Computer Based Analysis of Proteins and their Interactions with the Solvent and Surfaces

Martin Hoefling

SIGILL. UNIVERSITAT. LUDOVIC. MAXIM.

Munich 2007

Computer Based Analysis of Proteins and their Interactions with the Solvent and Surfaces

Martin Hoefling

Diploma Thesis
Faculty of Physics
Ludwig–Maximilians–University
Munich
submitted by
Martin Hoefling

at

Chair for Applied Physics
Prof. Dr. H. E. Gaub

Munich, December 20, 2012

First review by: Prof. Dr. H. E. Gaub

Second review by: Prof. Dr. E. Frey

Abstract

Interactions of proteins with their environment are of fundamental interest for understanding biological and hybrid systems consisting of biological and inorganic compounds. The environment of a protein is composed of solvent, other proteins and surfaces.

Protein-protein interactions are essential for many cellular processes. For transient complexes, electrostatic steering has an important contribution to the association of proteins. This contribution dependens on the physical properties of the solvent. Therefore, an interesting mechanism to fine-tune the complexation rates or to confer specificity would be to influence the physical properties of the solvating water. To test the influence on their environment for the facilitation of complexation, I examined water properties between approaching proteins using constrained molecular dynamics simulations of the solvated Barnase-Barstar complex.

Applying the Shannon Entropy formalism to the dipole angles of water with the vector connecting the centers of the two proteins showed significant deviations from a random distribution for all separation distances. Interestingly, at a separation distance of 10Å, a distance commonly associated with the Encounter Complex (EC) of the complexation process, the Shannon entropy has a pronounced local minimum. In simulations with mutants, which have been shown to destabilize the EC, this minimum is not observed.

Furthermore, the analysis of the water dipole vectorfields showed up stable dipole fields even at a separation distances of 20Å for the wildtype protein, but not for the mutant. The increased water ordering and the stable dipole fields at large distances suggest a significant contribution of water structure for binding processes and to electrostatic steering in analogy to the hydrophobic effect, even though the binding sites are not more hydrophobic than the rest of the protein surface.

Inducing ordered water might constitute a general mechanism, which facilitates the complexation of transient protein-protein complexes. Using this property, I tried to predict interface regions based on the surrounding water properties. Since water analysis in a trajectory is computationally expensive, a combination of parallelization and index optimizations are required. Initial results show a coincidence of ordered water with known interface regions. Yet it appears as if geometrical constraints stemming from the protein's surface shape are the predominant ordering factor and not the aminoacid composition at the binding site. Interestingly, a higher reliability of prediction on enzymes compared to the inhibitors was observed.

In biological circumstances, proteins interact not only with other proteins, but also with inorganic surfaces like bones. Compatibility with nonbiological surfaces and classification of protein-surface interactions is of increasing importance for nanotechnology and drug design. Yet, a physical understanding for these interactions is currently lacking. For a better characterization of the interaction of proteins with surfaces, I describe a general implementation of a force field for simulation of proteins on 111 oriented gold surfaces developed in cooperation with the Theoretical Nanoscience group of the University of Modena. Initial simulations of gold surfaces and nanocrystals show that the behavior of interactions with biomolecules can be described with a single set of van-der-Waals parameters combined with a dipole modeling of coulombic interactions in vacuum. First simulations in water are underway.

Contents

It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature.

Niels Henrik David Bohr

Chapter 1

Introduction

Transient protein-protein complexes regulate the cellular life. Therefore, the complexation and the recognition of the complex partners needs to be fine-tuned, specific and reliable. Protein-protein recognition can be divided into three steps: Free diffusion, steered diffusion and dewetting. Some proteins reach the diffusion limit for complexation[1, 2], despite the fact that the rotational degrees of freedom complicate the recognition of the respective binding sites[3]. This is only possible if the proteins are pre-aligned before contact. In order to enable a pre-orientation, the information about the position and orientation of the binding sites has to be encoded in the proteins architecture[4, 5] and needs to be transmitted through space. Such a design results in an association for nearly each collision. These effects are observable in the second regime of protein complexation, the steered diffusion[6]. Hence, I try to understand the principles of steering and the role of the solvent molecules involved in this process.

One way of encoding information within the protein architecture is the specific arrangement of charged amino acids in the vicinity of the binding site[4]. This leads to complementary electrostatic fields around and between the protein partners. These fields cause a pre-orientation of the proteins with an accompanying increase in complexation speed – an effect usually called electrostatic steering[2, 7]. The electric fields lead to an increase of probability to find the complexation partners in a preferred orientation.

Recent studies classifying conformational entropy during complexation[8] focus on the protein structure and not the separating solvent. Conformational entropy and thus flexibility of proteins shows an ambivalent behaviour during complexation[8]. A thorough understanding of the influence of the electrostatic fields on the solvating water is currently lacking. Yet, recent studies have shown that a structural modulation of water by solvated specimen can greatly influence the physical properties of the solvent[9–11]. Despite the potential modulation of the water properties by the approaching proteins, the importance of the solvating water in protein-protein recognition has commonly been underemphasized. Only recently, the key role of water in protein complexation has been recognized[12, 13]. The dewetting transition of constitutive complexes has lately been examined using molecular dynamics simulations[14, 15]. Furthermore, it has been shown that so-called de-hydrons are overrepresented at binding sites [13], which might facilitate the dewetting transition. Here, we are interested in the question how solvating water influences the recognition process of protein complexes at an early stage of complexation, the steered diffusion. How can water influence protein recognition? It has been shown that stable dipole fields form between two solvated, specifically oriented, charged aminoacids[16]. If similar dipole fields are also formed between approaching proteins, the overlapping, complementary dipole fields can lead to a weak pre-alignment of the proteins at long distances. Furthermore, an increased order of water forming dipole fields leads to an entropic gain during displacement of this water. Additionally, ordered water has a lower dielectric susceptibility[10], potentially leading to an improved electrostatic steering.

Here, I point out that indeed approaching proteins form stable dipole fields between each other even at protein-protein distances in the order of 20Å, exceeding the region commonly assumed for electrostatic steering[17]. Thus, long-range propagation of dipole-dipole interactions can facilitate protein complexation. Mutations that disturb complexation lead to a loss of the dipole field. Furthermore, it is possible to show by using the Shannon definition for entropy and information[18, 19], that the information content of the water between the proteins is higher than the information content at other places. This is in accordance with results on monomers [Hadas Lapid and Kay E. Gottschalk, submitted], which show that binding site water is more ordered than non-binding site water. Hence, information from the protein binding sites is transmitted to the solvent. In addition, the order of the water is rising with decreasing protein-protein distance, so that the entropic gain of displacing this water as well as the effect on electrostatic steering is increasing accordingly.

The results of water orientation between complexes suggest a further analysis of the water structure around the complex forming monomers itself[20]. A first aim here is to find a measure of water orientation and then try to use this measure for the prediction of interfaces. For this purpose an extensive analysis of the whole surrounding water is necessary creating additional difficulties coming from the complicated geometry of the protein surfaces[21–23]. To use the solvent properties for the prediction of interfaces, I analyzed various monomers for which at least one binding partner is known. Due to the high computational requirements parallelization and other computational optimizations become necessary for analysis and were implemented.

While interactions between biological subunits are designed by evolutionary processes, compatibility towards inorganic materials is a recent requirement of emerging nanotechnology. Studying biocompatibility of surfaces is of increasing importance for rational drug design, nanoelectrics and biomaterial science, yet lacking generic tools as are available for purely biological systems. In order to provide tools that can be used to design surface-compatible proteins, we became involved in the collaborative effort called PROSURF[1]. The target of PROSURF is to interdisciplinary develop a computational toolbox for protein surface docking (see flowchart Fig. 1.1). This involves knowledge on several fields, provided by the participating groups of PROSURF.

Studies of interactions between proteins and inorganic surfaces materials are also important in natural systems[24]. They can regulate crystal growth to prevent freezing in cells[25]. Recently it was revealed in combination of different biotechniques[26] that binding towards a given inorganic surface can be tuned to exploit protein-surface interaction specificity. However, the surface binding mechanism has not been understood. Which parts of the surface are determining protein binding? Is it possible to design surfaces to interact with specific proteins? There are very few theoretical considerations of these questions[24], and methods of general applicability are still absent.

Since parameter studies are computational cumbersome, we foremost limit the focus to (111) gold surfaces. Gold is a versatile surface for bio-electronics[27] and other biological applications[28].

Although biomolecules on gold have been simulated in various configurations[29], a generic implementation, reproducing experimental data is currently lacking. Either, water influence is neglected[30–32] or a very specific subset of molecules[33–35] is chosen. Some simulations only address (covalent) parts of the interactions[34, 36] and neglect other contributions to the interactions. Also, a vast number of studies consider monolayer assembly of alkanethiols[37–39] and other molecules[29, 40–42] on gold surfaces. A generic surface model, applicable to a variety of biomolecules

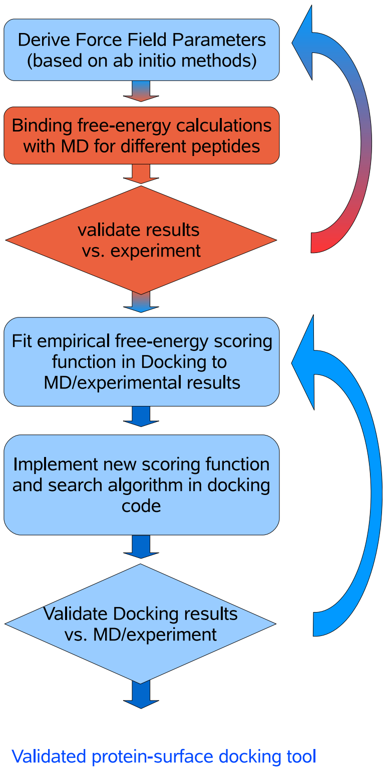


Figure 1.1: PROSURF workflow, the focus of our group is marked in red.

---

[1] http://www.s3.infm.it/prosurf/

as well as energetic information of adsorbed proteins[43] is still not available. Also, recent experiments[44, 45] examining adhesion of amino acids on inorganic surfaces, only surface concentrations and not adsorption energies could be measured.

Here, the idea emerged to extend available force fields for protein docking and molecular dynamics simulations to reproduce interactions between proteins and gold (111) oriented surfaces. To overcome the computational limits of Molecular Dynamics Simulations, a generic docking-like protocol allowing effective exploration of protein-surface orientation should be developed. The framework for these studies is the PROSURF project, trying to develop parameters in cooperation with the quantum chemical group of Prof. Elisa Molinari [1]. Experimental feedback from SPR[2] and other experimental setups is provided by Prof. Gideon Schreiber's group[3]. In the future, AFM[4] experiments are planned at our group. Our focus now is the simulation of peptides and proteins on gold to classify the interaction and provide a feedback to parameter derivation via ab initio methods and experimental results, as described in flowchart Fig. 1.1 on the previous page.

Molecular Dynamics Studies can be used to obtain free energy profiles during a binding process by evaluating the potential of mean force[46–49] with respect to a reaction coordinate such as distance to the surface. The results will provide the input for the development of a generic docking-protocol at the group of Dr. Rebecca Wade (EML, Heidelberg).

---

[1] University of Modena

[2] Surface Plasmon Resonance

[3] Weizmann Institute of Science, Rehovot, Israel

[4] Atomic Force Microscopy

God does not care about our mathematical difficulties. He integrates empirically.

Albert Einstein

Chapter 2

Theory

2.1 Molecular Dynamics Simulations

MD[1] simulations provide a bottom up tool for biological system analysis. The basic idea is to reduce complexity of ab initio theory to the relevant interactions, so that computation of time development is possible with recent computer systems, at least for smaller biological systems. In the optimal case, the time-dependant relativistic Schroedinger equation describes the evolution of a system in time. Unfortunately, the complexity at an ab initio level prohibits its usage on systems of scales larger than a few atoms. Thus, evaluation with classical mechanics is chosen, combined with an empirical parameter set describing the interactions (force fields).

Calculations are done by solving newtons equations of motion (2.1 on the following page) for each timestep and each interacting particle. Derivation of a potential function (2.2 on the next page) $V(r_1, \ldots, r_N)$, provides the forces on the particle in

---

[1] Molecular Dynamics

each step.

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, i = 1 \ldots N \tag{2.1}$$

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \tag{2.2}$$

The coordinates as a function of time can be seen as the trajectory of the system, by classical means. NMR spectroscopy or x-ray diffraction are used to provide initial coordinates.

2.1.1 Force Fields

Empirical parameter sets (force fields) define the behavior of the system during simulation. Force fields can lack some atoms or even introduce some virtual ones. There are three types of force fields:

- **All Atom** force fields include parameters for all atoms in a system.
- **United Atom** force fields merge hydrogens with its corresponding carbon atom in methyl and methylene groups
- **Coarse Grained** force fields do some severe modifications on the particle representations of building blocks

In my studies, I used the OPLS-AA[1][50–53] for trajectory calculations, some older trajectories were calculated with GROMOS96[54–56], an united atom force field derived from GROMOS-87[57, 58].

In the following, I will introduce the mathematical description of the force field mechanics. Force fields differ in parameters and in their potential function definitions.

---

[1] Optimized Potential for Liquid Simulations - All Atom

The potential can be expressed as a sum of three contributions:

- **Non-Bonded**: Lennard-Jones interaction or coulomb potential
- **Bonded**: Bond stretching (2-body), bond angle (3-body) and dihedral angle potentials (4-body)
- **Constraints and Restraints**: e.g. Position or distance restraints (see 2.2.1 on page 12)

Bonded

Bond stretching is often modeled with a harmonic potential (2.3). If anharmonicity is needed, e.g. for zero force at infinite distance, the Morse potential[59] can be used.

$$V_b(r_{ij}) = \frac{1}{2}k^b_{ij}(r_{ij} - b_{ij})^2 \tag{2.3}$$

The bond angle potential is evaluated analogous with θ_{ijk} as parameter. Proper dihedrals[1] are expressed as a cosine power series known as Ryckaert-Bellmans potential (2.4).

$$V_{rb}(\psi_{ijkl}) = \sum_{n=0}^{5} C_n(cos(\phi))^n, \text{with } \psi = \phi - 180^\circ \tag{2.4}$$

Non-Bonded

Nonbonded interactions can be divided into interactions of charged and uncharged atoms. They have a radial symmetry and are pairwise additive. Due to computational advantages, the Lennard-Jones (12-6) potential (2.5 on the next page) is preferred to the Lennard-Jones (exp-6) potential, also known as Buckingham poten-

---

[1] there are also improper dihedrals, preventing molecules from flipping to their mirror image, which are handled also in a harmonic potential $V_{id}(\xi_{ijkl}) = k_\xi(\xi_{ijkl} - \xi_0)^2$

tial (2.6) for uncharged atom interactions.

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \tag{2.5}$$

$$V_{Bh}(r_{ij}) = A_{ij}exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^{6}} \tag{2.6}$$

The coefficient matrices ($C_{ij}^{(x)}$ or B_{ij}, A_{ij}) are set up during runtime by combination rules (2.7) – here shown for the (12-6) potential form – from the coefficients ($C_{ii}^{(x)}$ and $C_{jj}^{(x)}$) of the two atoms.

$$C_{ij}^{(x)} = (C_{ii}^{(x)} + C_{jj}^{(x)})^{\frac{1}{2}} \tag{2.7}$$

Charge interactions are described by Coulomb potentials (2.8), known from electrostatics.

$$V_C(r_{ij}) = \frac{1}{1\pi\epsilon_0} \cdot \frac{q_i q_j}{\epsilon_r r_{ij}} \tag{2.8}$$

In both cases, the potential functions are modified to reach zero at a cutoff distance, which is necessary to prevent infinite potentials in systems with periodic boundary conditions. This cutoff distance, in combination with box vector lengths exceeding the cutoff distance at least twice, guarantees that no particle interacts with it's mirror image.

The original OPLS-AA Lennard-Jones parameters σ and ϵ used in my simulations can be converted into a C_n parameter set[60]. Lennard-Jones calculations for 1-4 pairs (connected over 2 atoms) need to be excluded for Ryckaert-Bellmans potential, other potentials need a special 1-4 interaction treatment.

Particle Mesh Ewald Method

Long range electrostatic calculations produce the most relevant part in computation time ($\mathcal{O}(n^2)$) and can be sped up by the PME[1] method[61–63] ($\mathcal{O}(n \log n)$). Although pure cutoff based electrostatics is faster[64] ($\mathcal{O}(n)$) and not always a loss of accuracy[65], it has been shown that PME leads to more realistic results in various applications[66]. Especially in reproducing the water structure, there are artifacts related to cutoff usage[67]. Similar effects can be found in DPPC bilayer simulations[68].

Calculation of electrostatic energy is a slowly converging sum (2.9), where P is the summation due to periodic images and C is the summation between all particles.

$$V_{el} = \frac{f}{2} \underbrace{\sum_{n_x} \sum_{n_y} \sum_{n_z}}_{\text{P}} \underbrace{\sum_{i}^{N} \sum_{j}^{N}}_{\text{C}} \frac{q_i q_j}{r_{ij,\mathbf{n}}} \tag{2.9}$$

Ewald suggested[61] to convert this single sum into a direct and reciprocal sum $V_{el} = V_{dir} + V_{rec} + V_0$. This decomposition allows to apply a short cutoff in the direct as well as in the reciprocal sum.

Particle-Mesh Ewald has been suggested by Tom Darden[63] to improve calculation of the reciprocal sum. Virtual particles are placed on a grid instead of summing up the wave vectors directly. The grid is then fourier transformed in k-space, where the energy can be calculated from a single sum.

In my simulations PME has been used to avoid artifacts from cutoffs exclusively. The additional computational costs for PME compared to a simple cutoff based approach is moderate, particularly when highly optimized FFT[2] algorithms and implementations come into play.

---

[1] Particle-Mesh Ewald

[2] Fast Fourier Transformation

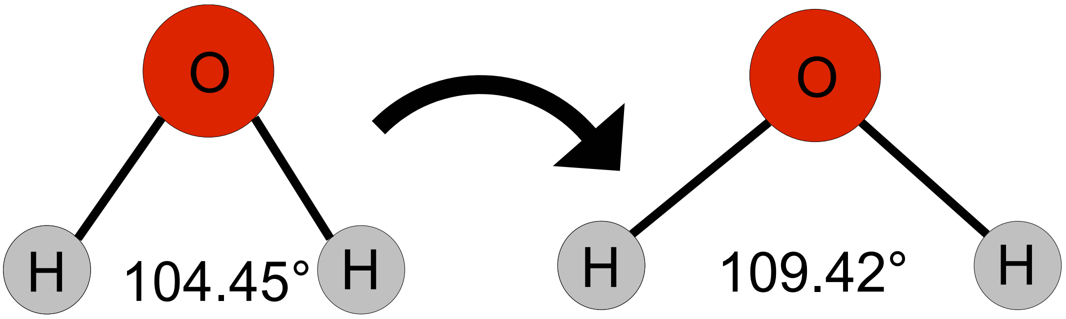


Figure 2.1: SPC-E water model increases the angle beetween H atoms to reproduce known water properties[69].

2.1.2 Water Models

Several water models with different complexities exist. The simplest model, TIP3P[58] reduces water to three charges, located at the water atom positions. Lennard-Jones interactions are calculated only from the oxygen atoms. An improvement of this model is the SPC-E[1][69, 70] water model, which additionally modifies the angle empirically from 104.45°, describing the natural water angle, to the new fixed value of 109.42° (Fig. 2.1) to achieve better compliance with known water properties. Other water models introduce additional point charges (TIP4P, TIP5P) or polarizability (GCPM, POL5/TZ).

Since water contributes the largest computational part in my system, I have chosen SPC-E water. It provides a high accuracy at lowest computational cost[71].

2.2 Constrained Simulations

Simulations with additional constraints provide a tool, to sample a specific subspace of a systems conformational space. The method I used in my simulations is described in section 3.4 on page 19. Constraint algorithms can also be used to constrain the bond lengths for all atoms or a specific subset.

---

[1] Single Point Charge - Extended

2.2.1 Constraints in Molecular Dynamics Simulations

Constraints in MD simulations are typically resolved by either LINCS[72], SHAKE[73] or SETTLE[74] (analytical version of SHAKE, used for rigid water models) algorithm. SHAKE has been applied to constrain the protein-protein distance in my simulations and thus shall be discussed further.

While restraints only add a potential, e.g. during simulations of AFM pulling, a set of constrained coordinates needs to fullfill one or more distance and/or angular constraints in each step. A transformation has to be found to replace the unconstrained coordinates with constrained ones. This can be achieved by solving a set of Lagrange Multipliers λ_r for R holonomic constraints (2.10) providing the necessary forces (2.11).

$$\psi_r(\mathbf{r_1}, \dots \mathbf{r_N}) = 0 \quad \text{with} \quad r = 1 \dots R \tag{2.10}$$

$$F_i = -\frac{\partial}{\partial \mathbf{r_i}} (V + \underbrace{\sum_{r=1}^{R} \lambda_r \psi_r}_{\text{constraint force}}) \tag{2.11}$$

These equations can be evaluated in a set of equations of the second degree in the leap frog or Verlet integration scheme used in most MD simulations. Due to numerical solvation of this set of equations, a tolerance needs to be set.

2.2.2 Constraint Force vs. Thermodynamic Force

Binding properties are described by a free energy profile or the PMF[1] along a reaction coordinate. Thermodynamic potentials have the dimension of energy and can be derived along the reaction coordinate in analogy to potentials in classical mechanics to obtain the thermodynamic forces[75].

An interesting approach is to derive the PMF from the constraint forces, which one obtains during constrained simulations. It can be shown[46], that the average

---

[1] Potential of Mean Force

constraint force $\langle f_\alpha^c \rangle_c$ during a simulation equals the negative thermodynamic force allowing to calculate the free energy or PMF along the reaction coordinate from constraint forces through integration. This allows the creation of a free energy profile along the unbinding path characterized through different distances from a set of constrained simulations at fixed COM-COM[1] distances, described in 3.4 on page 19. The concept behind is to show the correspondence between

$$\langle \frac{\partial H_c(\mathbf{Q}^0, \Gamma_c)}{\partial Q_c^0} \rangle_c = \langle f_\alpha^c \rangle_c \tag{2.12}$$

from a thermodynamical integration (2.13). Q^0 are the fixed reaction coordinates while Γ_c represents the remaining set of generalized, free coordinates. H_c and F_c are the Hamiltonian and the free energy of the constrained system.

$$\Delta F_c = \int_A^B dQ_\alpha^0 \langle \frac{\partial H_c(\mathbf{Q}^0, \Gamma_c)}{\partial Q_c^0} \rangle \tag{2.13}$$

Equation 2.13 defines the free energy profile along the reaction coordinate Q_α^0. The equivalence has been shown via Lagrange formalism which leads to an expression with Lagrange multipliers for the constraints[46]. Lagrange and Hamiltonian can be related (shown in detail in Ref. [46]).

Later on, with the description of Jarzynski[76, 77] equality in 1997, similar results emerged from restraints with a stiff spring approximation[47, 78].

2.3 Information and Entropy

The terms entropy, uncertainty, order and information lead often to confusion, when used as synonyms or describing different effects. Here, I intend to clarify terminology for my application. Initially, entropy was introduced in thermodynamics to describe non-reversible parts of a process. Later on the term “Entropy” was used by Shannon

---

[1] Center Of Mass

to describe the results of its information theory[18, 79]. The formalism of different entropy definitions are conceptually closely related. There exist several approaches to show correspondency of various definitions and formalisms[80, 81]. The actual measure of entropy and its meaning depends strongly on the definition. Two of the historically important considerations are sketched below.

2.3.1 Boltzmann-Gibbs Entropy

Entropy was introduced in thermodynamics to describe irreversible parts of a non-adiabatic process. In this context entropy is most often referred to as the Boltzmann-Gibbs entropy.

$$S = -k \cdot \sum_i (P_i \log(P_i)) \tag{2.14}$$

Equation 2.14 has been derived first by Boltzmann in “Lectures on Gas Theory”[82]. Here, P_i states the probability to find a particle in a given microstate. All P_i belong to the same macrostate of the system, corresponding to thermal equilibrium. With the assumption that all reachable states have the same probabilities P_i, it can be written in the known form $S = k \log(\Omega)$, where Ω is the number of reachable microstates. k is an arbitrary constant depending on units which can be chosen as 1 without loss of generality. In thermodynamics, the Boltzmann Constant $k_b = 1.380658 \cdot 10^{-23} \frac{J}{K}$ is used in most cases.

2.3.2 Entropy in Information Theory

Claude Shannon introduced the term “Entropy” in his information theory after discussion with John von Neumann[83, 84]:

My greatest concern was what to call it. I thought of calling it “information”, but the word was overly used, so I decided to call it “uncertainty”. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, “You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.”[83]

Entropy and information strongly depend on the point of view. A noisy channel allows you to transmit a certain amount of information normally measured in bits, depending on channel entropy[19]. On the other hand algorithms minimize needed “space” for information through maximizing entropy in lossless compression algorithms such as LZW[85][1].

Shannon’s lack of information or entropy formulation is exactly the same as Gibbs formulation in equation 2.14 on the preceding page, with different interpretation of the variables.

$$H = -k \cdot \sum_i (P_i \log(P_i)) \tag{2.15}$$

Again, k can be set to 1 without loss of generality. Shannon used H for uncertainty while P_i refers to the probability for the value of a specific information unit in the message, such as a bit. In binary systems, the logarithm to the base 2 can be used with $k = 1$ to receive uncertainty or lack of information in $[bits]$.

---

[1] Lempel-Ziv-Welch

2.3.3 Definitions

In the angular analysis of this thesis (section 4.1.3 on page 28) I used Shannon's formalism, applying it to angular distributions of water molecules. Order and entropy are not generally correlated. For my system, I define order as a more constant orientational behavior of water dipoles over time.

Uncertainty can be considered as lack of information. Therfore its negative gradient can be defined as information gain. One is tempted to see information figurative for a "communication process" between two protein partners. Here, I am neither talking about this kind of information nor am I able to define a channel capacity or equivalent for my system. I use (lack of) information to describe deviations from average behavior.

I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand ...

Lord Kelvin

Chapter 3

Model System

3.1 Structure and Function

Barnase-Barstar is an enzyme-inhibitor system whose first analyzes go back to Hartley in 1972[86]. Initial research focus was the expression and extraordinary reaction properties[87]. In 1993, a x-ray crystal structure with a resolution of 2.60Å was made available by Guillet et al.[88] which has also been used in these studies. In the 90's Buckle, Schreiber and Fersht did extensive analysis[2, 89, 90] of mutants under various external conditions, identifying potentially relevant residues for the binding process[91]. In recent studies, Barnase-Barstar has also been computationally examined by a variety of methods[6, 92–94].

Barnase is an extracellular ribonuclease from bacillus amyloliquefaciens with a sequence length of 110 amino acids. To prevent damage on intracellular DNA, Barnase is synthesized with its inhibitor called Barstar, which binds very tightly in the active site of Barnase. The RNAse activity is inhibited by sterical blocking of the active site of Barnase through the Barstar binding helix. Barstar has a sequence length of 90 amino acids, which results in moderate computational efforts for the whole system size of a few thousand atoms, depending on the force field. Here, the advantage of GROMOS96 united atom towards the OPLS all atom force field is about 30% savings in protein particle count and states no relevant computational contribution in simulations with large water boxes of up to 30000 water molecules.

In addition to its usage as model system of computational and molecular biologists, the extraordinary binding properties of the Barnase-Barstar complex is used in the "green biotechnology". Two applications are transgenic potato plants with self inhibition of fungal disease[95] and hybrid seed production[96] (Patent: DE69635181T, US6344602).

Binding between Barnase and Barstar is mediated via hydrogen bonds and salt bridges. Another advantage for computational studies is the fast kinetics of the process, due to its strong electrostatic steering[3, 7], which allows simulations on a more relevant timescale. The tight binding and fast kinetics near the diffusion limit behavior is explainable through strong electrostatic interactions[7].

Finally, the fact that Barnase-Barstar is a model-system itself is a reason for its usage, since many properties are already known and available for this system.

3.2 Encounter Complex

A point of special interest is a transient state during complexation, also referred to as Encounter Complex. It is thought as an intermediate state without contributing to the total binding energy[92, 97]. Encounter Complexes are stabilized mainly by polar and charged residues[98]. This behavior has also been found for the Barnase-Barstar complex through isothermal titration calorimetry[99]. In this work, the Encounter Complex identification is a side-product of the constrained simulations.

3.3 Mutations

Mutations are an effective tool to identify functionally or structurally important regions of a protein. By modifying charged and polar groups, the electrostatics of a protein can be tuned. For the Barnase-Barstar complex, several residues are known to stabilize the encounter complex[97, 99]. In contrast to experimental studies, introducing a mutation is very easy but the experiment (simulations) itself is very expensive. Therefore I mutated 2 residues on each monomer with an high contribution to EC formation to achieve maximum effect on a second simulation series.

3.4 Constrained Simulations

Despite computational advances, Molecular Dynamics is still limited to the timescale of several nanoseconds. Thus, AFM like simulations[1] differ from the experiments by several orders of magnitude in loading rate potentially leading to completely different observations[100, 101]. The thermal contribution to the process, e.g. unbinding or unfolding, is lowered since the timescale is too fast, compared to typical off-rates and positional relaxation times.

Although protocols like conformational flooding to speed up processes beyond the time scale of standard MD simulations[102] exist, it is still difficult to model slow processes. Conformational flooding adds a potential, destabilizing the current conformation in each step. It has been successfully applied to model conformational transitions, which are unlikely during Molecular Dynamics timescales[103][2].

Yet, choice of determining the PMF through constrainted simulations allows the system to reach rotational relaxation times τ_r during each simulation for unbinding processes like the Barnase-Barstar. A constraint between the centers of the complex constituents is the most accurate way to calculate the PMF compared to other techniques[49] such as umbrella sampling[48, 104]. For unbinding processes of pro-

---

[1] pulling with restraints

[2] yet, the transition duration itself needs to be in MD timescale

tein complexes, the reaction coordinate is well known and described by the distance between the proteins, particular for complexes without conformational change.

Each simulation at a constant separation distance might still not allow full relaxation of the system. Yet, one can estimate the additional relaxation time: First one can assume simulation durations of one single simulation is in the same magnitude as the 21 constrained simulations, due to limited parallelization capabilities of MD. 21 constrained simulations from 0Å to 20Å, each 5ns simulated time, compared to a single conformational flooding or pull simulation of for instance 10ns simulated time, result in $\frac{21\cdot 5ns}{20\mathring{A}}$ versus $\frac{10ns}{20\mathring{A}}$ and gives about one additional order of magnitude of relaxation time. This can be pessimistic if the energy landscape changes homogeneously from the starting point, but the time to reach a newly emerged global or local minimum at a specific distance is increased with the constraint method.

Beside the advantages of PMF calculations, the second focus is the water structure between constituents during complex formation. Analysis of water structure is more realistic in the constrained picture since water is relaxed during binding and unbinding in real systems. Pulling involves an additional non-trivial pressure on water between the constituents. Unconstrained simulations of the whole binding process event would yield ideal conditions during complexation process, however this results in simulation durations about one magnitude ($\approx 0.1\mu s$ see 3.1[1]) above typical simulations on low-end parallel systems. Those simulation times involve high-end supercomputing efforts or massive parallel usage of 100000 CPU cores[105].

Association rate approximation:

Largest simulation box with one complex:

Box volume: $(10nm)^3 = 1000 \cdot 10^{-24} l$

$\rightarrow 10^{21}\frac{1}{l}$ = complexes per l

Association rate constant[2]: $5 \cdot 10^{9}\frac{1}{M \cdot s}$

$$\rightarrow 5 \cdot 10^{9} \cdot 10^{21}\frac{1}{mol \cdot s} \approx 10^{7}\frac{1}{s} \rightarrow 100ns$$

$$[M] = \frac{[mol]}{[l]} \tag{3.1}$$

---

[1] optimistic, since dilution in MD is very low compared to real systems.

These efforts are not related to the expected small differences between the unconstrained and the constrained setup, but have been successfully applied to solve higher dimensional problems, such as protein folding[106]. For analyzing the water properties, complex structures with a gap from 4Å up to 20Å were examined, as shown in Figure 6.1a on page 64 and 4.5 on page 31. The lower limit was chosen corresponding to the thickness of one water layer whereas the upper limit marks the computational limit.

You should call it entropy…

…no one knows what entropy really is, so in a debate you will always have the advantage.

John von Neumann to Claude Elwood Shannon

Chapter 4

Constrained Simulations

4.1 Results

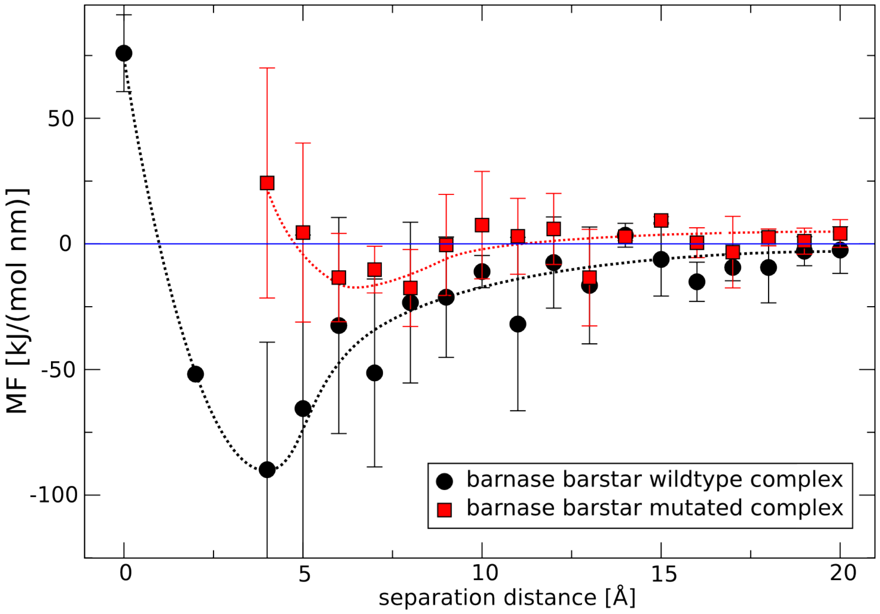
4.1.1 Force - Energy Profile

As described in section 2.2.2 on page 12, the PMF can be easily obtained by averaging the negative constraint force. An analysis of the mutated and wildtype trajectories shows the mean force (Fig. 4.1a on the following page) of a typical binding process and their integrated (PMF) form (Fig. 4.1b on the next page). The non-zero mean force at zero separation emerges from the nonrelaxed crystal structure[88] or from froce field inaccuracies. Since a total of four charged residues are mutated, the electrostatic interaction in total is also significantly lowered compared to the wildtype, which particulary can be observed on larger separation distances in Figure 4.1b. An estimate of the mutant association rate based on PARE[1][7, 107] shows a decrease in association by a factor of 55 compared to the wildtype complex. This corresponds to the tremendous decrease in potential depth within our currently sampled picture. The minimum of ΔF in our simulations at constant volume is $\approx 50\frac{kJ}{mol}$ while measures of ΔG at constant pressure yield $\approx 79\frac{kJ}{mol}$[99].

---

[1] `http://www.weizmann.ac.il/cgi-bin/USERbcges/PARE.cgi`

a) Mean Force during complexation



b) Potential of Mean Force during complexation

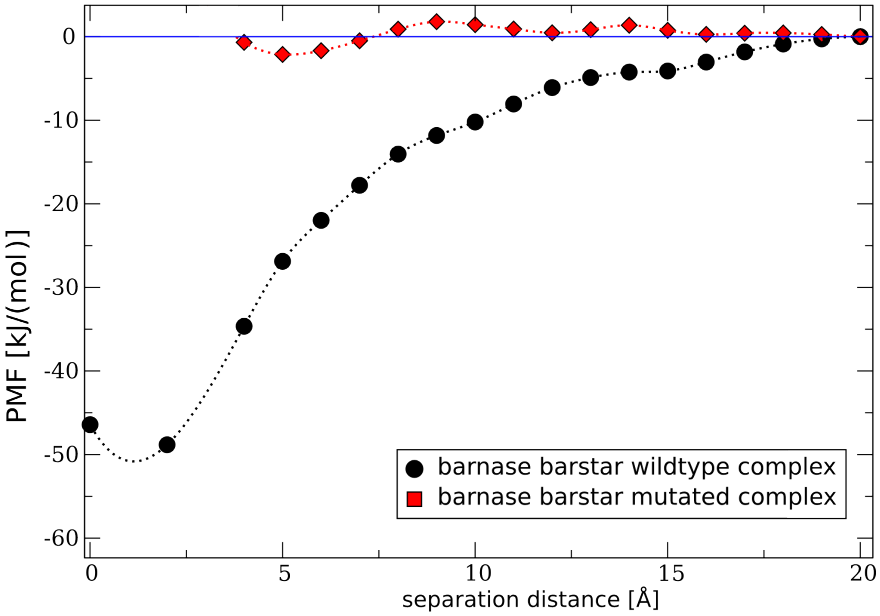
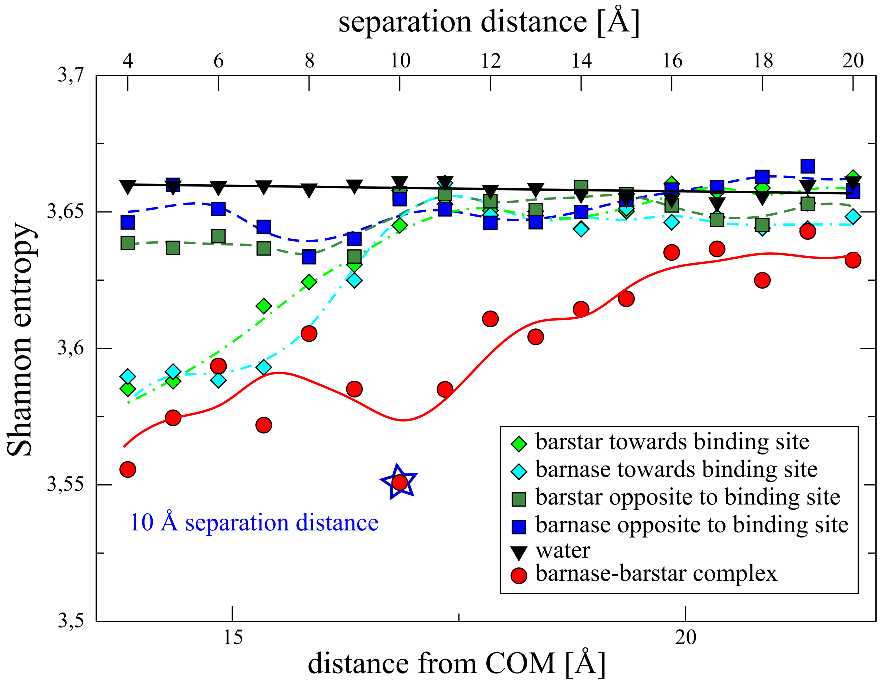


Figure 4.1: Fig. 4.1a and Fig. 4.1b show the mean force as well as potential of mean force along the complexation reaction coordinate. x-axis is the length of translation vector between the two constituents of the Barnase-Barstar complex. y-axis is the mean force (4.1a) or potential of mean force (4.1b). Particular at low separation distances, the poor sampling does not allow an accurate prediction of the potential of mean force. Both traces have been sketched with an cubic-spline.

4.1.2 Water Dipole Angle Analysis

Beside the potential of mean force, a main focus of my analysis is the solvent during the complexation process. Constrained simulations allow analysis of solvent properties at different discrete complexation states, defined by the separation distance.
My first analysis was the average angle of the dipole of the water molecules with respect to different vectors for four different cases: a) the simulation of water without solvated protein, b) for the solvated monomers, c) for the wildtype complex and d) a mutated complex. I mutated Lys27 and Arg59 on Barnase as well as Asp39 and Glu76 on Barstar to destabilize the encounter complex[97, 99]. For the simulations of the wildtype and the mutated protein, the proteins were simulated at different separation distances ranging from 4Å to 20Å. The proteins have been translated along the vector connecting the centers of mass (COM) to the desired separation distance. During the simulation, only the distance between the two proteins was constrained, while all other degrees of freedom were undisturbed.
A random distribution of the water dipole angle with respect to an arbitrary axis would yield an average angle of 90°. For a simulation of water without solvated protein, one observes this angle (black in Fig. 4.2a on the following page). Solvated proteins disturb this random distribution close to their surface. The water dipole angle with respect to the axis going from the COM of the protein through the center of the binding site is taken as reference vector (cyan and light green in Fig. 4.2a). If this reference vector is used, the disturbance is of higher extent and farther-reaching than the disturbance with respect to the vector in the opposite direction (blue and green in Fig. 4.2a). Proteins induce more order at their binding site than in the opposite direction. For this analysis, the distance from the analyzed water molecules to the COM of the single protein is taken as length scale (Fig. 6.1b on page 64).

a) Average angle of wildtype simulations



b) Average angle comparison of wildtype and mutant complexes

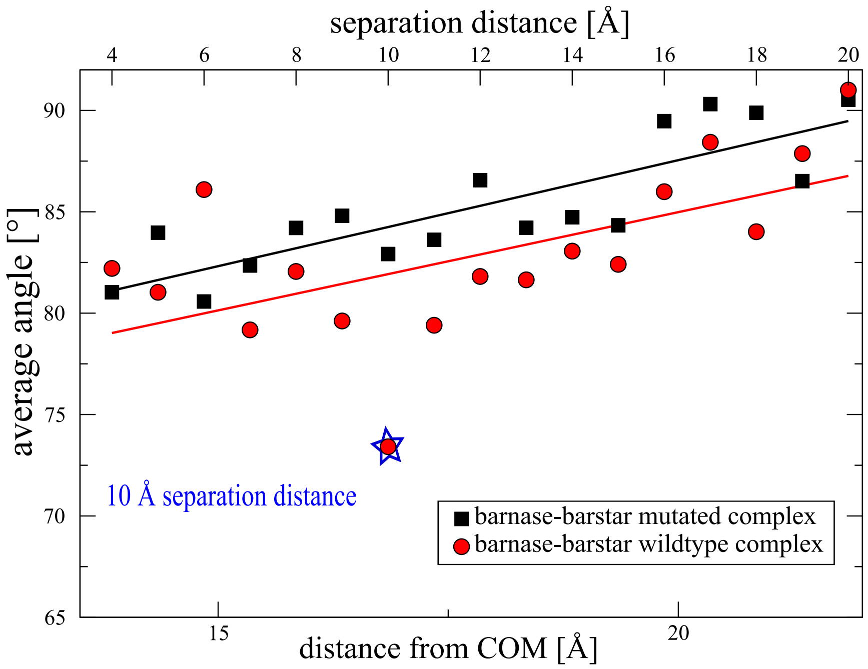
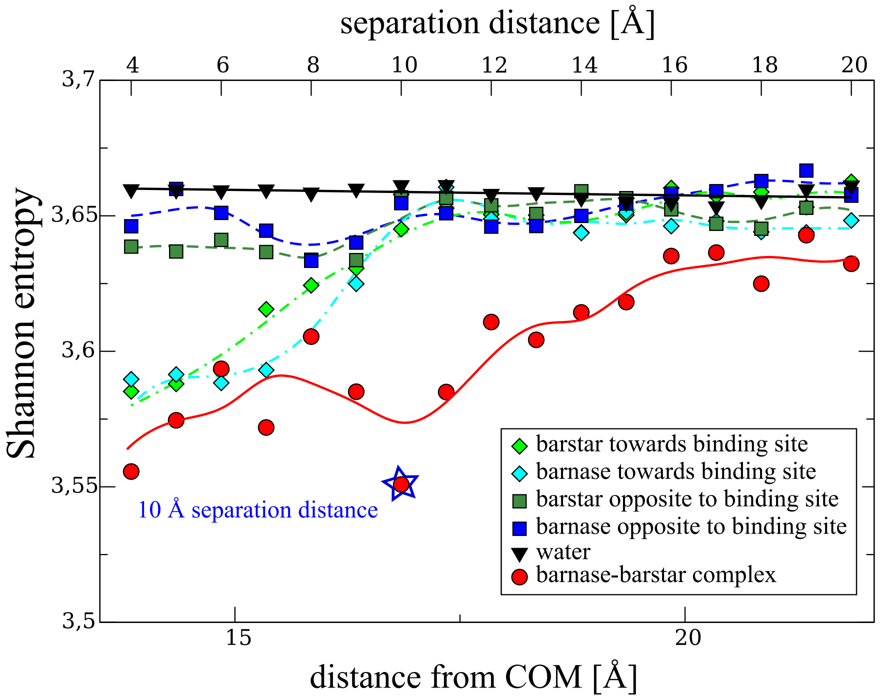


Figure 4.2: 4.2a shows the average angle between the dipole vector of water and different reference vectors. The distance from the COM is shown as x-axis for both figures. Separation distance has been plotted on the opposite x-axis. In 4.2a, a cubic spline over a running average of 3 points shows the trend. 4.2b shows a comparison of the average angles vs distance of wildtype and mutated complex. The line fit in 4.2b reveals a gap between wildtype and mutated complexes. The wildtype complex has a higher disturbance of the angular distribution. The special behavior at 10Å distance is only observed for the wildtype complex.

In case of the complex, the absolute distance (x-axis in 4.2a and 4.2b on the preceding page) corresponds to the distance to both COMs and hence to the middle of the vector connecting the centers of the two proteins. For the monomers, two vectors were analyzed, one going from the COM through the center of the binding site, the second one pointing towards the opposite direction. The offset to half of the displacement from the COM-COM is 11.85Å, corresponding to half of the distance, when no displacement is present. Thus, the displacement distance can be converted in absolute positions from the COMs via $x = 11.85\AA + \frac{1}{2} \cdot x_{sep}$.
The largest deviation from a random distribution is seen for simulations of the two wildtype proteins during complexation (Fig. 4.2a as well as Fig. 4.2b on the previous page). Here, the angle between the water dipoles of the central water molecules with respect to the axis connecting the COM of the two proteins has been analyzed (red in Fig. 4.2a and both traces in Fig. 4.2b on the preceding page). In this simulation, the water molecules in the middle of the two proteins were always exposed to the influence of both proteins at the same distance, but possibly with different rotational orientations during the simulation. The results show that the observed angular distributions for the complexing proteins, both wildtype and mutant, differ significantly from a random distribution. In both cases the disturbance is higher than the one caused by a single protein at the same distance. This demonstrates that the effects of the single proteins re-enforce each other during complexation. Interestingly, even at a protein-protein separation of 20Å the water between the two proteins is disturbed.
A particular difference from a random distribution with an average angle of 71° is observed at a displacement of 10Å for the wildtype protein, corresponding to half a COM-COM distance of 16.85Å as shown in Figure 6.1a on page 64. This deviation is not observed for the mutated complex (Fig. 4.2b on the preceding page).

a) Entropy of wildtypes



b) Entropy comparison of wildtype and mutant complexes

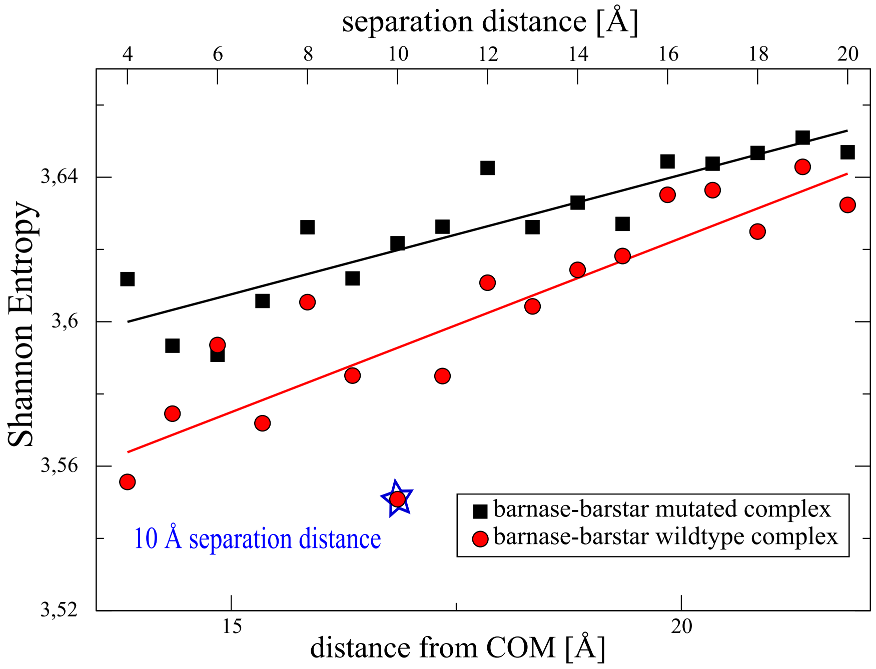


Figure 4.3: 4.3a shows the Shannon entropy of the respective angle distributions. The x-axis and opposite x-axis are labeled as in 4.2a and 4.2b on page 25, it can be converted as described in section 4.1.2 on the previous page. In 4.3a) a cubic spline over a running average of 3 points shows the trend. 4.3b shows a comparison of the distance dependent entropy for wildtype and mutant simulations. The line fit in 4.3b reveals a gap between wildtype and mutated complexes. The wildtype complex has a higher reduction in entropy. The special behavior at 10Å distance is only observed for the wildtype complex.

4.1.3 Water Entropy Analysis

To better describe the effect of the proteins on the water order, I used the Shannon entropy of the angle distribution as an additional measure for the order of the water molecules. The Shannon entropy is a measure for lack of information as described in section 2.3.2 on page 15. A low Shannon entropy is therefore an indication of a high information content, which can be correlated with a higher order of the water molecules. Shannon entropy contains information not only about the mean value of a distribution, but also about the shape. Hence, even if the mean value of a random and a non-random distribution are identical, the Shannon entropy can describe the differences between these distributions. The Shannon entropy is defined as

$$H = -\sum_{x} p(x) \log p(x) \tag{4.1}$$

where x is the quantity of interest – in this case the angle of the water dipole vector to the vector connecting the centers of the proteins – and $p(x)$ the probability distribution of this quantity (see also information theory formulation on page 15). Using this formalism to analyze the water structure around the monomers, one can observe better defined differences between the binding site and its opposite direction. The change in entropy of the solvent around monomers is strongly dependent on the direction of the reference axis. If the reference axis goes through the binding site, the first water layers show a high information content compared to non binding site water. Furthermore, the water dipoles have the highest order at small distances. Compared to the ambiguous behavior of the average angle especially at small distances, the Shannon entropy provides a clearer picture about the overall change in order.

Comparing the information content of the water around monomers with the water between the complex partners shows that at low distances, the information content of the water close to the monomers is similar to the information content of the water between approaching complex partners (Fig. 4.3a on the preceding page) for the binding site directions. The order of water molecules between the complex partners is only enhanced at distances exceeding two water layers. This indicates that the monomers order the first two water layers already to the maximum possible extend at their binding sites. In contrast, in the opposite direction to the binding site, the information content is not significantly different from pure water, despite the disturbed angle distribution (Fig. 4.3a). This underlines that the binding sites are designed to have the maximum impact on the solvating water.

Interestingly, not only the average angle, but also the Shannon entropy of the angular distribution (Fig. 4.3a) shows a local minimum at a separation distance of 10Å for the wildtype protein, emphasizing that the water molecules are highly ordered at this separation distance. This special behavior at 10Å separation distance does not occur in the mutated complex (Fig. 4.3b) as already observed for the average angle.

4.1.4 Volumetric Analysis

The anomaly in both angle distribution and information content is at a distance of 10Å. This separation distance is commonly associated with the Encounter Complex (EC)[97]. The EC is thought to be an intermediate state in the complexation process, at which the proteins are already in first contact, but still need to find the correct orientation for proper assembly. Hence, I tested whether this separation distance is indeed correlated with the first significant contact between the proteins. To show this, I analyzed the volume between the two wildtype complex partners along the vector connecting the centers of the proteins along the simulated trajectory.

Figure 4.4 on the next page shows various conformations of Barnase-Barstar during constrained MD-simulations. The contact points shown are stable over the whole simulation for most of the trajectories.

A high protein-occupied volume in the middle of the COM-COM vector indicates significant contact between the proteins. For the Barnase-Barstar complex, this analysis shows that the anomalous high information content coincides indeed with the first significant encounter(Fig. 4.5 on page 31 and Fig. 4.4 on the next page). Up

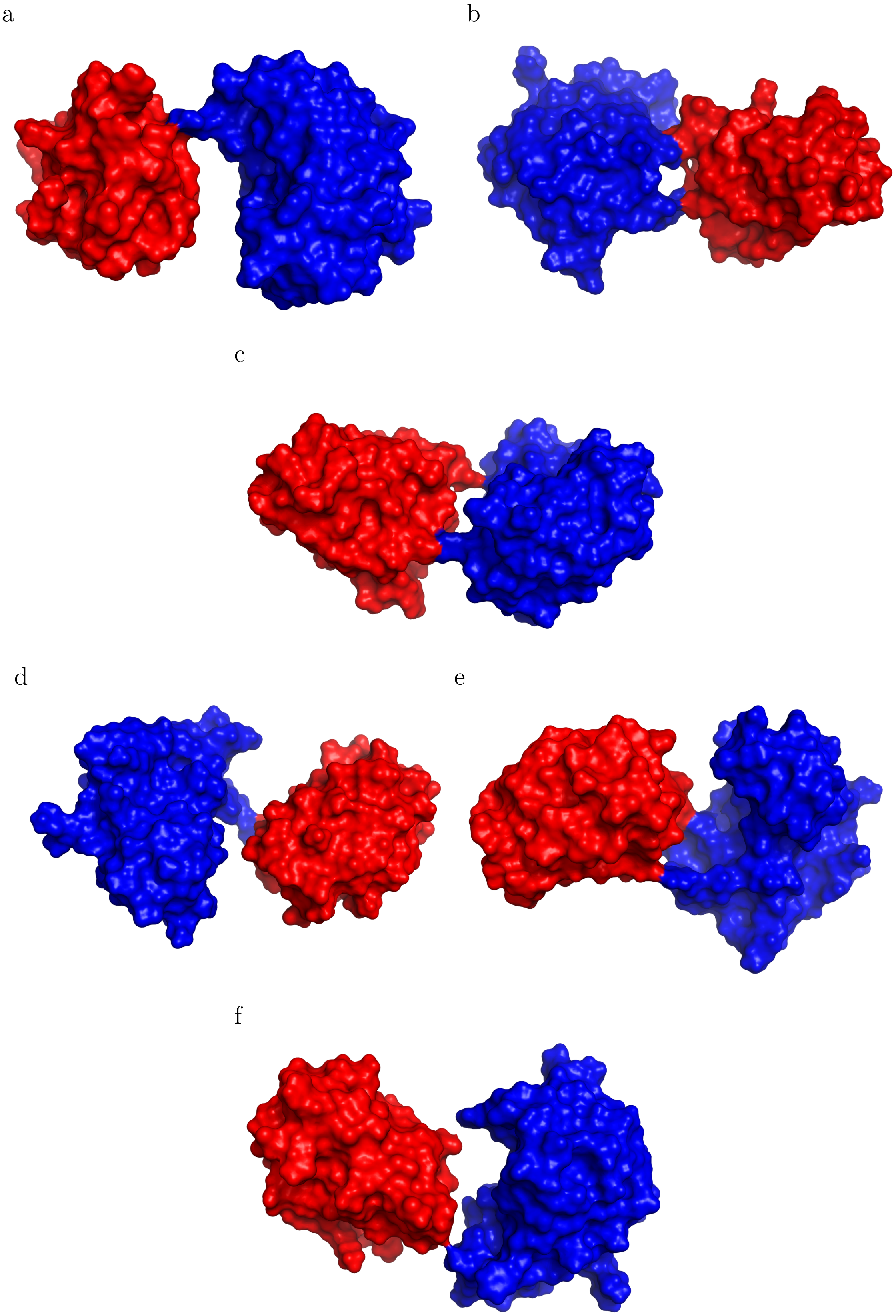


Figure 4.4: Figures a-f show typical conformations during a 10Å separated simulation. Encounter is typically happening at at least one contact point, which stays constant during simulation but also several multi contact states are present in the trajectories. Figures rendered with PyMOL[108][a].

---

[a] http://pymol.sourceforge.net/

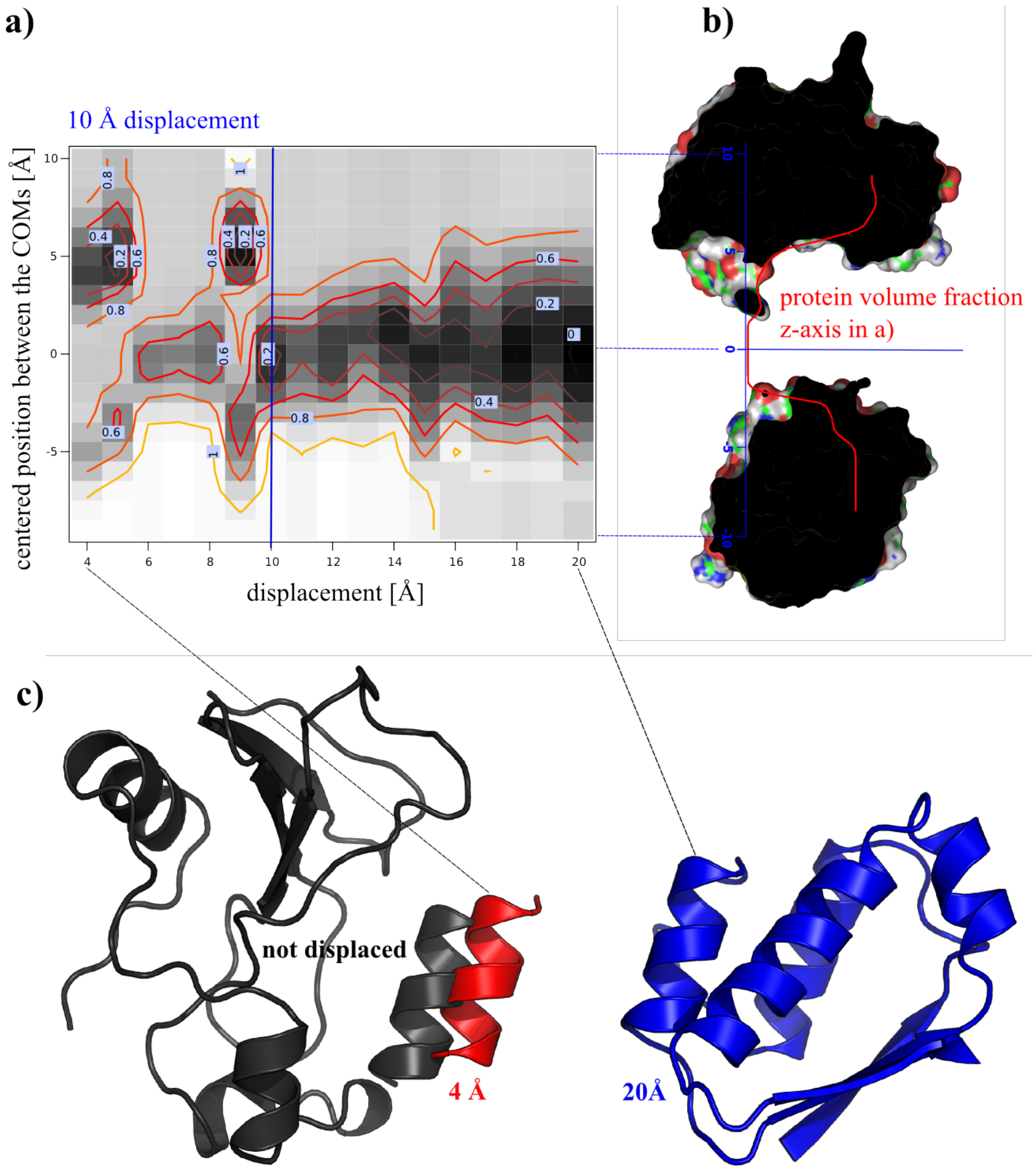


Figure 4.5: 4.5a is a contour plot of the volume occupied by protein between the two complex partners. The x-axis shows the COM-COM displacement. The y-axis refers to the volume occupied by proteins along the vector connecting the COMs. The colors indicate the percentage of volume occupied by protein; black means that no volume is occupied by protein, while white indicates that all the volume is occupied by protein. 4.5b shows a cut through the proteins together with the volume occupied by protein along the COM-COM connecting vector (red line). Each distance in the x-axis in 4.5a corresponds to one such line. Figure 4.5c shows the starting conformations of the complex(black), the minimum(red) and maximum displacement(blue), that have been simulated. To increase clearness, just the binding helix is shown for the minimum distance and the complex. The visualization was created with PyMOL[108][a].

---

[a] http://pymol.sourceforge.net/

to 10Å, nearly no volume is occupied by protein in the center of the vector, while at distances lower that 10Å, the proteins are in contact and only need to slide into the correct position.

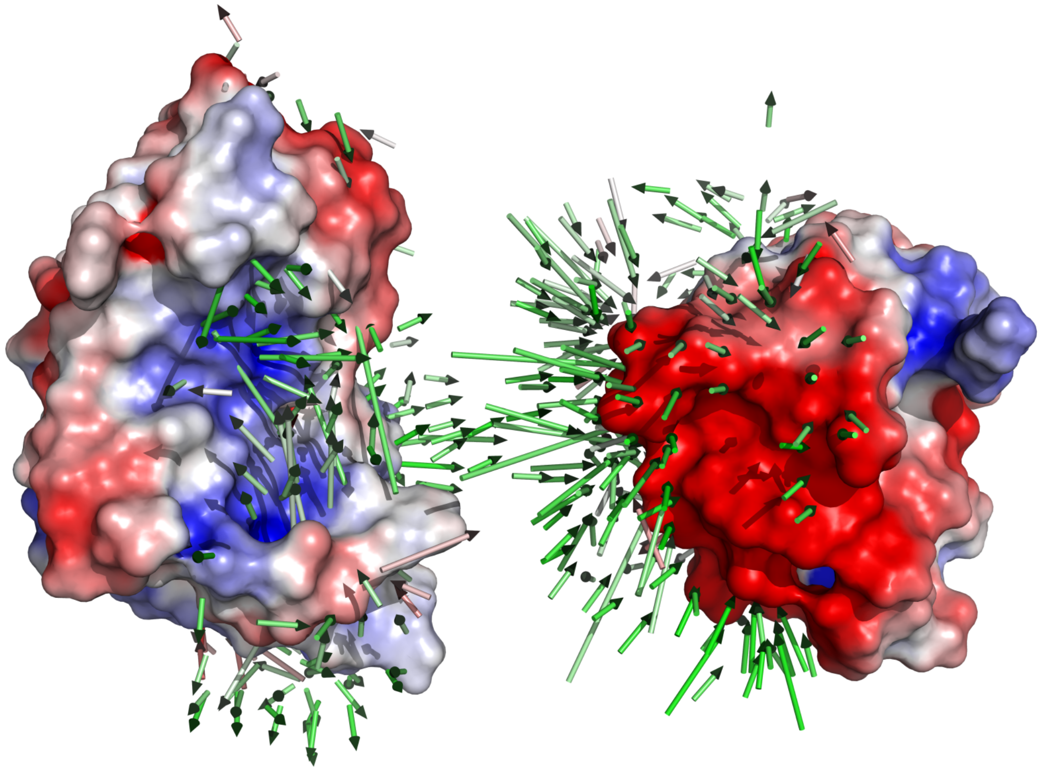
4.1.5 Dipole Fields

The next interesting question is whether or not the observed deviations from random distributions is associated with long-ranging, stable dipole fields between the complex partners. Overlapping dipole fields can lead to a pre-alignment of the proteins and therefore enhance complexation[16]. To test whether or not ordering of the water is caused by overlapping dipole fields, I further analyzed the spatial distribution of water dipoles between the separated proteins. Indeed, one can find that continuous dipole fields between both proteins are still present at a COM-COM separation of 20Å (Fig. 4.6a on the following page) for the wildtype, but not for the mutated protein (Fig. 4.6b). This is possibly due to the weaker electrostatic field between the mutated complex, which leads to a less defined pre-orientation. Also a clearly preferred orientation parallel to the fieldlines of the water dipole momentum can be observed for both cases in Figure 4.6.

4.1.6 Monomer Simulations

The results for the monomers in Fig. 4.3a on page 27 suggest that water might have a different order at interface regions. This suggests that the prediction of interfaces based on the solvent properties might be feasible. To test this I simulated and analyzed a set of 16 monomers with known complexation partner(s)[5]. The idea is to evaluate water molecule orientation towards the surface and map the results on the surface, to distinct between potential interface and non-interface regions. I tested various maximum radii. The following analysis was based on the water with a maximum distance of 5Å. This distance is associated with an increase of entropy in previous experiments (see monomers in Fig. 4.3a on page 27).

a) Wildtype complex



b) Mutant complex

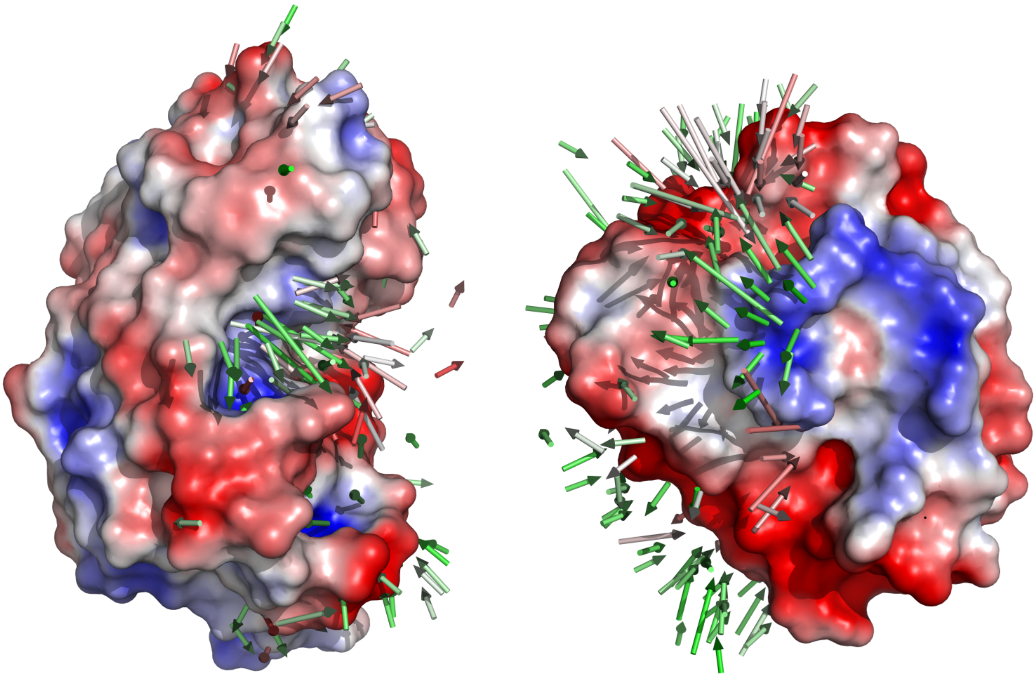


Figure 4.6: The water dipole field, averaged over a 20Å simulation is indicated with arrows. The color of the arrows corresponds to the alignment towards the fieldlines with green parallel and red antiparallel. The fieldlines are calculated from an electrostatic map, generated with APBS[109][a]. The surface color shows the electrostatic potential in units of $\frac{k_bT}{e_c}$ from -3 (red) to 3 (blue). Figures were created and rendered with PyMOL[108][b]

---

[a] Adaptive Poisson Boltzmann Solver

[b] http://pymol.sourceforge.net/

Surface Mapping

Here, I describe the first results from entropy surface mapping of angular distributions. After a 5ns simulation, the water atoms in every 5ps frame were assigned to the closest surface element and its atom. The Shannon entropy of distributions were mapped on the surface, in analogy to section 4.1.3 on page 28. For a detailed description of trajectory analysis see section 6.2.1 on page 69.

To give an overview of the preliminary results, I present exemplarily results of 5 out of the 16 monomers. The first ones are the constituents of the previously used Barnase-Barstar complex. Here, one can find a significantly lowered Shannon entropy in the binding pocket of Barnase (blue in Fig. 4.8a on page 36) as well as near the binding helix of its inhibitor Barastar (Fig. 4.8b). On Barnase, there exist surface patches with similar lowered entropy on the back (③ in Fig. 4.8b) of the binding site. This is weaker pronounced on the inhibitor Barstar.

The next system presented is a serine proteinase (PDB code 1CHG[111]) together with its inhibitor (PDB code 1HPT[112]). Here, surface mapping leads to an unambiguous picture since low entropy only occurs inside the binding pocket (① in Fig. 4.9a) and not near the cavities on the back side (Fig. 4.9a on page 37). Yet, this holds not for the inhibitor, where the surface mapping returns an ambiguous picture (Fig. 4.9b). The minimum is not directly located on the binding site (① in Fig. 4.9b) and also on the side view (② in Fig. 4.9b) a patch with low entropy can be observed.

Finally, again mapping on a serine proteinease called Subtilisin (PDB code 1SUP[113]) is shown. Here the picture is clearly ambiguous even for the enzyme (Fig. 4.10 on page 38). Although the binding site (① in Fig. 4.10) has a pronounced entropy minimum, the solvent at the back side (③ in Fig. 4.10) is even more ordered.

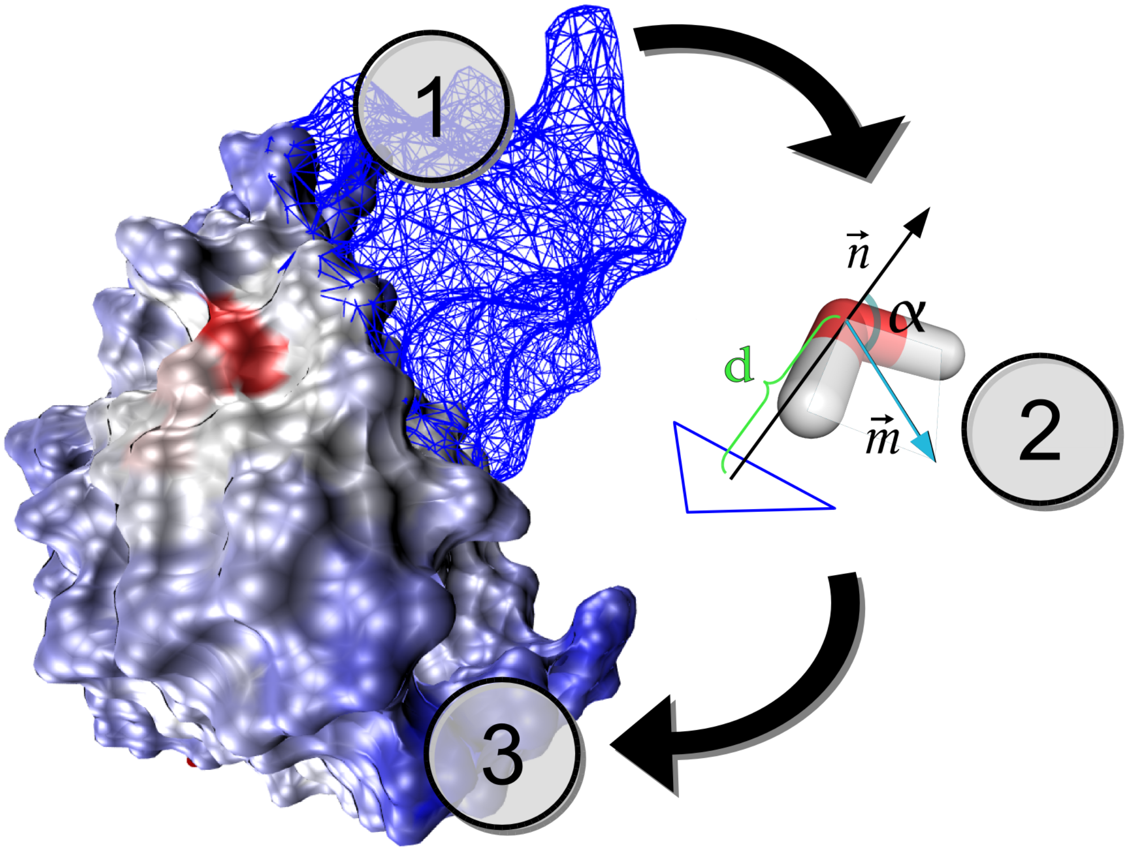


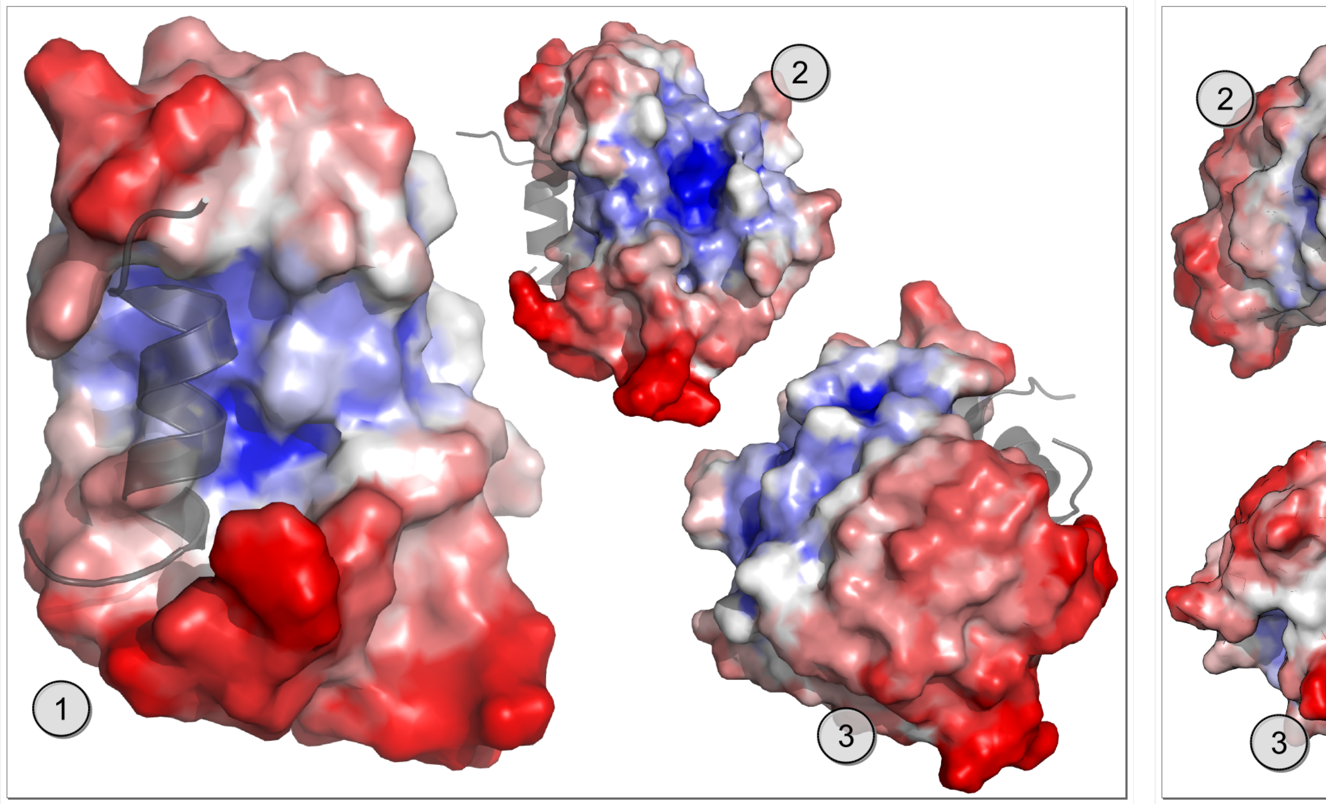
Figure 4.7: This figures shows the simplified surface mapping process.
In step ①, the surface of the protein is triangulated. Step ② calculates the distance and angle towards the closest surface element. Finally, the data is analyzed and mapped on the surface of the protein in step ③. Figure has been created with VMD[110][a], POV-Ray[b] and Inkscape[c].

---

[a] http://www.ks.uiuc.edu/Research/vmd/

[b] http://www.povray.org/

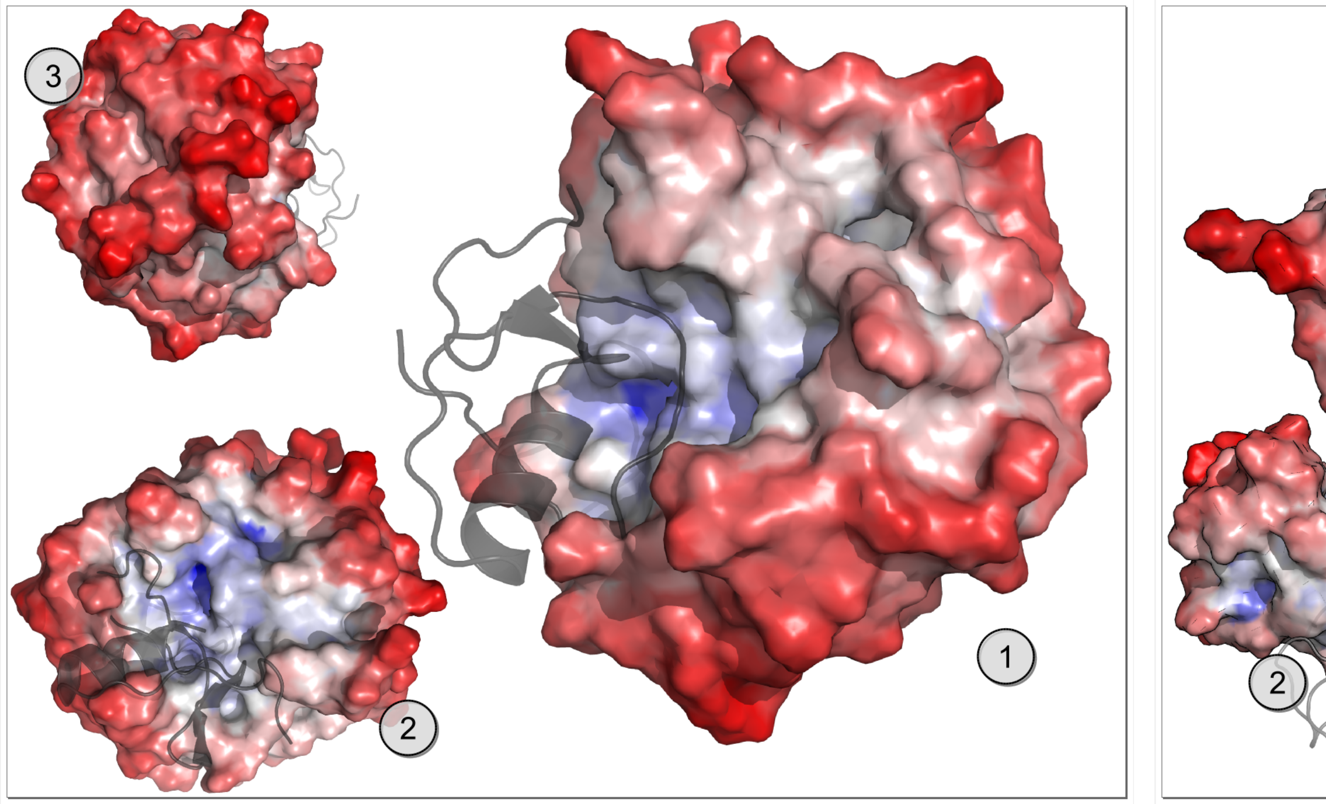
[c] http://www.inkscape.org



a) Barnase

b) Barstar

Figure 4.8: This figure shows the surface mapping of the water structure on Barnase and Barstar[88] monomer. ① shows the binding pocket of Barnase (4.8a) and the binding helix of Barstar (4.8b). Number ③ is the back view as well as number ② a side view on the binding site. The color scheme is blue for low and red for high entropy near surface.



a) Serine proteinase (1CHG)

b) Serine pro

Figure 4.9: Here, Hydrolase Zymogen an eucaryotic serine proteinase (PDB code 1CHG[111]) is shown (4.9a) together with its inhibitor (PDB code 1HPT[112]) in Figure 4.9b. Numbering and coloring scheme is consistent with Fig. 4.8 on the preceding page.

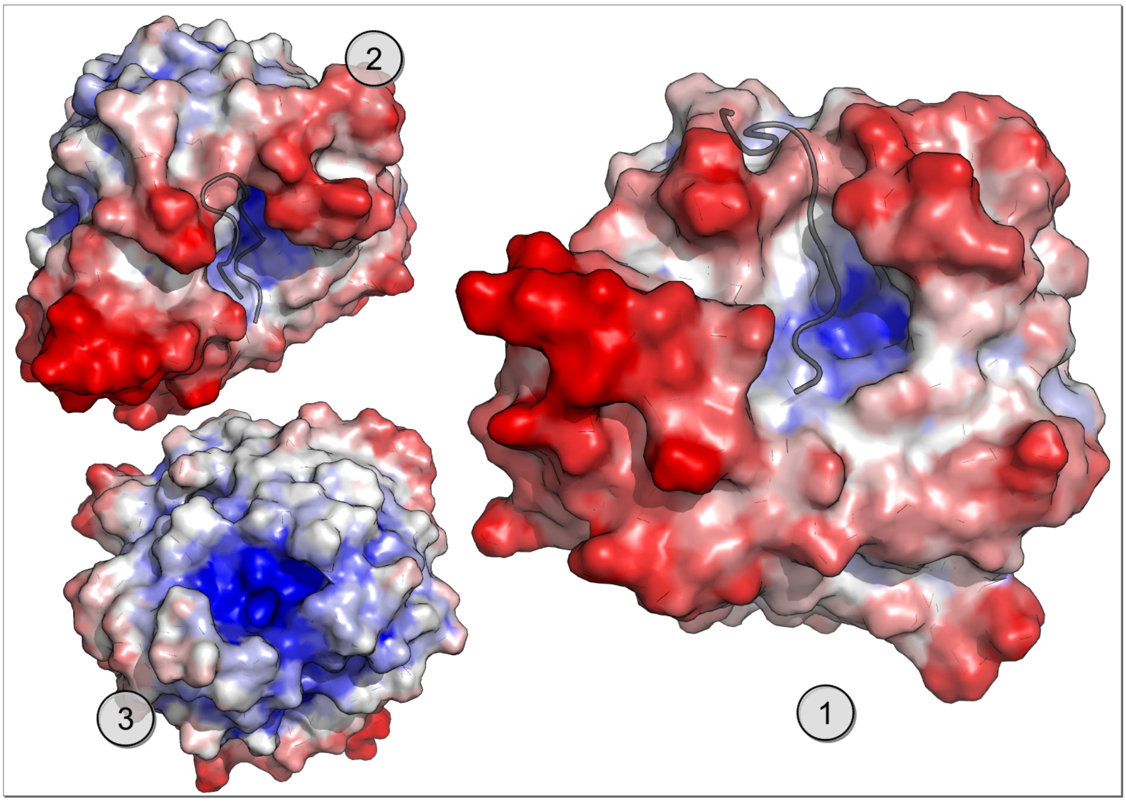


Figure 4.10: This subtilisin structure (PDB code 1SUP[113]) is a serine proteases from bacillus amyloliquefaciens. Numbering and coloring scheme is consistent with Fig. 4.8 on page 36. Figures 4.8, 4.9 and 4.10 were rendered with PyMOL[108][a] and assembled with Inkscape[b]

---

[a] http://pymol.sourceforge.net
[b] http://www.inkscape.org

Table 4.1: Octree benchmark results

Benchmark	`-O0`	`-O2`	`-O3 -ipo -static`
Octree range query in s	31.9	27.3	25.8
Linear range query in s	281	98	103
Octree nearest neighbor query in s	3.6	3.3	2.5
Linear nearest neighbor query in s	306	213	104

Index Structure Optimizations

In order to assign the water molecules to the appropriate surface patch, a nearest neighbor search has to be performed. Spatial searches, like nearest neighbor (NN) or range queries, are very expensive for large datasets. Thus, water assignment to specific surface patches and calculating the normals can be a serious performance bottleneck, when running over whole trajectories. Binary trees scale very well and allow a fast access on random data compared to linear searches.

A very efficient indexing scheme is the hierarchical octree structure[114, 115], widely used for collision detection problems[116, 117] in 3D engines and robotics. Unfortunately, there is no suitable free implementation available. Hence, I had to implement this procedure for my analysis. Here, I show some of the problem specific speedups achieved with my implementation; details can be found in section 6.2.3 on page 71.

Octree performance comparison

First, I checked performance with 10000 synthetically generated random points. The code was compiled witch the Intel Compiler Suite icc 9.1 (20070109) and switches: `-O3 -ipo -static -no-prec-div`. The expected complexities are $\mathcal{O}(\log n)$ for the octree and $\mathcal{O}(n^2)$ in case of linear search. The benchmark was conducted on an Intel Pentium M®Banias 1,5GHz CPU. For a 40000 point octree, 40000 range queries and the same number of nearest neighbor queries, several benchmarks were conducted (Table 4.1). The benchmark size is in the order of a typical system solvent molecule count of a single frame.

Remarkable is the speedup when doing NN queries of 40. A as an outlook in the parallel future of processors, the octree algorithm design is much better because of its gentle memory bandwith usage compared to linear searches. Further advantages are the capabilities of in-leaf storage of status, when doing multiple searches and possible memory optimizations to realign spatial close data to fit in the same processor cache line.

4.2 Discussion

4.2.1 Potential of Mean Force

Typical experiments[49] for an accurate reproduction of the PMF contain trajectories in the magnitude of 100 per separation distance compared to our sampling of less than 10 (see Table 6.1 on page 62). Particular in the low distance regime of Fig. 4.1a and 4.1b on page 23, additional simulations need to be conducted in order to achieve an accurate picture of the PMF allowing to dertermine the position and minimum of the potential energy. This is in particular the case for the mutant complex, where the mean force is not sampled in the low distance regime. Yet, despite the poor sampling in this distance regime, the generated PMF and the exerimental results[99] for the wildtype are in good agreement.

Beside an increased sampling, simulations of single mutants would yield an interesting dataset. Compared to the wildtype, no significant deviation in binding energy and off-rate could be found in mutants of specific residues[97, 99]. In the current results, one can clearly expect a different binding strength due to fourfold mutations. Thus, single mutations might reproduce the experimental findings of only a weak influence on binding strength of specifically Encounter Complex stabilizing residues[97, 99] together with storing influence on the EC.

4.2.2 Solvent Structure

The results of a constrained Barnase-Barstar wildtype and mutant complex show that water between approaching proteins is ordered. Furthermore this ordering increases with decreasing protein-protein distance. Since I do not actually analyze phase space variables, the order described here cannot directly be quantified in terms of the Gibbs entropy. Yet, formally there exists a close relationship between the Shannon entropy and the Gibbs definition of entropy[81]. This similarity between the Shannon entropy (derived from information theory) and the Gibbs entropy (derived from statistical thermodynamics) has been demonstrated in quantum mechanical systems[80].

The advantage of Shannon's formalism[18] is that it can easily be applied to any chosen variable and is not limited to actual phase space variables. Although quantitatively the analysis here does not provide a measure for the thermodynamic entropy of the water between two approaching proteins, qualitatively one can presume that the observed ordering is correlated with a decreased entropy of the solvent. This effect shows that mechanisms similar to the hydrophobic effect are also valid for protein-protein interactions. This is somewhat surprising, since the binding sites of transient complexes such as Barnase-Barstar are not more hydrophobic than the rest of the protein surface[4]. In the specific case of Barnase-Barstar, they are even charged and therefore less hydrophobic[7]. Hence, general principles observed for hydrophobic surfaces can also be applied to hydrophilic surfaces, provided that the intrinsic architecture of the surfaces induces ordering. Even for the monomers, a difference in behavior at close distances can be observed between the binding sites and other protein patches (Fig. 4.3a).

The low entropy of water at a separation distance of 10Å shows that this separation distance, at which the first significant contact takes place, has special properties. This is in accordance with the theory of Encounter Complex (EC) formation[17, 92, 94]. The assumption that the special behavior at 10Å separation distance is correlated with the formation of a stabilized encounter complex is supported by the simulation of the mutated complexes (Fig.4.2b and 4.3b). This alanine mutant (Lys27 and Arg59 on Barnase, Asp39 and Glu76 on Barstar) does not form a stabilized EC anymore [2], and it does not show any special behavior at this distance. As deduced from a variety of observations, the EC is regarded to be in a local energy minimum of two approaching proteins[17]. The major energy barrier for the following complexation is the displacement of the remaining water. According to the results, this energy barrier is initially lowered by a decrease in solvent entropy. This entropy gain may be counteracted by enthalpic effects. Entropy-enthalpy compensation is commonly observed in folding and association phenomena of proteins[99, 118, 119]. Still, the induced structure of the solvating water may help in lowering the energy-barrier associated with dewetting in analogy to the hydrophobic effect in protein folding.

The induced order of the water between the two approaching proteins has additional effects. Transient protein complexes are steered towards each others via their electrostatic field. This field decays in water with $\propto \frac{1}{\varepsilon}$, where ε is the dielectric constant of water, presuming that ε is isotropic and homogeneous. Importantly, it has been shown[10] that ordered water has a lower dielectricity constant ε than bulk water.

$$\chi_z = \lim_{\vec{E}_h \to 0} \frac{1}{\varepsilon_0} \cdot \frac{\partial Q_z}{\partial E_h} \tag{4.2}$$

E_h is an external electric field in z-direction while Q_z stands for the average polarization respectively χ_z for the susceptibility in field direction.

This difference in the susceptibility (4.2) can be expressed dependent of the relative depletion parameter $\frac{f}{m}$, where f is the reduction of H-bond exchange possibilities from m possibilities in bulk water. In our case, the H-bond exchange possibilities are reduced due to alignment of the water dipoles. For increasingly ordered water, the depletion parameter ($\frac{f}{m} \to 1$) leads to a lower susceptibility χ and thus to a lower dielectricity constant $\varepsilon = 1 + \chi$[10]. Therefore, the steering effect is increased by the observed solvent ordering.

In conclusion, constrained simulations showed that water between approaching proteins is increasingly ordered with decreasing protein-protein distances. This ordering facilitates complexation a) through entropic effects and b) trough a decrease of the dielectric constant ε of the water.

Here, I used the SPC water model which is very simple compared to various other water models. It has been shown[120] that together with the more complex TIP4P, SPC is the most accurate when reproducing structure parameters. Still, the generality of simulated structures compared to real water is arguable, since only few real experiments allowing a direct comparison of parameters exist. Additionally, most purely classical models fail in reproducing some parameters accurately. Another problem can be the short simulation time for reaching a specific conformation. Yet, the $2.5ns$ and $5ns$ simulations are in the order of $\tau_r = 5.56ns$, the rotational correlation time τ_r of Barnase[121]. τ_r refers to the free Barnase monomer. If there are additional electrostatic steering forces, τ_r should further decrease.

4.2.3 Monomer Simulations

Surface Mapping

Near the binding pockets, one can generally observe a higher order of nearby surrounding solvent. However, current surface mapping results do not allow a reliable prediction of interface patches.

The results shown in section 4.1.6 on page 34 are exemplary for all monomers. Current data clearly reveals that improvements of the analysis method are necessary to reach higher confidence in prediction. In general, one can find that surface mapping and thus future automated prediction of binding sites is much more reliable for the enzymes than the inhibitors[122]. Explanations for this observations are a typical smaller size of the inhibitor as well as a different surface geometry. Many enzymes have a pocket at the active site and are inhibited either by sterically blocking the binding site or through binding into a second pocket followed by a conformational change of the enzyme[123, 124]. Additionally, one can observe an obvious curvature dependency of the entropy which might overlay the structural differences near binding site regions and non binding site regions.

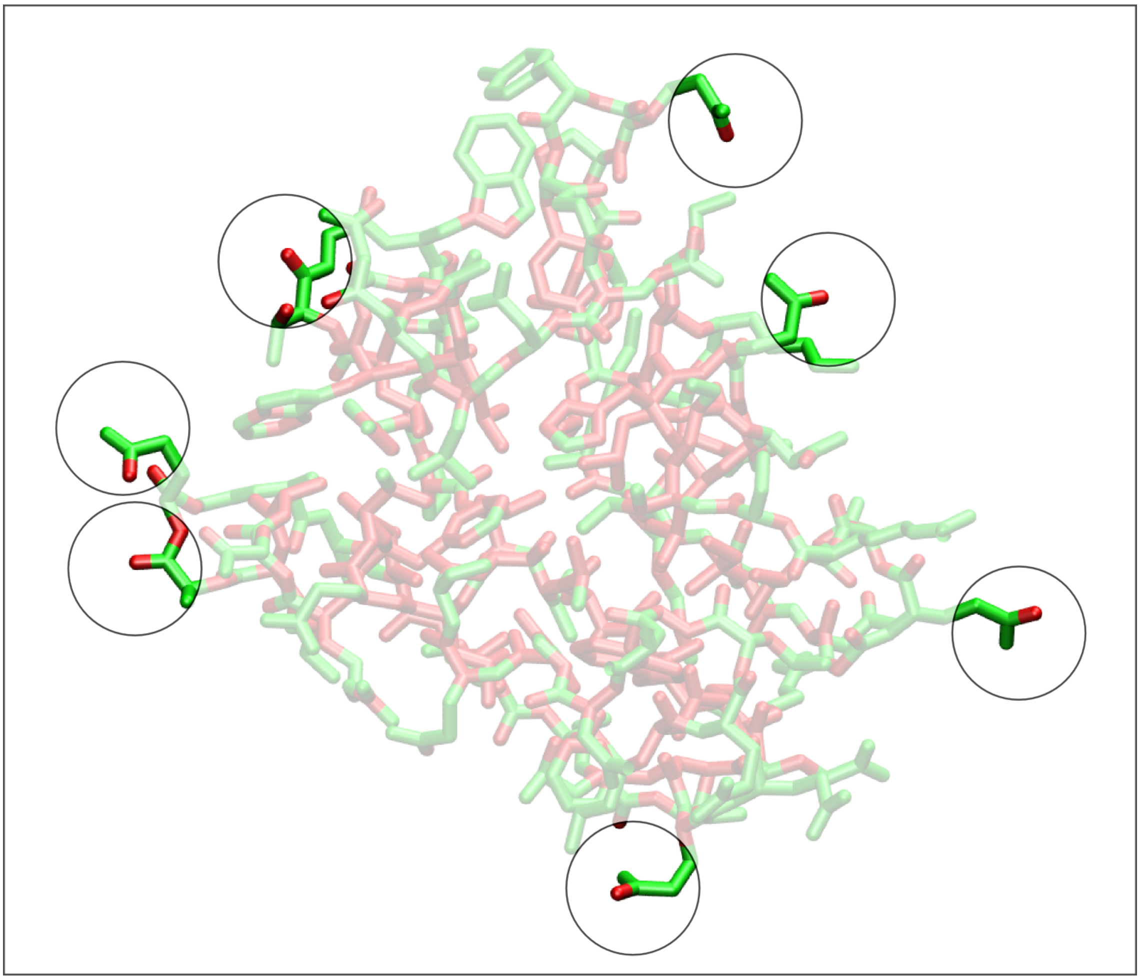


Figure 4.11: This figure illustrates triangulation errors. The protein is colored in green, if surface waters could be assigned. Some of the surface exposed oxygens are still red. Since assignment relies on the triangle-atom mapping from the triangulation programs, triangulation errors prevent assignment of waters. Figure rendered with VMD[110][a], POV-Ray[b] and Inkscape[c]

---

[a] http://www.ks.uiuc.edu/Research/vmd/

[b] http://www.povray.org/

[c] http://www.inkscape.org

The water analysis of Barnase and Barstar shows a preference of order near the binding site. Yet, the binding site is not the only site, at which ordered water is observed. In the eucaryotic serine proteinase (Hydrolase Zymogen), one observes the clearest picture for an enzyme. Its inhibitor's results are consistent with other inhibitors and their typical indefinite behavior of surface water order on the other inhibitors. The Subtilisin enzyme reveals a strongly expressed minimum on the back side. It is known, that Subtilisin has various additional ligand binding sites, differing in geometry and location from the known S1 binding pocket[122, 125]. However, the binding site of Subtilisin shows a local minimum of order in the binding pocket.

There are still computational problems to solve, such as triangulation problems[1], preventing analysis of some trajectories: There are several surface oxygens without assigned water atoms (see Fig. 4.11 on the preceding page) due to improper assignment of surface triangles to the atoms. Here improvements beside correcting atom assignment to the triangles are promising. A constant triangulation over time would allow to directly assign waters to the surface instead of the nearest atoms. Yet this will only be implemented in the future.

Software optimization

Beside classical optimizations as described in section 6.2.2 on page 70, which mostly implies detailed knowledge of underlying hardware, parallelization of software is of increasing importance due to the new emerging multicore processors. The tendency to slower but multiple cores forces one to make extensive use of parallelization techniques to exploit hardware and achieve shorter runtimes. Due to dependencies this is not trivial in most cases. Additionally, not every algorithm is suitable for parallelization. Evaluation of trajectories is a task which fortunately can be parallelized easily, if there is no dependency between frames. Normally, the most outer loop e.g. the loop over all the frames is parallelized but detailed knowledge of processor architecture, caches and memory IO can lead to other decisions in special cases. I used Message

---

[1] current triangulation via msms[22] or surf[126] fails in some cases, crashing the entire trajectory analysis.

Passing, a very common technique in HPC[1] to achieve speed improvements during analysis (discussed next section 4.2.3). Finally, index structure optimizations, as described in section 4.1.6 on page 39, can contribute to speedup through algorithmic improvements.

Explicit parallelization via MPI

Various parallelization techniques exist to make use of computational power of multiple processors. They can be divided either by their programming scheme into implicit and explicit techniques or by communication. Here I discuss OpenMP and MPI as examples for implicit and explicit techniques. MPI implementations either are optimized for a specific communication on HPC[2] specialized hardware such as NUMA Link from SGI, or implement various communication methods such as OpenMP, MPICH and LAM-MPI.

OpenMP as an implicit scheme is optimized for communication trough shared memory and runs only on one (logical) machine, yet there are recent developments as Cluster OpenMP[3] from Intel to overcome this limitation. Parallelization is done by pragmas, which mark parallel sections in the serial code.

Compared to implicit parallelization techniques like OpenMP, Message Passing Interface (MPI) programming enforces explicit parallelization inside the program trough communication. The program is written to run in parallel, several instances on the same or distributed machines communicate explicitly via message passing between processes. The implementation and scaling of communication depends strongly on the hardware[127, 128]. Communication works differently on a shared memory systems, ccNUMA[4] systems such as the SGI Altix 4700, Myrinet interconnect or Ethernet interconnect[129, 130]. MPI defines a specific set of library functions[131] a specific implementation has to provide. This guarantees flexibility and a minimal

---

[1] High Performance Computing

[2] High Performance Computing

[3] http://www.intel.com/cd/software/products/asmo-na/eng/329023.htm

[4] cache coherent Non Uniform Memory Access

porting effort of analysis software on different architectures, for example from our Opteron®/Gigabit ethernet based cluster on the Itanium®/ccNUMA Altix 3700Bx2. In my case, the MPI communication scheme was very simple. Each process determines its rank in the communicator and based on the communicator size, only every n-th frame is processed. The data is send back to the rank 0 process which processes and writes out analyzed data. OpenMP tends to be in advantage if one wants to parallelize a program after having written a serial version. Yet difficult portability and lack of implementations for distributed systems of OpenMP renders MPI as first choice for portability and performance.

4.3 Outlook

In the following section, I will give an overview over potential future projects and improvements. To test if change of order in water is a general complexation mechanism, the constraint simulations protocol is to be applied to other complexes. A major target are other complexes which are known to form Encounter Complexes during association such as TEM-BLIP[7]. Beside quantifying the order of water, analysis of other structural properties such as the RDF[1] might provide more insight into the complexation process. Additionally, the large dataset generated in these studies allows a structural analysis of different complexation states. Also other water models should be tested. Despite the fact that SPC together with TIP4P water is currently standard in MD simulations, newly emerged quantum mechanical modeled water with low additional computational cost[132] may provide refined insights into structural properties during complexation.

Considering the interface predictions, several approaches provide improved results. To reduce geometrical dependencies, simulations with surface shuffled aminoacids can provide normalization parameters. On the algorithmic part, switching to other

---

[1] Radial Distribution Function

SAS[1]/SES[2] algorithms might increase the stability of the analysis program. Additionally, a time independent constant triangulation would allow direct mapping on the surface elements instead of the nearest atoms. This could also significantly lower geometrical problems of curvature, avoided through constant triangulation density.

---

[1] Solvent Accessible Surface

[2] Solvent Excluded Surface

The more you see how strangely Nature behaves, the harder it is to make a model that explains how even the simplest phenomena actually work. So theoretical physics has given up on that.

Richard Feynman, The Strange Theory of Light and Matter

Chapter 5

Gold Surface Parameter Studies

5.1 Results

5.1.1 Force Field Parameterization and Modeling

Gold surfaces are a preferred target for single molecule experiments with biological subunits[43]. Here, I will give a summary over implementation of interactions and the first simulation results with two gold configurations and a quartz-lysine setup.
Stefano Corni and coworkers from the University of Modena suggested the following scheme for modeling van-der-Waals and coulombic interactions at the cost of moderate increase in electrostatic calculations:
Virtual sites in the planes of (111) aligned surfaces are used for van-der-Waals interactions, placed in the geometrical center of each triangle, leaving no van-der-Waals interactions on the gold atoms itself (Fig. 5.1a on page 51). To avoid problems[1] with edges and vertices of non planar gold geometries surfaces such as nanocrystals, the initial scheme was changed to virtual sites only placed on the (111) surface (see

---

[1] Electrostatics can increase to infinity with improper van-der-Waals parameters due to clashes with solvent molecules

Fig. 5.1a on the next page), leading to comparable results. Edges and vertices without repulsive terms from van-der-Waals interactions allows water molecules penetrate the gold.

Quantum chemical DFT[1] simulations show that all aminoacids can be treated with one set of σ_{ii} and ϵ_{ii} for nonbonded interactions, which are equivalent to parameterization through $C_{ii}^{(6)}$ and $C_{ii}^{(12)}$[2], combined in (2.5) (Stefano Corni et al., personal communication). Imidazole and histidine are the only exception and will be treated by a separately developed set of σ_{ii} and ϵ_{ii} in the near future.

In the current model, edges, vertices and bulk atoms carry van-der-Waals interactions themselves (see Fig. 5.2a), to avoid clashes with water at the edges. The separation in the three groups allows a convenient adaption of parameters in case of edge related studies in the future. The gold atoms and virtual sites are either frozen[3] or restrained in position.

While van-der-Waals interactions can be modeled with fixed virtual sites, the treatment of coulombic interaction is more complicated. As a metal, gold becomes polarized when charged particles approach the surface. To accurately describe those effects, an implementation of plarizability needs to be introduced. Initial suggestions of our group used mirror charges and a specific fine tuning of each pair of $C_{ii}^{(6)}$,$C_{ii}^{(12)}$ involving gold atoms to keep computational overhead as low as possible. Yet an easier model (Stefano Corni et al, personal communication) turned out as very effective in reproducing adsorption energies and distances of quantum chemical DFT calculations appropriately at the cost of a moderate increase in computation during MD simulation steps. Each gold atom is treated as a dipole with a distance $d = 0.7\AA$, charge $q = \pm 0.3e$ and a mass of $0.5u$ for the virtual charged particle (see Fig. 5.1b on the following page).

Stability under normal 2fs integration steps empirically suggests a fixed dipole distance of 0.7Å. The mass is also arbitrary chosen in the current model. The charge is

---

[1] Density Functional Theory

[2] historical reasons, since OPLS-AA force field parameters are in σ_{ii} and ϵ_{ii} notation

[3] fixed position

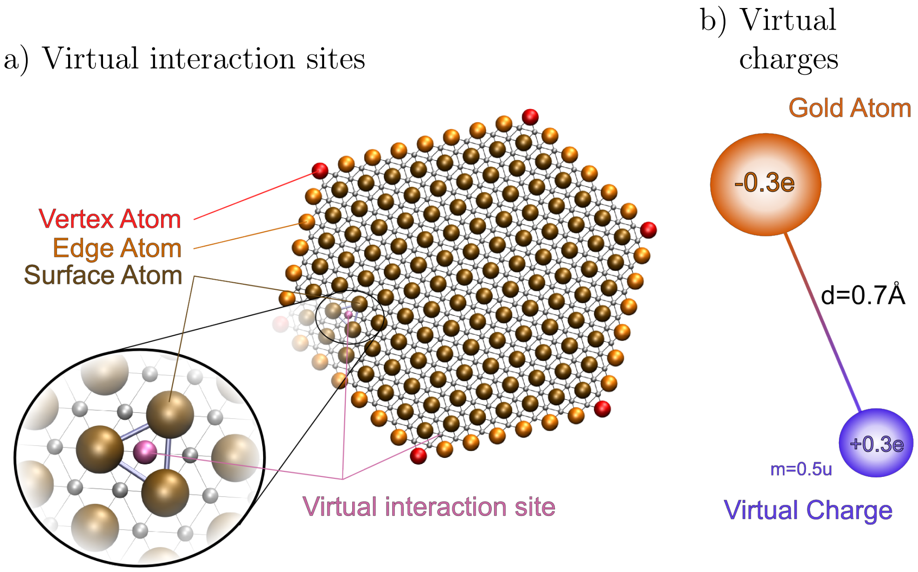


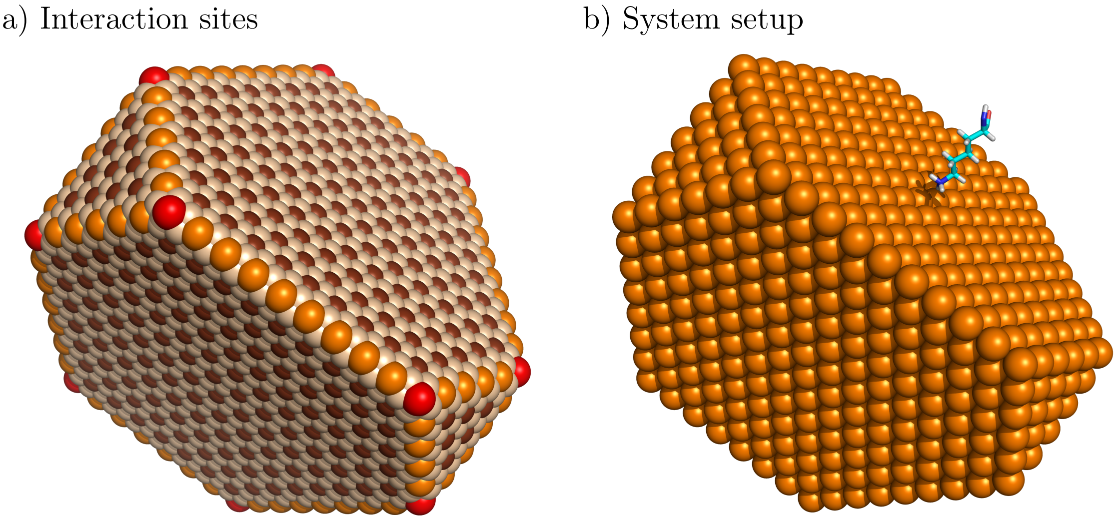
Figure 5.1: Figures 5.1a and 5.1b illustrate the additional interaction site placement. For each triangle of surface atoms, a virtual interaction site carrying the van-der-Waals parameters of the surface is placed in the geometrical center (Fig. 5.1a). For each gold atom, the charge is set to -0.3e and an additional particle with a mass of 0.5u and opposite charge is added with a constrained distance of 0.7Å (Fig. 5.1b). Figures were created with VMD[110][a], POV-Ray[b] and Inkscape[c].

---

[a] http://www.ks.uiuc.edu/Research/vmd/

[b] http://www.povray.org/

[c] http://www.inkscape.org



a) Interaction sites

b) System setup

Figure 5.2: Figure 5.2a shows the nanocrystal with colored interaction sites. Red gold atoms are treated as vertices, orange ones as edges and dark brown ones as surface. All surfaces on the crystal are (111) aligned. The virtual sites on the surface are indicated in light brown. The future setup is drawn in figure 5.2b (water not shown). Figures were created and rendered with PyMOL[108][a].

---

[a] http://pymol.sourceforge.net/

Table 5.1: This table shows components of the electric dipole moment in Debye[a]. The columns correspond to the nanocrystal in vacuum, solvated in water and with a F^- ion on it's surface.

	vacuum	water	vacuum and F^-
$\langle M_x \rangle$ [D]	-3.06	-2.20	29.78
$\langle M_y \rangle$ [D]	-0.59	-1.65	55.13
$\langle M_z \rangle$ [D]	-0.69	-2.24	-6.99
$\sqrt{\langle M_x^2 \rangle}$ [D]	20.02	34.65	35.32
$\sqrt{\langle M_y^2 \rangle}$ [D]	21.26	34.64	58.79
$\sqrt{\langle M_z^2 \rangle}$ [D]	21.28	35.39	20.25
$\sqrt{\langle \vec{M}^2 \rangle}$ [D]	36.14	60.44	71.51
$\langle \vec{M} \rangle$ [D]	3.20	3.55	63.06

---

[a] $1D = 3.33564 \times 10^{-30} C \cdot m$

adjusted to reflect physical properties such as adsorption energies and distances.

5.1.2 Initial Simulations

Water and Vacuum Simulations

As an initial step I examined the properties of a gold nanocrystal structure in vacuum and in water at 300K. We found out, that permanent dipole moment of the crystal is non zero ($36.14D$, see Table 5.1). We also considered a crystal with pre-oriented dipoles and found convergence within computational error in a few hundred picoseconds towards the unoriented crystal. In water, the permanent dipole moment is higher ($60.44D$, see Table 5.1) but also in the magnitude of 10-30 water molecule dipole moments of SPC[69] water ($2.27D$). Yet, the single components averages are in the maximum order of $\pm 3D$ for vacuum and water simulations, suggesting a non permanent direction of the dipole moment inside the crystal over time.

Adsorption Simulation

To test if charged molecules adsorb correctly on the surface, I simulated a fluorine in vacuum on the nanocrystal. Figure 5.3 on the following page shows the dipole orien-

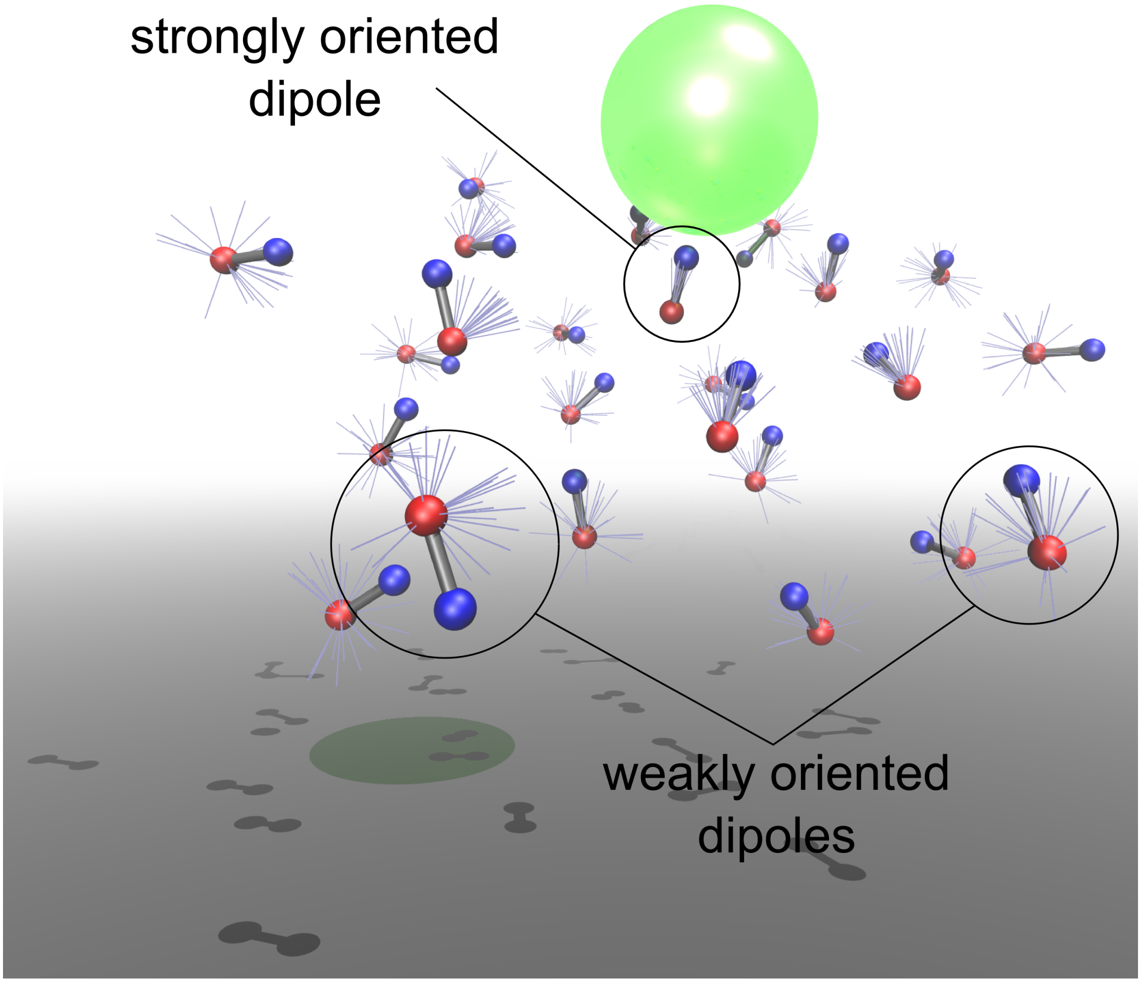


Figure 5.3: F^- on Au surface. All Gold atoms in 7Å distance to F^- are drawn. The current state of dipoles is represented with ball, and sticks model, while red corresponds to positive and blue to negative charges. 25 states in 5 ps distance are marked with thin lines from the gold atoms. Figure has been created with VMD[110][a] and rendered with POV-Ray[b]

---

[a] http://www.ks.uiuc.edu/Research/vmd/

[b] http://www.povray.org/

tation with the fluorine on a gold surface at 300K. Only dipoles in a distance of a few Å show significant preorientation. The average dipole moment of the crystal with a fluorine on the surface more than doubles to $71.51D$ (see Table 5.1 on page 53). The individual averaged components of the dipole moment are slightly lower, but in the same magnitude as the total dipole moment. This reveals the permanent orientation of the dipole momentum, also observable in $\langle \vec{M} \rangle$, and allows determination of the direction which indeed corresponds with the surface normal of the F^- adsorbed surface.

The system also shows the behavior of charged particles to adsorb on top of the gold atoms and not in the cavities between them. This is modeled through the placement of the van-der-Waals interactions on virtual sites.

5.1.3 Lysine-Quartz Simulations

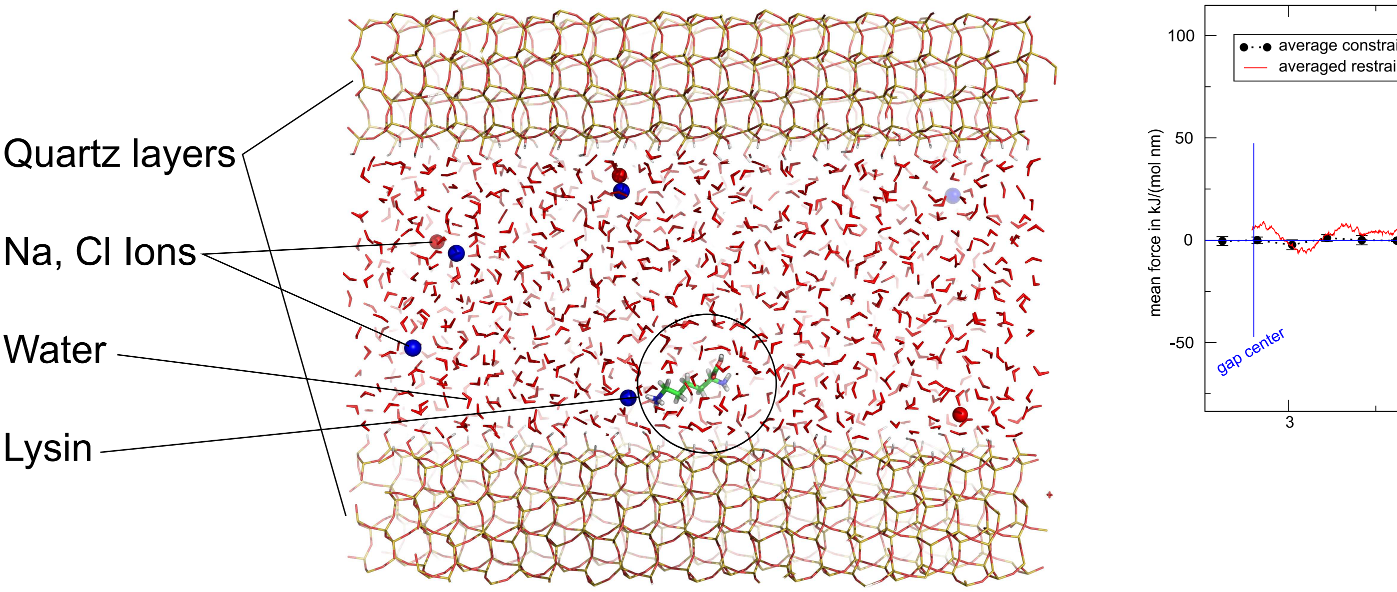
The next task was to find an appropriate setup for PMF calculation. Since nanocrystal geometries depend on the chrystal structure of the material, a geometry independent setup of one plane extending in the x-y direction was chosen.

A simple lysine-quartz model (Setup see Fig. 5.4a on the following page) with known force-field parameters was used, to test different methods for PMF calculations. Quartz parameters were obtained from Ref.[133]. I analyzed a long run, where lysine has been pushed slowly[1] ($v = 0.25\frac{nm}{ns}$, $k = 500\frac{kJ}{mol \cdot nm^2}$, $\Delta t = 5ns$) towards the surface (red in Fig. 5.4b and 5.4c on the next page) and compared it to several short runs ($\Delta t = 2.5ns$), where COM position of lysine has been constrained (black in Fig. 5.4b and 5.4c). In both cases, the constraint or restraint forces have been analyzed. The analysis shows two types of deviations:

During the approach in the pull run, the PMF suffers deviations from the hydrodynamic drag resulting in the offset in Fig. 5.4b. Additionally, despite the slow push speed, fluctuations are still dominant in the approach. The PMF calculated from constrained simulations allow a clearer picture at larger distances. Unfortunately at

---

[1] compared to typical MD simulation timespans



a) Lysine-quartz setup

b) Mean Force com

Figure 5.4: Figure 5.4a shows a cut through the system setup with Na$^+$ (red) and Cl$^-$ (blue) ions. The resulting mean force (Fig. 5.4b) and the PMF obtained through integration (Fig. 5.4c) from 60 distance constrained simulations (black) and a slow AFM push run (red) is also shown. The red AFM like curve is a running average over the heavily fluctuating values on fs timescales. Figure 5.4a was rendered with PyMOL[108][a].

---

[a] http://pymol.sourceforge.net

lower distances it shows no distinct extremum within the current sampling scheme. One can also observe, that extremum locations and thus repulsive forces are farther away from surface for the slow-push simulation.

5.2 Discussion

First results of the force field in vacuum and water are promisingly reproducing desired binding energies and conformations from quantum chemical experiments (Stefano Corni and Francesco Iori, personal communication). Comparison of the permanent dipole moment present in the nanocrystal is difficult, since few experimental results on gold-nanocrystal dipole moments are available. However, the averaged dipole moment $\langle \vec{M} \rangle$ is small and in the order of a single SPC water molecule for the uncharged setups (Table 5.1 on page 53).

Considering the electrostatic interactions, there are several parameters which can be further refined in this model. Charges and masses of the dipoles (Figure 5.1b on page 51) is adjustable. Additionally, the temperature can be easily modified by individual coupling to the dipoles with a different temperature. Experiments on adjusting the distance in Fig. 5.1b on page 51 lead to problems in electrostatic computations (Stefano Corni, personal communication).

Adsorption behavior of charges on the surface was analyzed with a F^- setup in vacuum. The position on top of the gold atom agrees with DFT[1] calculations (Stefano Corni and Francesco Iori, personal communication).

To evaluate different setups for calculating the distance dependent PMF, a lysine-quartz setup was used to analyze possible problems. A calculation of the potential of mean force was conducted for two different setups. The PMF allows to classify interaction of a peptide with the surface.

Comparison of the two methods shows advantage of constraint based methods[49]. The AFM based method includes contributions from hydrodynamic forces which

---

[1] Density Functional Theory

are difficult to estimate in small setups particularly near surfaces. Close to the surface, the AFM based method suffers from too short relaxation time available and delivers a different energy minimum position as the constrained based method. The current scheme for the constrained method reveals the poor sampling near the surface. Additional constrained simulations with lower distances, and between the currently sampled points are needed close to the surface to provide a clear picture of the PMF shape at low distances.

5.3 Outlook

The implementation model needs to be validated for larger peptide constructs. Here, the simulation focus will be the TEM-BLIP complex, for which experimental data is available (Gideon Schreiber and Dana Reichmann, personal communication (PRO-SURF)).

The major focus in near future will be simulations in comparison to wet-lab experiments to test and improve correspondency between both. Single molecule AFM experiments with single aminoacids and oligopeptides on gold surfaces are planned in our lab accompanied by SPR[1] experiments at the Weizmann Institute of Science. Combined with the derived parameter set for histidines, a complete set reproducing the same interaction order of aminoacids as determined in real experiments would be a milestone towards the ability to simulate protein gold interactions.

From the current results, modeling of the electronic structure of gold surfaces is the weakest point. Van-der-Waals interactions and orientational minimum were achieved conveniently through introduction of virtual interaction sites. Yet the properties of the virtual dipoles need to be examined in further studies of different systems in particular with respect to the influence of solvation. Water surface interactions and their modeling as well as water structure near surfaces itself might be crucial for the reproducability of adsorption properties of proteins.

The ability to produce reliable results from protein-surface simulations will allow studies with a completely different direction, comparable to the impact from the availability of generic force-fields for protein-protein studies.

---

[1] Surface Plasmon Resonance

Physics is becoming so unbelievably complex that it is taking longer and longer to train a physicist. It is taking so long, in fact, to train a physicist to the place where he understands the nature of physical problems that he is already too old to solve them.

Eugene Wigner

Chapter 6

Methods

6.1 Constrained Simulations

6.1.1 System setup

I used chain A and E from the Barnase-Barstar complex x-ray structure (PDB code: 1BGS[88]) as initial structure for the simulations. The second chain of the complex was displaced along its COM-COM axis in distances ranging from 4Å to 20Å (Fig. 4.5c). The usage of a cubic box allows the complex partners to freely rotate within the box in all directions without interacting with their periodic images. Additionally, each monomer, Barnase and Barstar, was simulated seperately in a water box. The resulting structures were solvated and ionized with Na^+ and Cl^- to achieve electric neutrality and a physiological salt concentration (150 mM). The molecules were first aligned along their principal axis and solvated with a minimum distance of 1.2Å to the cubic box boundaries in case of the complexes and a total box edge length of 68Å for the reference monomers.

Listing 6.1 on the next page shows a simplified version of the preparations for energy minimization of the resulting structures. In the first line, the structure is converted to the force field representation[1] and a topology is generated. Line two sets up the box and line three fills it with water. The next step creates a dummy binary topology, used to randomly replace water molecules with ions in step 5.

Listing 6.1: Solvating and ionizing a structure

```
1 pdb2gmx -f myStruct.pdb -o myStructFF.pdb -ff oplsaa -ignh
2 editconf -f myStructFF.pdb -princ -box $BOXSIZE -bt tric -o myStructBoxed.pdb
3 genbox -cp myStructBoxed.pdb -o myStructSolvated.pdb -cs
4 grompp -c myStructSolvated.pdb -f dummy.mdp -p topol.top
5 genion -random -s topol.tpr -o myStructIonized.pdb -np $POSI -nn $NEGI -g -rmin 0.8
```

I have written several scripts to fully automatize boxing and solvation, e.g by calculating charge counts needed for physiological ion conditions (150mM) and electroneutrality[2].

6.1.2 Mutants

The mutant complex was also derived from pdb code 1BGS and set up the same way as described above (6.1.1 on the preceding page). I mutated Lys27 and Arg59 on Barnase as well as Asp39 and Glu76 on Barstar to Ala. The mutated residues play an important role in encounter complex stability[97, 99].

6.1.3 MD Simulation

Simulations have been conducted with the GROMACS 3.3 suite[60, 134] on different architectures. Some jobs were run on our 12 core Opteron®cluster but the vast majority of runs was set up on the LRZ linux cluster[3], using the serial job queues on IA32 (e.g. Intel Pentium®4) and x86_64 (AMD Opteron®) architecture. Since maximum runtime for serial jobs is limited to 144 hours and a typical run duration is 1-2 month, scripts had to be implemented to split the simulation in several parts. To

---

[1] for instance missing H-Atoms are created

[2] though this should not be needed, since Particle Mesh Ewald calculations do not suffer from infinite potential problems due to background corrections

[3] http://www.lrz-muenchen.de/services/compute/linux-cluster/

Table 6.1: Conducted simulations from mutated and wildtype complexes.

separation [Å]	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
wildtype:																	
2.5 ns	8	9	9	9	8	10	-	10	-	6	-	6	-	-	-	-	-
5.0 ns	-	5	-	-	-	4	6	-	7	-	4	5	3	3	2	3	3
mutated:																	
5.0 ns	6	6	6	6	6	6	6	6	3	4	4	3	3	3	3	4	4

this end, I wrote an extensive Python[1] script to control the setup, continuation and failure handling of runs. Unfortunately current GROMACS versions (v3.3.1) do not support domain decomposition[2] in conjunction with intermolecular constraints for further speeding up simulations by usage of multiple cores. A total of 807.5 ns were simulated; a single (5ns + 0.5ns) requires one to two month depending on separation distances computations on a recent AMD Opteron®core. This leads to a total usage of about 20 CPU-years.

SPC-E[69, 70] water model and OPLS-AA[52] force field parameters were used in all simulations. All simulations were performed with 2fs time steps and using Berendsen temperature coupling to 300K. Fast Particle-Mesh Ewald[62] electrostatics was used with maximum grid spacing for the FFT grid of 1.2Å and cubic interpolation. The solvated complex has been energy minimized and relaxed in a 0.5ns position restrainted relaxation run. Several simulations either with 2.5ns or 5ns total length were conducted (Tab. 6.1). The distances between the monomers were constrained using the SHAKE[73] algorithm.

---

[1] `http://www.python.org`

[2] a parallelization technique, see Ref. [60]

Two different protocols were used, but showed no difference in results:

- A fast-cooling simulation scheme was used for the 2.5ns runs. The system was simulated at decreasing temperatures starting from 400K to 300K in 10K steps. At each temperature the system was simulated for 100ps. At 300K, a 2.5ns simulation was performed.
- The 5ns runs were energy minimized and equilibrated (position restrained relaxation) for 1ns at a temperature of 300K and then simulated for 5ns at the same temperature.

For both simulations schemes the last 2.5ns were evaluated with 5ps steps between the time frames.

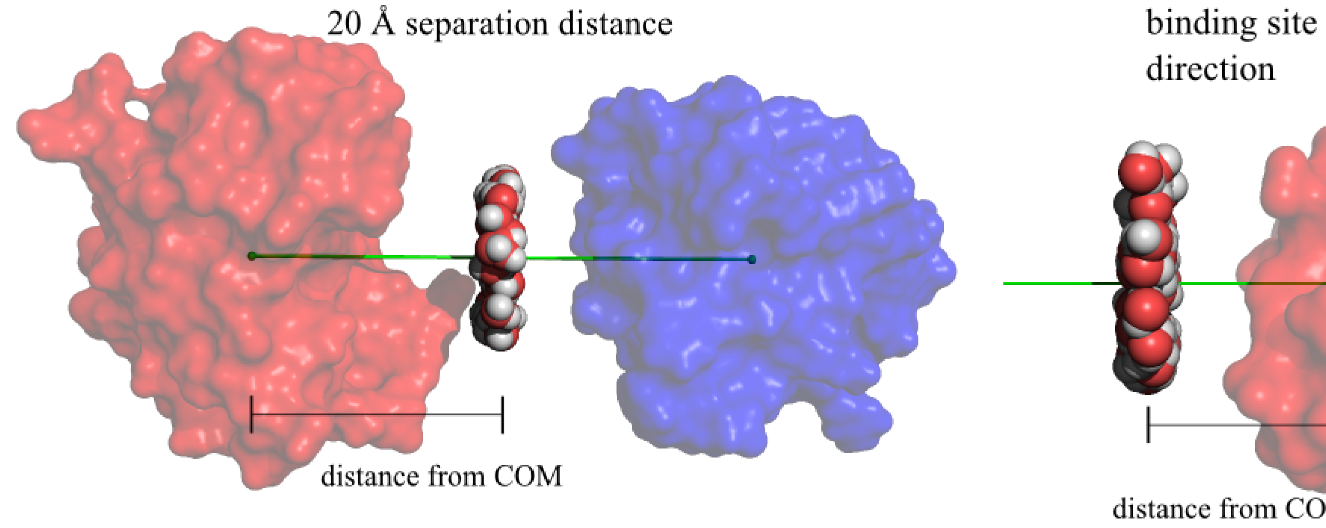
6.1.4 Analysis

Average angle distributions

Only the water in a 8Å cylinder around the COM-COM axis was considered from each frame (Fig. 6.1b on the next page and 6.1a). For the monomers, an artificial axis was set up in binding site or opposite direction through fitting to the complex structure. From these water molecules, the angle towards the axis and their position on the axis from the COMs was calculated and statistically interpreted. The closest 5000 water molecules[1] around the middle position were analyzed to obtain average dipole momentum and Shannon entropy. The Shannon entropy for each distribution was calculated according to equation (4.1 on page 28), with a sampling range of 180°, divided in 45 bins each ranging over 4°.

---

[1] from all 600 frames in the trajectory



a) Analyzed waters for complexes

b) Analyzed water for m

Figure 6.1: The location of the analyzed water molecules of the complex 6.1a and of the monomers 6.1b is shown as a superposition of 5 water frames. Only the waters in a cylinder with radius 8Å around the center of the axis are considered. 6.1a and 6.1b were rendered with PyMOL[108][a].

---

[a] `http://pymol.sourceforge.net`

Volumetric analysis

To calculate the volume occupied by the protein between the COMs I first created a surface with the msms[22, 23] program and a solvent radius of 1.4Å for 10 equidistant frames ranging over the 2.5ns. Only the triangulated surface inside the cylinder from both proteins was used to calculate the protein volume usage dependent from the position between the COMs (Fig. 6.2 on the next page). This was done by projection of each triangle on the cylinder's base plane and adding up the area while considering orientation and location of the triangles. The accuracy was limited by errors from triangulation and showed deviations of $\approx 5\%$ from the true values.

Dipole Fields

The analysis of the dipole fields has been done by sampling all water molecules in 3 dimensions in a specific area. The reference frame had to be set up for each frame separately. I used the COM-COM vector and the orthogonal part of the binding helix vector of Barnase as the direction for 2 of our 3 base vectors. According to this orthonormalized basis, the water molecules were assigned to a grid point for each frame. Then the total dipole vector at each grid point from all frames was calculated from average water orientation. Gridpoints with fewer than 300 assigned waters were omitted because of the too low water density. These gridpoints are close to a surface or in range of the fluctuating surface and do not provide a comparable dipole vector. Clipping of short vectors and concatenation of spatially and directional close vectors allows improved visualization of the vector field. For Figure 4.6, everything below 20% of the common maximum length was clipped and vectors not having a direction difference of more than 10° and a distance of 1.3 box edge lengths perpendicular to their direction were concatenated. The box edge length of 0.26Å and the grid dimension of 16x16x16 boxes covers approximately a cubic space between both COMs. Colors of the vector were chosen from green to red, according to the parallelism to the electrostatic potential gradient. The potential map was calculated with the apbs[109] program.

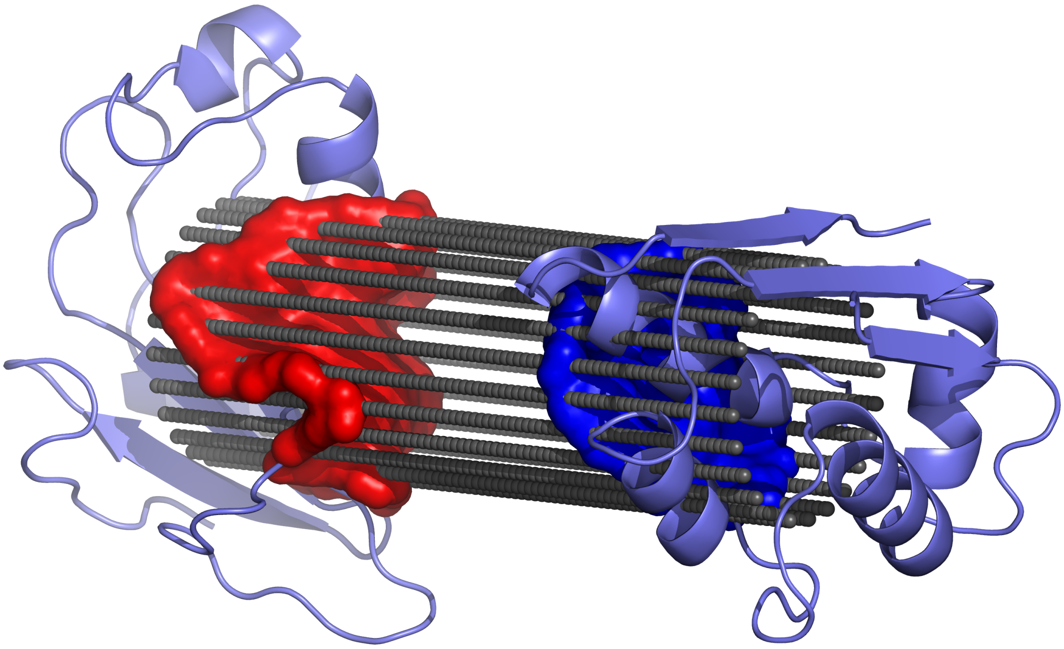


Figure 6.2: Figure 6.2 shows the relevant part of the protein surfaces for volumetric analysis inside the cylinder of 8Å. The cylinder between the center of masses is shown with grey spheres whereas the surfaces are shown in red and blue in volumetric representation. The figure has been created with PyMOL[108][a]

---

[a] `http://pymol.sourceforge.net`

6.1.5 Implementation Details

Analysis programs with low computational requirements were written in Python[1] with NumPy[2] and SciPy[3] extension modules. At the beginning, trajectories were stored as plain text PDB files and parsed from Python, however, this turned out to be very slow for large trajectories and/or large systems with water. Therefore, I wrote an wrapper interface with the ctypes[4] module (included in Python distribution since 2.5) to directly process compressed xtc trajectory files using the available shared libraries for gromacs trajectories.

Fieldline Visualization - a Gradient Walking Algorithm

Most molecular visualization programs allow isosurface plotting and coloring of surfaces with volumetric data. To directly compare dipole fields with a calculated electrostatic volumetric map, intuitively fieldlines would provide the best picture. Among most of the molecular visualization programs even two prominent ones, VMD and PyMOL, did not include fieldline visualization[5]. PyMOL was chosen as implementation target, because of its high rendering quality and versatile plugin architecture.

Support for fieldline visualization has been written as a PyMOL plugin completely in Python first. Python is an excellent rapid prototyping programming language suitable for proof of concept code. In the meantime, there is a fast C implementation in PyMOL[108] development version, which was inspired by my Python code. Here, the ideas behind the algorithm should be sketched.

As an initial point, I generated a 3D potential map from the structure by solving the Poisson-Boltzmann equation via the apbs[109] tool. The fieldlines are generated from these map file when walking along the gradient. From an initial point, the fieldline is constructed as a list of points. The points are calculated in parallel and antiparallel gradient directions. The gradient itself is formed from the weighted closest points in

---

[1] http://www.python.org

[2] http://numpy.scipy.org

[3] http://www.scipy.org

[4] http://docs.python.org/dev/lib/module-ctypes.html

[5] Oct 2006

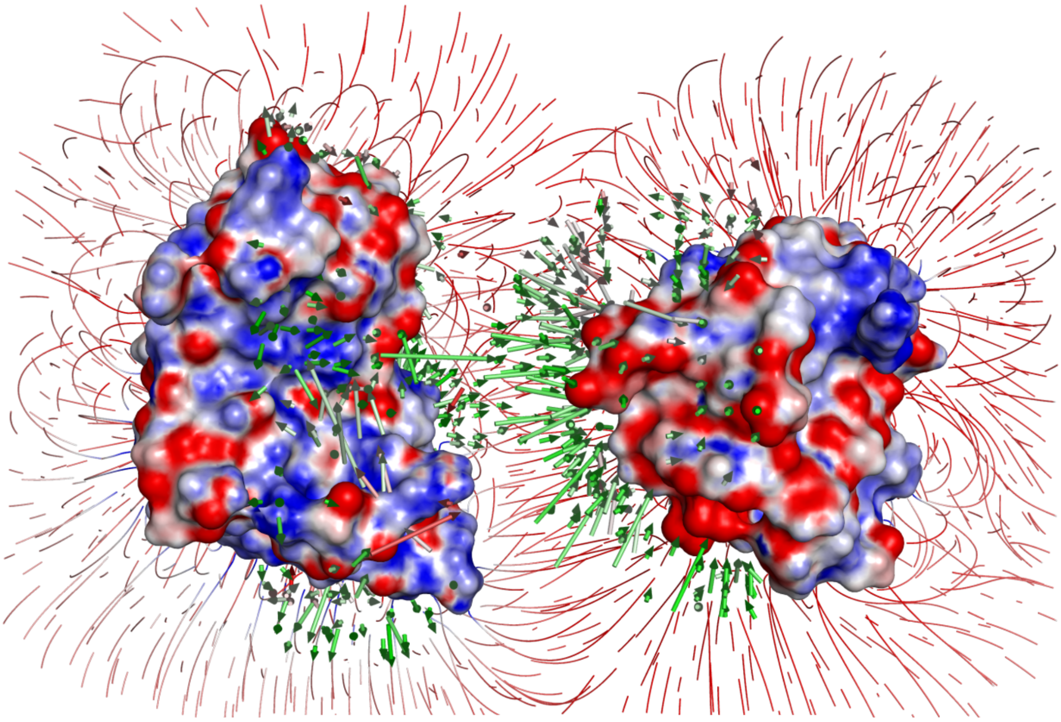


Figure 6.3: Electrical field lines and water dipole visualization of a 20Å separated Barnase-Barstar wildtype complex. Scene rendered with PyMOL[108]

each of the three spatial directions. The segment size stays constant along a fieldline, so that the gradient only provides the direction. As a termination condition, the angle between two of the subsequent segments was chosen, and showed to be very reliable. For subsequent fieldlines, the distance of each new point towards other fieldlines was checked, and in case of a too low distance, the current fieldline was terminated. The hookup points[1] were generated randomly and stored, so that fieldline visualization even of trajectory data was possible as movies.

---

[1] starting point for each field line calculation

The fieldlines were stored together with the potential at the fieldline points and loaded from PyMOL via Python's pickle/unpickle[1] functionality. The actual display in PyMOL makes use of Compiled Graphics Objects[2], which provide a limited subset of OpenGL functionality. A resulting fieldline visualization can be seen in Figure 6.3 on the preceding page.

6.2 Surface Mapping of Water Structure

Analysis of waters in a trajectory and mapping on surfaces involves computational optimizations. I first describe the basic ideas of the algorithm for trajectory (see Fig. 4.7 on page 35) processing, followed by a short overview of profiling tools used for optimizations of the code in my studies. Finally the concept and implementation of an octree is discussed.

6.2.1 Analysis Scheme

My implementation is based on the GROMACS[134] C libraries, and written completely in ANSI C. The program uses MPI library calls (see Section 4.2.3 on page 46), to allow execution in parallel environments.

During initialization, the rank[3] r and world communicator size[131][4] n of the process is determined, which allows only every $(i \cdot n + r)$ frame to be processed, $i = 1 \dots k$ where $k = ceil(\frac{frames}{processes})$. Data is send back to rank 0 at the end of each frame.

At the beginning of each frame, the coordinates are read and fed into an external program for surface triangulation of the solvent excluded surface (msms[22] from Michael Sanner or surf[126] from A. Varshney et al was used). The triangulated surface vertices and their normals were read in and stored in an octree. Two additional octrees for all water molecules in the frame and all protein atoms were created

---

[1] http://docs.python.org/lib/module-pickle.html

[2] http://pymol.sourceforge.net/newman/user/S0500cgo.html

[3] rank: e.g. process 3 of 7

[4] communicator size: total number of processes

(Water octree is used for RDF[1] calculation in future).

For every water atom, the nearest vertex is determined, the distance with respect to it and the angle towards the surface normal is computed and stored with the atom number, the vertex belongs to (see Fig. 4.7 on page 35). At the moment it is not possible to store it with the vertices directly on surface, since triangulation with the mentioned programs is not constant over time. The octree used for nearest neighbor searches is described in section 6.2.3 on the following page.

The generated (atomindex;distance;angle) tuples are read in and processed for visualization. For each atom, the angular distribution of waters is calculated and Shannon's entropy formalism is applied (4.1 on page 28). The obtained entropy values are written into the B-factor of a PDB file for visualization with PyMOL[108].

6.2.2 Profiling

Code profiling helps to find bottlenecks of implemented code and was mainly done via Intel's VTune® utility[2]. On AMD based computers, oprofile[3] kernel module was used in combination with AMD CodeAnalyst[4] environment. This solution is by at least one order slower, and allows only basic profiling. By looking at the produced callgraphs and how often each function has been called, one can focus on specific functions for optimizations. Profiling also allows to find the bottlenecks inside functions. VTune® reads several performance counters of the processor and allows detailed analysis of the time spend in each code section. This information e.g. lead to an optimization of memory allocation, since *malloc* turned out to be very time consuming. Replacing of divisions and alteration of some loop structures also allows speed improvements whereas the impact of optimizations can be strongly dependent on processor type and architecture itself.

---

[1] Radial Distribution Function

[2] `http://www.intel.com/software/products/vtune/`

[3] `http://oprofile.sourceforge.net/`

[4] `http://developer.amd.com/downloads.jsp`

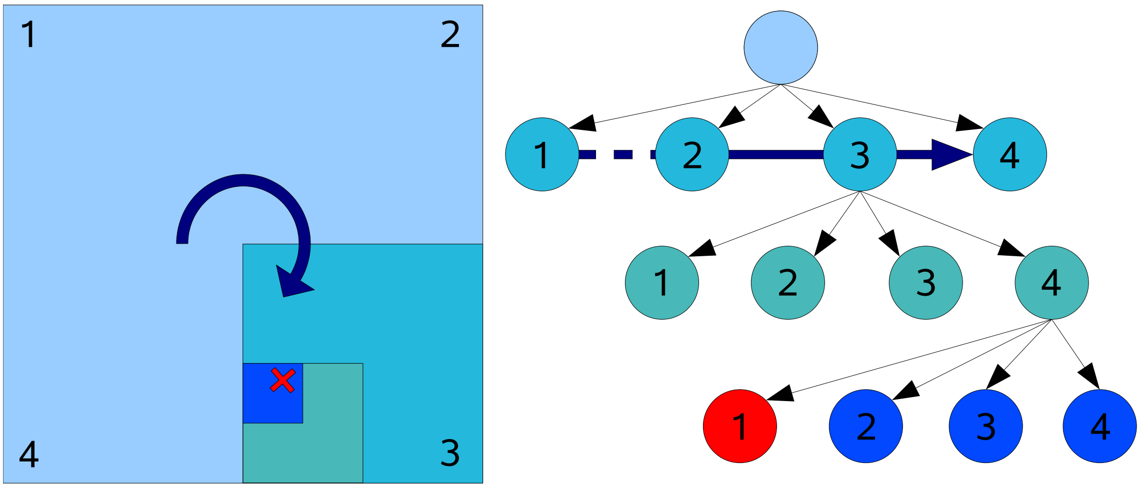


Figure 6.4: Figure 6.4 shows a schematic representation of a quadtree, which is the equivalent to an octree in two dimensions. The red x is stored in the red node in the hierarchical tree, while each node has four child nodes. Nodes without children are called leaves.

6.2.3 Spatial Indexing via Octrees

Octree Operation Mode

Here I describe the basic idea behind an octree and sketch its implementation. 3D space is divided up into 8 octants, and the data (or references to it) is stored in the leaves of a tree. If a node is already a leaf, which means a data point is already stored in it, it needs to be divided up and the data is pushed to deeper leaves. The octant is divided up until the two points are in different leaves. This implies that, an octree can never store two identical datapoints, luckily, this constellation is very rare in simulations.

Range queries return all elements in a certain range. While searching, we are descending to the maximum depth with a longer box edge than our query radius. In case of multiple dimensions, the shortest edge must fullfill this criteria. In the desired depth, a binary pattern is used, to search in all neighboring boxes. Only the distance of the leaves reachable from these boxes are checked. Nearest neighbor searches are derived from range queries, determining a close element by ascending the tree one level, and doing a range query with it's distance. Then the point with minimal distance is calculated from the result set of the range query.

Octree Implementation

My octree implementation is based on pointer connected structures and is written in plain ANSI C. Linear search can be faster for small data sets (< 500) due to compiler optimizations like loop unrolling on specific hardware. Here I show some exemplary code illustrating the octree node structure and filling algorithm.

Listing 6.2: File: octree.h octree node structure

```
1 struct octnode {
2         struct octnode *down[8];
3         float *coords;
4         struct octnode *up;
5         struct octant oct;
6         unsigned int is_leaf  :1;
7         unsigned int is_visited :1;
8 };
```

Listing 6.2 declares a C structure, the structure instances are initialized and freed with constructors and destructors not shown here. Line 2 is a pointer array to the 8 possible child nodes, Null pointers if no child node present. Line 6 marks the node as a leaf, if so float pointer in line 3 can point to float data. Data and structure has been separated to achieve smaller memory requirements for each node leading to a smaller overall memory usage of the structure. Secondly, a pointer on data allows storage of arbitrary tuples. Line 4 points on the parent node while line 5 inherits the position in parent octant as an octant structure (definition not shown here). Finally line 7 is a helper bit to speed up octree searches by avoiding redundancy visits of nodes and their children.

Listing 6.3: File: octree.c filling function

```
1  void insertCoords(struct octree *otree, float *basepntr){
2          struct octnode *node;
3          int i, recursed, octant, curbit;
4          float tmpcoord[3];
5          for(i=0;i<3;i++)
6                  tmpcoord[i]=basepntr[i]-otree->min[i];
7
8          recursed=1;
9          node=otree->rootnode;
10         float mypower=0.5;
```

insertCoords needs an octree structure together with a pointer on the first coordinate for octree insertion. In line 4-6, the coordinates are transformed to the octree's internal coordinate system. Line 9 initializes the node pointer to point on the root node of the octree.

Listing 6.4: File: octree.c filling function line 11-17

```
11         while (1)
12         {
13                 if (node->coords==NULL && node->is_leaf){
14                         node->coords=basepntr;
15                         otree->elements+=1;
16                         break;
17                 }
```

Line 11 opens the main loop, descending the tree until a suitable level for data point insertion is found. Line 13 checks, if the current octant is a leaf and does not already have data stored in it, if so line 14-16 store the data point and break the main loop.

Listing 6.5: File: octree.c filling function line 18-23

```
18                 if (node->is_leaf){
19                         node->is_leaf=0;
20                         otree->elements-=1;
21                         insertCoords(otree,node->coords);
22                         node->coords=NULL;
23                 }
```

If we are in a leaf node with already data stored in it as checked in Line 18, we need to split it up. Line 19 degrades the node to a normal node. Next line decreases the element count of octree, followed by a recursive call of *insertCoords* to reinsert the current coordinate one level deeper in line 21. Line 22 resets the coord pointer. This could be done more efficiently, by providing a deeper insertion point for reinsertion, but readability is favored since octree creation time is negligible.

Listing 6.6: File: octree.c filling function line 24-31

```
24          octant=0;
25          int tpow=1;
26          for (i=0;i<3;i++){
27              curbit=(int)(tmpcoord[i]/(otree->range[i]*mypower));
28              octant+=tpow*curbit;
29              tpow*=2;
30              tmpcoord[i]-=curbit*(otree->range[i])*mypower;
31          }
```

Starting from line 24, the rest of the main loop handles the case if we are not in a leaf node. First, the octant is set to zero in line 24. The variable *tpow* stores current dimensions bit. Line 26-31 loop over all dimensions. Line 27 determine the bit value for current dimension, line 28 adds the dimensions bit to the octant number if necessary. Line 29 increases the dimension bit value while line 30 decreases current coordinates if necessary.

Listing 6.7: File: octree.c filling function line 32-43

```
32          if (node->down[octant]==NULL){
33              node->down[octant]=octnode_ctor();
34              node->down[octant]->oct.top=octant_top(octant);
35              node->down[octant]->oct.left=octant_left(octant);
36              node->down[octant]->oct.front=octant_front(octant);
37          }
38          node=node->down[octant];
39          recursed++;
40          mypower*=0.5;
41      }
42      if (recursed>otree->depth) otree->depth=recursed;
43  }
```

Lines 32-37 create a new octant node and set its position in parent node, if pointer to child octant is a Null pointer. Line 38 recurses to the deeper node. Finally, the maximum depth is updated in octree structure, if the current recursion is deeper than the previous ones.

Listing 6.3 on page 73 to Listing 6.7 demonstrate principal octree handling. Search functions are more complex and contribute the majority to the octree code of about 1500 lines. Typical recursion depths are 8 for 40000 points, which illustrates the achievable speedups when searching or descending an octree.

6.3 Gold Simulations and Analysis

6.3.1 Force Field Implementation for Au Nanoparticles and Surfaces

The base for implementing interactions between 111-Au surfaces and aminoacids is the OPLS-AA force field, already available for GROMACS. Here I describe the modifications on existing files and the new gold topology files. For a detailed role, layout and usage of each file see chapter 5 in the GROMACS 3.3 manual[60].

Listing 6.8: File: top/atommass.dat

```
...
??? AU     196.00
??? AUC    0.5
??? AUI    0.0
```

Listing 6.9: File: top/vdwradii.dat

```
...
 ??? AU       0.142
```

The atommass.dat and vdwradii.dat (see listing 6.8 and 6.9) are used by GROMACS tools, which do not need explicit statement of the used force field. The van-der-Waals radius is used e.g. by the *genbox* utility to place solvent molecules or by *editconf* to calculate box boundaries. Atommasses are e.g. used for center of mass calculations but most tools like *g_traj* allow usage of a binary topology (.top) as secondary input to read out atommasses from the used force field.

Listing 6.10: File: top/ffoplsaa.atp

```
...
AUB      196.97     ; Gold in Bulk
AUC        0.5      ; Gold Dipole Charge
AUS      196.97     ; Gold on Surface
AUE      196.97     ; Gold at Edge
AUV      196.97     ; Gold at Vertex
AUI      196.97     ; Virtual Gold interaction site
```

The atomtypes of the force field are specified in ffoplsaa.atp (listing 6.10) file. Six additional atoms are needed to model interactions of the gold surface with other atoms. Even the virtual interaction sites AUI and dipole charges AUC have a mass greater than zero. In case of completely freezing the virtual interaction sites, their mass does not contribute to dynamics at all. If a restraint potential is used to model lattice vibrations, the virtual sites need a mass itself, to behave like gold atoms in a lattice. Mass parameters for AUI and AUC are purely artificial since molecular dynamics simulations will not be able to model the electron-gas properties of gold at all as MD simulates no explicit electron particles.

Listing 6.11: File: top/ffoplsaanb.itp

```
...
AUI        AUI    79    196.97      0.000      A    3.20000e-01   0.62000e+00
AUC        AUC    79      0.5       0.000      A    0.00000e+00   0.00000e+00
AUV        AU     79    196.97      0.000      A    3.20000e-01   0.62000e+00
AUE        AU     79    196.97      0.000      A    3.20000e-01   0.62000e+00
AUS        AU     79    196.97      0.000      A    0.00000e-00   0.00000e+00
AUB        AU     79    196.97      0.000      A    3.20000e-01   0.62000e+00
```

The van-der-Waals parameters of the additional atoms are declared in ffoplsaanb.itp (listing 6.11). At the moment, no special treatment for histidine/imidazole is implemented. Of course dipole charges have no van-der-Waals interaction with other atoms. Since van-der-Waals interactions with the surface are modelled with virtual sites, the surface gold atoms (AUS) do not contribute to the van-der-Waals interaction at all. Instead the van-der-Waals interaction is placed on the virtual sites (AUI), reproducing placement of charged particles on the top of gold atoms properly. The edges and vertex Atoms need to carry a van-der-Waals interaction to avoid intrusion of molecules at the edges. Their van-der-Waals parameters are purely artificial, since there is no experimental data available for modeling. DFT calculations are problematic as well as system size to model edges and vertices instead of surfaces increases and results in enormous computational efforts.

Listing 6.12: File: top/gold.itp

```
#include "gold_vertex.itp"
#include "gold_bulk.itp"
#include "gold_edge.itp"
#include "gold_surface.itp"
#include "gold_virtual.itp"

#ifdef GOLDPOSRES

#define GLDFX 1000
#define GLDFY 1000
#define GLDFZ 1000

#endif
```

Listing 6.12 shows the include topology file. Position restraints are defined here as they apply to all parts of the topology. Additionally, all sub-topologies are included for convenience reasons. Here I only show gold_surface.itp and gold_virtual.itp (Listing 6.13 and 6.14 on the following page) since edges, vertices and bulk gold atoms are modelled in the same fashion as the surface building blocks.

Listing 6.13: File: top/gold_surface.itp

```
[ moleculetype ]
; Name              nrexcl
GoldSurface              3
[ atoms ]
; nr        atype  resnr  residue  atom  cgnr  charge  mass      typeB  chargeB  massB
  1          AUS    1      AUS      AU    1      -0.3   196.97  ;  qtot   0
  2          AUC    1      AUS      AUC   2       0.3   0.5     ;  qtot   0

[ constraints ]
1 2 1 0.07

#ifdef GOLDPOSRES
[ position_restraints ]
     1    1 GLDFX GLDFY GLDFZ
#endif
```

Listing 6.14: File: top/gold_virtual.itp

```
[ moleculetype ]
; Name              nrexcl
GoldVirtualSite        3
[ atoms ]
; nr        atype  resnr  residue  atom  cgnr  charge  mass      typeB  chargeB  massB
  1          AUI    1      AUI      AUI   1       0.0   196.97  ;  qtot   0

#ifdef GOLDPOSRES
[ position_restraints ]
     1    1 GLDFX GLDFY GLDFZ
#endif
```

Each building block for a gold atom consists of two particles (listing 6.13). Both of them carry opposite charges leading to a total charge of zero. Their distance is constrained to 0.7Å and if position restraints instead of freeze groups are used, the gold atom(1) is restrained in all dimensions. The virtual sites (listing 6.14) carry no charges and are restrained in the same way as the gold atoms.

Beside the modifications on the force field, a conversion of the input gold structure is necessary. Additional charges need to be introduced and randomly placed as well as the existent gold atoms classified in bulk, surface, vertex and edge. Several extensive scripts are necessary, to convert arbitrary gold structures purely covered with 111 oriented surfaces to input structures and to create appropriate topologies of the system. Normally this is done by the GROMACS utility *pdb2gmx* which is only capable of handling building blocks for aminoacids and not yet arbitrary ones.

Appendix

List of Tables

List of Figures

Bibliography

1. S. H. Northrup and H. P. Erickson. Kinetics of protein protein association explained by brownian dynamics computer-simulation. *Proceedings of The National Academy of Sciences of The United States of America*, 89(8):3338–3342, **1992**.

2. G. Schreiber and A. R. Fersht. Rapid, electrostatically assisted association of proteins. *Nature Structural Biology*, 3(5):427–431, **1996**.

3. J. Janin. The kinetics of protein-protein recognition. *Proteins-Structure Function And Genetics*, 28(2):153–161, **1997**.

4. H. Neuvirth, R. Raz and G. Schreiber. Promate: a structure based prediction program to identify the location of protein-protein binding sites. *Journal of Molecular Biology*, 338(1):181–199, **2004**. doi:10.1016/j.jmb.2004.02.040.

5. K.-E. Gottschalk, H. Neuvirth and G. Schreiber. A novel method for scoring of docked protein complexes using predicted protein-protein binding sites. *Protein Engineering Design and Selection*, 17(2):183–189, **2004**. doi:10.1093/protein/gzh021.

6. R. R. Gabdoulline and R. C. Wade. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods*, 14(3):329–341, **1998**.

7. T. Selzer, S. Albeck and G. Schreiber. Rational design of faster associating and tighter binding protein complexes. *Nature Structural Biology*, 7(7):537–541, **2000**.

8. R. Grünberg, M. Nilges and J. Leckner. Flexibility and conformational entropy in protein-protein binding. *Structure*, 14(4):683–693, **2006**.

9. K. R. Gallagher and K. A. Sharp. A new angle on heat capacity changes in hydrophobic solvation. *Journal of the American Chemical Society*, 125(32):9853–9860, **2003**. doi:10.1021/ja029796n.

10. F. Despa, A. Fernandez and R. S. Berry. Dielectric modulation of biological water. *Physical Review Letters*, 93(22):228104, **2004**.

11. C. Yang and K. A. Sharp. Hydrophobic tendency of polar group hydration as a major force in type i antifreeze protein recognition. *Proteins*, 59(2):266–274, **2005**. doi:10.1002/prot.20429.

12. F. Despa. Biological water: Its vital role in macromolecular structure and function. *Annals of the New York Academy of Sciences*, 1066:1–11, **2005**. doi:10.1196/annals.1363.023.

13. I. Mihalek, I. Res and O. Lichtarge. On itinerant water molecules and detectability of protein-protein interfaces through comparative analysis of homologues. *Journal of Molecular Biology*, 369:584–595, **2007**. doi:10.1016/j.jmb.2007.03.057.

14. R. Zhou, X. Huang, C. J. Margulis et al. Hydrophobic Collapse in Multidomain Protein Folding. *Science*, 305(5690):1605–1609, **2004**.

15. P. Liu, X. Huang, R. Zhou et al. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature*, 437(7055):159–162, **2005**. doi:10.1038/nature03926.

16. J. Higo, M. Sasai, H. Shirai et al. Large vortex-like structure of dipole field in computer models of liquid water and dipole-bridge between biomolecules. *Proceedings of The National Academy of Sciences of The United States of America*, 98(11):5961–5964, **2001**.

17. T. Selzer and G. Schreiber. New insights into the mechanism of protein-protein association. *Proteins*, 45(3):190–8, **2001**. 0887-3585 (Print) Journal Article.

18. C. E. Shannon. A mathematical theory of communication. *Bell System technical Journal*, 27:623–656, **1948**.

19. C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication* (University of Illinois Press, Urbana), **1949**. ISBN 0-252-72548-4.

20. H. Lapid. *Water structure and dynamics at protein binding sites.* Master's thesis, Weizmann Institute of Science, Rehovot, Israel, **2006**.

21. T. J. Richmond. Solvent accessible surface-area and excluded volume in proteins - analytical equations for overlapping spheres and implications for the hydrophobic effect. *Journal Of Molecular Biology*, 178(1):63–89, **1984**.

22. M. Sanner, A. Olsen and J.-C. Spehner. Fast and robust computation of molecular surfaces. In *Proceedings of the 11th ACM Symposium on Computational Geometry*, pages C6–C7 (ACM, New York), **1995**.

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

24. J. J. Gray. The interaction of proteins with solid surfaces. *Current Opinion in Structural Biology*, 14(1):110–115, **2004**. doi:10.1016/j.sbi.2003.12.001.

25. Y. H. Dai and J. S. Evans. An energy-based mapping method for identifying the in-plane orientations of polypeptides and other macromolecules at crystalline interfaces. *Journal Of Chemical Physics*, 112(11):5144–5157, **2000**.

26. M. Sarikaya, C. Tamerler, A. K.-Y. Jen et al. Molecular biomimetics: nanotechnology through biology. *Nature Materials*, 2(9):577–585, **2003**. doi: 10.1038/nmat964.

27. G. Maruccio, A. Biasco, P. Visconti et al. Towards protein field-effect transistors: Report and model of prototype. *Advanced Materials*, 17(7):816–+, **2005**.

28. M. C. Daniel and D. Astruc. Gold nanoparticles: Assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chemical Reviews*, 104(1):293–346, **2004**.

29. J. Hautman and M. L. Klein. Simulation of a monolayer of alkyl thiol chains. *Journal Of Chemical Physics*, 91(8):4994–5001, **1989**.

30. S. Piana and A. Bilic. The Nature of the Adsorption of Nucleobases on the Gold [111] Surface. *Journal of Physical Chemistry B Condensed Matter Mater Surface Interfaces Biophysics*, 110(46):23467–23471, **2006**. doi:10.1021/jp064438i.

31. S. Rapino and F. Zerbetto. Modeling the stability and the motion of dna nucleobases on the gold surface. *Langmuir*, 21(6):2512–2518, **2005**.

32. R. Braun, M. Sarikaya and K. Schulten. Genetically engineered gold-binding polypeptides: structure prediction and molecular dynamics. *Journal of Biomaterials Science-Polymer Edition*, 13(7):747–757, **2002**.

33. R. J. Baxter, G. Teobaldi and F. Zerbetto. Modeling the adsorption of alkanes on an au(111) surface. *Langmuir*, 19(18):7335–7340, **2003**.

34. A. R. Bizzarri, B. Bonanni, G. Costantini et al. A combined atomic force microscopy and molecular dynamics simulation study on a plastocyanin mutant chemisorbed on a gold surface. *ChemPhysChem*, 4(11):1189–1195, **2003**.

35. J. Qian, R. Hentschke and W. Knoll. Superstructures of cyclodextrin derivatives on au(111): A combined random planting molecular dynamics approach. *Langmuir*, 13(26):7092–7098, **1997**.

36. G. M. Wang and A. C. Sandberg. Non-equilibrium all-atom molecular dynamics simulations of free and tethered dna molecules in nanochannel shear flows. *Nanotechnology*, 18(13):135702, **2007**.

37. J. J. Gerdy and W. A. Goodard. Atomistic structure for self-assembled monolayers of alkanethiols on au(111) surfaces. *Journal Of The American Chemical Society*, 118(13):3233–3236, **1996**.

38. S. Vemparala, B. B. Karki, R. K. Kalia et al. Large-scale molecular dynamics simulations of alkanethiol self-assembled monolayers. *Journal Of Chemical Physics*, 121(9):4323–4330, **2004**.

39. L. Z. Zhang, W. A. Goddard and S. Y. Jiang. Molecular simulation study of the c(4x2) superlattice structure of alkanethiol self-assembled monolayers on au(111). *Journal Of Chemical Physics*, 117(15):7342–7349, **2002**.

40. W. Mar and M. L. Klein. Molecular-dynamics study of the self-assembled monolayer composed of s(ch2)(14)ch3 molecules using an all-atoms model. *Langmuir*, 10(1):188–196, **1994**.

41. A. J. Pertsin and M. Grunze. Low-energy structures of a monolayer of octadecanethiol self-assembled on au(111). *Langmuir*, 10(10):3668–3674, **1994**.

42. H. Sellers, A. Ulman, Y. Shnidman et al. Structure and binding of alkanethiolates on gold and silver surfaces: implications for self-assembled monolayers. *Journal of the American Chemical Society*, 115(21):9389–9401, **1993**.

43. M. Sarikaya, C. Tamerler, D. T. Schwartz et al. Materials assembly and formation using engineered polypeptides. *Annual Review Of Materials Research*, 34:373–408, **2004**.

44. K. Goede, P. Busch and M. Grundmann. Binding specificity of a peptide on semiconductor surfaces. *Nano Letters*, 4(11):2115–2120, **2004**.

45. R. L. Willett, K. W. Baldwin, K. W. West et al. Differential adhesion of amino acids to inorganic surfaces. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(22):7817–7822, **2005**.

46. T. Mulders, P. Kruger, W. Swegat et al. Free energy as the potential of mean constraint force. *Journal Of Chemical Physics*, 104(12):4869–4870, **1996**.

47. S. Park and K. Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *Journal Of Chemical Physics*, 120(13):5946–5961, **2004**.

48. B. Roux. The calculation of the potential of mean force using computer-simulations. *Computer Physics Communications*, 91(1-3):275–282, **1995**.

49. D. Trzesniak, A. P. E. Kunz and W. F. van Gunsteren. A comparison of methods to compute the potential of mean force. *ChemPhysChem*, 8(1):162–169, **2007**.

50. W. L. Jorgensen. Optimized intermolecular potential functions for liquid alcohols. *Journal Of Physical Chemistry*, 90(7):1276–1284, **1986**.

51. J. Chandrasekhar, D. C. Spellmeyer and W. L. Jorgensen. Energy component analysis for dilute aqueous-solutions of Li+, Na+,F-, and Cl- ions. *Journal Of The American Chemical Society*, 106(4):903–910, **1984**.

52. W. L. Jorgensen, D. S. Maxwell and J. TiradoRives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, **1996**.

53. W. Damm, A. Frontera, J. TiradoRives et al. OPLS all-atom force field for carbohydrates. *Journal Of Computational Chemistry*, 18(16):1955–1970, **1997**.

54. H. Liu, F. Müller-Plathe and W. F. van Gunsteren. A force field for liquid dimethyl sulfoxide and liquid proporties of liquid dimethyl sulfoxide calculated

using molecular dynamics simulation. *Journal of the American Chemical Society*, 117:4363–4366, **1995**.

55. A. R. van Buuren and H. J. C. Berendsen. Molecular dynamics simulation of the stability of a 22 residue alpha-helix in water and 30 % trifluoroethanol. *Biopolymers*, 33:1159–1166, **1993**.

56. A. R. van Buuren, S. J. Marrink and H. J. C. Berendsen. A molecular dynamics study of the decane/water interface. *Journal of Physical Chemistry*, 97:9206–9212, **1993**.

57. W. F. van Gunsteren and H. J. C. Berendsen. *Gromos-87 manual.* Biomos BV, Nijenborgh 4, 9747 AG Groningen, The Netherlands, **1987**.

58. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura et al. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79:926–935, **1983**.

59. P. M. Morse. Diatomic molecules according to the wave mechanics. II. vibrational levels. *Physical Review*, 34:57–64, **1929**.

60. D. van der Spoel, E. Lindahl, B. Hess et al. *Gromacs User Manual version 3.3.* www.gromacs.org, **2005**.

61. P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annals of Physics*, 64:253–287, **1921**.

62. U. Essmann, L. Perera, M. L. Berkowitz et al. A smooth particle mesh ewald potential. *Journal of Chemical Physics*, 103:8577–8592, **1995**.

63. T. Darden, D. York and L. Pedersen. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *Journal Chemical Physics*, 98:10089–10092, **1993**.

64. A. Baumketner and J. E. Shea. The influence of different treatments of electrostatic interactions on the thermodynamics of folding of peptides. *Journal of Physical Chemistry B*, 109(45):21322–21328, **2005**.

65. R. S. Varekova, J. Koca and C. G. Zhan. Complexity and convergence of electrostatic and van der waals energies within pme and cutoff methods. *Internatinoal Journal Of Molecular Sciences*, 5:154–172, **2004**.

66. A. Cordomi, O. Edholm and J. J. Perez. Effect of different treatments of long-range interactions and sampling conditions in molecular dynamic simulations of rhodopsin embedded in a dipalmitoyl phosphatidylcholine bilayer. *Journal of Computational Chemistry*, 28(6):1017–1030, **2007**.

67. D. Van der Spoel and P. van Maaren. The origin of layer structure artifacts in simulations of liquid water. *Journal of Chemical Theory and Computation*, 2(1):1–11, **2006**. ISSN 1549-9618.

68. M. Patra, M. Karttunen, M. T. Hyvonen et al. Molecular dynamics simulations of lipid bilayers: Major artifacts due to truncating electrostatic interactions. *Biophysical Journal*, 84(6):3636–3645, **2003**.

69. H. Berendsen, J. Postma, W. van Gunsteren et al. *Interaction models for water in relation to protein hydration* (Springer), 1 edition, **1981**.

70. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, **1987**.

71. D. van der Spoel, P. van Maaren and H. J. C. Berendsen. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *Journal Of Chemical Physics*, 108(24):10220–10230, **1998**.

72. B. Hess, H. Bekker, H. J. C. Berendsen et al. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, **1997**.

73. J. Ryckaert, G. Ciccotti and H. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, **1977**.

74. S. Miyamoto and P. A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *Journal of Computational Chemistry*, 13:952–962, **1992**.

75. T. Fließbach. *Lehrbuch zur Theoretischen Physik 4. Statistische Physik (Spektrum Lehrbuch)* (Spektrum Akademischer Verlag), **1999**. ISBN 3827405564.

76. C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690–2693, **1997**.

77. C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5):5018–5035, **1997**.

78. S. Park, F. Khalili-Araghi, E. Tajkhorshid et al. Free energy calculation from steered molecular dynamics simulations using jarzynski's equality. *Journal Of Chemical Physics*, 119(6):3559–3566, **2003**.

79. C. E. Shannon. Communication in the presence of noise. *Proceedings of the I.R.E.*, 37:10–21, **1949**.

80. J.-B. Brissaud. The meanings of entropy. *Entropy*, 7(1):68–96 (electronic), **2005**. ISSN 1099-4300.

81. C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal Of Statistical Physics*, 52(1-2):479–487, **1988**.

82. L. Boltzmann. *Lectures on Gas Theory* (Dover Publications), **1995**. ISBN 0486684555.

83. M. Tribus and E. C. McIrvine. Energy and information. *Scientific American*, 225(3):179–&, **1971**.

84. J. Avery. *Information Theory and Evolution* (World Scientific Publishing Co. Pte. Ltd.), **2003**.

85. T. Welch. A technique for high-performance data compression. *Computer*, 17(6):8–19, **1984**.

86. R. W. Hartley. Production and purification of extracellular ribonuclease of bacillus-amyloliquefaciens (barnase) and its intracellular inhibitor (barstar) .1. barnase. *Preparative Biochemistry*, 2(3):229–&, **1972**.

87. R. W. Hartley. Barnase and barstar: two small proteins to fold and fit together. *Trends in Biochemical Sciences*, 14(11):450–454, **1989**.

88. V. Guillet, A. Lapthorn, R. W. Hartley et al. Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. *Structure*, 1(3):165–176, **1993**.

89. G. Schreiber and A. R. Fersht. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry*, 32(19):5145–5150, **1993**.

90. A. M. Buckle, G. Schreiber and A. R. Ferscht. Protein-protein recognition - crystal structural-analysis of a barnase barstar complex at 2.0-angstrom resolution. *Biochemistry*, 33(30):8878–8889, **1994**.

91. G. Schreiber, C. Frisch and A. R. Fersht. The role of glu73 of barnase in catalysis and the binding of barstar. *Journal of Molecular Biology*, 270(1):111–122, **1997**.

92. C. J. Camacho, Z. P. Weng, S. Vajda et al. Free energy landscapes of encounter complexes in protein-protein association. *Biophysical Journal*, 76(3):1166–1178, **1999**.

93. A. Spaar and V. Helms. Free energy landscape of protein-protein encounter resulting from brownian dynamics simulations of barnase : Barstar. *Journal Of Chemical Theory And Computation*, 1(4):723–736, **2005**.

94. A. Spaar, C. Dammer, R. R. Gabdoulline et al. Diffusional encounter of barnase and barstar. *Biophysical Journal*, 90(6):1913–1924, **2006**. doi:10.1529/biophysj.105.075507.

95. G. Strittmatter, J. Janssens, C. Opsomer et al. Inhibition of fungal disease development in plants by engineering controlled cell-death. *Bio-Technology*, 13(10):1085–1089, **1995**.

96. C. Mariani, V. Gossele, M. Debeuckeleer et al. A chimeric ribonuclease-inhibitor gene restores fertility to male sterile plants. *Nature*, 357(6377):384–387, **1992**.

97. C. Frisch, A. R. Fersht and G. Schreiber. Experimental assignment of the structure of the transition state for the association of barnase and barstar. *Journal of Molecular Biology*, 308(1):69–77, **2001**. doi:10.1006/jmbi.2001.4577.

98. F. B. Sheinerman, R. Norel and B. Honig. Electrostatic aspects of protein-protein interactions. *Current Opinion In Structural Biology*, 10(2):153–159, **2000**.

99. C. Frisch, G. Schreiber, C. M. Johnson et al. Thermodynamics of the interaction of barnase and barstar: Changes in free energy versus changes in enthalpy on mutation. *Journal of Molecular Biology*, 267(3):696–706, **1997**.

100. N. Jan. *Protein-Protein Interactions using Molecular Dynamics Simulations*. Master's thesis, Ludwig-Maximilians-University Munich, **2007**.

101. Y. J. Sheng, S. Y. Jiang and H. K. Tsao. Forced kramers escape in single-molecule pulling experiments. *Journal Of Chemical Physics*, 123(9):091102, **2005**.

102. Grubmüller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Physical Review. E. Statistical Physics, Plasmas, Fluids, And Related Interdisciplinary Topics*, 52(3):2893–2906, **1995**.

103. O. F. Lange, L. V. Schäfer and H. Grubmüller. Flooding in gromacs: accelerated barrier crossings in molecular dynamics. *Journal of Computational Chemistry*, 27(14):1693–1702, **2006**. doi:10.1002/jcc.20473.

104. G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in monte-carlo free-energy estimation - umbrella sampling. *Journal Of Computational Physics*, 23(2):187–199, **1977**.

105. V. S. Pande. Folding@home: Using desktop grid computing to overcome fundamental barriers in biomolecular simulation. *Abstracts Of Papers Of The American Chemical Society*, 230:U1295–U1295, **2005**.

106. N. J. Marianayagam, N. L. Fawzi and T. Head-Gordon. Protein folding by distributed computing and the denatured state ensemble. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(46):16684–16689, **2005**.

107. T. Selzer and G. Schreiber. Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. *Journal of Molecular Biology*, 287(2):409–419, **1999**.

108. W. DeLano. *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, Usa, **2002**.

109. N. A. Baker, D. Sept, S. Joseph et al. Electrostatics of nanosystems: Application

to microtubules and the ribosome. *Proceedings of The National Academy of Sciences of The United States of America*, 98(18):10037–10041, **2001**.

110. W. Humphrey, A. Dalke and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, **1996**.

111. S. T. Freer, J. Kraut, J. D. Robertus et al. Chymotrypsinogen - 2.5-a crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation. *Biochemistry*, 9(9):1997–&, **1970**.

112. H. J. Hecht, M. Szardenings, J. Collins et al. 3-dimensional structure of a recombinant variant of human pancreatic secretory trypsin-inhibitor (kazal type). *Journal Of Molecular Biology*, 225(4):1095–1103, **1992**.

113. T. Gallagher, J. Oliver, R. Bott et al. Subtilisin bpn' at 1.6 angstrom resolution: Analysis for discrete disorder and comparison of crystal forms. *Acta Crystallographica Section D-Biological Crystallography*, 52:1125–1135, **1996**.

114. D. Meagher. Geometric modeling using octree encoding. *Computer Graphics And Image Processing*, 19(2):129–147, **1982**.

115. K. Yamaguchi, T. L. Kunii, K. Fujimura et al. Octree-related data-structures and algorithms. *IEEE Computer Graphics And Applications*, 4(1):53–59, **1984**.

116. J. Edward and S. Ii. Fast collision detection with an n-objects octree. Technical report, The Ohio State University, Advanced Computing Center for the Arts and Design, **1993**.

117. I. J. Palmer and R. L. Grimsdale. Collision detection for animation using sphere-trees. *Computer Graphics Forum*, 14(2):105–116, **1995**.

118. K. Sharp. Entropy-enthalpy compensation: Fact or artifact? *Protein Science*, 10(3):661–667, **2001**.

119. T. Schindler and F. X. Schmid. Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry*, 35(51):16833–16842, **1996**.

120. J. Zielkiewicz. Structural properties of water: comparison of the spc, spce, tip4p, and tip5p models of water. *Journal of Chemical Physics*, 123(10):104501, **2005**. doi:10.1063/1.2018637.

121. D. S. Korchuganov, I. E. Gagnidze, E. N. Tkach et al. Determination of protein rotational correlation time from nmr relaxation data at various solvent viscosities. *Journal of Biomolecular NMR*, 30(4):431–442, **2004**. doi: 10.1007/s10858-004-4242-7.

122. K. P. Peters, J. Fauck and C. Frommel. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *Journal of Molecular Biology*, 256(1):201–213, **1996**.

123. B. K. Shoichet, W. A. Baase, R. Kuroki et al. A relationship between protein stability and protein frunction. *Proceedings of The National Academy of Sciences of The United States of America*, 92(2):452–456, **1995**.

124. C. M. Wells and E. Dicera. Thrombin is a Na+-activated enzyme. *Biochemistry*, 31(47):11721–11730, **1992**.

125. P. P. Wangikar, J. O. Rich, D. S. Clark et al. Probing enzymic transition state hydrophobicities. *Biochemistry*, 34(38):12302–12310, **1995**.

126. A. Varshney, F. P. Brooks and W. V. Wright. Linearly scalable computation of smooth molecular surfaces. *IEEE Computer Graphics and Applications*, 14:19–25, **1994**.

127. W. Gropp, E. Lusk, N. Doss et al. A high-performance, portable implementation of the mpi message passing interface standard. *Parallel Computing*, 22(6):789–828, **1996**.

128. G. Burns, R. Daoud and J. Vaigl. LAM: An Open Cluster Environment for MPI. In *Proceedings of Supercomputing Symposium*, pages 379–386. **1994**.

129. C. Kutzner, D. van der Spoel, M. Fechner et al. Improved gromacs scaling on ethernet switched clusters. *Recent Advances In Parallel Virtual Machine And Message Passing Interface*, 4192:404–405, **2006**.

130. C. Kutzner, D. V. D. Spoel, M. Fechner et al. Speeding up parallel gromacs on high-latency networks. *Journal of Computational Chemistry*, in print:NA, **2007**.

131. M. Snir and W. Gropp. *MPI: The Complete Reference* (MIT Press), **1998**.

132. A. G. Donchev, N. G. Galkin, A. A. Illarionov et al. Water properties from first principles: simulations by a general-purpose quantum mechanical polarizable force field. *Proceedings of The National Academy of Sciences of The United States of America*, 103(23):8613–8617, **2006**. doi:03.

133. D. van der Spoel, E. J. W. Wensink and A. C. Hoffmann. Lifting a wet glass from a table: a microscopic picture. *Langmuir*, 22(13):5666–5672, **2006**. doi: 10.1021/la053284f.

134. D. Van der Spoel, E. Lindahl, B. Hess et al. Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, **2005**.

Acknowledgements

During the past year, there were several people who helped me with my dayily struggle. Fist to mention is my tutor **Kay Gottschalk**. Although I should *blame* him for comments like “Can you *quickly* check this?” which lead to many of the new ideas, he also seriously helped me trough discussions, fair comments and opinions on my analysis. Additionally, he eased my life with really good coffee! **Jan Neumann** helped me finding even the simplest bugs in my code, while **Julia Schmitz** solved my ”Bayern Ticket“ reimburse problem with the administration. **Martin Tschöpe** not only shared his first name with me, often leading to confusion if both were present at the same time, he also demonstrated several times, that everything possible to write in 10 lines of Python is even more *compact* in one line of Perl. **Lorenz Rognioni** provided musical expertize in selecting ”Swiss Groove“ as our background webradio. When there was the time for mathematical struggles, **Steffen Mihatsch** always found illustrative examples. **Manuel Hammerich** found all bugs in my C code, althoug I did my best in hiding them even from experienced programmers.

Thanks also to **Jan Neumann, Lorenz Rognioni, Martin Tschöpe, Stefanie Daub, Sohaila Ouffata and Cornelius Kaschinski**, who carefully read the manuscript or parts of it.

Special thanks to **Stefano Corni** and **Francesco Iori** for the time in Modena, showing me what PROSURF is *really good for*, **Prof. Hermann Gaub** for making helpful suggestions well as all of **Gambicrew** for the discussion on various topics.

This thesis has been composed with LaTeX[1] and modules from the Comprehensive TEXArchive Network[2]. Figures were assembled with the Gimp[3], OpenOffice Draw[4] and Inkscape[5].

---

[1] `http://www.latex-project.org`
[2] `http://www.ctan.org`
[3] `http://www.gimp.org`
[4] `http://www.openoffice.org`
[5] `http://www.inkscape.org`

Statement

Herewith I declare that this diploma thesis was done completely by myself and only using the cited references.

Munich, December 20, 2012

Martin Hoefling