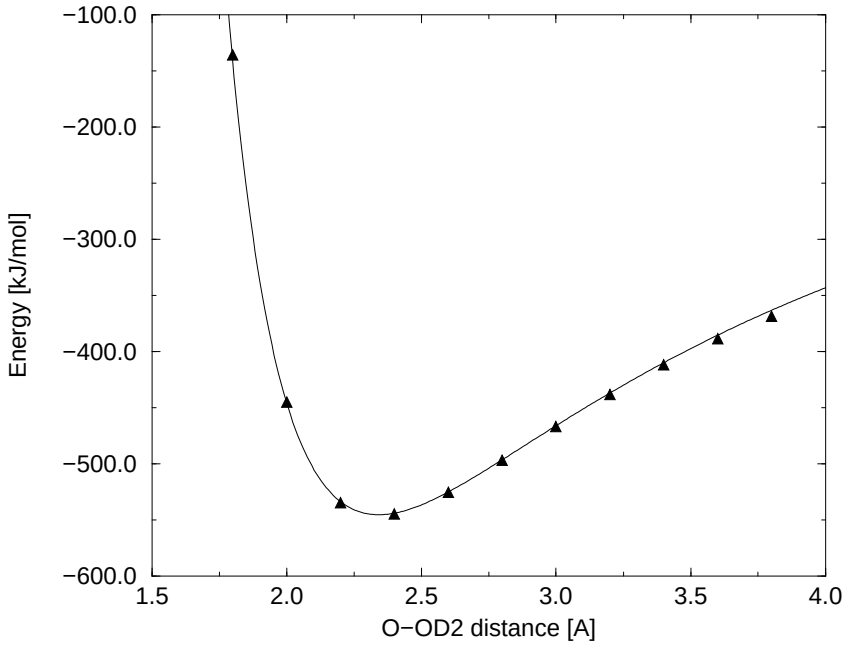

Figure 2.17: Plot showing the selected energy profile for the approach of the O atom of a doubly protonated carboxylic acid to the OD2 oxygen of an ionized aspartic acid (zero energy level corresponds to the separated molecules, P1 and $A1^-$). Fitted parameters are presented in Tables 2.13–2.14.

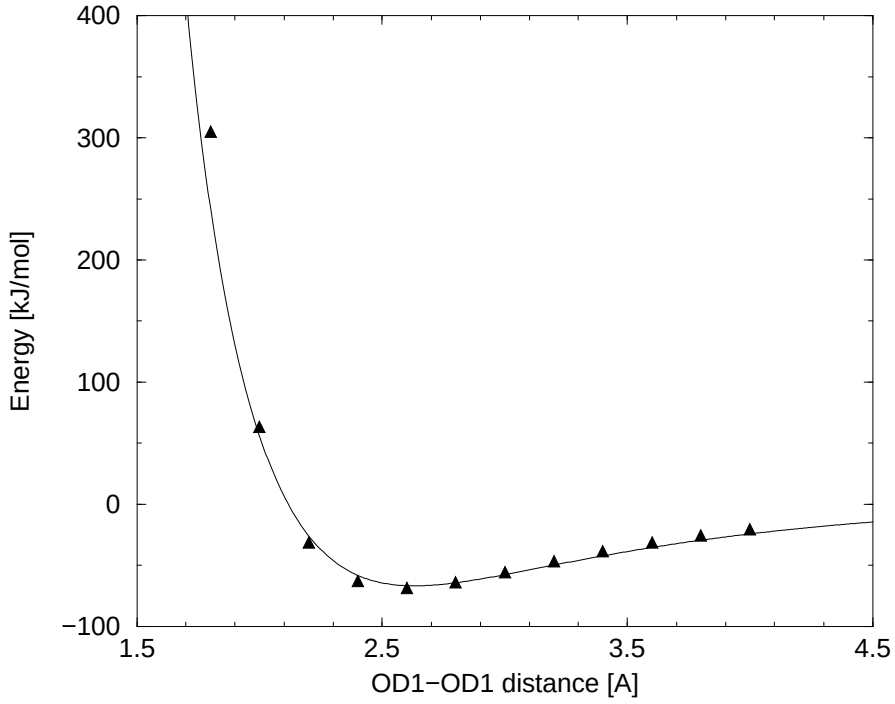

Figure 2.18: Plot showing the selected energy profile for proton transfer between OD1 oxygens of two ionized aspartic acids (zero energy level corresponds to the separated molecules, A1 and $A2^-$). Fitted parameters are presented in Tables 2.13.

2.4.7 Nonbonding Interactions between Oxygens of Neutral and Ionized Carboxylic Acids

It is still questionable which of the oxygens (OD1 or OD2) of Asp25 is protonated (see Fig. 1.3). In this parameterization the possibility of Asp25 to be protonated at OD1 oxygen is also taken into account. To achieve this, appropriate interactions between the OD1 oxygens of Asp25 and Asp125, with the hydrogen between them, have to be additionally modelled. These intermolecular interaction parameters (V^{HB} potential type, atom types ok- and ok*) were fitted to the DFT calculations for a complex of CH_3COOH (A1) and CH_3COO^- ($A2^-$) molecules (Fig. 2.18). The molecules were located so the OD1 oxygens of carboxylic acids and the hydrogen formed a linear hydrogen bond. The two CG-OD1-OD1 dihedral angles were set to 138° so while the OD1–OD1 distance was modified OD2 atoms could be put in a closer distance.

2.4.8 Proton transfer from Neutral Carboxylic Acid onto the Nitrogen of the $CH_3CO^-OHNHCH_3$ (I1) Molecule

The $RCO^-OHNH_2^+R$ (I2′) molecule is a consequence of a proton transfer onto the nitrogen of $CH_3CO^-OHNHCH_3$ (I1) molecule (see Figure 2.2). This transfer is modelled with three valence structures: I1 + A2, I1 + H^+ + $A2^-$, and I2′ + $A2^-$. However, the I2′ molecule is not stable. The optimization of I2′ molecule with the Gaussian'94 program [89] with the B3LYP [90, 91] functional and 6-31G(d,p) basis set was performed. It resulted in either two product molecules RNH_2 and RCOOH aprroximately 3.5 Å apart or in an I2 molecule

Table 2.13: Parameters for the hydrogen bond interactions ($V^{HB}_{a,b}$), (Eq. 2.25).

Type a	Type b	A_0 (kJ/mol)	α (Å^{-}1)	B_0 (kJ/mol*Å^6)
ow*	ok*	4520.	1.5610	1070.
ok-,o	hw	41400.	4.3310	–
ow-	h	55200.	4.3930	–
ok	h	143500.	4.8410	–
ok-	ok*	6250.	1.5830	1020.
ok	ok	70800.	2.9280	1380.
ok-	h	41400.	4.3310	–
om	ok*	66100.	2.7780	1630.
ow*	ok*	4520.	1.5610	1070.
om	h	2890.	1.4770	–
ok*	ha	2400.	1.7780	–
ok*	n	40500.	3.8780	1000.
ok-,ow-	hn	13400.	3.4280	–
n*	h,hw	30300.	3.1740	–
ow-	o	79700.	2.8690	4280.
n	ow-	42600.	2.8780	1430.
ch3	ow*	56100.	3.3250	9100.
na	ok*	84200.	3.6280	4800.
ch3	na	69600.	2.6170	8790.

Table 2.14: Parameters for nonbonding valence interactions ($V^{IB}_{a,b}$), (Eq. 2.12)

Type a	Type b	A (kJ/mol*Å)	α (Å^{-}1)	B (kJ/mol)
ok-	h+	5510.	1.5560	5200.
ow-	h+	5600.	1.4240	4100.
ok	h+	6980.	1.7760	8080.
om	h+	8700.	1.8240	6270.
n	h+	4560.	1.1600	4650.
cn	ow-	27100.	1.2120	7040.
cn+	ow-	5840.	1.2920	3370.
c	na	4720.	1.5180	1000.

since one of the hydrogens bonded to nitrogen drifted to form a covalent bond with the negatively charged oxygen of carboxylic group. So this proton transfer could not be parameterized in the same manner as the previous ones. The geometry of I2′ molecule was therefore taken as in the I3 molecule with one of the hydrogens removed from oxygen. Only the energy and ESP charges (Table 2.6) were calculated with DFT methods. The ground state of I2′ was modelled taking appropriate parameters as in I3 molecule except the parameters for bonds and angles that include O^- atom which were taken as in I1 molecule. The I2′ molecule was assigned the same valence types as I3 molecule and the nonbonding parameters were taken based on this valence types according to parameterization of the proton transfer onto N of I2 molecule. The parameters of P2* were taken the same as for P2 except for the charges on C and O^- which were set to 1 and -0.831, respectively. Ionic dissociation energies and parameters of ionic structure coefficients are listed in Table 2.9. Ionic dissociation energy for P2 → P2* was taken as in the fitted S → S* process.

2.4.9 Summary

This Chapter presented the DFT calculations performed to parameterize the potential energy surface in the AVB method for the reaction catalysed by HIV-1 PR. The modifications of the previously developed AVB model were also described. This parameterization allows to perform the MD/AVB simulations in which the reactions occurring in the active site of this enzyme are accounted for. In the MD/AVB dynamics two regions of the solvent and protein are distinguished: the AVB region described in a manner and with the parameters presented in this Chapter and the classical surrounding environment described with conventional force field. The energy of all interactions in the AVB region and its electrostatic and polarization interactions with the classical region are calculated within the AVB formalism. The energy of all atomic interactions inside the classical region, as well as the Lennard-Jones potential and bonding interactions across the border between the classical and AVB regions, are calculated with the conventional classical Gromos'96 force field [7]. The total potential energy surface of the system is a sum of both the classical and AVB energy terms. The MD/AVB simulations performed for the whole system are presented in Chapter III.

The reliability of the AVB model may be, of course, discussed. One question that can be asked is if the B3LYP functional and 6-31G(d,p)or 6-31G+(d,p) basis sets used for parameterization give reliable energetic description of the considered chemical reactions in vacuum. It has been checked and compared with various literature (e.g. [96], [95]) that the results of the AVB method concerning proton affinities and the energy of the nucleophilic attack are consistent with those computed with the MP2 method in similar bases. This results could be, however, confirmed further by using methods based on a higher basis sets and more precise treatment of electron correlation effects.

In the present description of the valence structures, formulae of elementary potentials of the classical force field have been adopted. In particular, the potential used to describe a single valence structure is a superposition of elementary nonbonding interactions and bonding terms characterizing appropriate bond lengths and all types of angles in molecules. It should be noted that it is possible to add more sophisticated elementary potentials, e.g. the mixed bonding potential terms as used in extensions to some classical force fields. Such approach may be a matter of discussion and is only limited by the effort and time one can assign for the parameterization procedure.

Regarding the parameters itself, some of them may be better optimized and differen-

tiated. For example a_w in Eq. 2.16 can be fitted for every angular deformation connected with the dissociable bond. Herein one value of a_w is assigned for the whole molecule.

Chapter III

Molecular Dynamics Simulations of the Reaction Mechanism Catalysed by HIV-1 Protease

3.1 Introduction

There are several questions regarding the mechanism of the reaction catalysed by HIV-1 PR that cannot be answered by experimental means and are also addressed in this Chapter. One regarding the location of protons in the active site has been described previously in Chapter I. Others are, e.g. the position of the water molecule in the cleavage site, the mobility of the active site environment, the stability of protein-substrate interactions, and the protonation state of the reaction intermediate.

Prior to molecular dynamics simulations of the modelled complex a special relaxation procedure had to be applied due to the addition of a water molecule to the active site and some alterations of the inhibitor. The modifications of the initial HIV-1 PR : MVT-101 crystal structure are described in Section 1.2.2. The active site region had to be gradually relaxed what involved a series of energy minimization runs. Afterwards, the standard thermalization was carried out. First, however, as a test performance calculations were carried out for a system in vacuum. Such effort was taken to check if the HIV-1 PR molecular dynamics simulations can be reliably performed for an isolated protein what would greatly lower the computational time in further MD/AVB calculations. The conclusions are presented in this Chapter. Next, the system was solvated and molecular dynamics was performed with explicit solvent and mobile ions. All atom names that appear in this Chapter are presented in Figures 1.3, 2.1, and 2.2. The word atoms is used meaning either the nuclei or their positions.

3.2 Equilibration and Thermalization of the HIV-1 PR : Substrate Complex

3.2.1 Energy Minimization

As a first step, a topology file for the modelled system with the classical interaction parameters was prepared. Atomic charges used for the minimizations were based on the Gromos'96 force field (B-version) [7], suitable for applications to macromolecules with

implicit solvent. However, several modifications of the parameters have been applied in order to neutralize strong electrostatic field which may be generated, e.g. by solvent-exposed charged side chains that are not surrounded by explicit water molecules, or to provide stronger interactions for pairs of charged amino acids. The first alterations concerned the single charged residues on the surface of the protein. Thus, the electrostatic charges of all terminal atoms of Glu21, Glu121, Lys41, Lys141, Lys55, Lys155, Lys70 and Arg114 were set to zero. Secondly, for some amino acids that are grouped in pairs or sets of three, or are situated inside the protein, the A-version of the Gromos'96 force field (suitable for applications to proteins with explicit solvent) was applied. It refers to Arg8, Asp25, Asp29, Arg108, Asp129 and Asp125 residues. In the HIV-1 PR : MVT-101 crystallographic complex, on which the calculations are based, two α-amino-N-butyric acids replace cysteines 67 and 95 in each monomer. Such residue was build on the basis of the alanine in the Gromos'96 topology file and the charge on the additional Cγ carbon was set to zero as in the united atom model. The modified inhibitor, a peptide consisting of 6 residues, is neutralized from one side by the acetyl group CH3-CO and from the other side by the CO-NH2 group. The CO charges were taken as in the peptide bond, for the CH3 group were set to zero within the united atom model, and for the NH2 group as in the glutamine and asparagine residues of the Gromos'96 charges.

Hydrogens were added with the HBUILD command of `CHARMm22` [41, 42] and the united atom model was used. After the structural and charge modifications the positions of hydrogens of crystallographic waters were optimized with 100 steps of the steepest descents method followed by 200 steps of the conjugate gradients minimization method. The dielectric constant ϵ was set to 4. All further minimizations were performed with the same procedure. The next step was the optimization of all the protein hydrogens' positions except those of Asp25, catalytic water (207), the water situated above and close to the modified inhibitor (301), and HN of the hydrolyzable peptide bond. Afterwards, the positions of all atoms of crystallographic waters (except 207 and 301) were optimized. Later, the active site atoms of Asp25 and 125, the substrate residues Met203 and 204, waters 207 and 301, and the residues of the flap region (numbers: 48 to 53 and 148 to 153) were relaxed with 50 steps of the conjugate gradients method. C_α atoms were constrained during this simulation. Further on, the bond and planar angle constraints (the harmonic force constant of 400 [kJ/mol·Å^2] as a starting value) were gradually decreased every 50 steps of the conjugate gradients minimization. So finally, with no constraints applied, the active site atoms could adjust themselves to the rest of the protein.

3.2.2 Thermalization Procedure

The prepared structure was then subject to thermalization in vacuum. The potential energy surface included the AVB parameterization in the active site region (see Chapter II) and Gromos'96 force field parameters for the rest of the enzyme. However, this thermalization attempt was not successful even though some alterations of the standard Gromos'96 force field parameters to neutralize the system have been applied as described in 3.2.1. When the temperature reached 200K in the NVT ensemble the modelled system was not stable and the molecular dynamics simulations in vacuum could not be continued. It was therefore decided to perform MD simulations with explicit water and ions.

Following such conclusions the protein was solvated in an octahedral box of SPC water molecules [97]. The width of the box was set to 7 Å from the border of the solvent accessible surface. Most (94 of 113) water molecules observed in the crystal structure were

preserved. It involved the waters buried inside the protein, and those situated close to the surface and hydrogen bonded to the amino acids on the surface of the enzyme. In total, the box was filled with 5572 water molecules. The parameters of the force field were changed to those used for the system with explicit solvent (A-version).

A 20ps MD simulation of the solvent at 300K temperature (with the protein fixed) was performed with a time step of 0.25fs. Velocities were generated every 5ps from the Maxwellian distribution at temperature of 100K. The relaxation time of the thermal bath was set to 0.01ps. The pair list was updated every 10 steps with a cut-off of 12 Å. The cut-off radius used in long range interactions and in the reaction-field calculations was set to 14 Å. The truncated octahedron periodic boundary conditions were used.

In order to neutralize the modelled system the next step involved adding the charged ions. Fourteen waters were replaced by ten Cl^- and four Na^+ ions. The ions were located at positions of the chosen water oxygens. They were added to neutralize the singly charged residues on the surface of the protein and the groups of three charged amino acids situated close to one another. After the addition of ions the overall charge of the system (excluding the aspartyl dyad of the active site) was equal zero.

Then in order to adjust both the solvent and ions to the protein crystallographic structure molecular dynamics simulations of the water and ionic environment were carried out. Ten 0.1ps runs at 300K temperature with a time step of 0.25fs with velocities reassigned each time at temperature of 100K were performed. Only the water and ion atoms were allowed to move. Next a 20ps dynamics of the solvent and ions at 300K with velocities reassigned every 5ps at temperature 100K was carried out.

Then the structure underwent the standard thermalization procedure with the potential energy surface including the AVB parameters for the active site and Gromos'96 force field for the rest of the system. The MD/AVB runs where performed for the first valence bond structure: $A1_c$ + H_2O_c + $A2_g^-$ + S_c (see Section 2.2.2). The solvent atoms were positionally restrained (to allow for some flexibility of the system) using a harmonic force constant of $2.5{\cdot}10^4$ [kcal/mol·Å^2] as in the standard exemplar input of the Gromos'96 package. A 5ps run in the NVT ensemble at 10K temperature with a time step of 0.5fs was performed. Then a 30ps run with velocities generated every 5ps and temperature increased by 50K from 50K to 300K also every 5ps was carried out with a time step of 0.5fs (see Table 3.1). The solute atoms were free to move, however, some additional constraints were applied to the active site atoms as listed in Table 3.1. The constraints were applied to conserve the hydrogen bonds in the active centre and to keep the position of the added lytic water between the aspartic acids. The distances Gly27:N–Asp25:OD1, Gly27:N–Asp125OD1, Gly127:N–Asp25:OD1, Gly127:N–Asp125:OD1 were restrained using the harmonic potential energy functions with a changeable force constant at a reference distance of 3 Å. The above hydrogen bonds and strong dipoles of the Thr26–Gly27 and Thr126–Gly127 pointing toward the negative charge of aspartyl dyad are said to be responsible for maintaining the position and planarity of aspartates. The other active site atoms listed in Table 3.1 were restrained using one-sided harmonic potential energy functions, i.e. the restraining function was applied only when the actual distance between the restrained atoms was larger than the reference distance.

The solute was heated to 300K during 35ps dynamics and the next step was to gradually free the solvent atoms and to allow the movement of the whole system. This was done by decreasing the harmonic force constant, applied for restraining the solvent atoms, by 2 to $1.25{\cdot}10^4$ [kcal/mol·Å^2] and by 4 to $0.625{\cdot}10^4$ [kcal/mol·Å^2] (two 10ps runs). The last step involved a 100ps simulation at 300K with all atoms free to move and the constraints

Table 3.1: Restrained atoms, reference distances and harmonic force constants applied during the thermalization of the solute atoms of HIV-1 PR

Temperature	Time	restrained distance	reference distance	force constant
[K]	[ps]		[Å]	[kJ/mol·Å^2]
10	0–5	Met203:C – H_2O 207:OW	2.8	1000
		Asp25:OD2 – H_2O 207:OW	2.8	1000
		Asp25:OD2 – Met203:O	2.8	500
		Asp125:OD1 – H_2O 207:OW	2.8	1000
		Gly27,127:N – Asp25,125:OD1	3.0	3600
50	5–10	Met203:C – H_2O 207:OW	3.0	1000
		Asp25:OD2 – H_2O 207:OW	2.8	1000
		Asp25:OD2 – Met203:O	2.8	500
		Asp125:OD1 – H_2O 207:OW	2.8	1000
		Gly27,127:N – Asp25,125:OD1	3.0	3600
100	10–15	Met203:C – H_2O 207:OW	3.0	1000
		Asp25:OD2 – H_2O 207:OW	2.8	1000
		Asp25:OD2 – Met203:O	2.8	500
		Asp125:OD1 – H_2O 207:OW	2.8	500
		Gly27,127:N – Asp25,125:OD1	3.0	1800
150	15–20	Met203:C – H_2O 207:OW	3.0	1000
		Asp25:OD2 – H_2O 207:OW	2.8	500
		Asp25:OD2 – Met203:O	2.8	500
		Asp125:OD1 – H_2O 207:OW	2.8	500
		Gly27,127:N – Asp25,125:OD1	3.0	1800
200	20–25	Met203:C – H_2O 207:OW	3.0	1000
		Asp25:OD2 – H_2O 207:OW	2.8	500
		Asp25:OD2 – Met203:O	2.8	500
		Gly27,127:N – Asp25,125:OD1	3.0	900
250	25–30	Met203:C – H_2O 207:OW	3.0	1000
		Asp25:OD2 – H_2O 207:OW	2.8	500
		Asp25:OD2 – Met203:O	2.8	500
		Gly27,127:N – Asp25,125:OD1	3.0	900
300	30–35	Met203:C – H_2O 207:OW	3.0	1000
		Asp25:OD2 – H_2O 207:OW	2.8	500
		Asp25:OD2 – Met203:O	2.8	500
		Gly27,127:N – Asp25,125:OD1	3.0	900
300	35–155	Met203:C – H_2O 207:OW	3.0	500
		Asp25:OD2 – H_2O 207:OW	2.8	500
		Asp25:OD2 – Met203:O	2.8	500
		Gly27,127:N – Asp25,125:OD1	3.0	900

decreased. The relaxation time of the thermal bath in the calculations where protein was allowed to move was set to 0.1 ps.

The root mean square (rms) deviation of the averaged thermalized structure (30 coordinates of the last 15 ps of thermalization taken at equal intervals) with the crystallographic structure taken as a reference was calculated. The rms value equals 1.78, 1.80 and 2.14 Å for the C_α carbons (408 atoms), backbone (1616), and the heavy solute (3146) atoms, respectively.

3.3 MD/AVB Simulations

The first aim was to check if the parameterization of the potential energy surface in the active site works within the system and allows to perform reliable molecular dynamics simulations. The first step of the reaction, i.e. the dissociation of the lytic water molecule and the nucleophilic attack of the resulting hydroxy anion on the peptide bond carbon involved only eight valence structures in the AVB region (1–6 and 13, 14 of the first proposed model – see Section 2.2.2). So a 100 ps molecular dynamics run of the whole system in the NVT ensemble at 300 K temperature and a time step for integration of 0.5 fs was performed. Figure 3.1 shows the rms deviations from the modified and minimized crystal structure and the averaged thermalized structure (30 coordinates of the last 15 ps of thermalization taken at equal intervals). The rms value was calculated every 2.5 ps (last 30 frames were taken at equal intervals and averaged coordinates were generated).

The proton transfer did not happen during this simulation but obviously the time was too short for the reaction to occur. However, the distance fluctuations in the active site are significant and in a larger time scale it may be possible that the proton is transferred onto Asp125 and a hydroxy anion is formed. Figure 3.2 shows the changes of H_2O207:HW2...Asp125:OD2, H_2O207:HW2–H_2O207:OW, H_2O207:OW...Asp125:OD2 distances during a 100 ps simulation.

The minimal, maximal, and mean values with the standard deviations of the above distances during the 100 ps simulation are shown below:

atoms	minimal distance [Å]	maximal distance [Å]	mean distance [Å]	stddev [Å]
H_2O207:HW2 – Asp125:OD2	0.97	2.17	1.47	0.13
H_2O207:HW2 – H_2O207:OW	0.91	1.96	1.09	0.06
H_2O207:OW – Asp125:OD2	2.17	3.03	2.50	0.10

If the nucleophilic attack is to occur, the change in hybridization of the peptide bond carbon (C) and the formation of the bond between OW and C atoms is a must. The C...OW distance in the 100 ps simulation varies between 2.6 and 3.2 Å. The [CA2 N C O] improper angle maximal and minimal values are $-2°$ and $34°$ with an average value of only $16 \pm 5°$. This values, however, are far too small for the formation of any of the tetrahedral intermediate.

As mentioned before (Section 1.2.4 and Figure 1.10) the crystal structure with MVT-101 inhibitor includes a water molecule (301) which links the inhibitor with the flaps via two pairs of hydrogen bonds. It is assumed that this water molecule helps bind the substrate in a conformation conductive to cleavage and is important for maintaining the structure of the complex. Thus it was left in the MD/AVB simulations with the substrate. During the 100 ps dynamics its hydrogen bonds with the enzyme atoms are conserved. The

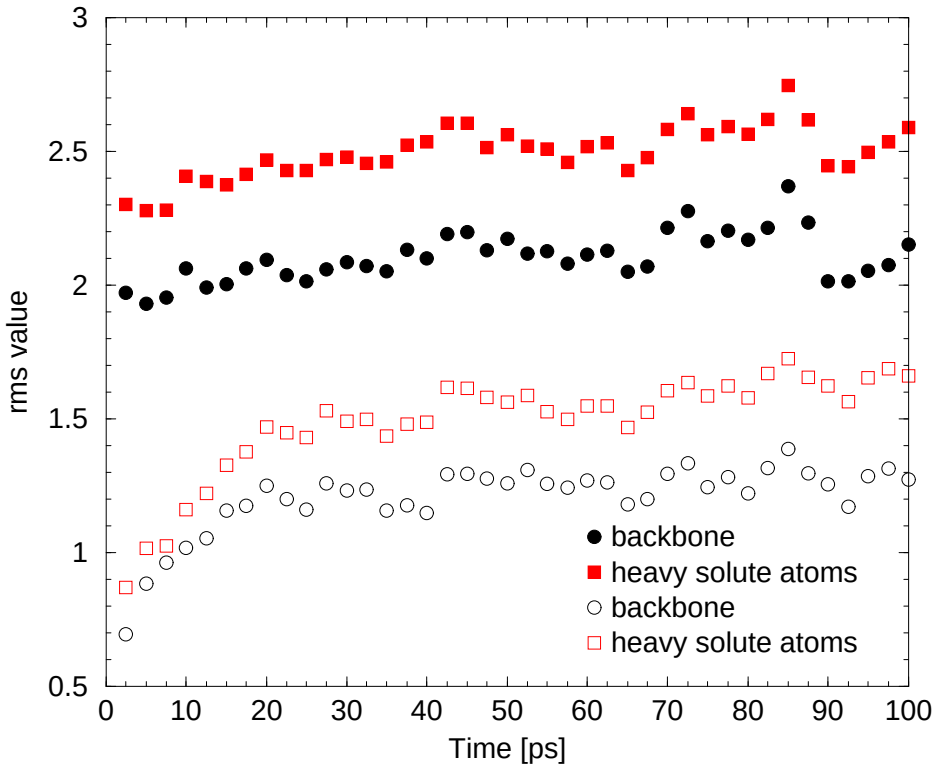

Figure 3.1: Changes of the rms value during the 100 ps of the MD/AVB simulation. Filled and opaque symbols show the rms deviation from the crystal and averaged thermalized structures, respectively. Circles refer to the main chain atoms (1616) and squares to heavy solute atoms (3146).

minimal, maximal, and mean values with standard deviations of the distances between atoms hydrogen bonded to this water molecule in the 100 ps MD/AVB simulation are shown below:

atoms	minimal distance [Å]	maximal distance [Å]	mean distance [Å]	stddev [Å]
H_2O 301:OW – Ile50:N	2.68	4.97	3.48	0.35
H_2O 301:OW – Ile150:N	2.56	4.07	2.96	0.22
H_2O 301:OW – Met[a]204:O	2.44	3.83	2.84	0.19
H_2O 301:OW – Ile202:O	2.45	4.81	2.95	0.31

[a]NLeu in the HIV-1 PR : MVT-101 complex

The changes of the Asp125:OD1 – Asp25:OD1 (Figure 1.3) distance were also traced in the simulation. This distance in all known crystal structures, native and with various inhibitors, varies between 2.6 and 3 Å. For example in the 3HVP native structure it equals 3.0 Å [13], in the complex with MVT-101 inhibitor [20] it reaches even 2.7 Å. In the studied case, the minimal distance between the "inner" oxygens is 2.93 Å and the maximal one 4.26 Å with the average value of 3.47 ± 0.18 Å over the 100 ps trajectory.

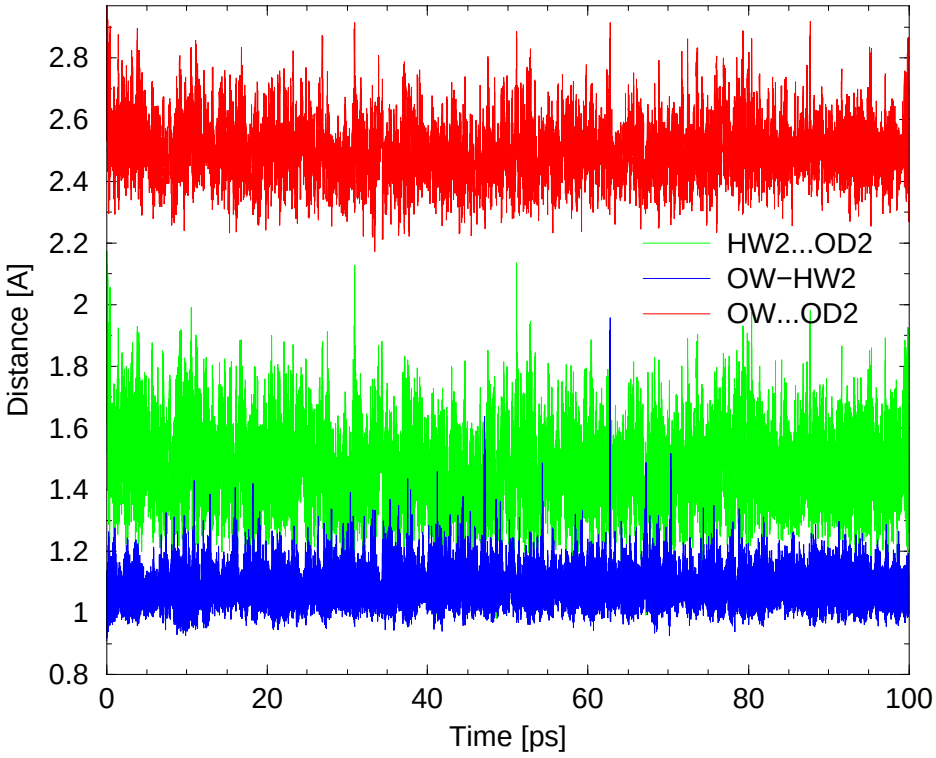

Figure 3.2: Distances between hydrogen bonded atoms of Asp125 (OD2) and the lytic water (OW and HW2) in the 100 ps of the MD/AVB simulation.

The short distance between the "inner" oxygens indicates that if these atoms do not share a proton (as in the studied case), they should form a hydrogen bond with a water molecule 207. Such hydrogen bond was found between Asp125:OD1 and HW1 of the lytic water with a mean distance of 3.47 ± 0.21 Å. However, this water molecule experiences large mobility also forming and breaking the hydrogen bonds with the other OD1 atom of aspartyl side chain (Asp25:OD1).

Also the planarity of the active site aspartates was reported during the simulation. Most probably, the almost coplanar conformation of the dyad is crucial for the enzymatic function and for the binding of a substrate or inhibitor. The minimal, maximal and mean values of the [Asp125:OD2 Asp125:OD1 Asp25:OD1 Asp25:OD2] dihedral angle are equal to -81, 30, and $-21 \pm 17°$, respectively. Thus on the average the dyad keeps its planarity in the simulations even though large fluctuations are observed.

The Asp25:Cγ–Asp125:Cγ distance was also recorded in the 100 ps simulation and its average value was equal 5.5 ± 0.2 Å.

A hydrogen bond was observed between Asp125:OD2 and Met204:N of the cleaved peptide bond. The average distance between these atoms in the 100 ps simulation equals 3.06 ± 0.25 Å with the values ranging from 2.30 to 4.85 Å.

The CA2...CA3 carbon minimal, maximal, and mean distances of the peptide bond carbons (see Figure 1.3) during the 100 ps simulation were found to be 3.4, 4.0 and 3.8 ± 0.1 Å, respectively.

The contributions of the eight valence structures were analysed during this simulation. Figure 3.3 shows the c_i^2 contribution coefficients of each structure at each time step of

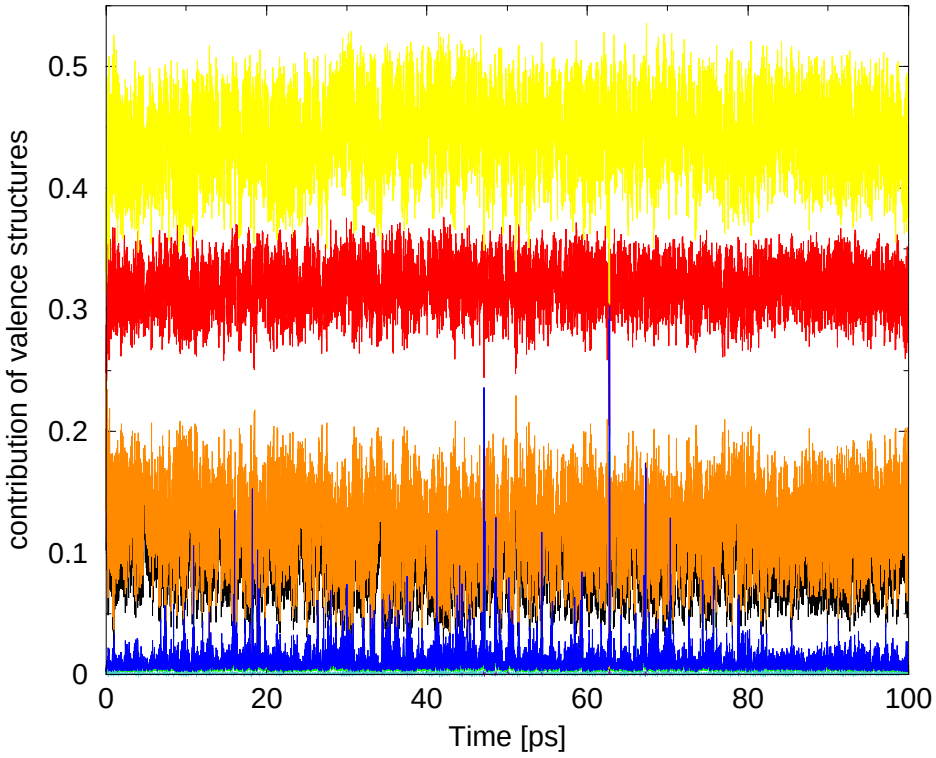

Figure 3.3: The plot shows the contributions (c_i^2) of valence structures (see text for colours) during the 100ps of the MD/AVB simulation.

the dynamics. The structures represent the model of Figure 2.1, are numbered as in Section 2.2.2, and coloured as denoted:

1. $A1_c$ + H_2O_c + $A2_g^-$ + S_c - black line
2. $A1_c$ + OH_g^- + H_w^+ + $A2_g^-$ + S_c - red line
3. $A1_c$ + OH_g^- + $A2_c$ + S_c - green line
4. $A1_c$ + H_2O_c + $A2_g^-$ + $S_c^\star$ - orange line
5. $A1_c$ + OH_g^- + H_w^+ + $A2_g^-$ + $S_c^\star$ - yellow line
6. $A1_c$ + OH_g^- + $A2_c$ + $S_c^\star$ - blue line
13. $A1_c$ + $A2_g^-$ + H_w^+ + $I1_c$ - turquoise line
14. $A1_c$ + $A2_c$ + $I1_c$ - violet line

In conventional molecular dynamics simulation, generally, only the first structure including whole molecular fragments would be accounted for. Herein, it can be seen that the molecular fragments of the enzyme active centre are preferentially in their ionized and polarized forms. Over the whole trajectory the main contribution comes from the 5^{th} structure that includes the ionized water molecule and the peptide bond with the C=O bond polarized. The first structure on the average gives only 10% contribution. The 2^{nd} and 4^{th} structures which give on average approximately 10 and 30% contribution, respectively, also include either an ionized water molecule or a polarized C=O bond.

In the next step eighteen valence structures of the first reaction model (see Section 2.2.2 were used allowing additionally for the proton transfer from the protonated Asp25 onto the peptide bond oxygen. A 100ps run was performed starting from the thermalized coordinates and velocities. The 50ps simulation yields an rms deviation value from the crystal structure of 1.82, 1.85, and 2.19 Å and from the averaged thermalized structure of 1.14, 1.14 and 1.52 Å for the trace, backbone and heavy solute atoms, respectively. After 100ps simulation these values are equal to 1.78, 1.79, and 2.24 Å for the crystal structure and 1.37, 1.37, and 1.78 Å for the averaged thermalized structure for the trace, backbone and heavy solute atoms, respectively. This rms deviations after 100ps of MD/AVB simulation, comparing to the ones with eight valence structures, are approximately 0.3 lower with respect to crystal structure and approximately 0.1 higher with respect to the averaged thermalized structure.

During this run the proton transfer also did not occur. Regarding the fluctuations of the distances they are of the same character as in the previous simulation. There is, however, a different contribution of the valence structures during this run. Due to one more possible ionized form, that of Asp25, the dominant structures are 8^{th} and 11^{th}, shown below, with the c_i^2 equal on the average 0.29 and 0.37, respectively.

8. $A2_g^- + H_w^+ + OH_g^- + S_c + H_a^+ + A1_g^-$
11. $A2_g^- + H_w^+ + OH_g^- + S_c^\star + H_a^+ + A1_g^-$

These dominant structures contain the same elements (i.e. the ionized or polarized molecular fragments) as in the previous calculations with eight valence structures. Additionally they include an ionized form of A1. There are also smaller contributions (approximately 8%) of structures 5^{th}, 9^{th} and 10^{th}.

5. $A1_c + OH_g^- + H_w^+ + A2_g^- + S_c^\star$
9. $OH_g^- + A2_c + S_c + H_a^+ + A1_g^-$
10. $H_2O_c + A2_g^- + S_c^\star + H_a^+ + A1_g^-$

3.4 Steered MD/AVB Simulations of the First Step of the Reaction

The reactions occurring in peptides and proteins happen most often on a time scale much larger than 1ns, i.e. the typical scale for present molecular dynamics simulations. The performance of 100ps of MD/AVB simulation was, of course, not sufficient for the reaction to occur by itself taking into account that the nucleophilic attack by water is usually the slow step of the cleavage process. Therefore, for an improved sampling of the configurational space a time-saving technique was applied which is based on a deformation of the potential energy surface in such a manner that the system is steered to cross the energy barrier. So, in the HIV-1 PR case, to gain the nucleophilic attack of the hydroxy anion in a shorter time scale than in reality external forces were applied to the oxygen nucleus OW of the lytic water molecule and to the peptide carbon nucleus C (see Figure 1.3). The distance between these nuclei served as a reaction coordinate (r). The force is a consequence of an external harmonic potential – $U = 1/2k(r - r_0 - vt)^2$ added to the Hamiltonian of the system (k being the stiffness of the harmonic spring). The atoms were driven towards each other along the C...OW line with the external force expressed as

$$\mathbf{F}_i = k(r - r_0 - vt)\frac{\delta r}{\delta \mathbf{r}_i} \tag{3.1}$$

where v is the velocity with which the harmonic spring is pulled, t is the time step in the MD/AVB simulations, and r_i is the position of C or OW atom. This force corresponds to the OW nuclei (more loosely atom) being pulled by a harmonic spring of stiffness k with its end moving with velocity v.

A 72ps steered MD/AVB simulations, i.e. 144000 steps with a time step equal to 0.5fs, were performed. The initial distance between the C and OW atoms was set to $r_0 = r = 3.15$ Å. The driving force constant of $k = 2000$ [kJ/mol·Å^2] was applied and the velocity of the harmonic spring of $v = -0.025$ Å/ps was set. The distance between C and OW atoms was changed up to 1.35 Å. The starting coordinates were taken after 50ps of molecular dynamics from the equilibrated ensemble. To keep the hydrogen bond between atoms Asp25:OD2 and Met203:O one-sided harmonic potential constraints (described in Section 3.2.2) of 500 [kJ/mol·Å^2] with a reference distance of 3 Å were applied. In the force driven simulation the first proton transfer occurred around the 52ps of the simulation and a transition state was formed (see Figure 3.4). During the transfer the OW…OD2 distance reaches its shortest value (around 2.4 Å) and when the RCO$^-$OHNHR (I1) molecular fragment is formed this distance grows to the value of 3.2 Å. R refers to the atoms of the model substrate.

Eight valence structures (1–6, 13, 14) were applied in the simulation of the first step of the reaction in the steered MD/AVB simulations. The structures represent the model of Figure 2.1, are numbered as in Section 2.2.2, and coloured as in Figure 3.5:

1. $A1_c$ + H_2O_c + $A2_g^-$ + S_c - black line
2. $A1_c$ + OH_g^- + H_w^+ + $A2_g^-$ + S_c - red line
3. $A1_c$ + OH_g^- + $A2_c$ + S_c - green line
4. $A1_c$ + H_2O_c + $A2_g^-$ + $S_c^\star$ - orange line
5. $A1_c$ + OH_g^- + H_w^+ + $A2_g^-$ + $S_c^\star$ - yellow line
6. $A1_c$ + OH_g^- + $A2_c$ + $S_c^\star$ - blue line
13. $A1_c$ + $A2_g^-$ + H_w^+ + $I1_c$ - turquoise line
14. $A1_c$ + $A2_c$ + $I1_c$ - violet line

The contribution of the structure coefficients changes during the process. In the beginning those valence structures which represent the system before the transition dominate, and in the final steps their contribution drops significantly as expected. This effect is presented in Figure 3.5.

Prior to the proton transfer the contributions of the 1^{st}, 2^{nd}, 4^{th}, and especially the 5^{th} structure are dominant in the AVB Hamiltonian. The 5^{th} structure contains a polarized C=O bond and a dissociated water molecule. Around the 50^{th}ps when OH^- starts to form a covalent bond with the peptide carbon, the contributions of the structures 1^{st}, 2^{nd} and 3^{rd} containing either a covalent form of H_2O or a covalent form of the peptide bond start to lower. The 5^{th} structure with all ionized and polarized forms grows to its maximal value (its c_i^2 reaches 0.7). The system is in a state where all the reactants are ionized and then other molecular fragments start to form. After the proton transfer, the nucleophilic attack, and the growth of the OD2…OW distance, the coefficients of the 1^{st}, 2^{nd} and 4^{th}

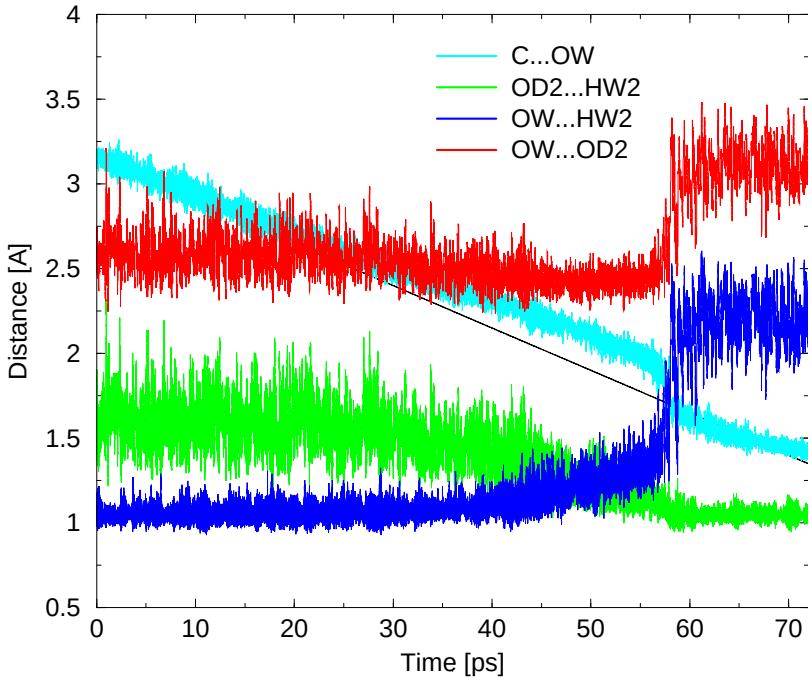

Figure 3.4: The plot shows the proton transfer from the catalytic water molecule to Asp125 in the force driven MD/AVB simulations. Dark blue, green, and red lines correspond to OW...HW2, OD2...HW2 and OW...OD2 distances, respectively. The black line shows the equilibrium length of the harmonic spring and the turquoise line the actual C...OW distance at each MD/AVB step.

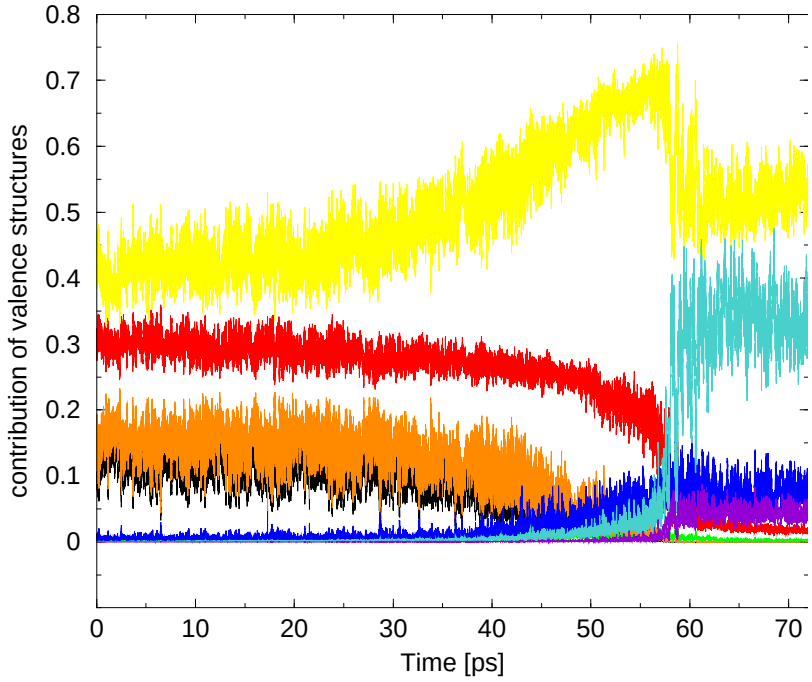

Figure 3.5: The contributions of the structure coefficients (c_i^2) at each time step of the force driven dynamics of the first step of the reaction (for colours see text).

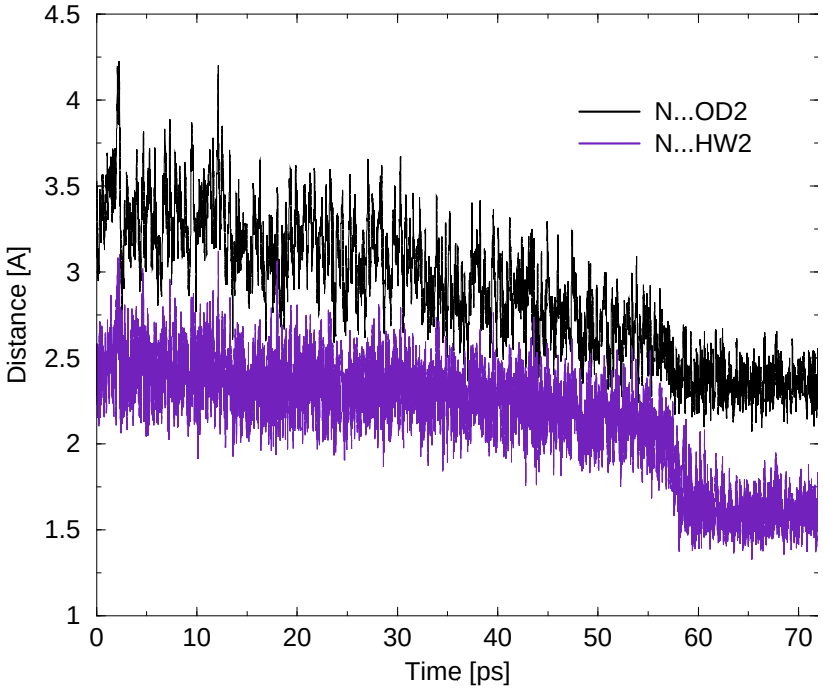

Figure 3.6: The plot shows the formation of the hydrogen bond between the peptide nitrogen and the OD2 atom of Asp125. Black and indigo lines correspond to N...OD2 and N...HW2 distances saved every 0.05ps of the 72ps simulation, respectively.

structures become negligible and the 13^{th} valence structure gains significant contribution. The c_i^2 of the 5^{th} valence structure drops to 0.35. Also, structures 6^{th} and 14^{th} yield higher contributions.

After the nucleophilic attack the I1 molecular fragment is formed with an sp^3 hybridization on the peptide carbon. The CA2...CA3 (see Figure 1.3) distance during and after the nucleophilic attack is shortened to an average of 3.2 Å. The [CA2 N C O] dihedral angle reaches approximately 45° and the planar angles around the C and N atoms gain the mean values as in the sp^3 hybridization.

Also a rotation of the C–N bond was postulated allowing the nitrogen lone pair to form a hydrogen bond with the protonated Asp125:OD2 atom. The distances between N...OD2 and N...HW2 atoms during the 72ps simulation are shown in Figure 3.6. It may be clearly seen that the hydrogen bond is formed around the 54^{th}ps of the molecular dynamics simulation.

Figure 3.7 shows the characteristic changes in the geometry of the active site atoms during the MD/AVB simulations.

To check if the proton transfer and the following nucleophilic attack did not occur accidentally, another simulation was performed from a different starting geometry taken after 42.3ps of the MD/AVB simulation presented in Section 3.3 and including eight valence structures. The character of Figure 3.4 and the geometric features of the transition state at equilibrium distance of C–OW (approximately 1.5 Å) are similar.

Also as a test case two calculations with different force constants were performed (1500 and 1000 [kJ/mol·Å^2]). The first reaction step occurred and the characteristic features of Figure 3.4 did not change qualitatively.

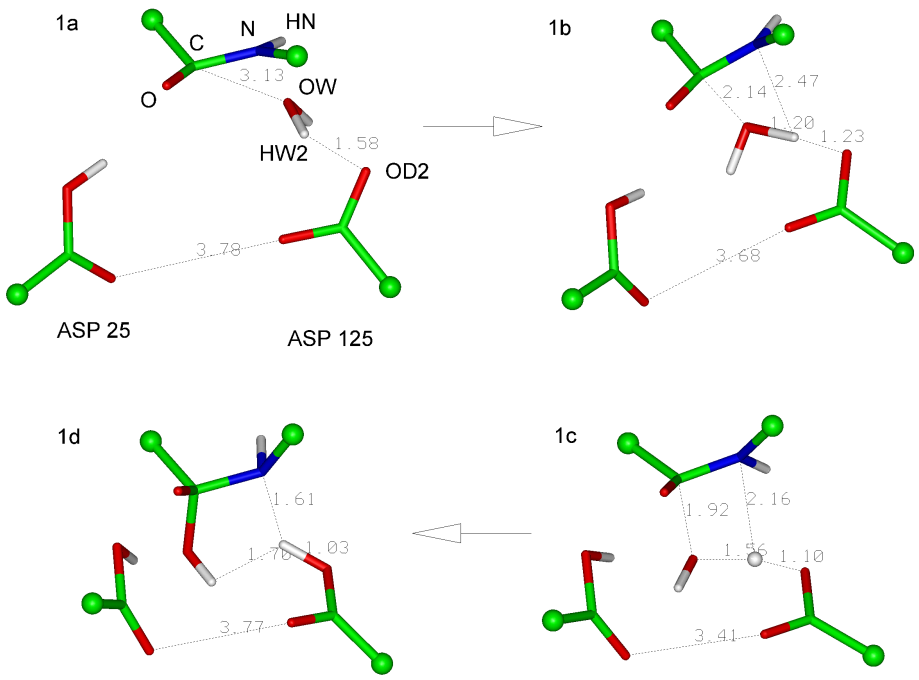

Figure 3.7: The plot shows four frames from the steered MD/AVB simulations of the HW2 proton transfer and the nucleophilic attack. Scheme 1a shows one of the starting geometries; 1b shows the position of the HW2 proton midway between OD2 and OW atoms; 1c shows the process of bond formation between the HW2 and OD2 atoms and between the hydroxy anion and the peptide carbon; 1d shows the geometry of the transition state just after the nucleophilic attack

Another calculation of the steered MD/AVB dynamics was carried out without any constraints on the Asp25:OD2 – Met203:O distance. This was performed to see if during the nucleophilic attack the above hydrogen bond is kept, what are the fluctuations of this distance, and therefore what is the mobility of Asp25. During this simulation the minimal and maximal distances between the oxygens were 3.0 and 5.4 Å, respectively, and the mean distance 4.3±0.3 Å. The Met203:O and Asp25:H distance was in the range 2.1 – 5.5 Å with an average value of 3.5 ± 0.4 Å. The hydrogen bond is therefore not stable over time and is created and broken during this simulation.

The above calculations were performed with eight valence structures. Next step involved the inclusion of the $RC(OH)_2NHR$ (I2), $RCO^-OHNH_2^+R$ (I2′) and $RC(OH)_2NH_2^+R$ (I3) molecular fragments in the description of the active site (R refers to atoms bonded to other fragments of the substrate). It resulted in the necessity of applying of twenty one valence structures altogether. This was done to check if other molecular structures are formed after or parallel to the nucleophilic attack and formation of RCO^-OHNHR (I1). The first nineteen out of the twenty one structures are the same as presented in Section 2.2.2 for the first model. Additional two structures include the I2′ molecular fragment (Figure 2.2) and are numbered 17′ and 18′:

17′. $A1_g^- + A2_g^- + H_a^+ + I2'_c$

18′. $A1_c + A2_g^- + I2'_c$

Another similar force driven simulation was performed with the starting geometry taken from the previous MD/AVB simulations (Section 3.3) with the C...OW starting distance of 2.95 Å. The driving force constant was set to 2000 [kJ/mol·Å^2]. The equilibrium length of the harmonic spring was shortened 0.0000125 Å every time step. 132000 steps, i.e. 66ps of the MD/AVB simulations were performed. No other constraints were applied. The changes of the distances for the proton transfer and the nucleophilic attack are shown in Figure 3.8 for the time steps of interest. The contributions of the structure coefficients (c_i^2) are presented in Figure 3.9. In the beginning of the process the 11^{th} structure gives main contribution. It consists of all possible ionic fragments that appear before the reaction. Just prior to the proton transfer its c_{11}^2 reaches the maximal value of around 0.6. Other structures present are 8^{th}, 5^{th}, 2^{nd}, and 10^{th} which also contain at least one of the ionic molecular fragments. After the proton transfer their contributions drop and c_i^2 of structures 15^{th}, $17'^{th}$, and 13^{th} increases. The 15^{th} structure includes the just formed I1 intermediate molecular fragment and ionized aspartates. Also the 13^{th} structure includes the I1 molecular fragment and one of the aspartates ionized. With twenty one structures after the formation of the I1 intermediate, another two proton transfers are possible (see Figure 2.1 and 2.2), one onto Met203:O and the other onto Met204:N of the peptide bond. It is questionable which molecular structure is formed first, I2 or I2', or if both are. The high contribution of the $17'^{th}$ structure suggests that geometry optimal for I2' molecular fragment is approached. The contributions of structures that include the I2 molecular fragment are negligible. Comparing this simulation with the previous one with only eight valence structures included, one may notice that also the structure containing all either ionized or polarized molecular fragments dominates. If other molecular fragments are included (especially I2'), Asp125 is not necessarily forming a strong covalent bond with the HW2 proton. The contributions of valence structures containing the protonated aspartate ($A2_c$) are not noticeable in the simulation. The acid most probably only accepts the HW2 atom for a short time and passes onto another proton acceptor (in this case the N atom).

3.5 Energy Checkpoint Between the AVB Parameterization and DFT Calculations

To check if the reaction site is represented reliably in the calculations and how well the parameterized potential energy surface describes the AVB region, additional test DFT calculations were performed. Ten randomly chosen geometries characterizing the system before and after the first proton transfer and the nucleophilic attack were taken. Only the coordinates of atoms presented in Figure 1.3 were extracted. Their energies in vacuum were calculated with the standalone version of the AVB program and averaged. Next, the difference between these averages was calculated. The same difference was obtained with the DFT calculations using the nonlocal hybrid functional B3LYP [90, 91] and 6-31G+(d,p) basis set implemented the Gaussian'94 program [89]. For the DFT calculations explicit hydrogen atoms are needed and they were added for the terminal carbons of the extracted atoms to form CH_3 groups. The difference between the obtained average energies is 117 ± 8 kJ and 145 ± 16 kJ for the AVB and the DFT calculations, respectively. The geometries following the first reaction step had the peptide bond deformed and were not optimized for the energy calculations. This is probably one of the reasons why DFT yields higher difference. The parameters which were subject to fit in the AVB method were obtained based on the DFT calculations for one or two molecular fragments at a time. In

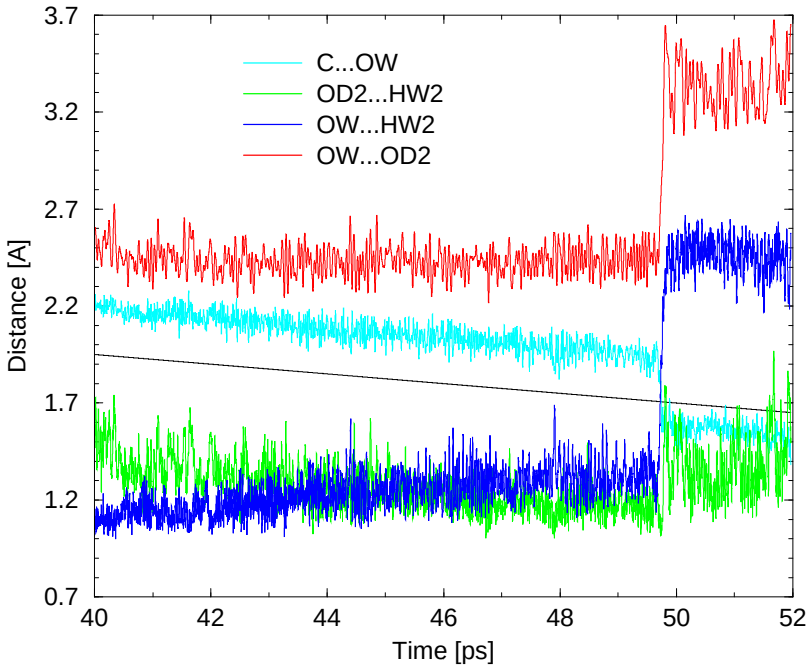

Figure 3.8: The plot shows a fragment of the dynamics of the HW2 proton transfer process in the steered MD/AVB simulations. Dark blue, green, and red lines present the OW...HW2, OD2...HW2 and OW...OD2 distances, respectively. The black line shows the equilibrium length of the harmonic spring and the turquoise line the actual C...OW distance at each MD step.

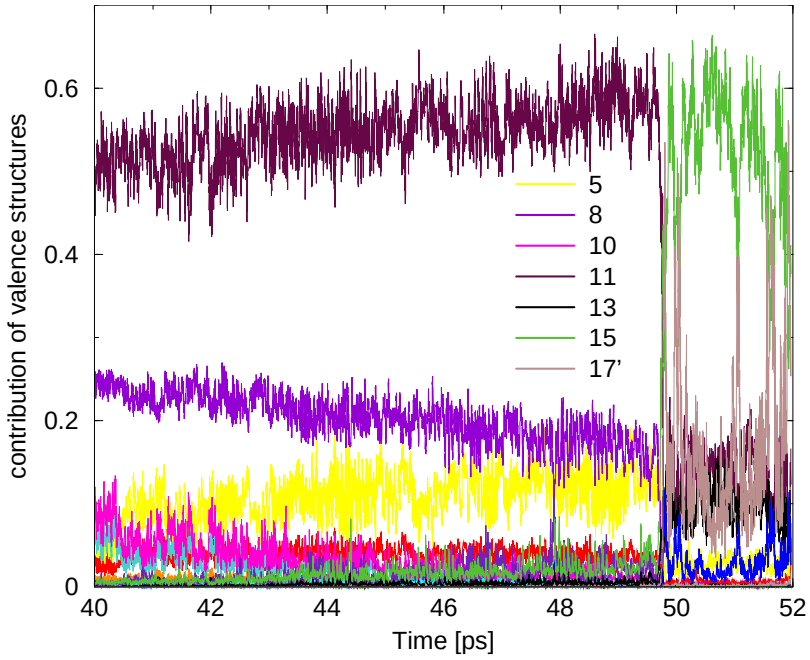

Figure 3.9: The contributions of the valence bond structures (c_i^2) in the fragment of the force driven MD/AVB simulation. The legend shows the numbers (see text) of the significant structures and their colours.

these energy difference calculations the whole set of molecular fragments was taken. So the agreement is acceptable for the purpose of the MD/AVB calculations, and it may be assumed that the parameterization reproduces the DFT calculations for the whole AVB region to a satisfactory degree.

3.6 MD/AVB Simulations Following the Nucleophilic Attack

After the steered MD/AVB simulations of the proton transfer and the nucleophilic attack a further run was carried out with no additional force and no constraints applied. The aim was to check if the obtained transition state geometry is not deformed, locally stable, and the reaction does not go back immediately. The geometry with the C...OW distance close to equilibrium was chosen. As a continuation of the previous steered MD/AVB a 30ps run of the whole system was performed with no constraints and using the combination of eight valence structures presented in Section 3.4. During such simulation the system conserves the geometry of the transition state. The mean C–OW distance was 1.51 ± 0.05 Å what is very close to the equilibrium distance of this bond in the $RCO^{-}OHNHR$ (R=CH_3) molecular fragment obtained with DFT calculations (see Chapter II and Table 2.1). The OD2...HW2, OW...HW2, OW...OD2 mean distances were 1.06 ± 0.03 Å, 2.3 ± 0.1 Å, and 3.2 ± 0.1 Å, respectively, showing that after the transfer the HW2 proton was still bonded to Asp125:OD2 atom. The main contribution belongs to the 5^{th} and the 13^{th} structure, their c_i^2 mean value was 0.43 ± 0.02 and 0.45 ± 0.03, respectively. Structures 6^{th} and 14^{th} gave also some small contributions. It is worth noting that the contributions of the first four structures, describing the state of the system before this reaction step were negligible.

3.7 MD/AVB Simulations of Further Steps of the Reaction

After the formation of the I1 molecular fragment there may be a few possible paths which lead to the breakage of the C–N bond. Three of them are accounted for in this study (see Figure 2.1 and 2.2).

3.7.1 Proton Transfer from Asp125 onto Peptide Nitrogen

First, a test was performed to check which molecular intermediate is formed following I1 and prior to I3. This is a continuation of a test performed in the steered MD/AVB simulation with twenty one structures "switched on". There are three possibilities: either a proton transfer from Asp25 onto Met203:O may occur forming the I2 molecular fragment or a proton transfer from Asp125 onto Met204:N forming I2', or both may happen simultaneously. The MD/AVB simulation was performed including molecular fragments: A1, $A1^{-}$, H_a^{+}, A2, $A2^{-}$, H_w^{+}, I1, I2, I2', and I3 (see Figures 2.1 and 2.2) which formed nine structures altogether. The structures are numbered as in Section 2.2.2 and the important ones during this simulation are coloured.

13. $A1_c + A2_g^{-} + H_w^{+} + I1_c$
14. $A1_c + A2_c + I1_c$
15. $A2_g^{-} + H_w^{+} + H_a^{+} + A1_g^{-} + I1_c$
16. $A2_c + H_a^{+} + A1_g^{-} + I1_c$

17. $A2_g^- + H_w^+ + A1_g^- + I2_c$

17′. $A2_g^- + H_a^+ + A1_g^- + I2'_c$

18. $A2_c + A1_g^- + I2_c$

18′. $A1_c + A2_g^- + I2'_c$

19. $A2_g^- + A1_g^- + I3_c$

The starting geometry was one of the geometries after the nucleophilic attack. The coordinates and velocities were taken after the 65ps of the steered MD/AVB simulations where eight valence structures were used. The starting C–OW bond distance was equal to 1.52 Å. A 50ps MD/AVB simulation with a time step of 0.5fs was performed with no constraints applied. A strong hydrogen bond was observed between the Met204:N and Asp125:OD2 atoms. The appropriate distances are shown in Figure 3.10.

In the simulations strong fluctuations of the N...OD2 and OD2...HW2 distances were observed. The HW2 proton quickly passes from the Asp125:OD2 atom and keeps a short distance, of a mean value of 1.08 ± 0.09, to the nitrogen atom. This happens due to the possibility of forming the I2′ molecular fragment after the inclusion of structures with this fragment (17′ and 18′). This could also be seen in the previous steered MD/AVB simulations with twenty one valence bond structures (see 3.4 and Figure 3.9) where the coefficients of structures including the I2′ molecular fragment were visible after the nucleophilic attack. After 23ps of the simulation while the hydrogen bond between the OD2 and N atoms was weakened and this distance grew significantly, the HW2 proton bonded covalently to nitrogen stayed close to this atom.

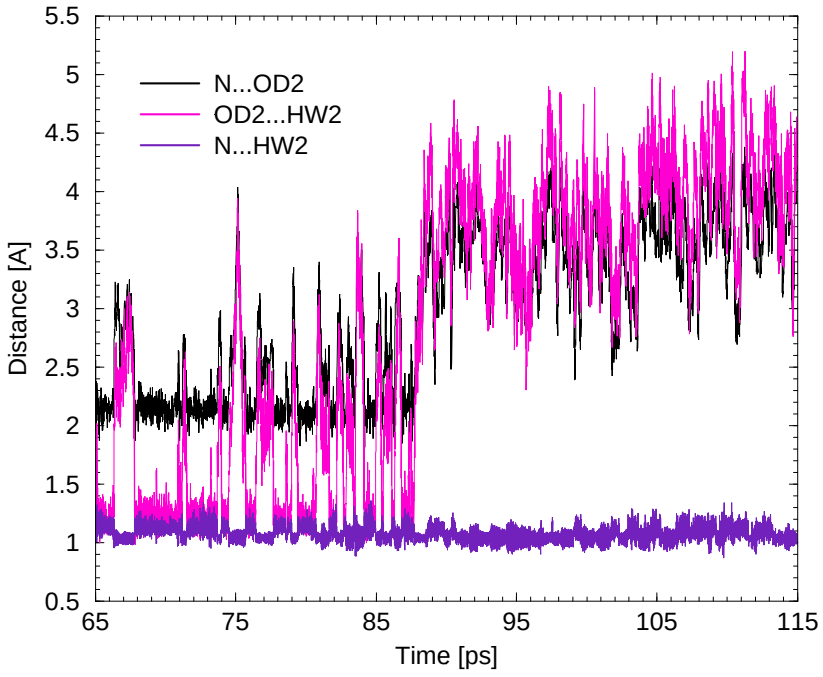

Figure 3.10: The plot shows the distances between the N...OD2, OD2...HW2, and N...HW2 atoms in the 50ps MD/AVB simulation following the nucleophilic attack where the I2′ molecular fragment is formed (see Figure 2.2).

Considering the valence structures, the main contribution is given by the 15^{th} and $17'^{th}$ structure. Again, the 15^{th} structure consists of ionized forms of aspartates. It confirms previous suggestions that the protons in the active centre prefer to be delocalized rather than to form covalent bonds. The I2′ molecular fragment was formed what could also be clearly seen while studying the geometries appearing in the trajectory. The contributions of structures including the I2 molecular fragment are negligible. Thus, it is rather preferential for the transition state to form a fragment similar to I2′. On the other hand, this molecule was not stable according to the DFT calculations. However, one has to bear in mind that I2′ in the protein is a part of a bigger fragment and must be considered as such. Also one must not neglect the influence of the protein environment as it can allow for stable geometries that are not such in vacuum.

Also the mobility of the C–N distance was traced during this simulation. It varies between 1.42 and 1.90 Å with a mean value of 1.61 ± 0.07 even though the breakage of that bond was not accounted for in the construction of the AVB structures for this simulation.

3.7.2 Inclusion of the C–N Bond Breakage

In further calculations the I2 molecular fragment was left out due to the negligible contribution of structures including it as observed in the previous simulations. An attempt was made to simulate the rupture of the C–N bond in the reaction path presented in Figure 2.2 and including the I1, I2′ and I3 intermediates. This simulation took into account the presence of 13 molecular fragments in the active site: A1, $A1^-$, H_a^+, A2, $A2^-$, H_w^+, I1, I2′, I3, $P1^*$, P1, P2, and P3 (see Figures 2.2) which formed eleven valence structures. Only a 15 ps MD/AVB simulation with these structures was carried out and further studies on that case are being performed. The starting geometry was the one just after the nucleophilic attack with the C–OW distance equal to 1.52 Å. No constraints were applied. In this simulation dominant structures were: 13^{th}, 15^{th}, and $17'^{th}$ numbered as in Section 2.2.2. However, the I3 molecular fragment has not yet been formed, so the breakage of the C–N bond could therefore not happen. The C–N distance in this short simulation grows as far as 1.78 Å with a mean value of 1.6 ± 0.1 Å.

In the next step the last possibility was accounted for which allowed for the breakage of I2′ molecular fragment. This corresponds to the right path of Figure 2.2. Two 50 ps MD/AVB simulations were performed starting from the geometry with covalent bonds as in I1, i.e. following the nucleophilic attack. One simulation was carried out without any constraints in the active site, the other kept the Asp25:OD2 and Met203:O atoms with one-sided harmonic potential and a force constant equal 500 [kJ/mol·Å^2]. This two processes did not lead to the C–N bond rupture even though the fluctuations of that bond reached as far as 1.8 Å. This process also needs further investigation.

3.8 Concluding Remarks

The current state of the MD/AVB simulations of the HIV-1 PR complexed with a model substrate was presented in this Chapter. Some calculations are still being performed. The parameterization of the AVB method presented in Chapter II works well within the MD/AVB model. This allows for simulation of the reactions occurring in the enzyme active centre.

The simulations show that in the active site the most preferable energetically and stabilized are the structures consisting of ionized or polarized molecular fragments. They

are not accounted for in conventional MD simulations while modelling, e.g. the Michaelis complex. Such MD simulations encounter also other difficulties in modelling the active site, and to avoid them approximate charge distributions are often chosen [16], geometric constraints are applied between the carboxylate moieties [96] or proton is positioned midway between the aspartyl groups [98]. One should also note that conventional MD methods do not model hydrogen bonds precisely beacuse most force fields do not include specific parameters for strong hydrogen bonding.

The present study shows that the proton transfer process and the nucleophilic attack are most probably concerted. The hydroxy anion is formed while the HW2 atom is accepted by the OD2 oxygen of Asp125. This happens when the OW of the water molecule approaches C at a distance of around 1.8 Å. During that process the hybridization of the peptide bond carbon is changed from sp^2 to sp^3. In most biochemical descriptions this process is divided into several steps.

According to the simulations, after the proton transfer and the nucleophilic attack processes, the $RCO^{-}OHNH_2^{+}R$ (I2') molecular fragment is preferentially formed (R denotes carbon atoms bonded to other fragments of the substrate). After the sp^3 hybridization is achieved by the intermediate, the HW2 atom is transferred in a short time scale onto the peptide bond nitrogen. However, there still remains the question of whether the C–N bond is ruptured with such protonation state of the reaction intermediate or another proton transfer onto the negatively charged oxygen of I2' is to occur.

It has also been noticed that with the inclusion of the water molecule in the active centre the hydrogen bond between Asp25:OD2 and Met203:O of the cleaved peptide bond is not so strong and is being formed and broken during the dynamics. Following the nucleophilic attack the proton transfer from Asp25:OD2 onto O does not occur. This result is in disagreement with previous studies [99], but they were performed with a different crystal structure and with a different model substrate.

On the basis of this theoretical model some other mechanisms may also be tested, e.g. ones that do not involve the lytic water molecule and a nucleophilic attack is performed directly by one of the aspartic acids. Only in the native X-ray complexes a distinct electron density of this water is observed. It may also be the H_3O^{+} ion. In the complexes with inhibitors this density has not yet been found. There may be few reasons for this fact. The HIV-1 PR inhibitors, being competitive, bind to the enzyme like protein chains do. Most of them mimic the geometry of the substrate's transition state analogue and contain e.g. hydroxyl groups which displace water from the active site interacting strongly with the aspartates. No X-ray complexes with substrates have been obtained due to the fact that they are cleaved before they are crystallized or solved. So it is still questionable if the mechanism with or without the water molecule is the right one. With small modifications of the AVB parameterization this model excluding the lytic water molecule may also be tested and preparations are done to do so.

Another test, yet to be done, is to perform the simulations including all the molecular fragments presented in Figures 2.2 and 2.2 in one MD/AVB dynamics. Such study would be more time consuming but would allow to deduce if the reaction occurs in sequential steps as in classical description of biochemical processes or is a continuous process involving contributions of all the valence structures all the time.

Summary

This thesis presents the computational modelling study of the protonation equilibria and the reaction mechanism of HIV-1 protease.

The first part concentrated on elucidating the protonation states of titratable residues in the free enzyme and in complexes with four different inhibitors. The ionization states were also studied for the protein complexed with a model substrate, its sequence being based on a real peptide cleaved by HIV-1 PR. The results show that the crucial aspartates in the active site of the enzyme may have different protonation states depending on the type of the bound ligand. The computed pK_as were also compared with the available experimental data. Two titration models were applied with two different dielectric constants. Their quality was discussed. Treatment of proton positions has been improved allowing the residues to adopt either of two neutral forms at each protonatable atom in the residue. To do so the modifications of the titration procedure have been proposed and applied.

In the second Chapter the bases of the AVB method were presented and a new approach to the formalism was proposed. The method was applied to HIV-1 PR what needed modifications of the procedure that was used previously with a different enzyme [3]. The DFT calculations leading to the extended parameterization of the AVB method were described and the full set of parameters obtained for the active site region was presented. The refined AVB method together with Gromos'96 package (MD/AVB molecular dynamics) allowed to simulate the reorganizations of chemical bonds in the active site. The calculations were too short for the whole reaction to take place since most biochemical enzymatic reactions occur on a time scale much larger than typical for MD simulations. Therefore, the steered MD/AVB calculations of the first step of the reaction were carried out. This simulation allowed the system to cross the energy barrier on a shorter time scale and the proton transfer and the nucleophilic attack could be observed. Simulations of the further steps of the reaction showed that one of the aspartic acids donates its proton to the nitrogen of the just formed reaction intermediate. This process follows the nucleophilic attack. Studies of the next steps of the reaction, especially the C—N bond rupture, require further simulations. Some of them are being performed.

Bibliography

[1] L. J. Hyland, T. A. Tomaszek Jr., 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.

[2] A. M. Silva, R. E. Cachau, H. L. Sham, and J. W. Erickson. Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.*, 255:321–346, 1996.

[3] P. Grochowski, B. Lesyng, P. Ba la, and J. A. McCammon. Density functional based parametrization of a valence bond method and its applications in quantum-classical molecular dynamics simulations of enzymatic reactions. *International Journal of Quantum Chemistry*, 60:1143–1164, 1996.

[4] A. Warshel. *Journal of American Chemical Society*, 102:6218, 1980.

[5] P. Ba la, P. Grochowski, B. Lesyng, and J. A. McCammon. Quantum dynamics of proton transfer processes in enzymatic reactions. Simulations of phospholipase A_2. *Phys. Chem.*, 102:580–586, 1998.

[6] P. Ba la, P. Grochowski, K. Nowiński, B. Lesyng, and J. A. McCammon. Quantum-dynamical picture of a multi-step enzymatic process. Reaction catalyzed by phospholipase A_2. *Biophysical Journal*, 79:1253–1262, 2000.

[7] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Hochschulverlag AG an der ETH Zurich and BIOMOS b.v., Zurich, Groningen, 1996.

[8] M. Geller, **J. Trylska**, and J. Antosiewicz. HIV-1 protease and its inhibitors. In *Theoretical and Computational Methods in Genome Research*, pages 237–254. Sandor Suhai, German Cancer Research Centre, Heidelberg, Germany, 1997.

[9] **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 Science*, 8:180–195, 1999.

[10] R. Bogacewicz, **J. Trylska**, and M. Geller. Ligand design package (Ludi-MSI) applied to known inhibitors of the HIV-1 protease. Test of performance. *Acta Poloniae Pharmaceutica*, 57:25–28, Nov. 2000.

[11] **J. Trylska**, P. Grochowski, and M. Geller. Parameterization of the Approximate Valence Bond (AVB) method to describe potential energy surface in the reaction catalyzed by HIV-1 protease. *International Journal of Quantum Chemistry*, 82:86–103, 2001.

[12] N. E. Kohl, E. A. Emini, W. A. Schleif, L. J. Davis, J. C. Heimbach, R. A. Dixon, E. M. Scolnick, and I. S. Sigal. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci., USA*, 85:4686–4690, 1988.

[13] A. Wlodawer, M. Miller, M. Jaskolski, B. K. Snthyanarayana, E. Baldwin, I. T. Weber, L. M. Selak, L. Clawson, J. Schneider, and S. B. H. Kent. Conserved folding in retroviral proteases. Crystal structure of a synthetic HIV-1 protease. *Science*, 245:616–621, 1989.

[14] M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. Kent, and A. Wlodawer. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Angstrom resolution. *Science*, 246:1149–1152, 1989.

[15] L. J. Hyland, T. A. Tomaszek Jr., G. D. Roberts, S. A. Carr, V. W. Magaard, H. L. Bryan, S. A. Fakhoury, M. L. Moore, M. D. Minnich, J. S. Culp, R. L. DesJarlais, and T. D. Meek. Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by ^{18}O isotope exchange. *Biochemistry*, 30:8441–8453, 1991.

[16] D. C. Chatfield and B. R. Brooks. HIV-1 protease cleavage mechanism elucidated with molecular dynamics simulation. *J. Am. Chem. Soc.*, 117:5561–5572, 1995.

[17] D. C. Chatfield, K. P. Eurenius, and B. R. Brooks. HIV-1 protease cleavage mechanism: A theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential. *Journal of Molecular Structure*, 423:79–92, 1998.

[18] 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.

[19] 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.

[20] M. Miller, M. Geller, S. M. Swanson, and J. Maizel. Analysis of the structure of chemically synthesized HIV-1 protease complexed with a hexapeptide inhibitor. Part I: Crystallographic refinement of 2 Angstrom data. *Proteins: Struct. Funct. Genet.*, 27:184–194, 1997.

[21] S. Kageyama, T. Mimoto, Y. Murakawa, M. Nomizu, H. Ford, T. Shirasaka, S. Gulnik, J. Erickson, K. Takada, H. Hayashi, S. Broder, Y. Kiso, and H. Mitsuya. In vitro anti-human-immunodeficiency-virus (HIV) activities of transition-state mimetic HIV protease inhibitors containing allophenylnorstatine. *Antimicrob. Agents Ch.*, 37:810–817, 1993.

[22] S. Erickson-Viitanen, R. M. Klabe, P. G. Cawood, P. L. O'Neal, and J. L. Meek. Potency and selectivity of inhibition of Human Immunodeficiency Virus protease by a small nonpeptide cyclic urea, DMP 323. *Antimicrob. Agents Chemother*, 38:1628–1634, 1994.

[23] P. Y. S. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. Rayner, Y. N. Wong, C.-H. Chang, P. C. Weber, D. A. Jackson, T. R.

Sharpe, and S. Erickson-Viitanen. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, 263:380–384, 1994.

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

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

[26] J. Antosiewicz, J. A. McCammon, and M. K. Gilson. The determinants of pK_as in proteins. *Biochemistry*, 35:7819–7833, 1996.

[27] A. M. Briggs and J. Antosiewicz. Simulation of pH-dependent properties of proteins. In K. B. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 13, pages 249–311. Wiley–VCH, New York, USA, 1999.

[28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[29] D. Poland. *Cooperative equilibria in physical biochemistry*. Clarendon Press, Oxford, 1978.

[30] C. J. F. Böttcher. *Theory of electric polarization*, volume I. Elsevier, Amsterdam, 1973.

[31] A. J. M. Garrett and L. Poladian. Refined derivation, exact solutions, and singular limits of the Poisson-Boltzmann equation. *Ann. Phys.*, 188:386–435, 1988.

[32] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules. Theory and applications. *Ann. Rev. Biophys. Chem.*, 19:301–332, 1990.

[33] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.*, 94:7684–7692, 1990.

[34] H.-X. Zhou. Macromolecular electrostatic energy within the nonlinear Poisson-Boltzmann equation. *J. Chem. Phys.*, 100:3152–3162, 1994.

[35] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.*, 62:187, 1991.

[36] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.*, 91:57–95, 1995.

[37] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.*, 157:671–679, 1982.

[38] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu, Zn superoxide dismutase. *Proteins: Struct. Func. Gen.*, 1:47–79, 1986.

[39] M. E. Davis and J. A. McCammon. Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradients methods. *J. Comp. Chem.*, 10:386–394, 1989.

[40] A. Nicholls and B. Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.*, 12:435–445, 1991.

[41] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[42] Molecular Simulations Inc. Waltham, MA. CHARMm Version 22. Polar hydrogen parameter set for CHARMm Version 22. Waltham, MA. 1992.

[43] 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.*, 110:1657–1666, 1988.

[44] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.*, 98:1978–1988, 1994.

[45] A. T. Brunger and M. Karplus. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins: Struct. Func. and Gen.*, 4:148–156, 1988.

[46] F. M. Richards. Areas, volumes, packing and protein structure. *Ann. Rev. Biophys. Bioeng.*, 6:151–176, 1977.

[47] M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.*, 9:327–335, 1988.

[48] Y. Nozaki and C. Tanford. Examination of titration behavior. *Methods Enzymol.*, 11:715–734, 1967.

[49] L. Stryer. *Biochemistry*. 2nd Ed., W. H. Freeman and Co., 1981.

[50] A. S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. On the calculation of pK_as in proteins. *Proteins: Struct. Func. Gen.*, 15:252–265, 1993.

[51] J. Antosiewicz and D. Porschke. The nature of protein dipole moments: Experimental and calculated permanent dipole of α-chymotrypsin. *Biochemistry*, 28:10072–10078, 1989.

[52] J. Antosiewicz and D. Porschke. Electrostatics of hemoglobins from measurements of the electric dichroism and computer simulations. *Biophys. J.*, 68:655–664, 1995.

[53] J. Antosiewicz. Computation of the dipole moments of proteins. *Biophys. J.*, 69:1344–1354, 1995.

[54] M. K. Gilson. Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins: Struct. Funct. Genet.*, 15:266–282, 1993.

[55] 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.

[56] J. R. Knowles. The intrinsic pK_a-values of functional groups in enzymes: improper deductions from the pH-dependence of steady-state parameters. *CRC Crit. Rev. Biochem.*, 4:165–173, 1976.

[57] K. Brocklehurst. A sound basis for pH-dependent kinetic studies on enzymes. *Protein Engineering.*, 7:291–299, 1994.

[58] S. Kuramitsu and K. Hamaguchi. Analysis of the acid-base titration curve of hen lysozyme. *J. Biochem.*, 87:1215–1219, 1980.

[59] O. Jardetzky and G. C. K. Roberts. *NMR in Molecular Biology*. Academic Press, New York, 1981.

[60] K. Wüthrich. *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, Inc., New York, 1986.

[61] K. Bartik, C. Redfield, and C. M. Dobson. Measurement of the individual pK_a values of acidic residues of hen and turkey lysozymes by two-dimensional ^{1}H NMR. *Biophysical Journal*, 66:1180–1184, 1994.

[62] J. A. Schellman. Macromolecular binding. *Biopolymers*, 14:999–1018, 1975.

[63] D. Bashford and M. Karplus. pK_as of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry*, 29:10219–10225, 1990.

[64] M. K. Gilson and B. H. Honig. The dielectric constant of a folded protein. *Biopolymers*, 25:2097–2119, 1986.

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

[66] E. G. Alexov and M. R. Gunner. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.*, 72:2075–2093, 1997.

[67] D. Bashford, D. A. Case, C. Dalvit, L. Tennant, and P. E. Wright. Electrostatic calculations of side-chain pK_a values in myoglobin and comparison with NMR data for histidines. *Biochemistry*, 32:8045–8056, 1993.

[68] D. Khare, P. Alexander, J. Antosiewicz, P. Bryan, M. Gilson, and J. Orban. pK_a measurements from nuclear magnetic resonance for B1 and B2 immunoglobulin G-binding domains of protein G: Comparison with calculated values for nuclear magnetic resonance and x-ray structures. *Biochemistry*, 36:3580–3589, 1997.

[69] F. C. Bernstein, T. F. Koetzle, T. F. Williams, G. J. B. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.

[70] S. Spinelli, Q. Z. Liu, P. M. Alzari, P. H. Hirel, and R. J. Poljak. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie*, 73:1391, 1991.

[71] 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.*, 39:3514–3525, 1996.

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

[73] M. Geller, M. Miller, S. M. Swanson, and J. Maizel. Analysis of the structure of HIV-1 protease complexed with a hexapeptide inhibitor. Part II: Molecular dynamics studies of the active site region. *Proteins: Struct. Funct. Genet.*, 27:195–203, 1997.

[74] Y.-X. Wang, D. I. Freedberg, T. Yamazaki, P. T. Wingfield, S. J. Stahl, J. 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.

[75] M. K. Gilson. Modeling protonation equilibria in biomolecules. In W. F. Van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors, *Computer simulations of biomolecular systems*, volume 3, pages 199–222. Kluwer/Escom, Dordrecht, 1997.

[76] T. Yamazaki, L. K. Nicholson, D. A. Torchia, P. Wingfield, S. J. Stahl, J. D. Kaufman, C. J. Eyermann, 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 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.

[77] Y.-X. Wang, D. I. Freedberg, T. Yamazaki, P. T. Wingfield, S. J. Stahl, J. 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.

[78] Robert C. Weast, editor. *Handbook of Chemistry and Physics*. CRC Press, 1975.

[79] W. E. Harte, Jr. and D. L. Beveridge. Prediction of the protonation state of the active site aspartyl residues in HIV-1 protease-inhibitor complexes via molecular dynamics simulation. *J. Am. Chem. Soc.*, 115:3883–3886, 1993.

[80] M. L. Moore, W. M. Bryan, S. A. Fakhoury, V. W. Magaard, W. F. Huffman, B. D. Dayton, T. D. Meek, L. Hyland, G. B. Dreyer, B. W. Metcalf, J. E. Strickler, J. Gorniak, and C. Debouck. Peptide-substrates and inhibitors of the HIV-1 protease. *Biochem. Biophys. Res. Commun.*, 159:420–425, 1989.

[81] T. Simonson, D. Perahia, and A. T. Brunger. Microscopic theory of the dielectric properties of proteins. *Biophys. J.*, 59:670–690, 1991.

[82] T. Simonson and D. Perahia. Internal and interfacial dielectric properties of cytochrome C from molecular dynamics in aqueous solution. *Proc. Natl. Acad. Sci. USA*, 92:1082–1086, 1995.

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

[84] D. Bashford and K. Gerwert. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.*, 224(2):473–486, 1992.

[85] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.*, 246:122–129, 1995.

[86] D. Qiu, P. S. Shenkin, F. P. Hollinger, and Still W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate born radii. *J. Phys. Chem.*, 101:3005–3014, 1997.

[87] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[88] R. Luo, M. S. Head, and M. K. Gilson. pK_a shifts in small molecules and HIV protease: Electrostatics and conformation. *J. Am. Chem. Soc.*, 120:6138–6146, 1998.

[89] M. J. Frish et al. *Gaussian 94, Revision E.2*. Gaussian, Inc., Pittsburgh, PA, 1995.

[90] A. D. Becke. Density functional termochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[91] R. G. Parr C. Lee, W. Yang. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37:785–789, 1988.

[92] K. R. Liedl T. Loerting. Toward elimination of discrepancies between theory and experiment: Double proton transfer in dimers of carboxylic acids. *J. Am. Chem. Soc.*, 120:12595–12600, 1998.

[93] W.-G Han and S. Suhai. Density functional studies on N-methylacetamide-water complexes. *J. Phys. Chem.*, 100:3942–3949, 1996.

[94] Y. Pan and M. A. McAllister. Characterization of low-barrier hydrogen bonds 4. Basis set and correlation effects: an *ab initio* and DFT investigation. *Journal of Molecular Structure (Theochem)*, 427:221–227, 1998.

[95] P. Burk, I. A. Koppel, I. Koppel, I. Leito, and O. Travnikova. Critical test of performance of b3lyp functional for prediction of gas-phase acidities and basicities. *Chem. Phys. Letters*, 323:482–489, 2000.

[96] W. F. van Gunsteren H. Liu, F. Müller-Plathe. A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. *J. Mol. Biol.*, 261:454–469, 1996.

[97] P. E. Smith and W. F. van Gunsteren. Consistent dielectric properties of the simple point charge and extended simple point charge water models at 277 and 300K. *J. Chem. Phys.*, 100:3169–3174, 1994.

[98] I. T. Weber and R. W. Harrison. Molecular dynamics simulations of HIV-1 protease with peptide substrate. *Protein Engineering*, 7:1353–1363, 1994.

[99] W. F. van Gunsteren, H. Liu, and F. Müller-Plathe. The elucidation of enzymatic reaction mechanisms by computer simulation: Human immunodeficiency virus protease catalysis. *Journal of Molecular Structure (Theochem)*, 432:9–14, 1998.