COMPUTATIONAL APPROACHES TO BIOLOGICAL AND SUPRAMOLECULAR SYSTEMS

COMPUTATIONAL APPROACHES TO PROTONATION AND DEPROTONATION REACTIONS FOR BIOLOGICAL MACROMOLECULES AND SUPRAMOLECULAR COMPLEXES

By

AHMED MOHAMMED

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements for the Degree

Master of Science

McMaster University

Master of Science (2013) McMaster University

(Chemistry & Chemical Biology) Hamilton, Ontario, Canada

TITLE: Computational Approaches to Protonation and Deprotonation Reactions for Biological Macromolecules and Supramolecular Complexes

AUTHOR: Ahmed Mohammed M.Sc. Physical Chemistry, Inha University

B.Sc. Chemistry, Assiut University

SUPERVISOR: Professor Paul W. Ayers

NUMBER OF PAGES: x, 74

ABSTRACT

Understanding and predicting chemical phenomena is the main goal of computational chemistry. In this thesis I present my work on applying computational approaches to study chemical processes in biological and supramolecular systems.

pH-responsive molecular tweezers have been proposed as an approach for targeting drug-delivery to tumors, which tend to have a lower pH than normal cells. In chapter 2 I present a computational study I performed on a pH-responsive molecular tweezer using *ab initio* quantum chemistry in the gas phase and molecular dynamics simulations in solution. The binding free energy in solution was calculated using Steered Molecular Dynamics. We observe, in atomistic detail, the pH-induced conformational switch of the tweezer and the resulting release of the drug molecule. Even when the tweezer opens, the drug molecule remains near a hydrophobic arm of the molecular tweezer. Drug release cannot occur, it seems, unless the tweezer is a hydrophobic environment with low pH.

The protonation state of amino acid residues in proteins depends on their respective pK_a values. Computational methods are particularly important for estimating the pK_a values of buried and active site residues, where experimental data is scarce. In chapter 3 I used the cluster model approach to predict the pK_a of some challenging protein residues and for which methods based on the numerical solution of the Poisson-

Boltzmann equation and empirical approaches fail. The ionizable residue and its close environment were treated quantum mechanically, while the rest of the protein was replaced by a uniform dielectric continuum. The approach was found to overestimate the electrostatic interaction leading to predicting lower pK_a values.

ACKNOWLEDGEMENTS

First I would like express my deepest gratitude for my advisor, Professor Paul W. Ayers for giving me the opportunity to work on my master's in his group. I'm grateful for his support, encouragement, and patience during a difficult time when I first came here. He was always there when I needed him, and he inspires his students by his suggestions, knowledge, and productivity.

I would like to thank Dr. Bain and Dr. Dumont who were my committee. Thanks to their help, encouragement, and constructive suggestions.

Many thanks go to the Ayers group. I would like to thank Dr. Steven Burger for his help with the research project. I would also like to thank Dr. Peter Limacher for his help and Guidance. I would also like to thank all the previous and current members of the Ayers group for their help and support: Rogelio Cuevas-Saavedra, Sandra Rabi, Carlos Cardenas, Lourdes Romero, Alfredo Guevara, James Anderson, Ivan Vinogradov, Debajit Chakraborty, Paul Johnson, Farnaz Heidarzadeh, Annie Liu, Cristina González, Eleonora Echegaray, Mathew Chan, Katharina Boguslawski, Pawel Tecmer, Pavel Kulikov. Thanks to all of them for being such good friends.

I would like to thank McMaster University and the department of chemistry and chemical biology for me the opportunity to work here. Thanks to the Natural Sciences and Engineering Research Council (NSERC) and McMaster University for financial support.

Last but not least, I would like to express my deepest gratitude for my family. Thank you for your love and support.

Table of Contents

Abstract ……..…….……..… iii

Acknowledgements…….……... …v

List of Figures…….……..…….…..ix

List of Tables……..... x

1 Background 1

1.1 Introduction 1

1.2 The Schrödinger Equation 3

1.3 Hartree-Fock Method 5

1.4 Density Functional Theory 8

1.5 Molecular Mechanics 11

1.6 Molecular Dynamics 14

1.7 Steered Molecular Dynamics 16

1.8 Free Energy Calculations 17

1.9 pK_a calculations 20

1.10 References 23

2 Drug Release by pH-Responsive Molecular Tweezers: Atomistic Details from Molecular Modelling 27

2.1 Introduction 27

2.2 Computational Details 30

2.2.1 Quantum Mechanics Calculations on the Switching Unit 30

2.2.2 Quantum Chemistry Calculations of the Binding Energy in the Gas Phase 30

2.2.3 Molecular Dynamics (MD) Simulations With Implicit Solvent 31

2.2.4 Steered Molecular Dynamics Simulations With Explicit Solvent 31

2.3 Results and Discussions 32

2.3.1 Relative stability of the switching unit conformers 32

2.3.2 Relative Stability of Tweezer Conformations In the Gas Phase 36

2.3.3 Gas-Phase Binding Energy 41
2.3.4 Stable Conformations of the Tweezer in Implicit Water Solvent 41
2.3.5 Drug Release from the Tweezer 42
2.4 Conclusions 46
2.5 References 48
3 Failures of Embedded Cluster Models for pK_a Shifts Dominated by Electrostatic Effects 53
3.1 Introduction 53
3.2 Computational Details 56
3.2.1 pK_a calculations 56
3.2.2 Free energy calculations 57
3.2.3 Protein model construction 57
3.3 Results 58
3.3.1 Aspartate 75 58
3.3.2 Aspartate 21 60
3.4 Discussion 61
3.5 Conclusions 63
3.6 References 65
4 Conclusions 70

List of Figures

Figure 2.1: Different conformers of the protonated and deprotonated switching unit *ortho*-py.... 33

Figure 2.2: Different conformers of the protonated tweezer..... 37

Figure 2.3: The half-open UW-conformer of the protonated tweezer. 38

Figure 2.4: Two snapshots from molecular dynamics simulation using explicit solvent model for the protonated tweezer 43

Figure 2.5: Two snapshots from steered molecular dynamics simulation using explicit solvent model for the protonated tweezer with the drug molecule. 45

Figure 3.1: Cluster model for Asp75 of bacterial barnase. 59

Figure 3.2: Cluster model for Asp21of fungal beta-cryptogein. 61

List of Tables

Table 2.1: Calculated and experimental ^{13}C chemical shifts for the switching unit ortho-py in the U shape..... 34

Table 2.2: Calculated and experimental ^{1}H chemical shift changes upon the protonation of the switching unit *ortho*-py..... 35

Table 2.3: Calculated ^{13}C chemical shifts for the fully-open (W-shaped) and the half-open (UW-shaped) conformations of the protonated tweezer. 39

Table 2.4: Calculated ^{1}H chemical shifts for the fully-open (W-shaped) and the half-open (UW-shaped) conformations of the protonated tweezer. 40

Table 3.1: Comparison between experimental and several pK_a prediction methods. 60

Chapter 1

Background

1.1 Introduction

Computational chemists strive to understand, quantify, and predict chemical phenomena by using efficient computer programs, based on fundamental principles from theoretical chemistry, to simulate atoms, molecules, and solids. Computational approaches allow us to characterize chemical phenomena in atomistic detail, and are therefore often complementary to experimental techniques.

In the last few decades, the types of systems that can be approached with computational chemistry method have rapidly increased in size and complexity, allowing computational chemists to enter new fields of chemistry, including materials' chemistry and (of primary relevance for this thesis) biological chemistry. The increased scope of computational chemistry is primarily due to two factors: the development of improved theoretical tools (mathematical methods, computer algorithms, and software) and the exponential growth in available computing power. Nowadays, computational chemistry is routinely used to predict molecular structure, energy profiles for chemical reactions, dipole and higher multipole moments, electronic charge distribution, vibrational frequencies, correlations between molecular structures and properties, molecular spectra, electronic and thermal properties of materials, and drug binding affinities.[1] [2] [3] This thesis presents my work using computational chemistry methods to simulate proton abstraction and addition reactions in a biological settings. Designing efficient and reliable computational approaches for simulating proton abstraction and addition reactions is important for drug delivery (chapter 2) and enzymology and drug design (chapter 3). To perform these simulations, a broad range of techniques were needed, ranging from quantum mechanical calculations of molecular electronic structure, to molecular dynamics simulations of thermodynamic properties, to continuum electrostatics computations to estimate the effects a molecule's surroundings on its (de)protonation energy.

Chapter two is a computational study of a pH sensitive molecular tweezer that is designed to release a chemotherapy drug in tumors, which are more acidic than the surrounding tissue. The goal of this study is to study pH sensitive molecular tweezers at atomic resolution and to assess whether it is a promising new drug delivery vehicle. Chapter 3 tests a cluster-model for determining the pK_a of amino acid residues in proteins. The pK_a of amino acid residues is determinative of protein structure, dynamics, and function and must be correctly assigned before numerical simulations of proteins can commence. The results of the cluster model are disappointing when electrostatic interactions between amino acids are important, suggesting that explicit simulations of electrostatic interactions and desolvation effects are needed to accurately determine protein pK_a values. Chapter 4 summarizes the results of the thesis.

1.2 The Schrödinger Equation

Molecules' structure, function, and dynamics are determined by the shape of the molecular potential energy surface (PES) which, at the most fundamental level, is determined by solving the Schrödinger equation. Solving the Schrödinger equation provides the quantum-mechanical description of the motion of nuclei and electrons. The time-independent Schrödinger equation for a molecule can be written in the form

$$\hat{H}\,\Psi(R_1, R_2, ..., R_N, r_1, r_2, ..., r_n) = E\,\Psi(R_1, R_2, ..., R_N, r_1, r_2, ..., r_n)\,, \qquad (1.1)$$

where $\hat{H}$ is the Hamiltonian operator (the sum of kinetic energy and potential energy operators for the system), Ψ is the wavefunction for nuclei and electrons in the molecule, R_I is the position for nucleus I and r_i is the position of electron i, and E is molecular energy associated with the wavefunction. The molecular Hamiltonian in atomic units for a molecule with n electrons and N nuclei can be written as

$$\hat{H} = -\frac{1}{2}\sum_{I}^{N}\frac{\nabla_I^2}{m_I} - \frac{1}{2}\sum_{i}^{n}\nabla_i^2 - \sum_{i}^{n}\sum_{I}^{N}\frac{Z_I}{r_{iI}} + \sum_{I<J}^{N}\frac{Z_I Z_J}{r_{IJ}} + \sum_{i<j}^{n}\frac{1}{r_{ij}}\ , \qquad (1.2)$$

where m_I and Z_I are the mass and atomic numbers of the I^{th} atom. r_{ij}, r_{IJ}, and r_{iI} are electron-electron, nucleus-nucleus, and electron-nucleus distances. The Schrödinger equation can be solved exactly only for one-electron atoms. For molecular systems and many-electron atoms, approximations must be used.

The Born-Oppenheimer Approximation is one of the most popular approximations; [4] with the Born-Oppenheimer approximation, the Schrödinger equation can be solved exactly for some one-electron molecules (and not just one-electron atoms). In addition, the Born-Oppenheimer approximation, or one of several similar approximations, is needed in order to define a molecular potential energy surface, which is one of the most fundamental concepts in chemistry.[5]

The Born-Oppenheimer approximation is based on the realization that nuclei are much heavier than electrons, and therefore move much more slowly. It is reasonable to assume, then, that electrons respond to the nuclear movement instantaneously and that the

nuclei do not move on the timescale of electronic motion. Therefore electrons can be considered moving while nuclei are fixed; this is the essence of the Born-Oppenheimer approximation.

The Born-Oppenheimer approximation enables us to separate the total molecular wavefunction into electronic and nuclear wavefunctions. The nuclear kinetic and potential energy terms do not affect the electronic Hamiltonian, which can be written as

$$\hat{H}_e = -\frac{1}{2}\sum_i^n \nabla_i^2 - \sum_i^n \sum_I^N \frac{Z_I}{r_{iI}} + \sum_{i<j}^n \frac{1}{r_{ij}} \ . \tag{1.3}$$

The eigenvalue of the electronic Hamiltonian is the electronic energy; adding the electronic energy to the nucleus-nucleus repulsion energy gives the potential energy surface. Most of theoretical chemistry is concerned with finding various approximations to the potential energy surface, and most of computational chemistry involves computing the potential-energy surface and using it to elucidate chemical phenomena.

1.3 Hartree-Fock Method

The electronic Schrödinger equation cannot be solved for many-electron systems, which motivates further approximations for the molecular wavefunction. One of the earliest and most fundamental approximations is the Hartree-Fock method. The Hartree-Fock method can be considered an approach that approximates the Hamiltonian by

replacing the problematic electron-electron repulsion potential with an nonlocal, but one-electron, effective potential operator. The resulting Hartree-Fock equations can be solved rather easily for systems with hundreds, and occasionally even thousands, of electrons.[6] In the Hartree-Fock approximation, the electronic wavefunction is a single Slater determinant containing the *n* occupied spin-orbitals.

To be more specific, in Hartree-Fock theory, the electronic Hamiltonian is replaced by the Fock operator,

$$\hat{F}\left[\left\{\phi_j\right\}\right](1) = \hat{H}^{core}(1) + \sum_{j=1}^{n/2}\left[2\hat{J}_j(1) - \hat{K}_j(1)\right], \tag{1.4}$$

where $\hat{F}\left[\left\{\phi_j\right\}\right](1)$ is the one-electron Fock operator, comprising the one-electron core Hamiltonian $\hat{H}^{core}(1)$,

$$\hat{H}^{core}(1) = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}}, \tag{1.5}$$

where $\hat{J}_j(1)$ is the Coulomb operator, and $\hat{K}_j(1)$ is the exchange operator.

Solving the one-electron Schrödinger equation with the one-electron Fock operator as a Hamiltonian gives one-electron wavefunctions for the system called the Hartree-Fock spin-orbitals,

$$\hat{F}(1)\phi_i(1) = \varepsilon_i\phi_i(1). \tag{1.6}$$

The Slater determinant of the occupied spin-orbitals is the many-electron wavefunction in the Hartree-Fock Approximation.

The major deficiency of the Hartree-Fock approximation is the absence of electron correlation: in Hartree-Fock theory, electrons move in an effective one-body potential, and the detailed interelectronic motions (induced by the two-electron repulsion term in the exact Hamiltonian) is neglected. There are many post-Hartree-Fock methods for including electron correlation. One of the most affordable of these is Møller-Plesset perturbation theory.[7] [8] [9]

In perturbation theory the Hamiltonian is written as

$$\hat{H} = \hat{H}^0 + \lambda \hat{H}^1 \ , \tag{1.7}$$

where $\hat{H}^0$ is the model Hamiltonian and λ is an arbitrary parameter. And the perturbation expansion of wavefunctions and energies are

$$\Psi_i = \Psi_i^0 + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} + ... \ , \tag{1.8}$$

$$\varepsilon_i = \varepsilon_i^0 + \lambda \varepsilon_i^{(1)} + \lambda^2 \varepsilon_i^{(2)} + ... \ . \tag{1.9}$$

Inserting these equations into the Schrodinger equation and equating terms, order-by-order, in λ gives the perturbative equations. This allows us to approximate the wavefunction and energies of a difficult-to-solve Hamiltonian (like the true electronic Hamiltonian) from an easier-to-solve model Hamiltonian (like the Hartree-Fock Hamiltonian). Møller-Plesset perturbation theory is perturbation theory where the model

Hamiltonian is the Hartree-Fock Hamiltonian. Like all post Hartree-Fock methods, Møller-Plesset perturbation theory is computationally demanding, and cannot be applied to large molecules.

1.4 Density Functional Theory

Density functional theory (DFT) is a method that approximates the effects of electron correlation (like post Hartree-Fock methods) with a computational cost that is similar to, and sometimes even less than, the Hartree-Fock approximation. It is widely used to study organic reactions, biological processes, and materials. DFT is based on the Hohenberg-Kohn theorem, which indicates that the ground-state electron density fully specifies all molecular properties, including the exact electronic energy, molecular potential energy surface, and many-electron wavefunction.[10] [11] [12] The electron density is a simpler mathematical object than the many-electron wavefunction because it is a function of only three variables: the three Cartesian coordinates x, y, and z. The electronic density is also a simpler conceptual object, since it (unlike the many-electron wavefunction) is a physical observable.

In DFT, the electronic energy is therefore written as a functional of the electron density,

$$E\left[\rho(r)\right] = E_{elec} \ , \tag{1.10}$$

where E_{elec} is the exact electronic energy. The energy functional can be written as a sum of two terms

$$E\left[\rho(r)\right] = \int V_{ext}(r)\,\rho(r)\,dr + F\left[\rho(r)\right], \qquad (1.11)$$

where the first term represents the interactions of electrons with an external potential $V_{ext}(r)$ typically due to nuclei. The second term is the sum of the kinetic energy of electrons and electron-electron interactions. The ground-state energy is determined by variationally minimizing the energy with respect to the electron density.

In DFT, the problem of solving the electronic Schrödinger equation is replaced by the problem of approximately determining the Hohenberg-Kohn function, F [ρ(r)], for which no explicit mathematical equation is known. Kohn and Sham [13] suggested that approximating F[ρ(r)] as the sum of three terms

$$F\left[\rho(r)\right] = E_{KE}\left[\rho(r)\right] + E_{H}\left[\rho(r)\right] + E_{XC}\left[\rho(r)\right], \qquad (1.12)$$

where $E_{KE}\left[\rho(r)\right]$ is the kinetic energy of a system of non-interacting electrons that has the same density as the real system, $E_{H}\left[\rho(r)\right]$ is classical self-repulsion energy of the electron density with itself, and $E_{XC}\left[\rho(r)\right]$ contains everything else, including the potential energy contributions from exchange and electron correlation, as well as the difference between the kinetic energy of the real and non-interacting systems. Only the last term—the exchange-correlation energy functional—is unknown. Most of the current research in DFT is related to designing improved approximations to E_{XC}.

The Kohn-Sham expression for the energy for a system of *N* nuclei and *n* electrons is

$$E\left[\rho(r)\right] = \sum_{i=1}^{n} \int \psi_i(r) \left(-\frac{\nabla^2}{2}\right) \psi_i(r) dr + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + E_{XC}\left[\rho(r)\right] - \sum_{A=1}^{N} \int \frac{Z_A}{|r - R_A|} \rho(r) dr \ .$$

(1.13)

Kohn and Sham also showed that the density can be obtained from a set of one-electron orthonormal orbitals:

$$\rho(r) = \sum_{i=1}^{n} \left|\psi_i(r)\right|^2 \ , \qquad (1.14)$$

which are the solutions of the one-electron Kohn-Sham equations

$$\left\{-\frac{\nabla_1^2}{2} - \left(\sum_{A=1}^{N} \int \frac{Z_A}{r_{1A}}\right) + \int \frac{\rho(r_2)}{r_{12}} dr_2 + V_{XC}\left[r_1\right]\right\} \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \ . \qquad (1.15)$$

Solving the Kohn-Sham equations is similar in cost to solving the Hartree-Fock equations, but density-functional theory (approximately) includes the effects of electron correlation.

DFT methods generally offer a good combination of accuracy and computational cost. They are widely used for large systems. Unfortunately, there are hundreds of approximate exchange-correlation functionals in the literature, and making a good choice

is not always easy. Given a good functional (which can be chosen by validation with experiment or by calibration with a high-accuracy "benchmark" quantum chemistry method), DFT is a reliable and robust computational tool. One of the biggest advantages of DFT is its ability to calculate analytical gradient of molecular energy with respect to the positions of the nuclei. That allows the molecular geometry to be optimized efficiently. Moreover, by using identities from linear algebra, conventional DFT calculations can be performed with computational costs that grow linearly with the size of the system. This allows one to perform DFT calculations on very large (even hundreds of thousands of atoms) systems.

1.5 Molecular Mechanics

The disadvantage of quantum chemistry methods is two-fold: the number of particles in the system is equal to the number of electrons plus the number of atomic nuclei, and the electrons must be treated quantum-mechanically. Chemists often do not think of molecules in terms of electrons, preferring to think of atoms linked by bonds—the venerable ball-and-spring model of molecular structure. The ball-and-spring model is the basis for a more computationally efficient model called molecular mechanics, where electrons are not treated explicitly, and atoms are treated as classical particles that interact using classical potentials. Molecular mechanics allows one to treat systems that are orders of magnitude larger than the systems that can be treated by quantum mechanics.

Molecular mechanics is especially important in condensed phases and for macromolecules, where the number of atoms is large and the movement (not just the optimal position) of atoms is important. Molecular mechanics techniques allow one to routinely treat thousands of atoms, and even million-atom simulations are possible (using supercomputers).[14] [15]

The essential approximations in molecular mechanics are represented by the force field, which is an expression for the molecular potential energy as a function of the nuclear positions. Different types of force fields are appropriate for different types of problems, and several popular commercial force fields are available for molecular modelling. Different force fields have different forms and values for parameters, but most force fields have the same general form, based on chemical intuition:

$$\begin{aligned} V &= V_{bonded} + V_{non-bonded} \\ V_{bonded} &= V_{bond} + V_{angle} + V_{dihedral} \\ V_{non-bonded} &= V_{van\ der\ Waals} + V_{electrostatic} \end{aligned} \quad . \tag{1.16}$$

The force field clearly represents a simplified model of interactions within a chemical system. Usually the bond and angle terms are modeled by harmonic oscillators, and torsion contribution are modeled by periodic functions. Van der Waals terms are usually modeled using the empirical Lennard-Jones potential.[16] Coulomb interactions between fixed atomic point charges are used to model electrostatic terms. In this study, we used

the Generalized Amber Force Field (GAFF); [37] its potential energy function has the form

$$V\left(r^N\right) = V_{bond} + V_{angle} + V_{dihedral} + V_{van\ der\ Waals} + V_{electrostatic}$$
$$= \sum_{bonds} \frac{1}{2} k_b \left(b - b^0\right)^2 + \sum_{angles} \frac{1}{2} k_\theta \left(\theta - \theta^0\right)^2 + \sum_{torsions} \frac{1}{2}\left[V_n \left(1 + \cos(n\omega - \gamma)\right)\right]$$
$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \varepsilon_{i,j} \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2\left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\},$$

(1.17)

where V is the potential energy, a function of positions (**r**) of N particles. K_b is the force constant for stretching and b^0 is the equilibrium bond length. K_θ is the force constant of bending and θ^0 is the equilibrium bond angle. V_n is the torsional barrier height and n is periodicity, which is the number of maxima in a full rotation, ω is the value for torsional angle, and γ is the phase angle. $\varepsilon_{i,j}$ is Lennard-Jones potential well-depth and r^0_{ij} is the equilibrium distance between atoms i and j. q_i and q_j are partial charges of atoms i and j; ε_0 is the dielectric constant. The values of these parameters depend on the atom, atom type, and its environment. There are standard techniques for parameterizing force fields, but parameterizing force fields is truly an art: no parameterization can be trusted without validation against more accurate calculations or experimental results.

1.6 Molecular Dynamics

In 1913, Einstein and Stern suggested that molecules, and quantum-mechanical oscillators in general, have a residual energy at the absolute zero and called it the zero-point energy.[18] This means that although molecules (and, indeed, all systems with bound ground states) have a structure with minimum potential energy, they rarely adopt that structure all the time: instead they fluctuate around it. That is, even in their ground state, small systems undergo structural fluctuations. Macroscopic properties of the system (such as work, free energy, and entropy) can be measured from the microscopic properties of its particles. Such systems should be modeled as an ensemble that represents all possible configurations of the system.

Molecular dynamics (MD) is a computational approach to statistical mechanics. Equilibrium and dynamic properties of systems, which cannot be calculated analytically, can be estimated using MD. In MD simulations, successive configurations of a system of interacting particles are produced by integrating Newton's laws of motion. The time evolution of positions and velocities of the particles of the system can be known.

As the MD simulation progresses, statistical data are collected. The statistical mechanical average for the position vector of a particle A can be calculated from the relation

$$\langle r_A \rangle = \frac{1}{n\Delta t} \sum_{i=1}^{n} r_A (t_i) \,, \tag{1.18}$$

where $r_A(t_i)$ is the position of atom *A* at time t_i. Similarly, the total kinetic energy of the atoms is calculated as the sum of the kinetic energies of individual particles

$$\langle E_k \rangle = \frac{1}{2n\Delta t} \sum_{i=1}^{n} \sum_{j=1}^{N} m_j \left| \mathbf{v}_j (t_i) \right|^2. \tag{1.19}$$

Although the basic idea of MD simulations is simple, running an MD simulation is challenging. Setting the initial conditions is not trivial and trajectories are sensitive to initial condition. Choosing the suitable numerical integrator for Newton's laws of motion can be challenging. Sometimes including high frequency mode is necessary, which requires using a very small timestep (Δt = 1 fs or less) which increases the computational cost needed to simulate a system for a long enough time to converge the time-averages one uses to compute observable properties.[19] There are standard techniques for avoiding most of these problems, but MD simulations for large systems requires insights from the computational chemist.

A more fundamental problem with MD simulations comes from the ergodic hypothesis. The ergodic hypothesis states that over a long period of time all accessible microstates are equiprobable; if the ergodic hypothesis is not true, then the time average cannot be used to calculate the ensemble average, and the entire idea of molecular dynamics is invalid. It is not known whether the ergodic hypothesis is actually mathematically true for chemical and biological systems treated with conventional molecular-mechanics force fields, but chemical systems can exhibit nonergodic behavior during a (short) MD simulation because the full phase space (position and momentum

space, together) may not be explored due to thermodynamic bottlenecks that separate different regions of phase space. Large systems are particularly problematic, since they diffuse slowly.

1.7 Steered Molecular Dynamics

Adequately sampling phase space is essential for computing thermodynamic properties with MD, but adequate sampling is increasingly difficult for large systems. This is especially true when different thermodynamic states are separated by barriers on the potential energy surface that are much higher than the typical kinetic energies of the atoms in the system. Steered molecular dynamics (SMD) is a relatively new approach in which time-dependent external forces are applied to the system to pull it along a certain degree of freedom, thereby forcing the system to sample new regions of phase space. The response of the system to the pulling-force is analyzed, and can be used to gain qualitative insight and to compute quantitative properties (like the free-energy difference between the starting and ending states). Like conventional MD simulations, SMD simulations reveal transitions between molecular structures in atomic detail. Unlike conventional MD simulations, SMD simulations can model processes that occur on timescales of milliseconds or longer like the dissociation and association of macromolecules. In biology SMD is used to simulate biological processes such as binding and dissociation of substrates and proteins, conformational changes of protein,

and DNA stretching. SMD is also ideal for studying the response of molecules to external forces, so it is used to study mechanical unfolding and stretching for proteins and other polymers, as well as the mechanical DNA-unzipping.

One of the biggest advantages of SMD simulations is that it allows one to simulate systems that are not in thermodynamic equilibrium. Jarzynski showed that the free energy difference between two states can be calculated from the exponential averages of irreversible work.[20] Through time series analysis methods, the potentials of mean force can be calculated from multiple SMD simulations. SMD simulations are particularly powerful when only qualitative results are required, rather than quantitative ones. For those cases, such as finding pathways for binding, SMD is a useful tool even when it has a large error bar. Another advantage of SMD simulations is that they mimic the atomic force microscopy, so direct comparison with experiment is available.[21] [22]

1.8 Free Energy Calculations

One of the primary objectives in computational chemistry is to compute free-energy differences. This allows one to ascertain the relative stability of different molecules and conformations, leading to a quantitative understanding of chemical kinetics and equilibrium. Of primary importance in this thesis are free-energy differences of biologically relevant molecules like proteins (chapter 3) and supramolecular drug-delivery vehicles (chapter 2). In biochemistry, free-energy differences are necessary for

computing ligand binding affinity (e.g., drug affinity), for pK_a prediction (proton affinity), and to characterize protein and DNA (un)folding.

In chemistry, it is usually the Gibbs free energy, *G*, that is of interest. The Gibbs free energy is the thermodynamic potential of isothermal-isobaric ensemble (NPT), that is the ensemble in which the number of particles *N*, the system pressure *P*, and temperature *T* are fixed. The Gibbs free energy is defined as

$$G = H - TS \ , \tag{1.20}$$

where *H* is the enthalpy and *S* is the entropy. When the volume, instead of the pressure, is constant one is in the canonical NVT ensemble, and the free energy is the Helmholtz free energy,

$$\begin{aligned} A &= H - TS - PV \\ &= U - TS \end{aligned} \ , \tag{1.21}$$

where *U* is the internal energy of the system.

Depending on how one designs the MD simulation, either *G* or *A* can be determined straightforwardly. Specifically, the free energy can be computed from the partition function. For example, the Helmholtz free energy can be written as

$$A(N,V,T) = -k_B T \ln Z(N,V,T) \ , \tag{1.22}$$

where k_B is Boltzmann constant and Z(N,V,T) is the canonical partition function, which is calculated from the relation

$$Z(N,V,T) = \sum_{\upsilon} e^{-\frac{E_\upsilon}{k_B T}} , \tag{1.23}$$

where the summation is done over all the points in phase space υ. Another way to calculate the partition function is to integrate over the classical phase space. The partition function can be written as

$$Z(N,V,T) = \int\int dp^N dr^N \exp\left(\frac{-E\left(p^N, r^N\right)}{k_B T}\right) , \tag{1.24}$$

where r^N is the position and p^N is the momentum. In conventional MD approaches, the phase-space averages are simplified by integrating with respect to the momentum degrees of freedom analytically; the remaining integration over conformational space is approximated by a time-average (assuming the validity of the ergodic hypothesis).

Sampling can be performed using MD or Monte Carlo simulations. Because the ensembles are (very) incompletely sampled by practical simulation techniques, accurate estimation of partition function and absolute free energy is impossible. Fortunately, for most applications were are only interested in the free energy difference between two states, and free-energy differences between states that are not too dissimilar can be estimated with reasonable precision. Traditionally, free energy differences have been

calculated using thermodynamic integration or umbrella sampling. In this thesis, we will use the more modern method of steered molecular dynamics to calculate the free energy differences.

1.9 pK_a calculations

The free energy difference between an acid (or a base) in its protonated and its deprotonated state is so important and fundamental to chemistry and biochemistry that it has inspired a unique nomenclature and associated numerical and formal computational techniques. The pK_a of an acid is defined as

$$\mathrm{pK}_a \equiv -\log_{10} K_a \ , \tag{1.25}$$

where the acid-dissociation equilibrium constant, K_a, is defined as

$$\Delta G_{\text{dissociation}} = -RT \ln\left(K_a\right) \ . \tag{1.26}$$

The pK_a value of a protein residue determines the protonation state of that residue at a certain pH. If the pK_a of the residue is less than seven, the residue tends to be deprotonated in aqueous solution at standard temperature and pressure; if the pK_a of the residue is greater than seven, the residue tends to be protonated. Whether a residue is protonated or not is important for protein function and stability.

In the past few decades, many computational approaches have been developed to predict protein residues' pK_a values. Most of these approaches fall into one of three categories. The first approach is based on solving the linearized Poisson-Boltzmann equation (LPBE) using finite difference methods.[23] [24] The protein is assigned a low dielectric constant and is surrounded by a solvent with high dielectric constant, usually water. Distribution of counterions in the solvent is described by Boltzmann distribution. Fixed atomic charges are assigned to protein atoms and the electrostatic potential of protein and solvent are calculated. The drawback of these methods is that it is formally invalid to use a concept from macroscopic continuum electrostatics (the dielectric constant) to compute a property that based on atomic details like (de)protonation.

The second approach is based on statistical fitting of an empirical model, rather than equations derived from classical physics. These methods have the advantage of being extremely fast and the disadvantage of lacking physical basis. The pK_a value of the protein residue is calculated by adding environmental perturbations to the pK_a value of the model residue

$$pK_a = pK_{a,\text{model}} + \partial pK_a \ , \tag{1.27}$$

where ∂pk_a is the sum of all perturbations from environmental factors: charge-charge interactions, hydrogen bonding, and desolvation effects.[27]

The third approach is based on calculating the difference in free energy for the protonated and non-protonated states using either QM methods or MM methods.[28] [29] pK_a values can be directly computed from the free energy difference through the relation

$$pK_a = \frac{\Delta G_{\text{dissociation}}}{2.3026RT} \quad . \tag{1.28}$$

The disadvantage of these methods is their computational expense. Rigorous methods based on long-timescale molecular dynamics are computationally impractical because each protein has many ionizable residues, suggesting that one should consider only a portion of the enzyme, i.e., a molecular cluster embedded in the protein environment. Chapter 3 studies a cluster model for computing protein pK_a's and compares it to cheaper and more conventional approaches based on the linearized Poisson-Boltzmann equation and statistical models.

1.10 References

[1] Leach, A. R. Molecular Modelling: Principles and Applications, Pearson Education Limited, second edition, 2001.

[2] Cramer, C. J. Essentials of Computational Chemistry: Theories and Models, John Wiley & Sons Inc., Hoboken, NJ, second edition, 2004.

[3] Schlick, T. Molecular Modeling and Simulation: An Interdisciplinary Guide, Springer: New York, second edition, 2002.

[4] Born, V. M.; Oppenheimer, R. “Zur quantentheorie der molekeln." [On the Quantum Theory of Molecules]. Annalen der Physik **1927**, 389, 457–484.

[5] Cramer, C. J. Potential Energy Surfaces, in Essentials of Computational Chemistry: Theories and Models, John Wiley & Sons Inc., Hoboken, NJ, second edition, 2004; pp 6-10.

[6] Szabo, A.; Ostlund, N. S. The Hartree-Fock Approximation, in Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory. Mineola, New York: Dover Publishing: 1996: pp 108-230

[7] Møller, C.; Plesset, M. S. "Note on an Approximation Treatment for Many-Electron Systems." Phys. Rev. **1934**, 46, 618–622.

[8] Head-Gordon, M.; Pople, J. A.; Frisch, M. J. "MP2 energy evaluation by direct methods." Chem. Phys. Lett. **1988**, 153, 503–506.

[9] Cramer, C. J. Perturbation Theory, in Essentials of Computational Chemistry: Theories and Models. John Wiley & Sons Inc., Hoboken, NJ, second edition, 2004; pp 216-224.

[10] Ayers, P. W.; Yang, W. Density Functional Theory, in P. Bultinck, H. de Winter, W. Langenaeker, J.P. Tollenaere (Eds.) Computational Medicinal Chemistry for Drug Discovery, Dekker, New York, 2003, pp. 571-616.

[11] Hinchliffe, A. Density Functional Theory, in Modelling Molecular Structures, John Wiley & Sons Ltd, Baffins Lane, Chichester, West Sussex, second edition, 2001; pp 218-229.

[12] Hohenberg, P.; Kohn, W. "Inhomogeneous Electron Gas." Phys. Rev. **1964**,136, B864-B871.

[13] Kohn, W.; Sham, L. J. "Self-Consistent Equations Including Exchange and Correlation Effects." Phys. Rev. **1965**, 140, A1133- A1138.

[14] Leach, A. R. Empirical Force Field Models: Molecular Mechanics, in Molecular Modelling: Principles and Applications, Pearson Education Limited, second edition, 2001; pp 165-252.

[15] Cramer, C. J. Molecular Mechanics, in Essentials of Computational Chemistry: Theories and Models, John Wiley & Sons Inc., Hoboken, NJ, second edition, 2004; pp 17-68.

[16] Lennard-Jones, J. E. "On the Determination of Molecular Fields" Proc. R. Soc. Lond. A **1924**, 106, 463–477

[17] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. “Development and testing of a general amber force field.” J. Comput. Chem. **2005**, 26, 114.

[18] Einstein, A.; Stern, O. Einige “Argumente fur die Annahme einer molekularen Agitation beim absoluten Nullpunkt.” Ann. Phys. **1913**, 345, 551-560.

[19] Schlick, T. Molecular Dynamics: Basics, in Molecular Modeling and Simulation: An Interdisciplinary Guide, Springer: New York, second edition, 2002; pp 425-461.

[20] Jarzynski, C. “Nonequilibrium equality for free energy differences.” Phys. Rev. Lett. **1997**, 78, 2690-2693.

[21] Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten. K. “Steered molecular dynamics investigations of protein function.” J. Mol. Graph. Model. **2001**, 19, 13–25.

[22] Isralewitz, B.; Gao, M.; Schulten, K. “Steered molecular dynamics and mechanical functions of proteins.” Curr. Opin. Struct. Biol. **2001**, 11, 224–230.

[23] Davis, M. E.; McCammon, J. A. “Electrostatics in biomolecular structure and dynamics.” Chem. Rev. **1990**, 90, 509–521.

[24] Gilson, M. K.; Rashin, A.; Fine, R.; Honig, B. “On the calculation of electrostatic interactions in proteins.” J. Mol. Biol. **1985**, 184, 503–516.

[25] You, T. J.; Bashford, D. “Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility.” Biophys. J. **1995**, 69, 1721–1733.

[26] Gilson, M. K.; Honig, B. H. “Electrostatic fields in the active sites of lysozymes.” Nature **1987**, 330, 84–86.

[27] Li, H.; Robertson, A. D.; Jensen, J. H. "Very Fast Empirical Prediction and Interpretation of Protein pK_a Values." Proteins **2005**, 61, 704–721.

[28] Simonson, T.; Carlsson, J.; Case, A. D. “Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models.” J. Am. Chem. Soc. **2004**, 126, 4167–4180.

[29] Li, H.; Robertson, A. D.; Jensen, J. H. “The determinants of carboxyl pK_a values in turkey ovomucoid third domain.” Proteins **2004**, 55, 689-704.

Chapter 2

Drug Release by pH-Responsive Molecular Tweezers: Atomistic Details from Molecular Modelling

2.1 Introduction

Dynamic molecular tweezers [1] [2] are two-armed molecules that switch from the closed (arms together) to open (arms apart) form when stimulated by an external trigger like light or a change in ion concentration.[3] These nanomechanical devices can hold guest molecules between their arms using non-covalent interactions and then release the molecules when triggered by environmental cues. This leads to a host of applications related to molecular recognition and sensing.[4] Our specific interest is in studying

whether molecular tweezers that are sensitive to hydronium ion concentration (pH) can be used as vehicles for drug delivery.

Drug targeting research is motivated by the observation that if a drug could be carried to the specific pathogen or type of cell that is implicated in a disease, then the effective dose of the drug could be reduced, alleviating side effects.[5] This is especially important when the drug is toxic, as is typically the case in chemotherapy. Supramolecular systems with chemical labels have been studied as drug carriers, and some success has been achieved bringing drugs to specific parts of the human body. However, the drug release at the desired site still represents a challenge.[6] The concept is simple: a stimulus from the drug carrier's environment will trigger a change in the carrier's molecular structure, releasing a drug. For example, a molecular tweezer with an acid-sensitive linkage will cleave in a mildly acidic environment like an endosome or a tumor.[7] [8] Unfortunately these systems are often either too labile or too stable to offer adequate control over the release rate of the substrate.[9] There are indications that supramolecular frameworks that bind and release drugs using noncovalent interactions will have reasonable affinities and response properties.[4] [10] The molecular tweezers we study were designed based on this principle.

Anne Petitjean's group recently designed a pH-responsive molecular tweezer that can release an aromatic substrate when it switches from the closed (U-shaped) to the open (W-shaped) conformation.[11] The change in conformation was confirmed by NMR experiments and emission fluorescence suggests that in the closed U-shape conformation,

the two naphthalene arms of the tweezer can bind an aromatic substrate using noncovalent π-π stacking interactions. When a change in pH triggers a change to the open W-shape conformation, the two arms are far from each other and the substrate can be released into solution.

In this study, we use quantum mechanical (QM) methods along with molecular dynamics (MD) simulations to investigate the switching ability of this tweezer. Quantum mechanics provides efficient tools to understand the intramolecular interactions within the tweezer and the intermolecular interaction between the substrate and the tweezer. MD simulations provide more insight into the mechanism and dynamics of molecular switching and the substrate release process.

A few previous studies of molecular tweezers have been reported, mostly using quantum chemistry methods or molecular mechanics approaches with continuum solvation.[12] - [22] The results of these studies indicate that dispersion-corrected density functionals are required to obtain sensible results because the host-guest binding interaction in most molecular tweezer complexes, including those considered here, is usually dispersion and, specifically, π-π stacking.[16] - [22] Several of the more recent studies indicate that the B97D functional,[23] [24] which we will use in this work, gives good results,[20] [22] comparable to the most modern -D3 dispersion models.[25] We will also use molecular dynamics simulations with both continuum solvation and explicit solvation; to our knowledge explicit solvent molecules have only been considered in one previous calculation on molecular tweezers.[18]

2.2 Computational Details

2.2.1 Quantum Mechanics Calculations on the Switching Unit

All quantum chemistry calculations were performed using the *Gaussian 09* program, [26] using Density Functional Theory (DFT).[27] [28] For the different conformers of the switching unit, the methoxyphenylpyridinemethoxyphenyl triad, the B3LYP functional [29] was used with the 6-31G(d) standard basis set.[30] In experiment, conformational changes in the tweezer are usually deduced from NMR spectra. We simulated NMR chemical shifts using the functional B97-2 [31] with the 6-311++G(2d,p) basis set, using geometries obtained by the same method.

2.2.2 Quantum Chemistry Calculations of the Binding Energy in the Gas Phase

For the tweezer including the naphthalene arms, dispersion interactions have to be included, so we optimized the geometry using the B97D functional with the TZVP/TZVPFIT basis set.[32] The counterpoise correction was used to reduce the basis set superposition error.[33] The binding energy was computed using the formula

$$E_b = E_{twz\text{-}sub} - [E_{twz} + E_{sub}] , \quad (2.1)$$

where E_b is the energy of binding, $E_{twz\text{-}sub}$ is the energy of the system with the substrate inside the tweezer, E_{twz} is the energy of the tweezer alone, and E_{sub} is the energy of the substrate alone.

2.2.3 Molecular Dynamics (MD) Simulations With Implicit Solvent

All molecular mechanics calculations were performed using the *Amber10* software.[34] [35] Atomic charges for the molecular mechanics force field were calculated at the HF/6-31G(d) level of theory using Restrained Electrostatic Potential (RESP) fitting.[36]

Minimizations were carried out by the *Amber* module *Sander* to relax the system prior to MD simulations. The QM optimized geometry is initially optimized by 250 steps with the steepest descent algorithm, followed by a 250 steps of minimization by the conjugate gradient algorithm with no restraints. No cutoff was used and periodic boundary conditions were turned off.

Molecular dynamics simulations were performed under isothermal/isobaric (NPT) conditions with the Parameterized Generalized Amber Force Field (GAFF) on the minimized geometries.[37] The Langevin thermostat was used to maintain the temperature at 300 K. The generalized Born implicit solvation model was used.

2.2.4 Steered Molecular Dynamics Simulations With Explicit Solvent

We used the steered molecular dynamics technique to calculate the free energy of binding, ΔG, using Jarzynski relationship.[37] [38] The free energy difference between the two equilibrium states is related to the work W done on the system by the inequality

$$\Delta G \leq W \tag{2.2}$$

And the binding free energy is calculated from the relation

$$\Delta G = \langle W \rangle - \sigma_w^2 / 2k_B T, \tag{2.3}$$

where σ_w^2 is the variance of work, k_B is Boltzmann's constant, and T is the absolute temperature.

2.3 Results and Discussions

2.3.1 Relative stability of the switching unit conformers

We started our study by optimizing the geometry of the switching unit (see Figure 2.1) at the B3LYP/6-31G(d,p) level of theory. When the pyridine derivative is deprotonated, the closed U-shaped conformer of the tweezer is 5.6 kcal/mol more stable than the fully open W-shaped conformer. Conversely, when the nitrogen atom is protonated, the U-shaped conformation becomes 4.9 kcal/mol less stable than the W-shaped conformer. Converting from the U-conformer to the W-conformer requires crossing a modest barrier of 2.8 kcal/mol, indicating that after the nitrogen atom is protonated in the U-conformer, the W-conformer is thermally accessible.

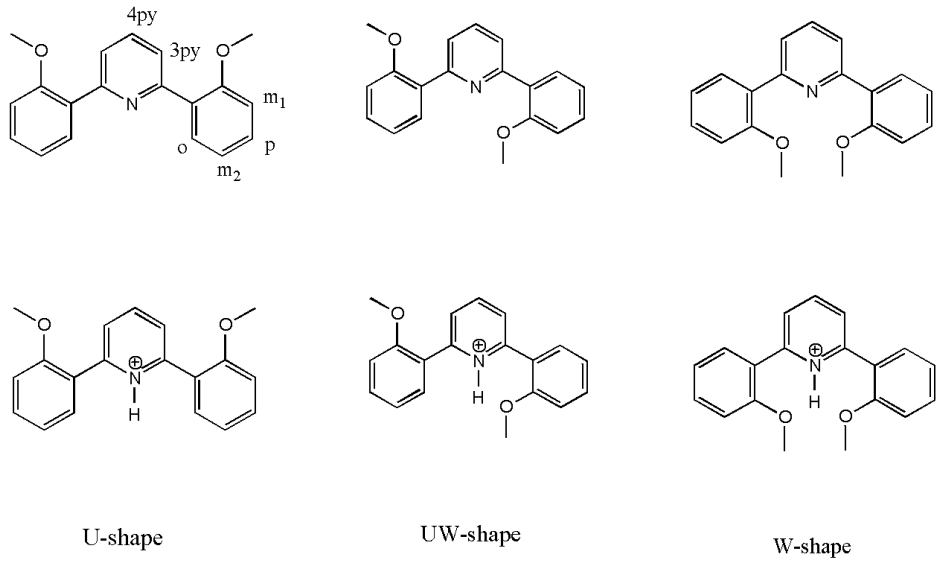

Figure 2.1: Different conformers of the protonated and deprotonated switching unit *ortho*-py.

The experimental evidence for the conformational rearrangement of the tweezer is based on NMR studies of the switching unit. The calculated and observed ^{13}C chemical shifts are reported in Table 2.1 for the *ortho* derivative of the U-conformer. The agreement between the computed and measured data supports the experimental assignment of the closed U-conformer to the deprotonated tweezer. Table 2.2 shows the calculated and the observed difference in ^{1}H chemical shifts between the deprotonated (U-shaped) and protonated (W-shaped) units. Only one hydrogen atom has the opposite order, but the difference is within the expected error of the DFT method.

Table 2.1: Calculated and experimental ^{13}C chemical shifts for the switching unit ortho-py in the U shape. Calculations were done at the B972/6-311++G(2d,p) level of theory.

	^{13}C chemical shifts (ppm)		
Atom[a]	Calc.	Exp.[b]	Error
CH3	53.69	56.48	-2.79
C3Py	124.00	123.95	0.05
Co	135.05	132.34	2.71
C4Py	135.25	136.02	-0.77
C-O	158.69	157.92	0.77
Cm_2	121.06	121.89	-0.83
Cm_1	108.64	112.19	-3.55
Cp	130.40	130.51	-0.11
C-C	132.13	130.39	1.74
C-C	158.21	156.24	1.97

[a] See Figure 2.1 for atom labels

[b]Experimental data were obtained from Ref. [11]

Table 2.2: Calculated and experimental ^{1}H chemical shift changes upon the protonation of the switching unit *ortho*-py. The deprotonated unit is in the U shape and the protonated one is in the W shape. Calculations were done at the B972/6-311++G(2d,p) level of theory.

	Change in ^{1}H chemical shifts (ppm)	
Atom[a]	Calc.	Exp.[b]
o-H3py	-0.03	0.5
o-H4py	0.79	0.8
o-Hp	0.61	0.3

[a] See Figure 2.1 for atom labels

[b]Experimental data were obtained from Ref. [11]

2.3.2 Relative Stability of Tweezer Conformations In the Gas Phase

Further evidence for the U-shaped conformation of the deprotonated tweezer comes from the experimental X-ray structure for the full tweezer.[11] Starting from the X-ray structure, we optimized the tweezer's geometry at the B97D/TZVP/TZVPFIT level of theory. The optimized geometry was close to the starting X-ray structure. As expected, the U-conformer was preferred. Key stabilizing interactions in the U-conformer are (1) the hydrogen bonds between the nitrogen atom of the pyridine ring and the hydrogen atoms of the anisole (methoxybenzene) rings, (2) hydrogen bonds between the oxygen atoms of the anisole rings and the facing hydrogens of the pyridine ring, and (3) the π-π stacking interaction between the two naphthalene rings that form the arms of the tweezer. In the W conformation, the π-π stacking interaction and all hydrogen bonds are lost. Moreover, the electrostatic repulsion resulting from the partial negative charges on the nitrogen and the two anisole oxygen atoms, which are towards each other in the W conformer, make the W shape unfavorable for the non-protonated tweezer.

Since we did not have a reference X-ray structure for the protonated tweezer, we performed a two-dimensional potential energy surface (PES) scan over the dihedral angles that connected the pyridine to the two anisole moieties; we then optimized the stable conformers identified from the PES scan, obtaining the structures in Figure 2.2. Optimization of the minima revealed that the UW-conformer is most stable, with the U-conformer .8 kcal/mol higher in energy. Surprisingly, the least-stable conformer was the W-conformer, which was 1 kcal/mol above the UW conformer. These energies are within

the error of DFT methods, but the results seem plausible when the structures are studied. The W-conformer is unstable because it brings the two anisole oxygen atoms into close proximity. The U conformer features a steric and electrostatic clash between the proton on the pyridine ring and hydrogen atoms on the anisole ring, which forces one of the arms of the tweezer to twist. The UW features a hydrogen bond to stabilize the pyridinium ion.

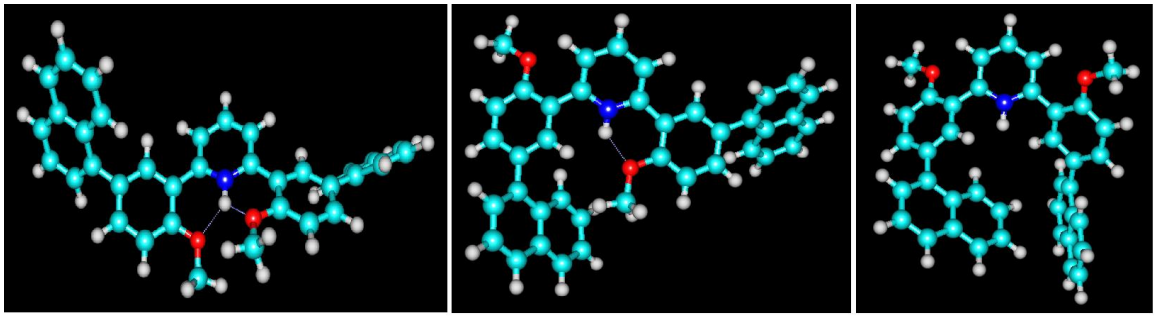

Figure 2.2: Different conformers of the protonated tweezer optimized at the B97D/TZVP/TZVPFIT level of theory.

To see if the fully-open and the half-open conformations of the protonated tweezer have a different chemical shifts, we calculated the ^{13}C chemical shifts for both the W- and UW-conformers. Results are shown in Table 2.3. The values are similar which makes the observed shifts inconclusive for distinguishing between the two conformations. In Table 2.4 the ^{1}H chemical shifts for both conformations are close for most of the atoms and support our results. In the experimental study, the loss of mutual shielding of naphthalene rings was interpreted as switching to the W-shape. However, this behaviour is also expected from the UW-shape, since upon rotation of a single arm of the tweezer, atoms of both arms will be deshielded. Based on our calculations, we believe that the experimental spectrum was misassigned, and that the most stable conformer of the protonated switching unit and of the protonated tweezer is not the open W-form, but the half-open UW-form.

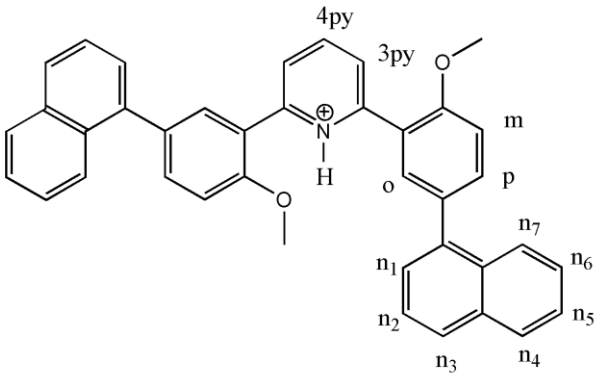

Figure 2.3: The half-open UW-conformer of the protonated tweezer.

Table 2.3: Calculated ^{13}C chemical shifts for the fully-open (W-shaped) and the half-open (UW-shaped) conformations of the protonated tweezer. Shifts were calculated at the B97D/6-311++G(2d,p) level of theory. Geometries were obtained at B97D/TZVP/TZVPFIT. For the UW-conformer, we show shifts for the arm that did not switch.

	^{13}C chemical shifts (ppm)		
Atom[a]	W	UW	Diff.
4py	139.51	141.16	1.65
3py	119.52	122.24	2.72
o	127.93	128.59	0.66
m	108.28	112.03	3.75
p	138.88	136.33	2.56
n_1	125.12	127.03	1.92
n_2	124.62	126.00	1.38
n_3	129.40	128.72	0.68
n_4	128.47	129.00	0.53
n_5	125.75	124.81	0.94
n_6	125.29	125.32	0.03
n_7	122.36	122.62	0.25

[a] See Figure 2.3 for atom labels.

Table 2.4: Calculated ^{1}H chemical shifts for the fully-open (W-shaped) and the half-open (UW-shaped) conformations of the protonated tweezer. Shifts were calculated at the B97D/6-311++G(2d,p) level of theory. Geometries were obtained at B97D/TZVP/TZVPFIT. For the UW-conformer, we show shifts for the arm that did not switch.

	^{1}H chemical shifts (ppm)		
Atom[a]	W	UW	Diff.
4py	8.16	8.15	0.01
3py	7.97	8.42	0.45
o	7.99	7.92	0.07
m	7.09	7.02	0.07
p	7.97	7.83	0.14
n_1	7.49	7.58	0.10
n_2	7.68	7.67	0.01
n_3	8.11	8.05	0.06
n_4	8.01	8.08	0.06
n_5	7.66	7.76	0.10
n_6	7.64	7.80	0.15
n_7	7.90	7.86	0.03

[a] See Figure 2.3 for atom labels.

2.3.3 Gas-Phase Binding Energy

Our calculations support the experimentalists' assertion that this molecular tweezer is flexible and pH-sensitive. These calculations do not assess, however, whether the tweezer can bind a drug molecule. To assess this, we considered the interaction energy between the tweezer and quinizarin in the gas phase. Quinizarin is an aromatic molecule (it will have favorable π-stacking interactions with the naphthalene rings) that is an anticancer drug. It was used in the experimental study.[11] The gas-phase binding energy of quinizarin was -22.0 kcal/mol for the protonated tweezer and -19.8 kcal/mol for the deprotonated tweezer.

2.3.4 Stable Conformations of the Tweezer in Implicit Water Solvent

To make a final assessment of whether the tweezer can work for drug delivery, we need to perform calculations in solution. We did this by parameterizing the Amber force field for the tweezer, as described in the computational methodology section. The gas-phase binding energy of quinizarin to the deprotonated tweezer, computed with the force field, was -22.9 kcal/mol and the geometry was similar to the one computed using DFT. Further support for the parameterization arose from the fact that the U-conformer of the deprotonated tweezer was 3.5 kcal/mol more stable than the W-conformer, which is also consistent with the trend from the DFT calculations, where the U-conformer is more stable by 8.5 kcal/mol.

After validating the force field model, we performed calculations on the U- and W-conformations of the deprotonated tweezer with implicit (generalized Born) water solvation. Within 50 ns, the W-conformer converted to the U-conformer. The simulation that started with the U-conformer was stable, with the U-conformer persisting until the end of the simulation. We also performed simulations on the protonated tweezer, also starting from the U-conformer and W-conformer. Within 50 ns, both the U- and W-conformers converted to the UW-conformer that our gas-phase calculations had predicted to be most stable. This is in contradiction to the NMR data from the experimental study, which was interpreted as supporting the W-conformer. As explained in the previous subsection, we are skeptical of that spectral assignment, and believe the half-open conformation is more stable. Our data does strongly support the experimental assertion that the (de)protonation of the pyridine ring induces a conformation shift in the tweezer.

2.3.5 Drug Release from the Tweezer

To test the ability of the tweezer to release the substrate, an MD simulation in explicit solvent was performed on the U-conformer of the protonated tweezer with the drug molecule inside. Figure 2.4 shows that after 35ns, the tweezer has switched to the UW-conformer and the substrate is exposed to the solvent. The substrate remains close to the tweezer, however, because the hydrophobic quinizarin molecule prefers to be near the hydrophobic naphthalene arms, rather than completely exposed to solvent. This

arrangement is favorable because separation of the hydrophobic molecules in water disrupts the hydrogen bonding between the polar molecules.[39] [40]

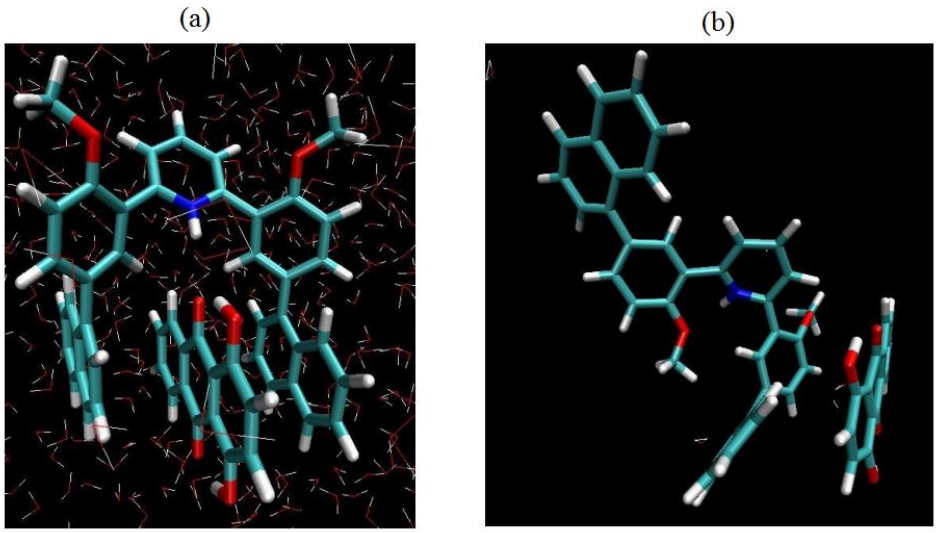

Figure 2.4: Two snapshots from molecular dynamics simulations using explicit solvent model for the protonated tweezer with the drug molecule inside. (a) The starting geometry for the U shaped conformer of the protonated tweezer with the quinizarin molecule between the two arms. (b) A snapshot after 35 ns shows the drug molecule released into the solution and the tweezer switched to the UW-conformation.

Finally, we studied the dissociation energy of the drug molecule from the protonated tweezer, over a timescale of 5 ns, using steered molecular dynamics in explicit water. Representative starting and ending structures for the simulations are given in Figure 2.5. The mean amount of work required to pull the drug molecule out of the tweezer was 9.5 kcal/mol. The simulation was repeated 12 times and the average of the work was taken. The binding free energy was calculated according to equation (2.3) as -8.3 kcal/mol. As can be seen from Figure 2.5, a large part of this work is associated with pulling the quinizarin molecule away from the hydrophobic tweezer arms and into the aqueous environment. The difference in solvation free energy for quinizarin in aqueous environment and a hydrophobic environment (modeled as *n*-hexane to mimic the hydrophobic part of a lipid membrane) is 3.8 kcal/mol; the difference in solvation free energy between water and benzene is 5 kcal/mol. This suggests that if quinizarin were released from the tweezer into a hydrophobic environment, the binding free energy would reduce to around 4 kcal/mol, which is clearly thermodynamically feasible.

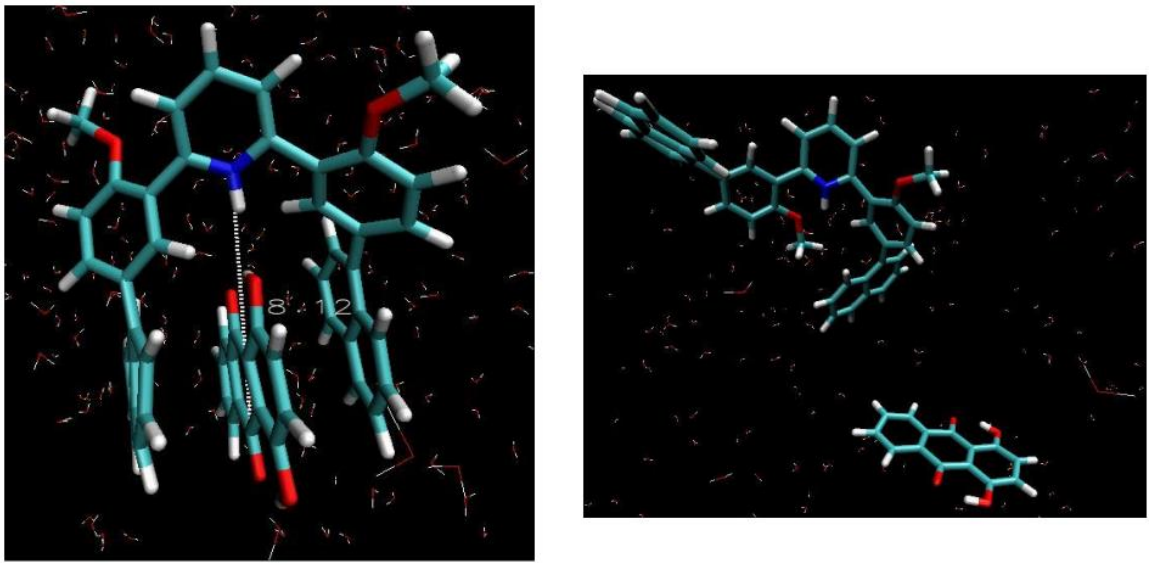

Figure 2.5: Two snapshots from steered molecular dynamics simulations using explicit solvent model for the protonated tweezer with the drug molecule inside. (a) The starting geometry for the U shaped conformer of the protonated tweezer with the quinizarin molecule between the two arms. (b) A snapshot after 5 ns shows the drug molecule released into the solution.

2.4 Conclusions

We have performed a systematic computational study of a new pH-responsive molecular tweezer of potential use for delivering hydrophobic chemotherapy drugs to tumors. Our results support experimental evidence that the tweezer has potential as a drug carrier, and provide a new insight: the conformation of the protonated tweezer, which is the conformation that would release drug molecules in the acidic environment of a tumor or endosome, is the half-open UW-conformation rather than the fully-open W-conformation. The main reason for this preference is probably the unfavorable electrostatic clashes between two anisole oxygen atoms in the W conformation. The UW-conformation is fully capable of releasing a drug molecule, however, as is apparent from the molecular-dynamics simulations of the noncovalent complex between the tweezer and the cancer drug quinizarin.

Our calculations support several key features of the tweezer: (1) it is flexible with small barriers to conformational changes; (2) the preferred conformation is dependent on the protonation state of the pyridine ring that occupies the hinge of the tweezer; (3) the tweezer is able to hold an aromatic substrate between its two arms with noncovalent interactions. Molecular dynamics simulations allowed us to observe the dynamics of the pH-induced conformational switching and the drug-release process.

The binding free energy from steered molecular dynamics was -8.3 kcal/mol, which is higher than expected. This is probably due to the reluctance of the hydrophobic

drug molecule to relinquish its contact with the hydrophobic arm of the tweezer. The binding energy in a physiological environment might be much less because there are hydrophobic environments in the cell in which the quinizarin molecule would be much more stable; in such environments the binding energy of the tweezer is probably closer to 4 kcal/mol. It is also worth noting that the barrier to conformational change, the pH at which the tweezer opens, and the drug-binding energy can all be tuned by adding electron withdrawing/donating substituents to the tweezer. The methodical computational study we have performed here provides a protocol by which new, substituted, tweezer molecules could be computationally tested for desirable properties, providing guidance and insight that can reduce the number of tweezers that need to be synthesized and experimentally tested.

2.5 References

[1] Steed, J. W.; Awood, J. L. Supramolecular chemistry, Wiley, Chichester, 2000.

[2] (a) Chen, C. W.; Whitlock, H. W. J. J. Am. Chem. Soc. **1978**, 100, 4921–4922. (b) Zimmerman, S. C. Top. Curr. Chem. **1993**, 165, 71–102. (c) Klärner, F. G.; Kahlert, B. Acc. Chem. Res. **2003**, 36, 919–932. (d) Harmata, M. Acc. Chem. Res. **2004**, 37, 862–873.

[3] (a) Petitjean, A.; Khoury, R. G.; Kyritsakas, N.; Lehn, J.-M. J. Am. Chem. Soc. **2004**, 126, 6637–6647. (b) Landge, S. M.; Aprahamian, I. J. Am. Chem. Soc. **2009**, 131, 18269–18271. (c) Skibinski, M.; Gómez, R.; Lork, E.; Azov, V. Tetrahedron. **2009**, 65, 10348–10354. (d) Legouin, B.; Uriac, P.; Tomasi, S.; Toupet, L.; Bondon, A.; van de Weghe, P. Org. Lett. **2009**, 11, 745–748.

[4] (a) Müller, B. K.; Reuter, A.; Simmel, F. C.; Lamb, D. C. Nano Lett. **2006**, 6, 2814–2820. (b) Petitjean, A.; Lehn, J.-M. Inorg. Chim. Acta. **2007**, 360, 849–856. (c) Anslyn, E. V. J. Org. Chem. **2007**, 72, 687–699. (d) Phillips, M. D.; Fyles, T. M.; Barwell, N. P.; James, T. D. Chem. Commun. **2009**, 43, 6557–6559. (e) Leblond, J.; Petitjean, A. Chemphyschem. **2011**, 12, 1043-1051.

[5] Rajendran, L.; Knolker, H.-J.; Simons, K. Nat. Rev. Drug Discovery **2010**, 9, 29–42.

[6] Duncan, R. Nat. Rev. Cancer **2006**, 6, 688–701.

[7] Guo, X.; Szoka, F. C. Acc. Chem. Res. **2003**, 36, 335–341.

[8] (a) Ulbrich, K.; Subr, V. Adv. Drug Delivery Rev. **2004**, 56, 1023–1050. (b) Oh, K. T.; Yin, H.; Lee, E. S.; Bae, Y. H. J. Mater. Chem. **2007**, 17, 3987–

4001. (c) Yessine, M.-A.; Leroux, J.-C. Adv. Drug Delivery Rev. **2004**, 56, 999–1021.

[9] (a) Gullotti, E.; Yeo, Y. Mol. Pharmaceut. **2009**, 6, 1041–1051. (b) Drummond, D. C.; Daleke, D. L. Chem. Phys. Lipids **1995**, 75, 27–41. (c) Walker, G. F.; Fella, C.; Pelisek, J.; Fahrmeir, J.; Boeckle, S.; Ogris, M.; Wagner, E. Mol. Ther. **2005**, 11, 418–425.

[10] (a) Schneider, H.-J. Angew. Chem., Int. Ed. **2009**, 48, 3924–3977. (b) Liu, F.; Urban, M. W. Prog. Polym. Sci. **2010**, 35, 3–23.

[11] Leblond, J.; Gao, H.; Petitjean, A.; Leroux, J.-C. J. Am. Chem. Soc. **2010**, 132, 8544–8545.

[12] Klarner, F. G.; Panitzky, J.; Preda, D.; Scott, L.T. J. Mol. Model. **2000**, 6, 318-327.

[13] Brown, S. P.; Schaller, T.; Seelbach, U. P.; Koziol, F.; Ochsenfeld, C.; Klarner, F. G.; Spiess, H. W. Angewandte Chemie-International Edition **2001**, 40, 717-720.

[14] Klarner, F.G.; Kahlert, B. Acc. Chem. Res. **2003**, 36, 919-932.

[15] Chang, C. E.; Gilson, M. K. J. Am. Chem. Soc. **2004**, 126, 13156-13164.

[16] Parac, M.; Etinski, M.; Peric, M.; Grimme, S. J. Chem. Theory Comp. **2005**, 1, 1110-1118.

[17] Parchansky, V.; Matejka, P.; Dolensky, B.; Havlik, M.; Bour, P. J. Mol. Struct. **2009**, 934, 117-122.

[18] Kessler, J.; Jakubek, M.; Dolensky, B.; Bour, P. J. Comput. Chem. **2012**, 33, 2310-2317.

[19] Risthaus, T.; Grimme, S. J. Chem. Theory Comp. **2013**, 9, 1580-1591.

[20] Graton, J.; Le Questel, J. Y. ; Legouin, B.; Uriac, P.; van de Weghe, P.; Jacquemin, D. Chem. Phys. Lett. **2012**, 522, 11-16.

[21] Grimme, S. Chemistry-a European Journal **2012**, 18, 9955-9964.

[22] Graton, J.; Legouin, B.; Besseau, F.; Uriac, P.; Le Questel, J. Y.; van de Weghe, P.; Jacquemin, D. J. Phys. Chem. C **2012**, 116, 23067-23074.

[23] Becke, A. D. J. Chem. Phys. **1997**, 107, 8554-8560.

[24] Grimme, S. J. Comp. Chem. **2006**, 27, 1787-1799.

[25] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. J. Chem. Phys. **2010**, 132, 154104.

[26] Frisch, M. J.; Trucks , G. W.; Schlegel, H. B. *et al*, Gaussian 09, Revision C.01, Gaussian, Inc., Wallingford CT, 2009.

[27] Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. Chem. Rev. **2012**, 112, 289-320.

[28] Ayers, P. W.; Yang, W. Density Functional Theory, in: Bultinck, P.; de Winter, H.; Langenaeker, W.; Tollenaere, J.P. (Eds.) Computational Medicinal Chemistry for Drug Discovery, Dekker, New York, 2003, pp. 571-616.

[29] (a) Becke, A. D. J. Chem. Phys. **1993**, 98, 5648-5652. (b) Lee, C.; Yang, W.; Parr, R. G. Phys. Rev. B **1988**, 37, 785-789. (c) Vosko, S. H.; Wilk, L.; Nusair, M. Can. J. Phys. **1980**, 58, 1200-1211. (d) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. J. Phys. Chem. **1994**, 98, 11623-11627.

[30] (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. J. Chem. Phys. **1971**, 54, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. J. Chem. Phys. **1972**, 56, 2257. (c)

Hariharan, P. C.; Pople, J. A. Theor. Chem. Acc. **1973**, 28, 213. (d) Frisch, M. J.; Pople, J. A.; Binkley, J. S. J. Chem. Phys. **1984**, 80, 3265.

[31] Wilson, P. J.; Bradley, T. J.; Tozer, D. J. J. Chem. Phys. **2001**, 115, 9233-9242.

[32] (a) Schaefer, A.; Horn, H.; Ahlrichs, R. J. Chem. Phys. **1992**, 97, 2571-2577. (b) Schaefer, A.; Huber, C.; Ahlrichs, R. J. Chem. Phys. **1994**, 100, 5829-5835. (c) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. Chem. Phys. Lett., **1995**, 240, 283-289. (d) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Theor. Chem. Acc. **1997**, 97, 119-124.

[33] (a) Boys, S. F.; Bernardi, F. Mol. Phys. **1970**, 19, 553. (b) Simon, S.; Duran, M.; Dannenberg, J. J. J. Chem. Phys. **1996**, 105, 11024-11031.

[34] Case, D. A.; Darden, T. A.; Cheatham, III, T. E. *et al.* AMBER 10, University of California, San Francisco, 2008.

[35] (a) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Phys. Commun. **1995**, 91, 1-41. (b) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr., Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. Computat. Chem. **2005**, 26, 1668-168.

[36] Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. J. Phys. Chem. **1993**, 97, 10269-10280.

[37] Wang, J; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. J. Comput. Chem. **2004**, 25, 1157-1174.

[38] Jarzynski, C. Phys. Rev. Lett. **1997**, 78, 2690.

[39] Pratt, L. R.; Chandler, D. J. Chem. Phys. **1980**, 73, 3434-3441.

[40] Chandler, D. Nature **2005**, 437, 640-647.

Chapter 3

Failures of Embedded Cluster Models for pK_a Shifts Dominated by Electrostatic Effects

3.1 Introduction

Accurate prediction of protein residues' pK_a values, which determine the protonation state of ionizable residues at a given pH, is very important since these residues are involved in intraprotein, protein-solvent, and protein-ligand interactions. Consequently, the protonation state of a protein makes significant contributions to protein stability, solubility, folding, binding ability and catalytic activity.[1] [2] Therefore it is

impossible to build a computational model for a protein, or even to obtain a qualitative understanding of protein structure and function, without first determining the protonation state of the protein. While the protonation state can be determined empirically, in principle, experimental measurements (mostly using NMR) are difficult for large proteins, buried residues, and active-site residues. This has motivated the development of computational methods to predict pK_a values for protein residues.[1] [3] - [16] Unfortunately, these methods are often unreliable for buried residues.

The most commonly used methods for protein pK_a prediction are based either on the numerical solution of Linearized Poisson-Boltzmann equation (LPBE) or on fast empirical methods built from quantitative structure-property relationships (QSPR). (The main challenge for developing empirical methods is that the specific properties that determine protein pK_as are still not clear.[17] [18]) For computational expediency, methods based on both of these approaches usually model the protein with a molecular mechanical force field, immersed in a dielectric continuum. The dielectric constant of the solvent is set to 80, and for the protein a value between four and twenty is used. The shift in pK_a for the residue is calculated by comparing the electrostatic energy of its protonated and non-protonated forms. Then the shift is added to a model pK_a value. The structure of the protein is assumed to not undergo any change upon (de)protonation, which is often a poor assumption, especially for buried residues.[19] [20] Methods based on these traditional approaches often fail to predict the pK_a of buried residues and those which have large pK_a shift.

Typical molecular mechanics models for electrostatic energies are limited in accuracy by the assumption that atoms of a certain type always have the same charge.[1] [21] - [23] This unrealistic simplification of electrostatic interactions can be overcome using flexible-charge molecular mechanics models [24] - [30] or, ideal, quantum mechanics (QM).[31] However, due to its computational cost, QM methods cannot be applied to the whole protein. The cluster approach—wherein the ionizable residue and its immediate environment are taken out of the protein and treated quantum mechanically—hurdles this obstacle. Since it was introduced two decades ago, the cluster approach has been successfully used to model enzyme reactions and the number of applications for it is growing.[32] - [34]

The cluster approach has been used by Li and coworkers to determine the pK_a of protein residues.[35] In their approach, the part of the protein that includes the ionizable residue and its immediate environment is treated by QM, while the rest of the protein is neglected. The solvent is described by polarized continuum solvation model. The method successfully predicted the pK_a of five residues in the turkey ovomucoid third domain (OMTKY3). The approach was also applied successfully to predict pK_a values for small organic molecules [36] - [40] and protein residues near metal atoms. However all the residues were on surface of the enzyme, where the interactions are dominated by hydrogen bonding.

We wondered whether the cluster method would work when electrostatic and empirical pK_a prediction methods fail. To this end, we have tested the cluster method for

two aspartate residues—one of which is buried in the protein interior, and one of which is solvent-exposed—in which the pK_a of the carboxyl group is strongly affected by nearby charged residue(s).

3.2 Computational Details

3.2.1 pK_a calculations

We followed the method developed by Li and coworkers to calculate the pK_a for a protein residue.[24] Using the reaction,

$$HA(aq) + C_2H_5COO^-(aq) \xrightleftharpoons{\Delta G} A^-(aq) + C_2H_5COOH(aq), \qquad (3.1)$$

the pK_a value of a carboxyl group in the protein, HA, can be determined from the equation

$$pK_a = 4.87 + \Delta G/1.36\ , \qquad (3.2)$$

where 4.87 is the experimental pK_a value for propionic acid at 298 K and 1.36 is 2.303*RT* at T = 298 K. ΔG is the change in the standard free energy of reaction (3.1) in kcal/mol. To compare the accuracy of the quantum mechanical cluster model to more conventional approaches, we used two continuum electrostatic methods, the web-based version of KARLSBERG+ [41] and the MCCE code.[42] - [44] To represent the empirical methods, we used the popular PROPKA program.[45]

3.2.2 Free energy calculations

All quantum mechanics calculations were performed with the Gaussian 09 program.[46] Solvation free energies calculations were performed with the Gaussian 03 program.[47]

The free energy of a molecule is given by

$$G = E_{\text{ele}} + G_{\text{sol}} \ . \tag{3.3}$$

The electronic energies, E_{ele}, were computed at the MP2/6-31+G(2d,p) level, using structures that were optimized at RHF/6-31+G(d) level of theory. The solvation free energies, G_{sol}, were calculated by the polarizable continuum model at the IEF-PCM/RHF/6-31+G(d) level of theory for the same geometries, using the keywords RADII=UAHF, RET=100, TSNUM=240, and SCFVAC. To account for conformational flexibility in the cluster, all possible conformers of the acid were considered and the final free energy was computed by Boltzmann weighting all conformers within 1 kcal/mol of the most stable conformer. I.e.,

$$G = -RT \ln\left[\sum_{\text{conformers}} \exp\left(-G_i / RT\right) \right] . \tag{3.4}$$

3.2.3 Protein model construction

For each model, the coordinates for the atoms were taken from the PDB files downloaded from the PDB database. The files were stripped of metal atoms, cofactors, inhibitors, and solvent molecules. Atomic coordinates for the model for Asp75 was taken

from pdb code: 1a2p bacterial barnase.[48] Atomic coordinates for the model for Asp21 were taken from pdb code: 1beo, which is fungal beta-cryptogein.[49] Hydrogen atoms were added using the web-based program molprobity software.[50] Where the residue was truncated, additional hydrogen atoms were added manually to satisfy the valence. The protein backbone was fixed and all the hydrogen atoms were optimized at HF/3-21G(d) level of theory; then the side chain (CH_2COO^- or CH_2COOH) was optimized. One oxygen atom of the carboxyl group was fixed. The positions of the neighboring protons of OH groups were also minimized at the same level of theory.

3.3 Results

3.3.1 Aspartate 75

Figure 3.1 shows the cluster model for Asp75. The model includes the immediate chemical environment of the ionizable residue. Three conformers were found for the protonated form. There are two hydrogen bonds between two hydrogens of Arg83 and the two oxygens of the carboxyl group of Asp75. A pK_a of -4.63 pH units was obtained by the method developed by Li and coworkers, which is much lower than the experimental value (error = -7.4.) Negative results were also obtained by PROPKA and KARLSBERG+, see Table 3.1. Only the MCCE code gave a reasonable prediction, 3.6. One possible explanation is that these methods overestimate the stabilization from the charge-charge interaction between the carboxylate group of Asp75 and two nearby

positively-charged amides: Arg83 and Arg87. These interactions favor the ionized (non-protonated) form, leading to a lower pK_a value. Another possible explanation, suggested by Li and coworkers, is that the position of the Arg83 in the crystal structure is different from its position in solution. To address this possibility, the geometries of Arg83 were optimized at the same level, and the pK_a was recalculated. This refined calculation did not improve the results. The solvation free energy was calculated for different values of dielectric constants, 4 and 20, but this did not improve the results either.

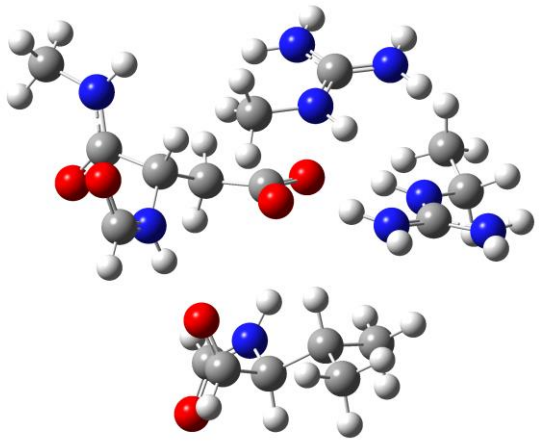

Figure 3.1: Model compound for Asp75 of bacterial barnase. The model contains the ionizable group and its immediate chemical environment

Table 3.1: Comparison between experimental and several pK_a prediction methods.

PDB Code	Residue	Exp. pK_a	Current study	PROPKA[a]	MCCE[a]	KARLSBERG+
1a2p	Asp75	3.1	-4.6	-1.3	3.6	-8.2
1beo	Asp21	2.5	-3.7	1.4	3.6	0.3

[a] Results were obtained from Ref. [64]

3.3.2 Aspartate 21

The model for Asp21 is represented in Figure 3.2. The pK_a shift is primarily due to electrostatic interactions between the carboxyl group and the positively-charged amide group of Lys62. A pK_a of -3.72 pH units was obtained by the method developed by Li and coworkers, see Table 3.1. Lower pK_a values were obtained by the rest of the methods (except for MCCE), which suggests that the explanation for these poor results is similar to the explanation we deduced for Asp75: nearby positively-charged residues cause these methods to overstabilize the ionized form of Asp21.

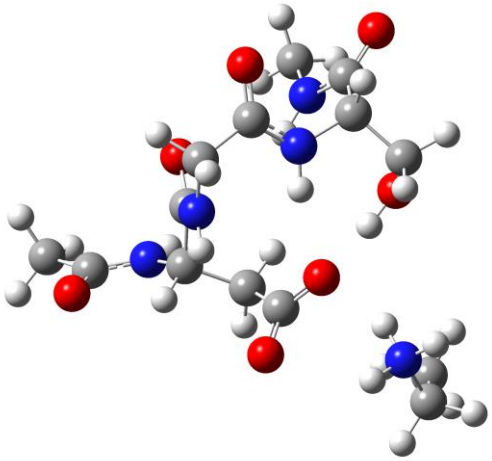

Figure 3.2: Model compound for Asp21of fungal beta-cryptogein. The model contains the ionizable group and its immediate chemical environment

3.4 Discussion

The calculated pK_a values for both aspartate residues we considered were far lower than the experimentally measured values for most of the methods. We attribute this to overestimation of Coulombic interactions between nearby positively-charged residues and the carboxyl group, causing the models to overestimate the stabilization of carboxylate anion. Not even the trends are right: the experimental pK_a value of Asp75 is higher than that of Asp21, but the methods predict the opposite order. This is consistent with our hypothesis: Asp75 has two nearby positively-charged Arg residues, while Asp21 has only one nearby positively-charged Lys residue.

Asp75 is a buried residue. Prediction of pK_a for buried residues is extremely challenging, partly because desolvation is a key determinant of pK_a. The hydrophobic environment stabilizes the neutral form of Asp75, which increase the pK_a of the residue. Desolvation effects were found to raise the pK_a shifts by 4-6 pH units.[51] - [54] A study by Laurents *et al.* reported that present-day continuum models underestimate the contribution of desolvation effect.[55] The inability of continuum solvation models to adequately capture desolvation effects helps explain why Asp75 is erroneously predicted to be more acidic than Asp21 by the computational approaches we consider.

Although Asp21 is a surface residue, the cluster method still underestimates its pK_a value. Again, the likely culprit is the implicit solvation model, the polarized continuum model, which probably fails to effectively screen the charges, giving too large a shift in pK_a values. This observation is consistent with the well-known, but typically ignored, observation that the dielectric screening—a concept from macroscopic electrostatics—is an inappropriate model for short-range electrostatic interactions in the protein environment.[56] - [61] Also, it is known that pK_a shifts are affected by slight changes in protein structure, including those induced by crystal-packing forces.[60] [62] [63] Especially for buried residues, minimization of the protein using a method like QM/MM is necessary to obtain the right structure.

Our results indicate that one needs to develop better continuum models for describing electrostatic interactions in proteins. In addition, explicit water molecules need to be included, as also suggested by a benchmark study.[64] The response of the

neighboring protein residues, due to protein flexibility, to (de)protonation also needs to be included. There may also be other important effects. Indeed, the failure of most of these methods can be attributed to our lack of a deep understanding of what properties of proteins affect the pK_a and the absence of quantitative models for how these properties affect the pK_a.

3.5 Conclusions

The goal of this work was to use a cluster model, similar to those that are widely used to model enzymatic chemical reactions, to predict the pK_a values of amino acid residues in proteins. Results from the cluster model were compared to those from more traditional (and less computationally demanding approaches) including those based on the numerical solution of the Poisson-Boltzmann equation (MCCE and KARLSBERG+) and those based on empirical fitting (PROPKA). We tested the approaches on two aspartate residues whose pK_a-shift is induced by nearby positively-charged residues. One of the residues, Asp75, was buried. The other, Asp21, was solvent-exposed.

The cluster approach predicted pK_a values that were far too low. This contradicts earlier findings from Li and coworkers, where the cluster model gave excellent results. We attribute the inadequacy of the cluster model for our cases to the overestimation of the stabilization effect due to nearby positively charged residues and the underestimation of the hydrophobic effect by the continuum solvation model. It is also possible that

(presumably slight) differences in the structure of the protein in solution, compared to the crystal structure, accounts for part of the discrepancy. Finally, the change in protein geometry upon (de)protonation may be important.

One obvious solution is to increase the size of the cluster model and (ideally) introduce explicit solvent molecules. This is not practical because the number of conformers grows exponentially as the size of the cluster model increases, causing a proportionate increase in computational cost. Approaches based on QM/MM method, where the protein environment is modeled with a molecular mechanics force field, are more promising. Such models would provide a better description of the hydrophobic environment of the ionizable residue, but the computational cost is much higher. QM/MM approaches are not yet routinely practical, but they could be used for difficult residues in important proteins.

3.6 References

[1] Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Biochimica Et Biophysica Acta-Proteins and Proteomics **2006**, 1764, 1647-1676.

[2] Schlick, T. Molecular modeling and simulation: An interdisciplinary guide, Springer, New York, 2002.

[3] Burger, S. K.; Ayers, P. W. J. Comput. Chem. **2011**, 32, 2140-2148.

[4] Burger, S. K.; Ayers, P. W. Proteins **2011**, 79, 2044-2052.

[5] Fogolari, F.; Brigo, A.; Molinari, H. J. Mol. Recognit. **2002**, 15, 377-392.

[6] Bashford, D.; Karplus, M. Biochemistry **1990**, 29, 10219–10225.

[7] Yang, A. S.; Gunner, M. R.; Sampogna, R.; Sharp, K; Honig, B. Proteins **1993**, 15, 252–265.

[8] Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. J. Mol. Biol. **1994**, 238, 415–436.

[9] Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K.; McCammon, J. A. J. Comput. Chem. **1996**, 17, 1633–1644.

[10] Sham, Y. Y.; Chu, Z. T.; Warshel, A. J. Phys. Chem. B **1997**, 101, 4458–4472.

[11] Delbuono, G. S.; Figueirido, F. E.; Levy, R. M. Proteins **1994**, 20, 85–97.

[12] Warshel, A.; Sussman, F.; King, G. Biochemistry **1986**, 25, 8368–8372.

[13] Kollman, P. Chem. Rev. **1993**, 93, 2395–2417.

[14] Mehler, E. L.; Guarnieri, F. A. Biophys. J. **1999**, 77, 3–22.

[15] Sandberg, L.; Edholm, O. Proteins **1999**, 36, 474–483.

[16] Matthew J. B.; Gurd, F. R. N. Methods Enzymol. **1986**, 130, 413–436.

[17] Forsyth W. R.; Antosiewiez, J. M.; Robertson, A. D. Proteins **2002**, 48, 388–403.

[18] Edgcomb, S. P.; Murphy, K. P. Proteins **2002**, 49, 1–6.

[19] Swails, J. M.; Roitberg, A. E. J. Chem. Theory Comp. **2012**, 8, 4393-4404.

[20] Meng, Y. L.; Roitberg, A. E. J. Chem. Theory Comp. **2010**, 6, 1401-1412.

[21] Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Biophys. J. **2008**, 95, 1080-1088.

[22] Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z. J. Am. Chem. Soc. **2009**, 131, 8636-8641.

[23] Lee, L. P.; Cole, D. J.; Skylaris, C.-K.; Jorgensen, W. L.; Payne, M. C. J. Chem. Theory Comp. **2013**, 9, 2981-2991.

[24] Mortier, W. J. **1987**, 66, 125-143.

[25] Mortier, W. J.; Ghosh, S. K.; Shankar, S. J. Am. Chem. Soc. **1986**, 108, 4315-4320.

[26] Mortier, W. J.; Vangenechten, K.; Gasteiger, J. J. Am. Chem. Soc. **1985**, 107, 829-835.

[27] Chelli, R.; Procacci, P.; Righini, R.; Califano, S. J. Chem. Phys. **1999**, 111, 8569-8575.

[28] Nistor, R. A.; Polihronov, J. G.; Muser, M. H.; Mosey, N. J. J. Chem. Phys. **2006**, 125.

[29] Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. J. Chem. Phys. **2013**, 138.

[30] van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. J. Phys. Chem. A **2001**, 105, 9396-9409.

[31] Helgaker, T.; Jørgensen, P.; Olsen, J. Modern electronic structure theory, Wiley, Chichester, 2000.

[32] Siegbahn, P. E. M.; Himo, F. J. Biol. Inorg. Chem. **2009**, 14, 643-651.

[33] Siegbahn, P. E. M.; Borowski, T. Acc. Chem. Res. **2006**, 39, 729-738.

[34] Siegbahn, P. E. M.; Blomberg, M. R. A. Chem. Rev. **2000**, 100, 421-437.

[35] Li, H.; Robertson, A. D.; Jensen, J. H. Proteins **2004**, 55, 689-704.

[36] Lim, C.; Bashford, D.; Karplus, M. J. Phys. Chem. **1991**, 95, 5610–5620.

[37] Chen, J. L.; Noodleman, L.; Case, D. A.; Bashford, D. J. Phys. Chem. **1994**, 98, 11059–11068.

[38] Richardson, W. H.; Peng, C.; Bashford, D.; Noodleman, L.; Case, D. A. Int. J. Quantum Chem. **1997**, 61, 207–217.

[39] Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. J. Phys. Chem. A **1997**, 101, 10075–10081.

[40] Li. H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. J. Phys. Chem. B **2002**, 106, 3486–3494.

[41] Kieseritzky, G.; Knapp, E. W. Proteins **2008**, 71, 1335-1348.

[42] Song, Y.; Mao, J.; Gunner, M. R. J. Comp. Chem. **2009**, 30, 2231-2247.

[43] Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. Biophys. J. **2002**, 83, 1731-1748

[44] Alexov, E.; Gunner, M. R. Biophys. J. **1997**, 74, 2075-2093.

[45] Li, H.; Robertson, A. D.; Jensen, J. H. Proteins, **2005**, 61, 704-721.

[46] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. *et al*, Gaussian 09, Revision C.01, Gaussian, Inc., Wallingford CT, 2009.

[47] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. *et al*, Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford CT, 2004.

[48] Oliveberg, M.; Arcus, V. L.; Fersht, A. R. Biochemistry **1995**, 34, 9424–9433.

[49] Gooley, P. R.; Keniry, M. A.; Dimitrov, R. A.; Marsh, D. E.; Gayler, K. R.; Grant, B. R. J. Biol. NMR **1998**, 12, 523–534.

[50] Davis, I. W. *et al*. Nucleic Acids Research **2007**, 35, W375-W383.

[51] Mehler E. L.; Fuxreiter, M.; Garcia-Moreno, B. Biophys. J. **2002**, 82, 1748.

[52] Mehler, E. L.; Fuxreiter, M.; Simon, I.; Garcia-Moreno, E. B. Proteins **2002**, 48, 283–292.

[53] Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; Garcia-Moreno. B. Biophys. J. **2000**, 79, 1610–1620.

[54] Lambeir, A. M.; Backmann, J.; Ruiz-Sanz, J.; Filimonov, V.; Nielsen, J. E.; Kursula, I.; Norledge, B. V.; Wierenga, R. K. Eur. J. Biochem. **2000**, 267, 2516–2524.

[55] Laurents D. V.; Huyghues-Despointes, B. M. P.; Bruix, M; Thurlkill, R. L.; Schell, D.; Newsom, S.; Grimsley, G. R.; Shaw, K. L.; Trevino, S.; Rico, M.; Briggs, J. M., Antosiewicz, J. M.; Scholtz, J. M.; Pace, C. N. J. Mol. Biol. **2003**, 325, 1077–1092.

[56] Berkowitz, M. L.; Bostick, D. L.; Pandit, S. Chem. Rev. **2006**, 106, 1527-1539.

[57] Mehler, E. L.; Guarnieri, F. Biophys. J. **1999**, 77, 3–22.

[58] Sandberg, L.; Edholm, O. A. Proteins **1999**, 36, 474–483.

[59] Matthew, J. B.; Gurd, F. R. N. Methods Enzymol. **1986**, 130, 413–436.

[60] Nielsen, J. E.; Vriend, G. Proteins **2001**, 43, 403–412.

[61] Schutz, C. N.; Warshel, A. Proteins **2001**, 44, 400–417.

[62] Nielsen J. E.; McCammon, J. A. Protein Sci. **2003**, 12, 313–326.

[63] Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. Biophys. J. **2002**, 83, 1731–1748.

[64] Stanton, C. L.; Houk, K. N. J. Chem. Theory Comp. **2008**, 4, 951-966

Chapter 4

Conclusions

This thesis develops and tests computational models for protonation and deprotonation reactions in biological and supramolecular systems.

Chapter 2 of the thesis presents a systematic study of a supramolecular structure based on a molecular tweezer that has been proposed as a drug carrier. The goal of our study was to elucidate the dynamics of interaction within the tweezer and between the tweezer and substrate.

Our study supports the experimental evidence that the molecule is pH-responsive. We showed that the molecule is flexible with a small barrier to rotation. We also showed that the shape of the molecule depends on the protonation state of its pyridine ring. Upon

protonation the molecule switches from the closed (U-shaped) to the open (UW-shaped) conformation. Quantum mechanics calculations showed that the molecule is capable of holding an aromatic substrate between its two arms through noncovalent π-π stacking interactions. Molecular dynamics simulations allowed us to observe at atomistic details the dynamics of shape-switching and substrate-releasing processes. The binding free energy of quinizarin, an anticancer drug, to the tweezer was calculated by steered molecular dynamics. The binding free energy is moderate, demonstrating that the molecule has a potential as a drug carrier to tumoral tissues, which tend to have lower pH than healthy cells. The flexibility of the molecule, the response to pH change, and the binding free energy are all tunable by introducing electron withdrawing/donating groups to the tweezer. The computational approaches we established for this molecule can be used to assess the effects of adding substituents to the tweezer, thereby streamlining the drug development process.

Chapter 3 studies the cluster model approach for predicting protein residues pK_a values. Predicting the pK_a values for buried and active site amino acid residues in proteins is important since these residues are involved in intraprotein, protein-solvent, and protein-ligand interactions. Therefore, protein conformation and stability depends on its protonation state. Computational models for the protein pK_a's are especially important for buried amino acids, as experimental data for pK_a values of buried protein residues is scarce.

The two residues that were chosen for this study are particularly challenging, and traditional methods, based on numerical solution of the Poisson-Boltzmann equation and empirical fitting, fail. We treated the residue and its immediate chemical environment using quantum mechanics, while the rest of the protein was replaced by a continuum dielectric. This method had been previously applied successfully to predict pK_a values of surface residues of turkey ovomucoid third domain. However, for the residues we studied, the cluster model predicts pK_a values much lower than those observed experimentally. The inadequacy of the cluster model can be attributed to the overestimation of the stabilizing effect of electrostatic interactions and the underestimation of the hydrophobic effect by the continuum solvation model.

The study highlights the importance of protein flexibility in the simulation. Including the bulk of the protein is important for good results, which is possible through the hybrid QM/MM methods, since increasing the size of the cluster is computationally too demanding. The determinants of pK_a shifts in protein need more investigation. The study also, along with other studies, indicates the need for developing better (de)solvation models.

In addition to the work reported in this Master's thesis, I have worked to calculate nonlinear optical properties of molecules using the finite field (FF) method. These properties are important for the development of optical devices, studying long range interactions, and characterizing solvation and the interactions between molecules and ionic reagents.

FF methods are straightforward, easy-to-implement techniques for calculating electric response properties. They are cheaper than other methods (e.g., response theory and the sum over states method). However, one of the major drawbacks of FF methods is the dependence of results on the chosen field strength. In particular, choosing a field strength that is too small gives inaccurate results due to round-off errors; choosing a field strength that is too large gives inaccurate results because higher-order responses become dominant.

We showed that for a large data set of 120 molecules, for the second hyperpolarizability, choosing a field sequence smaller than two gives better results than using a value of 2, as universally used in the literature. We also showed that Richardson extrapolation is superior to polynomial fitting for refining the calculated response properties and only two steps of refinements were necessary to reach the maximum precision. The results of this study have already published in the Journal of Computational Chemistry.

An open question so far is, how an optimal field strength should be chosen for a particular molecule. We developed a simple protocol to predict the optimal field strength to calculate the second hyperpolarizabilities. The protocol depends only on the maximum nuclear distance within the molecule and was applied successfully for a wide range of molecules.

Our findings are currently being extended to include lower derivatives, i.e. dipole polarizabilities and the first hyperpolarizabilities. We will also extend the approach to different levels of theory.

In summary, this thesis focused on studying protonation reactions of biochemical relevance, both in terms of drug delivery (chapter 2) and protein structure and function (chapter 3). Chapter 3 points out limitations in the cluster model for protein pK_a's (i.e., don't use that method for buried residues) and Chapter 2 elucidates the binding mechanism of a pH-sensitive molecular tweezer.