

A Simulation Analysis of Protein Hydration

DISSERTATION

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(Dr. rer. nat)

vorlegt von
Dipl.-Phys. Nikolai Smolin

aus der Ortschaft Komsomolskij, Bezirk Sowjetskij,
Gebiet Tjumen, Russland

eingereicht beim
Fachbereich Chemie
der Universität Dortmund

Dortmund 2006

Erstgutachter:

Prof. Dr. R. Winter

Zweitgutachter:

Prof. Dr. A. Geiger

Tag der mündlichen Prüfung:

16.05.2006

Danksagung

An dieser Stelle möchte ich mich der Aufgabe widmen, all jenen zu danken, die mir während meiner Promotion mit Rat und Tat Seite standen und somit großen Einfluss auf das Gelingen dieser Arbeit hatten.

Zuerst danke ich Herrn Prof. Dr. Roland Winter für seine großartige Unterstützung meiner Forschungsarbeit.

Herrn Prof. Dr. Alfons Geiger und Herrn Dr. Dietmar Paschek danke ich für ihr besonderes Interesse und die vielen Ratschläge zu den Computersimulations-Methoden. Herrn Prof. Dr. Alfons Geiger danke ich für die freundliche Übernahme des Zweitgutachtens.

Für die zahlreichen Ratschläge auf dem Gebiet der Perkolationstheorie abseits der üblichen Wege möchte ich an dieser Stelle ganz besonders herzlich unseren Kooperationspartnern Herrn Dr. Ivan Brovchenko und Frau Dr. Alla Oleinikova danken.

Nicht verschwiegen werden sollen auch alle übrigen wissenschaftlichen und nichtwissenschaftlichen Mitarbeiter des Lehrstuhls für Physikalische Chemie I, die durch ihre Kollegialität und Hilfsbereitschaft ein wunderbares Arbeitsklima geschaffen haben.

Mein ganz besonderer Dank gilt aber meiner Ehefrau Natalia und meiner Mutter, die mich auf meinem Weg begleitet, bestärkt und unterstützt haben.

Contents

Danksagung	III
Publications	VI
1 Introduction	1
2 Techniques and Methods	9
2.1 Monte Carlo Simulation	9
2.2 Molecular Dynamics Simulation	11
2.2.1 Forcefields.....	11
2.2.2 Non Bonded Interactions	12
2.2.3 Bonding Interaction.....	13
2.2.4 Constraint Dynamics.....	14
2.2.5 Periodic Boundary Conditions	14
2.2.6 Equation of Motion and Integrator.....	15
2.2.7 Manostat and Thermostat Using Weak Coupling.....	15
2.3 Protein Structure Analysis	17
2.3.1 The Root Mean Square Deviation.....	17
2.3.2 Volume and Surface.....	17
2.3.3 The Ramachandran Plot	19
2.4 Time Correlation Functions.....	20
2.5 Elements of Percolation Theory	21
3 Modelling Fully Hydrated Staphylococcal Nuclease	23
3.1 Introduction and System Setup.....	23
3.2 Structural Properties of SNase at Ambient Conditions	29
3.3 Water-Protein Interactions at Ambient Conditions	31
3.3.1 Water-Density Distribution at the Protein Surface	35
3.3.2 Water-Water H-Bonds Near the Protein Surface	38
3.3.3 Water-Polygons at the Protein Surface	39
3.4 Temperature Effect.....	44
3.5 Pressure Effect.....	52
3.6 The Temperature Induced Unfolding Pathway of SNase.....	62
3.7 Structural and Dynamic Properties of Hydration Water	68
3.8 Cosolvent Effects.....	82
3.9 Discussion and Conclusions	90

4 Percolation of Hydration Water in Biosystems	95
4.1 Introduction and System Setup.....	95
4.2 Formation of Spanning Water Networks.....	99
4.2.1 2D Percolation Transition of Water in Lysozyme Powders.....	99
4.2.2 Hydrophilic Spherical Surfaces.....	107
4.2.3 Single Lysozyme Molecule.....	113
4.3 Properties of Spanning Water Networks at Protein Surfaces.....	118
4.4 Discussion and Conclusions.....	141
5 Summary	148
6 Zusammenfassung	149
References	151

Publications

1. Smolin, N. & Winter, R. Molecular dynamics simulations of staphylococcal nuclease: the properties of water at the protein surface. *J. Phys. Chem. B* **2004**, *108*, 15928-15937.
2. Oleinikova, A.; Smolin, N.; Brovchenko, I.; Geiger, A.; Winter, R. Formation of spanning water networks on protein surfaces via 2D percolation transition. *J. Phys. Chem. B* **2005**, *109*, 1988-1998.
3. Smolin, N.; Oleinikova, A.; Brovchenko, I.; Geiger, A.; Winter, R. Properties of spanning water networks at protein surfaces. *J. Phys. Chem. B* **2005**, *109*, 10995-11005.
4. Oleinikova, A.; Brovchenko, I.; Smolin, N.; Krukau, A.; Geiger, A.; Winter, R. The percolation transition of hydration water: from planar hydrophilic surfaces to proteins. *Phys. Rev. Lett.* **2005**, *95*, 247802.
5. Brovchenko, I.; Krukau, A.; Smolin, N.; Oleinikova, A.; Geiger, A.; Winter, R. Thermal breaking of spanning water networks in the hydration shell of proteins. *J. Chem. Phys.* **2005**, *123*, 224905.
6. Smolin, N. & Winter, R. A molecular dynamics simulation of SNase and its hydration shell at high temperature and high pressure. *Biochim. Biophys. Acta* **2006**, *1764*, 522-534.
7. Mitra, L.; Smolin, N.; Ravindra, R.; Royer, C.; Winter, R. Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution – experiments and theoretical interpretation. *PCCP* **2006**, *8*, 1249-1265.

1 Introduction

Proteins influence a number of very important physiological and biochemical functions of life. A large variety of processes in living organisms significantly depends on protein activity. In many of those only the structure of a protein dominates its function (α -keratin in hair), protein dynamics is important to many others. All biological processes, which involve protein motion, are based on the protein dynamics. Muscle contraction, for example, is based on the combined action of actin and myosin. The ability to change the conformation is also essential for the function of many transport proteins, proteins involved in signal transduction, proteins in the immune system, and numerous enzymes [1]. In many enzymes, conformational changes serve to enclose the substrate, thereby preventing its release from the protein and ideally positioning it for the protein to perform its function, as in lysozyme. Dynamics plays an important role not only in the functional, native state of many proteins, but also the mechanism by which a protein reaches that native conformation, the protein folding process, which is a highly dynamic process.

The large part of the current knowledge about proteins is derived from experimental data (especially X-ray crystallography and Nuclear Magnetic Resonance (NMR)). There is currently no experimental technique that allows monitoring of protein conformational changes at atomic resolution as a function of time at time-scales of nanoseconds. Computer simulation techniques provide the only possibility to obtain dynamic information on proteins at atomic resolution in the picosecond to microsecond time range.

Proteins can be denatured by various chemical and physical treatments like high or low temperature, high pressure, extreme pH, co-solvents (urea, guanidine hydrochloride, etc.). The mechanism of unfolding may differ, however, for the different treatments. The denaturation by urea and GuHCl is accompanied by the binding of the co-solvents to the protein molecules [3], in combination with non-specific effects due to modifications of the solvent [4]. The action of an agent, whether it is structure stabilisation or destabilisation, precipitation or solubilisation, is defined only by the balance between the affinities of the protein for water and the particular agent [3]. Thus, the stabilization of a protein towards a particular denaturant is possible by either increasing its interaction with water or decreasing its interaction with the denaturant in the native compared to the denatured state. The effect of pressure on the stability of proteins is due to the volume of the protein-solvent system being smaller in the

unfolded state of the macromolecule than in its folded state. The activation volumes for both folding and unfolding are usually positive, leading to a decrease in both the folding and unfolding rate as a function of pressure. The destabilization by pressure results from the fact that the activation volume for folding is significantly larger in magnitude compared with that for unfolding [4]. Silva and Weber describe the pressure effect on the stability of globular proteins at high pressure as the infiltration of the protein by water [5]. At higher temperatures high entropy states are favoured. At temperatures above the unfolding temperature of the protein, ΔG becomes negative because the unfavourable entropy ($-T\Delta S$) overcompensates the enthalpy part (ΔH). ΔS and ΔH themselves, however, are also dependent on temperature. They both increase with temperature, but the entropy term $-T\Delta S$ increases faster [6].

Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. They function as catalysts, transport and store other molecules such as oxygen, provide mechanical support and immune protection, generate movement, transmit nerve impulses, and they control growth and differentiation. Proteins are large polypeptides consisting of many repeating units, which are called amino acids. Proteins are very complicated molecules. With 20 different amino acids that can be arranged in any order to make a polypeptide of up to thousands of amino acids long, their potential for variety is extraordinary. This variety allows proteins to function as exquisitely specific enzymes that compose a cell's metabolism. An *E. coli* bacterium, one of the simplest biological organisms, has over 1000 different proteins working at various times to catalyse the necessary reactions to sustain life. Amino acids are the building blocks of proteins. An amino acid consists of a central carbon atom, called α carbon, linked to an amino group, a carboxylic acid group, a hydrogen atom, and a distinctive group R (Figure 1.1). The group R is referred to as a side chain. The twenty amino acids found in biological systems are usually classified by properties of the side chain into four groups: acidic, basic, hydrophilic (polar), and hydrophobic (nonpolar). Proteins are linear polymers formed by linking the carboxyl group of one amino acid to the amino group of another amino acid with a peptide bond. These bonds are formed via a dehydration synthesis reaction between the carboxyl group one amino acid with the amino group of the next amino acid (Figure 1.1).

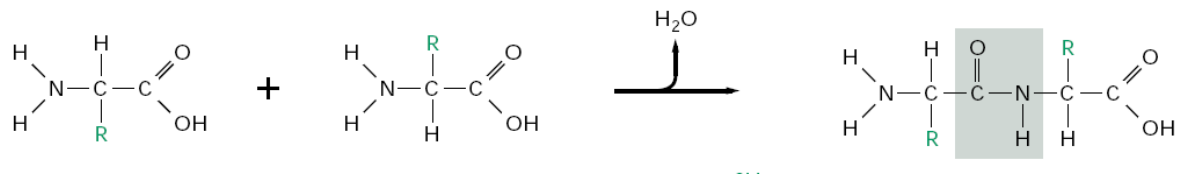


Figure 1.1: Peptide bond formation. Amino acids are commonly joined together by an amide linkage, called a peptide bond. Peptide bond: the four atoms in each grey box form a rigid planar unit. There is no rotation around the C-N bond.

Proteins exhibit multiple structural levels. The basic level is its primary structure. A protein's primary structure is simply the order of its amino acids. The protein secondary structure refers to certain common repeating structures found in proteins. There are two types of secondary structures: α -helix and β -sheet (Figure 1.2). An α -helix is a tight helix formed within the polypeptide chain. The polypeptide main chain makes up the central structure, and the side chains extend out and protrude away from the helix. The CO-group of one amino acid (n) is hydrogen bonded to the NH-group of the amino acid four residues away ($n+4$). In this way every CO- and NH-group of the backbone is hydrogen bonded. α -helices are most commonly made up of hydrophobic amino acids, because hydrogen bonds are generally the strongest attraction possible between such amino acids. α -helices are found in almost all proteins to various extents. β -sheets are the other type of secondary structure. They can be either parallel or anti-parallel. The tertiary structure is the full 3-dimensional folded structure of the polypeptide chain. The quaternary structure is only present if there is more than one polypeptide chain (Figure 1.3).

The contrasting distribution of polar and nonpolar residues reveals a key facet of protein architecture. In an aqueous environment, protein folding is driven by the strong tendency of hydrophobic residues to be excluded from water. Recall that a system is more thermodynamically stable when hydrophobic groups are clustered rather than extended into the aqueous surrounding. The polypeptide chain therefore folds so that its hydrophobic side chains are buried and its polar, charged chains are on the surface.

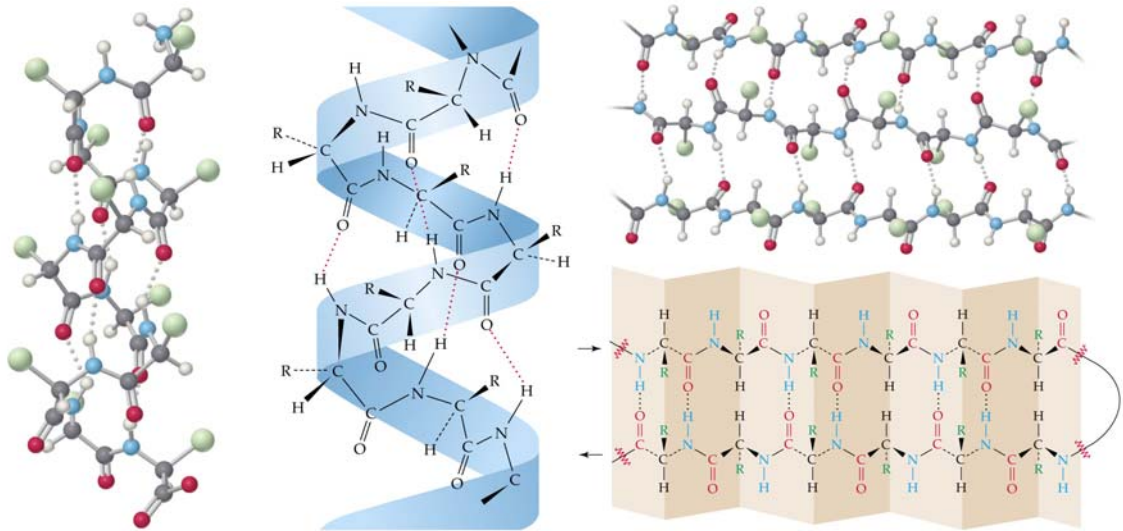


Figure 1.2: Left panel: structure of the α -helix. A side view of a ball-and-stick version depicts the hydrogen bonds (dashed lines) between NH and CO groups. Right panel: structure of the β sheets. Hydrogen bonds between NH and CO groups connect each amino acid on an adjacent strand, stabilizing the structure. Adapted from [7].

Interactions are the key means by which molecules interact with one another – enzymes with their substrates, hormones with their receptors, antibodies with their antigens. The strength and specificity of interactions are highly dependent on the medium in which they take place, and the majority of biological interactions take place in water. Two properties of water are especially important biologically:

1. Water is a polar molecule. The water molecule is not linear and so the charge distribution is asymmetric. The water molecule is thus an electrically polar structure.
2. Water is highly cohesive. Water molecules interact strongly with one another through hydrogen bonds. These interactions are apparent in the structure of ice. Networks of hydrogen bonds hold the structure together; similar interactions link molecules in liquid water and account for the cohesion of liquid water. The highly cohesive nature of water dramatically affects the interactions between molecules in aqueous solution.

The polarity and hydrogen-bonding capability of water make it a highly interacting molecule. Water is an excellent solvent for polar molecules; the reason is that water greatly weakens electrostatic forces and hydrogen bonding between polar molecules by competing for their attractions. The existence of life on Earth depends critically on the capacity of water to dissolve a remarkable array of polar molecules that serve as fuels, building blocks, and

catalysts. High concentrations of these polar molecules can coexist in water, where they are free to diffuse and interact with one another.

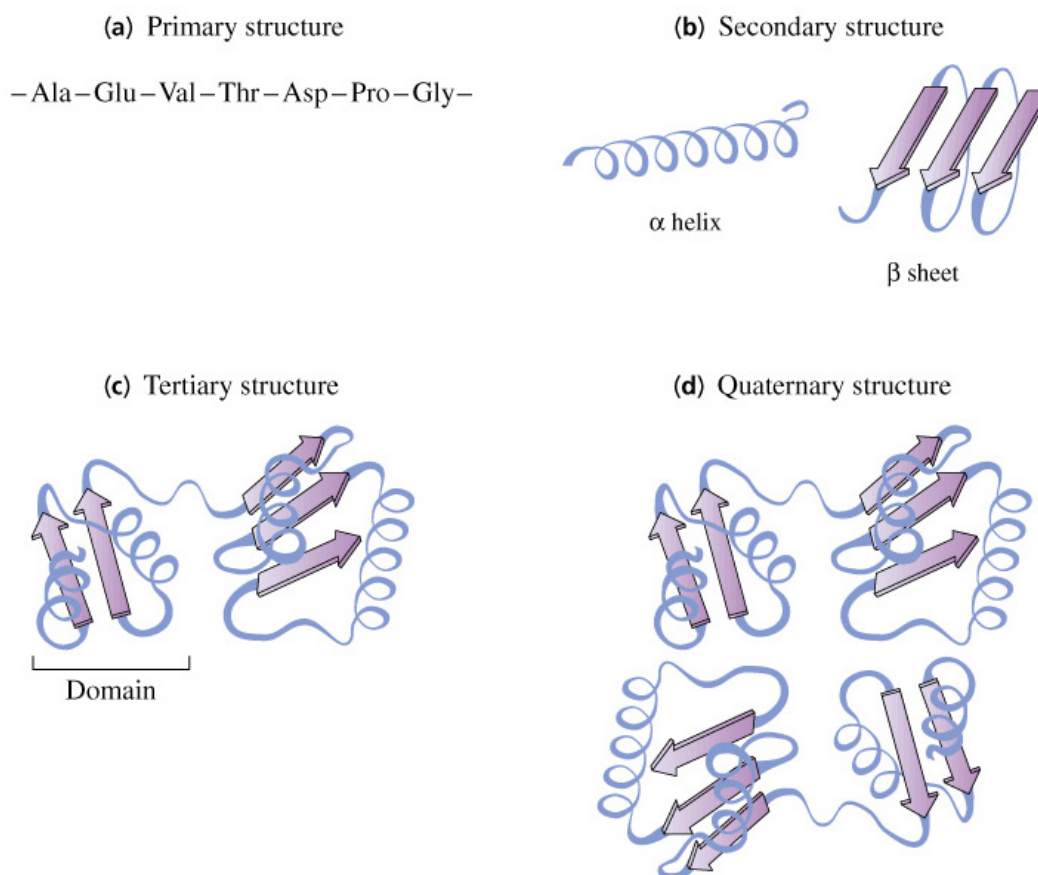


Figure 1.3: Schematic illustration of the various structural forms characteristic of proteins. The amino acid sequence of a protein's polypeptide chain is called its primary structure (a). Different regions of the sequence form local regular secondary structures, such as α -helices or β -sheets (b). Packing such structural elements into one or several compact globular units called domains forms the tertiary structure (c). The final protein may contain several polypeptide chains arranged in a quaternary structure (d). By formation of such tertiary and quaternary structure amino acids far apart in the sequence are brought close together in three dimensions to form a functional region, an active site. Adapted from [8].

The question of protein hydration, i.e., the interaction of protein molecules with water, and of the involvement of water molecules in protein reactions has been a focus of attention for the good part of the century [9]. Yet, the number of water molecules that interact with one protein molecule has been an elusive quantity, notwithstanding measurements by transport and NMR techniques. The reason for this lies in the essence of the physical phenomenon of

protein hydration. The fact is that there is no rigid shell of water around a protein molecule, but rather there is a fluctuating cloud of water molecules that are thermodynamically affected more or less strongly by the protein molecule. Hydration water is an expression of the sum of all the thermodynamic perturbations of water molecules by a protein molecule.

There are several definitions of hydration water. Rupley and Careri [10] suggested that the hydration shell could be defined as the water associated with the protein at the hydration end point, namely until a level of hydration is reached beyond which further addition of water produces no change and only dilutes the protein. One can assume that this shell represents a monolayer coverage of the protein surface. Water outside the hydration shell, i.e. in the bulk, is perturbed to a significantly smaller extent.

Water near a boundary forms one or two layers of specifically ordered water molecules. The properties of this “bound water” (orientational ordering, diffusivity, etc.) differ noticeably from the properties of bulk liquid water. Near the boundary, the first water monolayer forms a hydrogen bonded network, whereas the second water layer connects the specific first layer with the rest of the water.

Hydration water plays an important role in protein functions [10-12]. Local and orientational ordering of water molecules near hydrophobic and hydrophilic parts of protein surfaces were extensively studied during past decades both experimentally [13-17] and by computer simulations (see, for example, Refs. 18-21). Dynamical properties of water are strongly affected by the protein surface and their changes are sensitive to the particular chemical structure of the protein [22-27]. Structural and dynamical properties of proteins depend on the hydration level [28-33]. In particular, about one “monolayer” of water is required for restoring the full internal dynamics of proteins and their function [10-12].

Experimental studies on protein powders indicate the onset of some biological functions only when an *infinite* water network spans the protein surface [10, 34-38]. It is generally accepted that enzymes require internal flexibility for their biological activity, but what kind of motions are required is not yet clear [39]. There seems to be no direct correlation between the enzymatic activity and the global dynamics of proteins [40]. However, the biological activity of proteins could be driven by the dynamics of hydration water [41]. In fact, the effect of solvent fluctuations (in the enthalpy, volume and dipole moment) on proteins may be of fundamental importance, as fluctuations permit conformational motions. These motions may be “slaved”, or they may be “non-slaved” in the case of the protein motions being independent of solvent fluctuations [42]. Slaved motions have therefore rates that are proportional to the fluctuation rate of the solvent. In a recent study on myoglobin [42] it has been found that the

activation *enthalpy* of protein conformational and vibrational dynamics is controlled by the activation enthalpy of the solvent. Hence, the fluctuations of the amino acid residues and their hydration shell are coupled to and dominated by the surrounding solvent thermal bath. On the contrary, the protein and its hydration shell control the activation *entropy* through the shape of the energy landscape [42]. The prevalence for slaved motions, such as opening and closing of channels, ligand binding and enzyme catalysis, highlights the importance of the solvational environment of cells for the function of proteins.

Because the onset of enzymatic activity coincides with the formation of an infinite (spanning) network of the hydration water via a 2D percolation transition [10], a qualitative change of some dynamic properties of the hydrated protein should be expected at the percolation threshold.

The rationale of this work was to gain insight into the structural and dynamic properties of hydration water and staphylococcal nuclease (SNase) by using computer simulation methods. In order to understand the initial stages of pressure-induced protein unfolding, such as changes in hydration properties and reorganization of surface groups, the properties of the hydration water at the protein surface at different pressures were studied in detail. To yield a molecular interpretation of the different terms contributing to the partial protein volume and its temperature dependence and hence a better understanding of the experimental data, molecular dynamics computer simulations on SNase were also carried out at different temperatures and compared with the experimental data. Also a high-temperature molecular dynamics simulation of SNase was performed in order to reveal its temperature-induced unfolding pathways. Furthermore, the effects of urea and glycerol as co-solvents on the hydration of SNase were investigated over a wide concentration range.

In order to clarify the role of water in the appearance of biological function of proteins, the study of structural and dynamic properties of proteins should be carried out with respect to the formation of collective as well as individual *spanning* water networks. This study should be accompanied by the comparison of various properties of the hydration water networks above and below the percolation threshold. The specific properties of a *spanning* network of hydration water have not been studied yet, contrary to the various properties of local water networks [12-21]. To analyse the properties of the hydration water at the percolation transition threshold, the first computer simulation study of the percolation transition of water in a model lysozyme powder and on the surface of a single lysozyme molecule were performed. In order to develop an appropriate method to study the formation of a spanning network in finite, closed systems, such as the surface of a single protein, we studied

additionally the clustering of water on the surface of hydrophilic spheres of several sizes. This method was then applied to locate the percolation transition of water at the surface of a single lysozyme molecule.

2. Techniques and Methods

Computer simulations came into fashion among scientists when hardware became affordable. The computer experiments are done to study systems of interest to physicists, chemists and biologists. Molecular simulations are about solving the statistical mechanical equations by numerical techniques. Computer simulations solve the equations of motion of many particle systems over a time scale to yield information about the behaviour of a system.

This chapter gives a general overview of the computer simulation methods. The molecular dynamics (MD) method was developed over the last decades from a technique to study the dynamics and structure of solid spheres and Lennard-Jones particles to a powerful method to study different types of systems at atomic resolution [43]. In the field of biophysics and biochemistry, a large body of MD studies on proteins in vacuum and in solvents are available. The development of this particular use of molecular dynamics was significantly stimulated in the 1980's, when a number of general-purpose force fields for water (TIP3P, TIP4P, SPC/E), proteins and DNA, e.g., AMBER [44], CHARMM [45], GROMOS [46], and OPLS [47], became available.

2.1 Monte Carlo Simulation

The Monte Carlo method was developed by von Neumann, Ulam, and Metropolis at the end of the Second World War to study the diffusion of neutrons in fissionable material. The Metropolis Monte Carlo method aims to generate a trajectory in the phase space, which samples a chosen statistical ensemble. Monte Carlo simulations are a class of algorithms which use random number generators to sample the space of microscopic states. The ensemble averages are estimated by averaging over the set of states that were generated by the Monte Carlo procedure. The Metropolis method is a frequently used method to achieve appropriate sampling. It repeatedly tries to change the conformation a little bit (moving a few atoms) and then decides whether to accept the new conformation or to retain the previous one. Such a prospective change in the conformation is called a 'move'. The acceptance criterion should be chosen in such a way that the conformations are sampled with the desired probability (usually the Boltzmann distribution).

The method of generating a new conformation (on the basis of the previous one) is stochastic and should satisfy ‘detailed balance’: The probability of generating the state j when in the state i should be the same as the probability of generating the state i when in the state j . Furthermore, ‘ergodicity’ is required: One should be able to reach each microscopic state by one or more of these sampling steps.

If the sampling procedure satisfies detailed balance and ergodicity, the conformations are sampled with uniform probability. A popular acceptance criterion for achieving Boltzmann statistics is [48]:

$$P(\text{accept}) = \min(1, \exp(-\beta\Delta E)) \quad (2.1)$$

where P is the probability to accept a move with a change in energy ΔE , $\beta = 1/kT$, k is Boltzmann’s constant, T the temperature, and

$$\Delta E = E(\text{new}) - E(\text{previous}) \quad (2.2)$$

A convenient way to implement this in a computer program is:

1. Compute the change in energy ΔE .
2. If $\Delta E < 0$, accept the new conformation always.
3. If $\Delta E > 0$, compute a random number r from a uniform distribution on the interval $[0, 1]$ and compare it to $\exp(-\beta\Delta E)$. If $r < \exp(-\beta\Delta E)$, accept the new conformation. Otherwise, retain the previous conformation.

It is essential that, when a move is rejected, the previous conformation is taken into account once more in the statistics (for instance, in the averages of physical quantities). Not doing so would result in biased statistics and incorrect ensemble averages. Note that a move which leads to an excluded volume overlap should be handled as if $\Delta E = \infty$ and should therefore always be rejected. The Metropolis procedure described above generates a sequence of conformations, some identical to the previous one, some slightly altered. Ensemble averages can be calculated simply by averaging over the generated conformations. No weights are needed. One of the drawbacks of the method is that subsequent conformations are very much alike or identical (if the move was rejected). It takes a number of steps to obtain a conformation that is uncorrelated. Usually, such a conformation is artificial, and a number of steps is needed to reach a conformation that is uncorrelated to the initial conformation. This is called ‘equilibration’. The conformations that are generated during equilibration should not be

included in the statistics; they must be discarded. There is a lot of freedom in choosing the Monte Carlo moves, provided they satisfy detailed balance and ergodicity. This freedom of choice should be used to optimise the efficiency of the sampling: the phase space (the space of microscopic states) should be traversed as fast as possible.

2.2 Molecular dynamics simulation

2.2.1 Force fields

The Hamiltonian H of a classical system is given as the sum of the potential and the kinetic energy functions. It only depends on the particle positions (\mathbf{q}) and velocities (\mathbf{v}) (conjugate momenta \mathbf{p} , respectively). With the choice of performing our simulation in Cartesian space, the momenta \mathbf{p} depend on the particle velocities \mathbf{v} ; the classical Hamiltonian \mathbf{H} is now the sum of a kinetic and potential energy part $V(\mathbf{r})$:

$$H = H(\mathbf{q}, \mathbf{p}) = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}). \quad (2.3)$$

The kinetic part of the Hamiltonian uses the particle masses (m) and momenta of particles as parameters. The potential energy depends on the intermolecular interactions, which are position dependent. The intermolecular interactions are described by various forcefields, defined to fit certain transport/thermodynamic properties of interest and are discussed in detail later. It gives an expression for the potential energy as a function of particle coordinates (\mathbf{r}). This expression consists of different terms.

The negative of the derivative of the potential energy $V(\mathbf{r})$ with respect to the coordinates \mathbf{r}_i equals the force $\mathbf{f}_i(\mathbf{r})$ acting on a particle i

$$\mathbf{f}_i(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}_i} \quad (2.4)$$

From Newton's classical equation of motion, the positions and time are related according to the following equations

$$\mathbf{v}_i = \frac{d\mathbf{r}_i}{dt}, \quad \dot{\mathbf{v}}_i = \mathbf{a}_i = \frac{d\mathbf{v}_i}{dt} \quad (2.5)$$

where \mathbf{a}_i is the acceleration acting on the particle. The connection to the potential energy and force is given by equation 2.4 and

$$\mathbf{f}_i = m_i \mathbf{a}_i . \quad (2.6)$$

Besides some technical issues, which we explain later, the motion of a particle can be calculated from the force field and one starting configuration. The potential energy of a system is the sum of its intermolecular and intramolecular contributions. The interaction between atoms of different molecules is described by the intermolecular interaction potential energy. The interaction within the atoms of a single molecule is called the intramolecular interaction energy. The intramolecular interaction energy is thus divided into bond angle energies, bond torsion energies, dihedral angle energies and the non-bonded interaction energies between atoms of a single molecule, torsion and dihedral interactions.

For biomolecules, the potential energy corresponds to a set of semi-empirical functions that model the effective interactions between the particles. There has been a considerable effort to develop reliable interaction functions or force fields for proteins. Typical examples are AMBER [44], CHARMM [45], OPLS [47], and GROMOS [46]. In general, these force fields depend on the functional forms and parameters used to model each interaction. The interactions are modelled by two types of physical terms: Non-bonded interactions ((fixed) charge-charge and van der Waals interactions) and bonded interactions (stretching, bending and torsions). In addition, non-physical terms to restrain and constrain the system may also be included. It is justified to argue that the quality of the interaction functions, together with the time length of the simulation, are key points that determine the accuracy of the data generated.

2.2.2 Non-bonded Interactions

The separation of a forcefield into distinct mathematical terms in molecular dynamics simulations is reasonable usually for computational convenience, although quantum mechanics is a better way to describe intermolecular interactions. Each non-bonded interaction can be modelled using a Coulombic potential and a Lennard-Jones potential [43, 48, 49, 50]. The non-bonded interactions are divided into electrostatic and dispersion/repulsion (induced dipole, Pauli repulsion) contributions:

$$E_{\text{non-bonded}}(i, j) = E_{\text{LJ}}(i, j) + E_{\text{Coulomb}}(i, j) \quad (2.7)$$

$$E_{\text{LJ}}(i, j) = 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \quad (2.8)$$

$$E_{\text{Coulomb}}(i, j) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (2.9)$$

where (partial) charges are q_i and q_j , ε_0 is the vacuum permittivity, the Lennard-Jones parameters are σ_{ij} (LJ-radius) and ε_{ij} (LJ-energy).

2.2.3 Bonding Interactions

Throughout this thesis, all systems are simulated using the AMBER94 force field [51]. The AMBER94 force field was developed to describe biomolecular systems in the condensed phase [51].

The bonded terms E_{bonded} are:

Bond-stretching E_{bond} between two covalently bonded atoms:

$$E_{\text{bond}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 \quad (2.10)$$

with r is the bond length between bonded atoms, and the parameters K_r and r_{eq} are defined for each type of pair atoms.

Bond-angle bending E_{angle} :

$$E_{\text{angle}} = \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 \quad (2.11)$$

with θ is the angle between the atom triplet $i-j-k$, where atoms $i-j$ and atoms $j-k$ are covalently bonded. K_θ and θ_{eq} are defined for each type of atom triplets.

Dihedral-angle torsion E_{dihedral} :

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos(n\varphi - \gamma)) \quad (2.12)$$

to set interactions for the quadrupole of atoms $i-j-k-l$. The angle φ is defined as the angle between planes $i-j-k$ and $j-k-l$. V_n , γ and n are set for each type of atom quadruplets.

2.2.4 Constraint Dynamics

The constraint dynamics is handled by the SHAKE algorithm [52]. The SHAKE method allows one to consider atomic connectivity without harmonic bonds. Valence bonds vibrate at high frequency and impose a small integration time-step to a simulation. This method allows us to use larger time steps (e.g., from 1 fs to 2 fs). Consider two atoms bonded to each other at a fixed distance, a . The equality is usually written down in the form of a holonomic constraint:

$$r_k^2 - a_k^2 = 0. \quad (2.13)$$

In a constrained system, the coordinates of the particles are not independent of each other, and the equations of motion in each of the coordinate directions are related. The constraint forces lie along the bonds at all times. For each constrained bond, there are two equal, but opposite forces on the two atoms that comprise the bond, and therefore no energy is put into the system.

A comprehensive introduction to SHAKE is provided in the textbook of Allen and Tildesley [43].

2.2.5 Periodic Boundary Conditions

In view of the fact that the system size is small, a cluster of particles will have a lot of unwanted boundaries with its environment (vacuum). By using periodic boundary conditions, one avoids real phase boundaries.

There are several possible shapes for space filling unit cells. Some, like a rhombic dodecahedron and the truncated octahedron [53] are closer to a sphere than a cube and are therefore more economical for studying a (approximately spherical) macromolecule in solution, since fewer solvent molecules are required to fill the box given a minimum distance between macromolecular images.

In order to establish periodic boundary conditions for a cubic simulation box, the cubic simulation box is replicated in all the space to form an infinite lattice. During the simulation, when a molecule leaves the simulation box, its image enters the simulation box from the opposite side of the box, thus keeping the density in the box constant. The simulation box forms a convenient axis system for measuring the coordinates of the N molecules. To avoid

self-interaction, the box has to be larger than two times the cutoff. A very good pictorial description of the method can be found in the textbooks [43, 48, 49, 50].

2.2.6 Equations of Motion and Integrator

The motion of the atoms due to the forces acting on them follows Newton's equations of motions. The integration of Newton's equations of motion is done using the leap-frog algorithm, which is a descendent of the Verlet-Method [43]. The leap-frog method calculates velocities and positions with a shift of one half of the a time-step Δt :

$$\mathbf{v}_i\left(t_n + \frac{1}{2}\Delta t\right) = \mathbf{v}_i\left(t_n - \frac{1}{2}\Delta t\right) + \Delta t \frac{\mathbf{f}_i(t_n)}{m_i} + O(\Delta t^3) \quad (2.14)$$

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i + \mathbf{v}_i\left(t_n + \frac{1}{2}\Delta t\right)\Delta t + O(\Delta t^4) \quad (2.15)$$

Here \mathbf{r}_i , \mathbf{v}_i and \mathbf{f}_i are the position, velocity and force acting on particle i . The positions are accurate to an order of $O(\Delta t^3)$ of the simulation time-step.

2.2.7 Manostat and Thermostat Using Weak Coupling

The thermodynamic variables temperature T and pressure p are straightforward to access in a molecular dynamics simulation. For the temperature, one has to sum up the kinetic energy of all particles divided by the number of degrees of freedom:

$$T = \left\langle \frac{1}{3Nk_B} \sum_N mv^2 \right\rangle, \quad (2.16)$$

where N is the number of sites, k_B the Boltzmann constant and the angle brackets denote ensemble averaging.

For the pressure, the situation is more complex. Two terms contribute to the pressure:

$$p = \frac{1}{3V} \left(\left\langle \sum_N mv^2 \right\rangle + \left\langle \sum_N \mathbf{fr} \right\rangle \right) \quad (2.17)$$

where \mathbf{f} is the total force acting on a particle and \mathbf{r} is its position vector. The first term is the ideal gas contribution, and the second the so-called virial, which extends over all particle-particle interactions. The virial (the ensemble average $\langle W \rangle$, respectively) becomes an additional summand in the ideal-gas equation:

$$p = \frac{N k_B T}{V} + \frac{\langle W \rangle}{V} \quad (2.18)$$

Molecular dynamics potentials are often pair potentials. The virial part can also be transformed into pair interactions:

$$\sum_{i < j} \mathbf{r}_{ij} \mathbf{f}_{ij} \cdot \quad (2.19)$$

Such a treatment is better since it eliminates the calculation of the total force on an atom. A very comprehensive treatment can be found in the book by Allen and Tildesley [43].

The micro-canonical (NVE) ensemble is native to MD simulation and is still common practice. However, it is more realistic to simulate at constant temperature and pressure. For example, the Berendsen weak coupling method [53] rescales particle velocities by a factor s_T , which is computed from a first order decay of the actual temperature T towards the set temperature T_0 :

$$s_T = \left(1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right)^{\frac{1}{2}} \quad (2.20)$$

The parameter τ_T (the relaxation time constant) is chosen significantly larger than the integration time step Δt . If the relaxation time constant is chosen equal to the integration time step, one would brute the system to the fixed temperature, causing a mistake in the dynamics. The Berendsen algorithm conserves the total momentum and not the total energy.

The pressure is controlled using an equivalent of equation, but now scales particle positions and box sizes by a factor s_P . Usually, one uses isotropic scaling of the pressure p to the target p_0 :

$$s_P = \left(1 + \frac{\Delta t \kappa_P}{\tau_P} (p - p_0) \right)^{\frac{1}{3}} \quad (2.21)$$

Where τ_P as the pressure coupling constant and κ_P the system's compressibility, which is an experimental input parameter.

2.3 Protein structure analysis

2.3.1 The root mean square deviation

Conformations of proteins can be compared using the root mean square deviation (RMSD). The root mean square deviation measures the difference between the atoms positions of two two conformations of the protein. Simply, the root mean square deviation describes the “distance” between two conformations of a biomolecule (or group of selected atoms). First, the two conformations must be aligned. The smaller the deviations, the more spatially equivalent the two protein conformations are.

The root mean square deviation is defined as

$$RMSD = \sqrt{\frac{1}{N_a} \left(\sum_{a=1}^{N_a} (r_a - r_{a0})^2 \right)} \quad (2.22)$$

where r_a , and r_{a0} are the positions of the a -th atom in the given and the reference protein conformation.

In the study of globular protein conformations, one customarily measures the similarity in the three-dimensional structure by the root mean square deviation of the atomic coordinates after an optimal rigid body superposition.

2.3.2 Volume and surface

There are several definitions of protein surfaces discussed in the literature. Figure 2.1 schematically shows three types, namely: the van der Waals surface, the solvent accessible surface and the molecular surface. The van der Waals surface is defined by the intersections of the atomic van der Waals spheres that correspond to each atom of the protein. The solvent-accessible surface is defined as the surface that is traced out by the center of a probe sphere, which represents the solvent molecule H_2O , as it rolls over the van der Waals surface of the protein [55]. According to Richards [56], the molecular surface of a protein has two components, namely that part of the van der Waals protein surface which contacts the rolling probe sphere, and a re-entrant surface, which corresponds to a series of patches formed by the

interior-facing domain of the probe when it simultaneously contacts more than one atom on the protein surface.

For each type of protein surface one can define a corresponding protein volume, which is enclosed by the corresponding surface. The van der Waals volume, V_{vdW} , of the protein is the volume enclosed within the van der Waals surface of the protein. The volume enclosed within the molecular surface is also denoted solvent-excluded volume here [57], V_{SE} , of the protein, and equals the sum of the van der Waals volume and the volume of voids, V_{void} , which results from imperfect internal packing of the protein (note that a different definition is sometimes used, taking into account the centre of a spherical probe molecule). The partial volume, V_{P} , of a protein in solution is defined as the volume of the solution minus the volume of the solvent in the absence of solute. V_{P} can be represented by the sum of the solvent excluded volume, V_{SE} , of the protein molecule and by the changes in the solvent volume, ΔV_{h} , resulting from its interaction with the surface of the protein. The hydration contribution, ΔV_{h} , to the partial volume of the protein molecule reflects the interaction volume associated with the hydration of solvent accessible protein residues.

The molecular surface is chemically important because it represents the interface of the protein to the solvent. The analysis of the solvent-accessible surface furnishes information about the hydration shell because it is directly related to the possible number of water molecules at the protein surface.

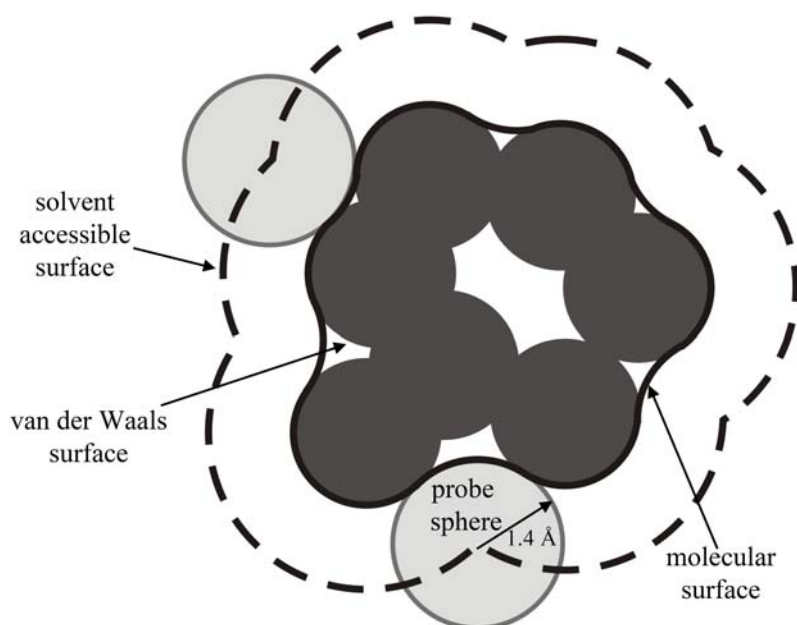


Figure 2.1: Schematic diagram of the solvent accessible, molecular, and van der Waals surfaces of a protein.

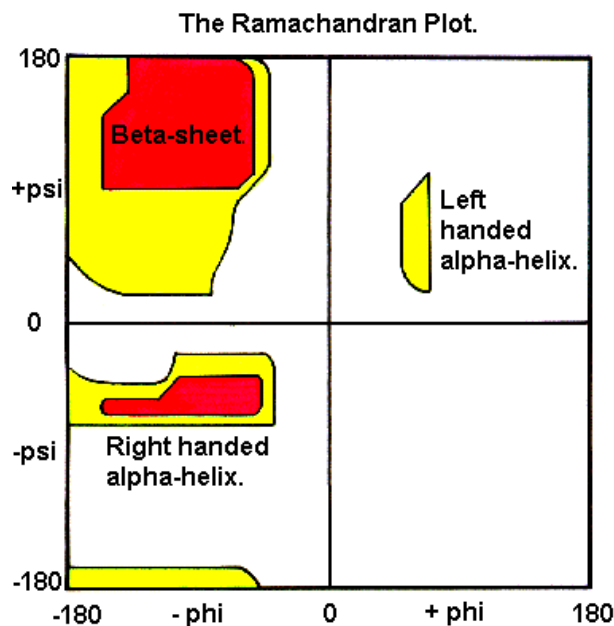


Figure 2.3: Ramachandran plots showing allowed combinations of the conformational angles ϕ and ψ defined in Figure 2.2. Colored areas show sterically allowed regions. The areas correspond approximately to conformational angles found for the usual right- and left-handed α -helices, and β -sheet.

2.4 The Time Correlation Functions

The dynamics of hydrogen bonds were investigated by calculating the relaxation of two different hydrogen bond time correlations functions, a continuous hydrogen bond correlation function $S_{\text{HB}}(t)$ and an intermittent hydrogen bond correlation function $C_{\text{HB}}(t)$.

The construction of these time correlation functions is based on two hydrogen bond population variables, $h(t)$ and $H(t)$: $h(t)$ equals one when a particular water—water or protein—water pair is hydrogen bonded at time t according to the adopted definition, as discussed before, and zero otherwise, whereas $H(t) = 1$ if the water—water or protein—water pair remains contentiously hydrogen bonded from $t = 0$ to time t , and zero otherwise. The continuous time correlation function $S_{\text{HB}}(t)$ is defined as

$$S_{\text{HB}}(t) = \langle h(0)H(t) \rangle / \langle h \rangle \quad (2.23)$$

where $\langle \dots \rangle$ denotes an average over all water—water or protein—water pairs. Clearly, $S_{\text{HB}}(t)$ describes the probability that an initially hydrogen bonded all water—water or protein—water pair remains bonded at all times up to t . The associated integrated relaxation time τ_{HB} (the

time integral of $S_{\text{HB}}(t)$) can be interpreted as the average lifetime of a hydrogen bond between two water molecules or between a protein and a water molecule.

The correlation function $S_{\text{HB}}(t)$ describes the dynamics of hydrogen bond breaking, since it depends on the continuous presence of a hydrogen bond. The intermittent hydrogen bond correlation function $C_{\text{HB}}(t)$ does not depend on the continuous presence of a hydrogen bond and is defined as

$$C_{\text{HB}}(t) = \langle h(0)h(t) \rangle / \langle h \rangle \quad (2.24)$$

Hence, the correlation function $C_{\text{HB}}(t)$ describes the probability that a hydrogen bond is intact at time t , given it was intact at time zero, independent of possible breaking in the interim time. $C_{\text{HB}}(t)$ relaxes to zero, since the probability that a specific pair of molecules is hydrogen bonded in a macroscopic system at equilibrium is negligibly small.

A similar approach for calculation of the protein-water hydrogen bonds lifetime and the average residence time of water within the hydration shell was used.

2.5 Elements of the Percolation Theory

Imagine a large array of squares. We imagine this array to be so large that any effects from its boundaries are negligible. Now a certain fraction of squares are filled whereas the other squares are left empty. We now define a cluster as a group of neighbour filled squares. Percolation theory deals with the number and properties of these clusters. We call p the probability of a site filled, that means that if we have N squares, and N is a very large number, then pN of these squares are occupied, and the remaining $(1-p)N$ of these squares are empty. For some p_c value, one cluster extends from top to bottom or from left to right of the lattice, and one says that this cluster percolates through the system. A large part of percolation theory deals with the peculiar phenomena of percolation properties near that concentration p_c , where for the first time a percolating cluster is formed [59].

Theoretical and computer simulation studies of clustering are based on the definition of a cluster. A recent review of various definitions of clusters in lattices and continuous models can be found in [60]. Due to the highly directional character of the intermolecular potential for hydrogen-bonded molecules, two water molecules are considered to belong to the same cluster if a continuous H-bond network connects them.

The occurrence frequency of water clusters of various sizes S is described by the cluster size distribution n_s , which at the percolation threshold obeys a power law, $n_s \sim S^{-\tau}$, with the

critical exponent $\tau = 187/91 \approx 2.05$ and $\tau \approx 2.2$ in the case of 2D and 3D percolation, respectively [59]. The mean cluster size $S_{\text{mean}} = \Sigma n_S S^2 / \Sigma n_S S$, calculated excluding the largest cluster, diverges at the percolation threshold in an infinite system or passes through a maximum when approaching the threshold in a finite system. The fractal dimension d_f of the largest cluster at the percolation threshold is lower than the Euklidean dimension of the system, and in the 2D case equals to $d_f^{2D} = 91/48 \approx 1.896$, [59] whereas $d_f^{3D} \approx 2.53$ in 3D systems [61]. We evaluated d_f from the function $m(r)$, which describes how the mass distribution of the largest cluster scales with distance r :

$$m(r) \propto r^{d_f} \quad (2.26)$$

We used these cluster properties to study the percolation transition of hydration water in all model systems. In the protein powders we also calculated the spanning probability R to find an infinite cluster in the periodic system, which spans (“wraps”) the simulation box at least in one direction. The value of R at the *true* percolation threshold does not depend on the system size, but it is sensitive to the spanning rule used, and to the system dimensionality. For the rule used in the present paper (spanning the cell with periodic boundaries, sometimes also called “wrapping” the system, either in x , y or z direction), the values of R at the *true* percolation threshold given in the literature for simple 2D and 3D lattices, range from about 0.5 to almost 1 [62-65].

3 Modelling Fully Hydrated Staphylococcal Nuclease

3.1 Introduction and System Setup

Water plays an important role for understanding the structure, dynamics and functionality of biological macromolecules, such as proteins. A variety of structural and dynamical properties of the hydration water of proteins are reviewed and compared with those of the bulk and with corresponding experimental results in refs. [24, 33]. The behavior of water at the protein surface has been a subject of a number of molecular dynamics simulations as well [14, 28, 32, 66-78]. Despite of these data already available, there is still a lack of knowledge on the ordering of water molecules and the structural characteristics of the H-bond network at the protein surface.

One of the important structural features of water is the presence of an H-bond network, which can be characterised by a distribution of hydrogen-bonded polygons of water molecules [79-83]. The protein-water interface is characterised by the occurrence of H-bonds formed between the protein and surrounding water and an H-bond network of water in the interfacial region of the protein. For enzymatic activity, transfer of protons along chains of hydrogen-bonded water molecules on the surface of the protein, with long-range proton movement over H-bond networks of water in the hydrated shell, is an important issue. As for a further example of the structural organization of the solvent within the hydration shell, there is a significant increase in heat capacity when proteins are unfolded or when hydrophobic compounds are dissolved in water, and this change in heat capacity is a linear function of the area of the hydrophobic surface. The large and positive heat capacity change is generally attributed to the extra heat needed to "melt" the ordered water structure near hydrophobic groups exposed to water. In general, the structural characteristics of the H-bond network around the protein is expected to be strongly depended on the chemical properties of the amino acids (a. a.) and atoms located at the protein surface.

The water structure has been shown to be very dependent on the chemical properties and the curvature of the surface near which the water molecules is located. The structure of liquid water at hydrophobic and hydrophilic surfaces has been the subject of molecular dynamics simulations [84, 85], and the results show that the hydration structure of large hydrophobic surfaces may be very different from that of small hydrophobic molecules. Recently, molecular dynamics simulation was also used for investigating the structure of

water droplets on graphite [86]. Moreover, the behavior of water has been investigated by Gibbs ensemble Monte Carlo simulations in cylindrical and spherical nanopores as a function of the radius of the pores and the strength of the water-substrate interaction (from hydrocarbon-like to metallic-like) [87-89]. In the computer simulations it was also observed that orientation preferences for water molecules in the first hydration shell around Lennard-Jones solutes are loosely reminiscent of the structure of clathrates [90]. Zichi and Rossky [91] presented a detailed geometrical analysis of the solvent in an aqueous solution containing two nonpolar atomic solutes. The analysis revealed the clathrate-like tendency for solvating water to form convex hydrogen bonded networks around non-polar solutes. Hydrophobic hydration of amphipathic peptides [92] was studied in terms of the molecular orientation of water relative to the solute surfaces. These works were concerned only with the orientational distribution of water molecules in the first hydration shell. Head-Gordon [93] investigated the structural properties of water around hydrophobic groups in terms of water polygons, whose correlations can directly be compared with clathrate arrangements. Recent neutron scattering experiments [94] revealed that the changes between the structural differences of water around hydrophobic amino acids and bulk water at room temperature is analogous to the changes that are observed between ambient temperature water and supercooled water, respectively. Molecular dynamics simulation was also used for a structural analysis of the hydration shells near leucine and glutamine amino acid residues in terms of hydrogen-bonded water polygons [18], and it was found that the hydration structure around the leucine side chain is more ordered than water near the glutamine side chain, while it is similar and less ordered near their backbones. Water rings forming clathrate-like arrangements have been found near the hydrophobic surfaces of protein [13, 17] and DNA [95] as well.

The focus of our present study is to investigate in very detail the structural properties of native staphylococcal nuclease (SNase), the water-protein H-bond network compared to the bulk water behavior, and the water-water H-bond network at the protein surface. The bacterium *Staphylococcus aureus* secretes a thermostable nuclease known as Staphylococcal nuclease (SNase, also known as thermonuclease). SNase is a calcium-dependent enzyme that catalyses the hydrolysis of both DNA and RNA. SNase is a small globular protein of 149 amino acid residues which forms three α -helices and a five-stranded β -barrel (Figure 3.1). The protein has a large positive charge (12 e) at pH 7.0 and ~20 % of the surface is covered by hydrophobic residues. Some structural, thermodynamic and dynamic properties of native, mutated, truncated and denatured SNase were studied in experimental work [96-107], in theoretical studies [108, 109] and by MD simulations [104, 110-114]. To our knowledge, no

simulations have been reported on the structural properties of water at the protein surface although a large body of experimental work on the hydration properties of the protein has been carried out, which calls for a molecular level understanding of the properties of water at its surface [105]. In particular, there is a strong need to quantify structural changes that occur in successive layers of the hydration shell, thus characterizing the range, or persistence length, of the structural changes that have long been believed to occur at the surface of a. a. side chains. To this end, this work focuses on the location of water and on the protein-water H-bonds network as a function of amino acid sequence and residue type, the structural properties of water (density, coordination number and number of water-water H-bonds) as a function of distance from the protein atoms, on the H-bond structure of the water in terms of "non-short-circuited" rings (e.g., "clathrate-like behavior") of hydrogen bonded water molecules, and on the correlation functions of water polygons near the protein surface.

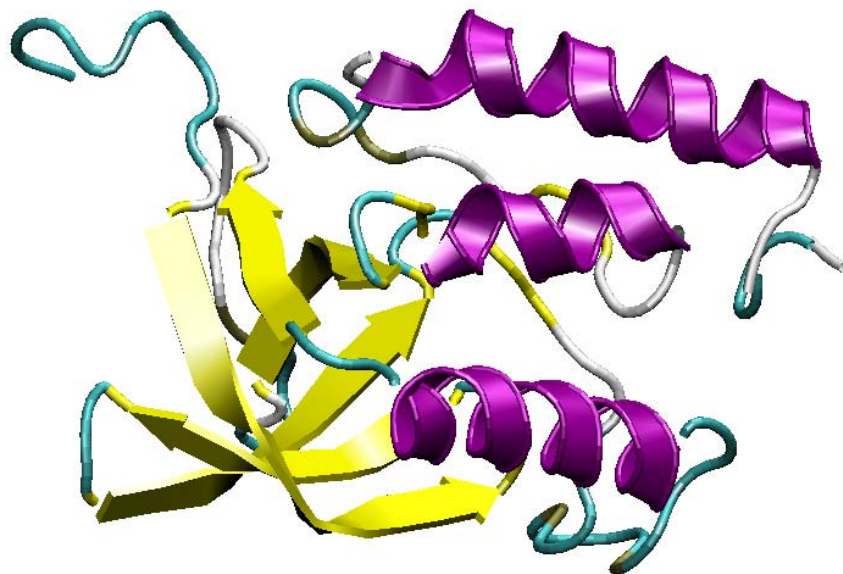


Figure 3.1: Schematic drawing of the native state structure of SNase taken from the coordinates of Hynes and Fox (1stn.pdb). The diagram was prepared using the program VMD. SNase is a small globular protein of molecular mass 16.8 kDa which contains 149 amino acid residues and no disulfide bonds. The backbone of SNase forms a five-stranded β -barrel and three α -helices.

The use of molecular dynamics simulation to examine protein unfolding and denaturation offers the possibility of obtaining detailed insight into factors influencing protein stability. In order to induce denaturation experimentally or by simulation, the native protein is subjected to a perturbation. In simulation studies, this has most frequently been a temperature increase, although pH, solvent, and pressure changes have also been applied. After application of the perturbation, the evolution of the system with time is followed. However, a main problem encountered with MD simulations is the limited, nanosecond time scale on which they can be performed with presently available computing power. If the time scale of the denaturation response is much longer than the nanosecond time scale, a significant denaturation is unlikely to be seen in the simulation.

To construct the native protein, crystallographic heavy atom coordinates were obtained from the Protein Data Bank [115] (PDB, www.pdb.org), entry 1STN [116]. Residues 1-6 and 142-149 are disordered in the crystal. The starting structure of the protein for the simulation was completed by adding coordinates for residues 1-6 (taken from PDB set 2SNS) and constructing residues 142-149 with AMBER 6.0 [117]. For removing bad contacts and adopting our system to the force field, energy minimization was carried out with the Steepest Descent and Conjugate Gradient methods. Eighty-five water molecules were identified crystallographically in 1STN and were included in the starting set. The MD simulations were performed using AMBER 6.0, the all-atom force field by Cornell et al. [51] and the particle mesh Ewald (PME) [118] was used for the calculation of electrostatic interactions. All protein atoms were explicitly included in the simulations. The protein structure was solvated by repeated overlays of an equilibrated cubic volume of TIP3P [119] water molecules to fill a truncated octahedron box. The dimensions were chosen such that the minimum distance between the edge of the box and the protein was more than 12 Å in the starting structure. Since in the protonation state given, the total charge of SNase is +12 e, twelve chloride counterions were added to provide a neutral simulation cell.

Equilibration was performed by placing a harmonic position restraining force constant equal to 104 kJ·mol⁻¹ on all solute atoms followed by 15 ps of constant volume MD during which only solvent molecules were allowed to move and the temperature was raised from 100 to 300 K during the first 4 ps. The resulting system was then equilibrated through 15 ps of *NVT*-MD at 300 K, in which the counterions and water molecules were restrained and the protein was able to move. The initial setup was then finished by a short, 3 ps *NVT*-MD run (without restrains) at 300 K to yield the initial configuration that contains the protein, twelve

chloride ions, and 8604 water molecules in a truncated octahedron box. This initial configuration was used as the starting point for the simulations of the protein in pure water.

The simulation was performed at *NPT* conditions and at residue-based cutoff of 10 Å for van der Waals interactions. The temperature was kept constant by a Berendsen thermostat [53] (weak coupling) with a coupling time of 0.5 ps. In the constant-pressure simulations, the pressure was kept constant by weak coupling to an external bath [53] with a relaxation time of 1.0 ps. The time step was 2 fs, and the bonds involving hydrogen are constrained using the SHAKE algorithm [52], with a relative tolerance of 10^{-5} . The simulation at 300 K and 1 bar was continued for 7.2 ns. For analysis purposes, the trajectory from 5.2 to 7.2 ns was used, and coordinates were saved every 0.2 ps.

A series of MD simulations were performed on SNase in water at constant temperature and pressure, namely at $T = 300\text{-}400$ K and $p = 1\text{-}5000$ bar. The simulations were continued for 7 to 13 ns. For the analysis, the trajectory from the last 2 ns was used.

Furthermore, a MD simulation at high temperature denaturing conditions ($T = 500$ K) was performed. In order to avoid artefacts caused by the proximity of portions of the protein upon unfolding with periodic images of it, the simulation box was reorganized according to the procedure described in Ref. 120. Briefly, during the simulation run the shortest distance between any protein atom and the image generated by the periodic boundary conditions of any other protein atom was monitored. When this distance becomes shorter than 10.0 Å, a sufficiently large water layer was added around the simulation box. A comparable procedure was applied in the unfolding simulation of an intrinsically unstable annexin domain [121]. This reorganization occurs at 7.8 ns, when the protein already lost most of the tertiary contacts and secondary structure elements, and has a small influence on the structural properties of the protein only, which still, however, has to be taken into account in the discussion of the data. Such reorganization of the simulation box may also slightly change the pathway from a highly disordered tertiary structure toward a random coil structure with some fluctuating helical structures (see below).

In the simulations of SNase in presence of co-solvents, the OPLS parameters for urea [122] and for glycerol from Ref. 123 were used. Firstly, equilibrated boxes with different concentrations of urea and glycerol were prepared. Urea and glycerol systems were constructed by randomly replacing water molecules with urea or glycerol, respectively. Next, the protocol (see above) for solvation and equilibration of the protein as for pure water simulations was repeated. Finally, four urea systems (mole fraction from 0.029 to 0.184) and three glycerol systems (mole fraction from 0.013 to 0.061) were constructed.

The trajectories were analyzed using the AMBER analysis modules PTRAJ and CARNAL for the calculation of the root mean square deviations (RMSd), the root mean square fluctuations (RMSf), and the radius of gyration (R_g). The program DSSP [124] was used for the calculation of the solvent accessible surface area (SASA) and the determination of the secondary structure. The volumes of the protein were calculated with the programs Mol_Volume [125] and MSMS [57]. A 1.4-Å radius for a probe sphere was used for all calculations of volume and SASA. All snapshots were prepared using the program VMD [126]. To determine the standard errors of the mean values of fluctuating physical properties derived, the usual statistical analysis was used [127, 128].

The protein-water and water-water hydrogen bonds were analyzed in detail. For the analysis of hydrogen bonds, the following geometric criterion was used: The proton-donor to heavy-atom acceptor distance must be less than 2.5 Å, and the hydrogen-bond angle must be greater than 120°. To detect "non-short-circuited" rings of hydrogen-bonded water molecules, the definition of Rahman and Stillinger was used [79]. These are polygons with three or more sides and no pairs, whose vertices are linked by a hydrogen-bond path shorter (in number of bonds, not geometrical length) than the minimal path within the polygon itself. The ring position is defined as the centre of mass of the assembly of water molecules forming the polygon. Polygon counts are enumerated as follows: For every water molecule all H-bond connections were determined; if a given water molecule has fewer than two H-bonds with the other water molecules, it can never serve as a polygon vertex. Firstly, all triangles, i.e. H-bonded chains that are closed in three steps, were determined. All these triangles were counted because any three molecules mutually bonded to form a triangle cannot be short-circuited. Next, all tetragons, i.e. polygons that are closed in four steps and non-short-circuited (no H-bonds between opposite vertexes), were found. This procedure continues for successively larger polygons. In the pentagon's case, one has to check that any three or four water molecules in a given pentagon are not involved in triangles or tetragons, respectively. Similarly, regarding hexagons, additionally, it was made sure that any five water molecules are not involved in any pentagons. This counting of the polygons at nonagons (9-sided polygons) was terminated. For example, for typical ices (hexagonal and cubic), the size of the polygons of the hydrogen bonded water molecules is six; in some others ices, polygons with sizes from four to eight were found [129]. For clathrate-like structures to be evident, the distribution along the hydrophobic surface is expected to be dominated by pentagons, with significant depletion of hexagons and larger polygons.

3.2 Structural Properties of SNase at Ambient Conditions

The root mean square deviations (RMSd) of the backbone atoms from the crystallographic structure, the radius of gyration (R_g), the solvent accessible surface area (SASA) and the volume were monitored as a function of time (Figure 3.2) up to a simulation time of 7.2 ns. The RMSd of the complete backbone from the crystal structure stabilized after ~ 1.5 ns at its final average value of $2.2 (\pm 0.3)$ Å. As expected, the main deviations are localized in the N-terminal region. Omitting this region from the calculation results in a RMSd value for the heavy backbone atoms of $1.6 (\pm 0.1)$ Å. The radius of gyration stabilized at $15.70 (\pm 0.1)$ Å, which is 2 % larger than the R_g -value of SNase in the crystal structure [116]. The solvent accessible surface area is $9417 (\pm 160)$ Å², a 7 % increase as compared to the X-ray crystal structure [116]. The SASA may be divided between residues into: 19.86 (± 0.6) % non-polar, 24.12 (± 0.6) % polar neutral, 40.27 (± 0.8) % positively charged and 15.75 (± 0.5) % negatively charged amino acid residues, respectively (Table 3.1). The volume of the protein increases from the start value of 33400 Å³ to 33650 Å³ (volume enclosed within the solvent accessible surface).

Table 3.1: Solvent accessible surface area (SASA) and average fractional solvent accessibility (f_{sa}) for all residues and for the various types of amino acid (a. a.) residues of fully hydrated SNase. Standard deviations are given in parentheses.

types of residues	no. of a. a. residues	% in a. a. sequence	SASA (Å ²)	% of SASA	f_{sa}
all residues	149	100	9417 (160)	100	0.218
nonpolar	62	41.61	1871 (60)	19.86 (0.6)	0.125
polar, neutral	33	22.15	2271 (70)	24.12 (0.6)	0.233
positively charged	33	22.15	3792 (74)	40.27 (0.6)	0.365
negatively charged	21	14.09	1483 (50)	15.75 (0.6)	0.238

The time evaluation of secondary structure and the root mean square fluctuations of all protein backbone atoms are given in Figure 3.3. For calculating RMSf, the last two nanoseconds of the simulation were used. It was found that α -helices make up $23.6 (\pm 1.3)$ % and β -strands $25.5 (\pm 1.2)$ % of all secondary structure of SNase, in very good agreement with the X-ray crystal structure data (22.1 % α -helices, 26.2 % β -strands) [116]. The rest are turns, loops and random, non-ordered structures and other helices. As expected, the most stable

parts of the secondary structure of the protein are the three α -helices and the five β -strands. Fluctuating parts of the protein are turns and bends between the α -helices and β -strands as well as the N-terminal region (Figure 3.3).

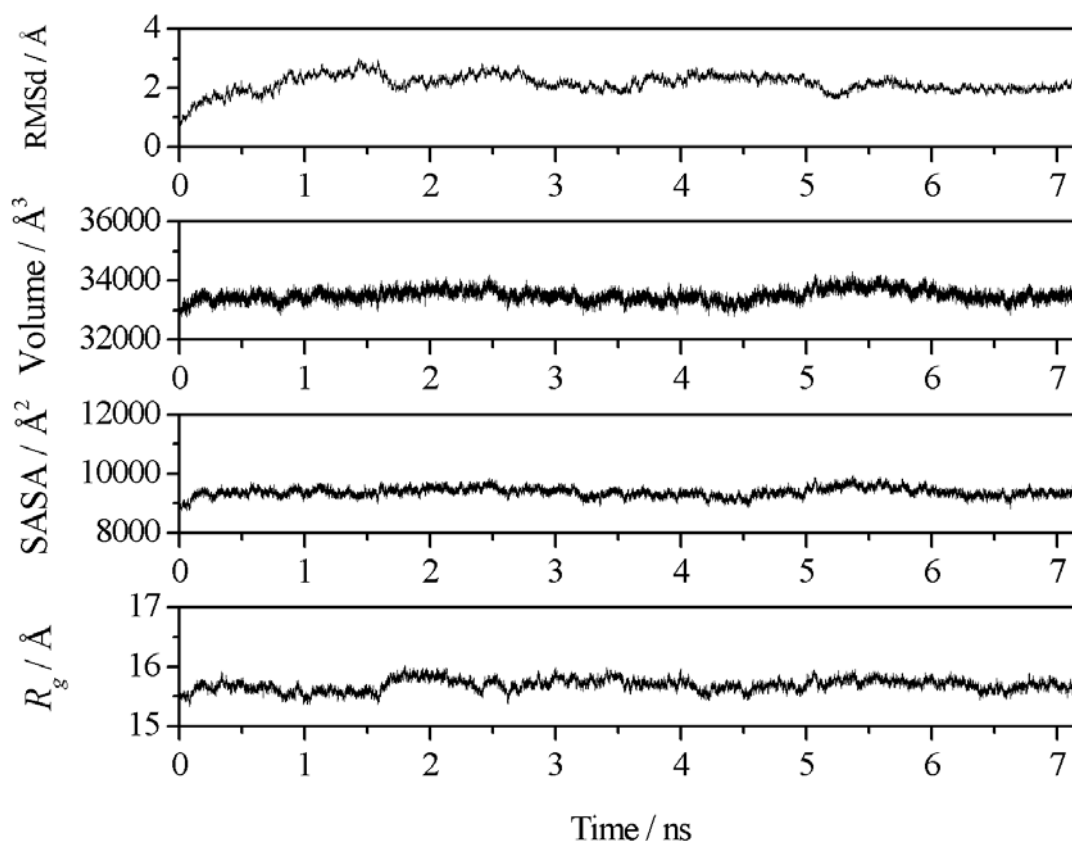


Figure 3.2: Time evolution of global structural properties of SNase from MD simulation: radius of gyration (R_g), the root mean square deviations (RMSd) from the starting structure, which was computed over all protein backbone atoms, the solvent accessible surface area (SASA), and volume of the protein.

To explore the degree of solvent exposure of each residue within the amino acid sequence, their time-averaged fractional solvent accessibilities were calculated [56], $f_{sa} = A_{\text{protein}}/A_{\text{free}}$, where A_{protein} is the solvent accessible surface area of a particular residue in the presence of the other surrounding residues of SNase, while A_{free} is the corresponding surface area if no other a. a. residues are present, i.e., of the free residue. Figure 3.4 shows the fractional solvent accessibilities of all residues along the amino acid sequence of SNase. Typically, f_{sa} values vary from 0 to ~ 0.5 . It was found out that the average fractional solvent accessibilities for positively and negatively charged amino acids are 0.365 and 0.238 respectively, while the f_{sa} value for neutral polar and non-polar groups are 0.233 and 0.125,

respectively (Table 3.1). As expected, the non-polar groups make up the smallest population of water exposed residues.

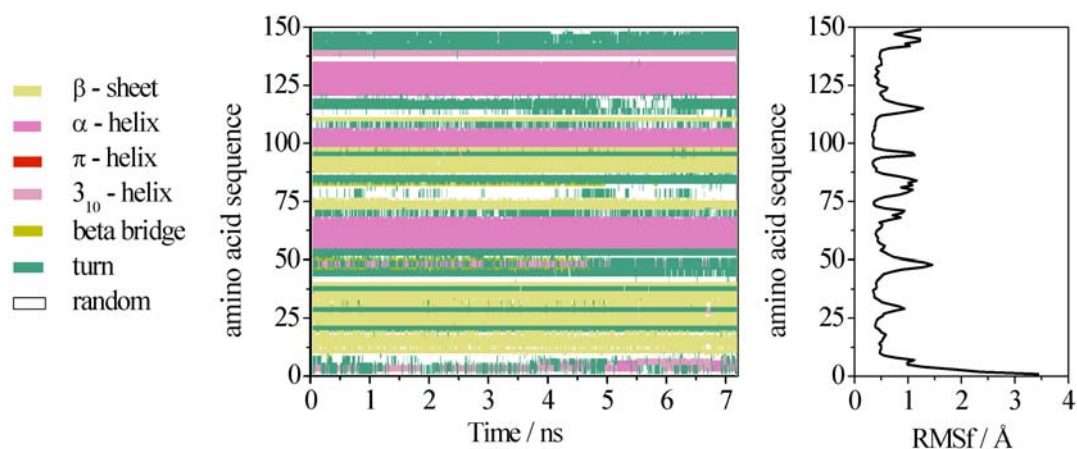


Figure 3.3: Left: Secondary structure of SNase as a function of simulation time (DSSP classification [104]). Right: Root mean square fluctuations (RMSF) of the protein backbone atoms calculated over the last 2 ns of the simulation as a function of residue sequence number. Clearly, small conformational fluctuations (sub-states) are visible on the 1-ns time scale for the folded state of the protein.

3.3 Water-Protein Interactions at Ambient Conditions

In order to focus on the structural properties of the water molecules that interact with the protein surface, two strongly bound internal water molecules, one water molecule bridging two a. a. (VAL₁₀₄ and TRP₁₄₀) and a second water molecule residing in an internal pocket, were not included in the analysis. For determining the hydration properties of each a. a. residue, the average coordination number, N , of each residue of SNase (Figure 3.4b) were calculated. The coordination number calculated corresponds to the number of water molecules of closest approach, i.e. that are located within a shell of 3.5 Å around any heavy atom (carbon, nitrogen, oxygen and sulfur) of the particular a. a. residue. Figure 3.4b shows that the level of hydration is particularly high around polar/charged residues. Only few non-polar residues have more than one water in their first hydration shell and probably only because of neighboring polar residues; their maximum coordination numbers may reach ~ 4 . Most of the hydrating water molecules thus arrange near positively and negatively charged a. a. residues. In Table 3.2, the time-averaged coordination numbers for the backbone and side chain atoms

are summarized. As the side-chain atoms of the polar a. a. residues are more exposed to the solvent, their average water coordination numbers are highest. The average number of water molecules within a shell of 3.5 Å around the whole protein is 455 (± 12). From comparison of Figure 3.4a and b, it was noted that the fractional solvent accessibility for some residues with relatively high values of N is very small or even close to zero, and the hydration level of the different a. a. residues depends significantly on the local solvent accessible surface area. To provide a more reliable comparison of the hydration level of the different types of a. a. residues, an SASA-weighted average coordination number, N_{weighted} , for each residue along the a. a. sequence of the protein were calculated (Figure 3.4c) by multiplication of N with the ratio of the SASA of a given a. a. residue to the SASA of the whole protein. Hence, high values of N_{weighted} in fact reveal large levels of hydration, while low values correspond to cases when the SASA of an a. a. residue is small or even close to zero. If the average coordination number is relatively high and the residual SASA very small or even close to zero due to the local protein topography, as often in the case of non-polar a. a. residues, all hydration water may be hydrogen bonded to neighboring water exposed to (mostly polar) residues. Comparison of Figure 3.4c, 3.5a and 3.5c exhibits that, for polar a. a. residues, decreasing values of N_{weighted} with respect to N are due to the fact that only polar atoms are accessible to the water molecules since they are able to form strong H-bonds.

Table 3.2: Time-average of the water coordination number of the various types of amino acid. residues of fully hydrated SNase as well as the corresponding values of their backbone/side chain atoms. Standard deviations are given in parentheses.

Type of residues	Coordination number				
	Total	Backbone		Side chain	
		Polar atoms	Apolar atoms	Polar atoms	Apolar atoms
Nonpolar residues	102.2 (5.5)	79.4 (4.2)	24.8 (2.1)	2.0 (0.8)	39.1 (3.6)
Polar, neutral	146.0 (7.4)	48.8 (3.3)	11.2 (2.3)	95.0 (6.3)	49.5 (4.1)
Positively charged	220.7 (9.2)	58.1 (3.8)	17.2 (2.2)	137.3 (7.5)	108.9 (6.7)
Negatively charged	125.2 (6.5)	35.1 (2.9)	12.0 (1.9)	110.7 (5.5)	69.5 (4.2)

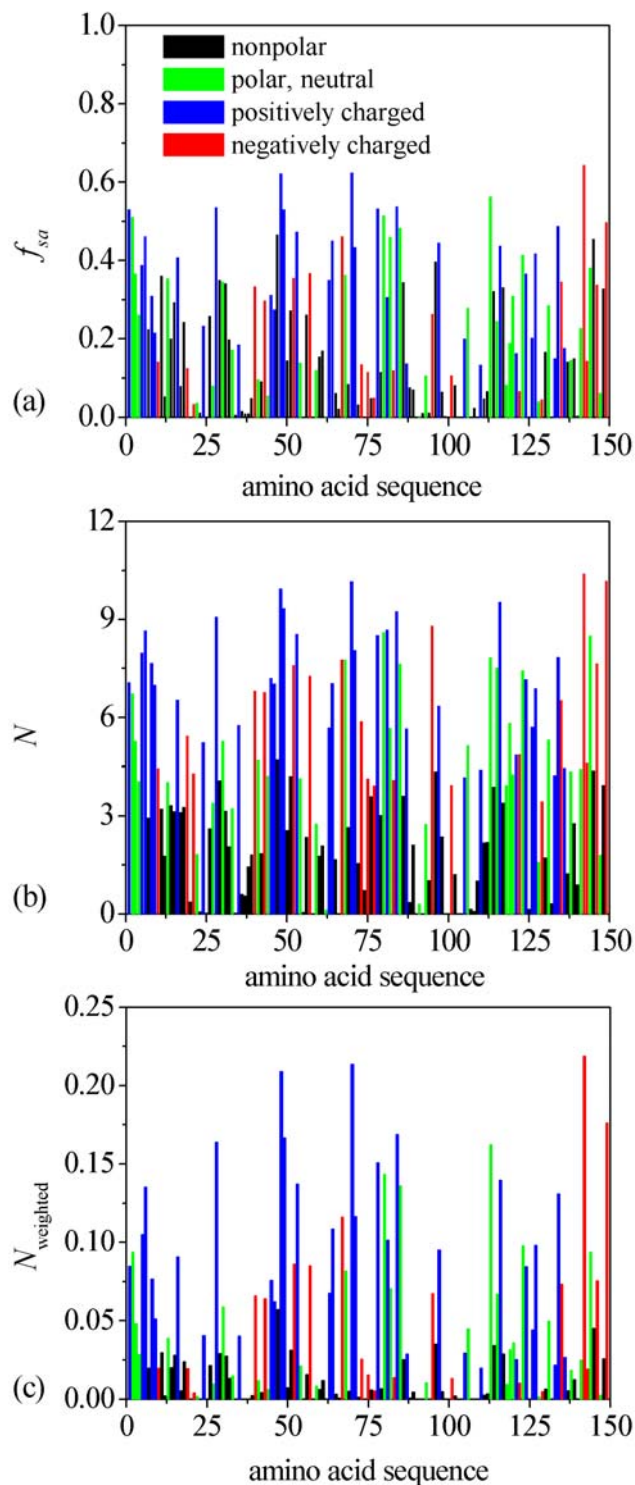


Figure 3.4: (a) Fractional solvent accessibilities (f_{sa}) as a function of residue sequence number and type of a. a. residue. (b) The average coordination number of residues (number of water molecules closer than 3.5 Å to any heavy atom of the amino acids) along the a. a. sequence of SNase. (c) The SASA-weighted average coordination number of residues (the average coordination number multiplied by the ratio of the SASA of a given a. a. residue to the SASA of the whole protein) along the a. a. sequence of SNase.

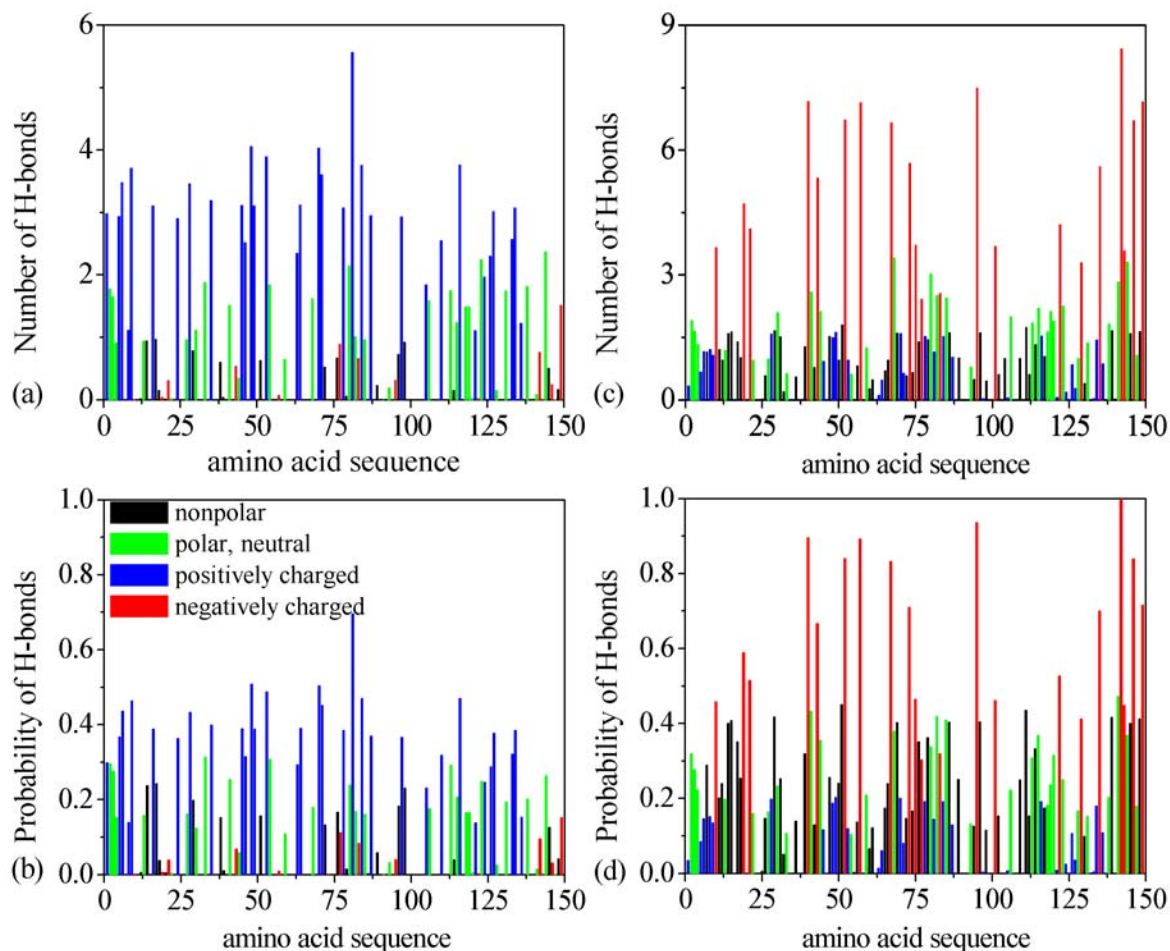


Figure 3.5: (a) The average number of protein-water hydrogen bonds when the protein atoms serve as a donor, and (c) when the protein atoms act as a acceptor, as a function of residue sequence number. (b) The average probability of protein-water hydrogen bonds (the number of H-bonds divided by the number of possible H-bonds) when the protein atoms serve as a donor, and (d) when the protein atoms act as an acceptor, as a function of a. a. residue sequence number.

Figure 3.5a and 3.5c display the average number of protein-water hydrogen bonds for each residue as a function of the amino acid sequence and residue type for the situation when atoms from the protein act as donor (donor H-bond) and when polar atoms of the protein act as acceptor (acceptor H-bond) of hydrogen bonds, respectively. The time-averaged number of donor H-bonds and acceptor H-bonds of our protein amounts to $149 (\pm 6)$ and $230 (\pm 7)$, respectively. It can be clearly seen (Table 3.3) that the largest numbers of H-bonds essentially occur at the side chains of the negatively charged (ASP, GLU and C-terminal) and positively charged (LYS, ARG, HIS and N-terminal) a. a. residues, which is due to the fact that they are largely exposed to the solvent. Average numbers of hydrogen bonds for polar, negatively and positively charged a. a. residues are 2.75, 5.50 and 3.73. Maximal numbers of protein-water hydrogen bonds for negatively charged and positively charged a. a. residues reach values of

~ 7 and ~ 4 , respectively. To determine the relative propensity of an a. a. residue to form hydrogen bonds with water, the average probability (the number of donor or acceptor H-bonds divided by the maximal possible number of donor or acceptor H-bonds, respectively) of protein-water hydrogen bonds (Figure 3.5b and Figure 3.5d) was calculated. Clearly, the probability of forming hydrogen bonds for negatively and positively charged a. a. residues of the protein is highest (since these residues are exposed to water) whereas that of the other types of a. a. residues seldom exceeds 50 % while they serve to keep the protein secondary structures (α -helices and β -strands).

Table 3.3: Time-averaged number of protein-water hydrogen bonds of various types of a. a. residues of fully hydrated SNase as well as the corresponding values of their backbone/sidechain atoms. Standard deviations are given in parentheses.

type of residues	No. of H-bonds					
	total		backbone		side chain	
	donor H-bond	acceptor H-bond	donor H-bond	acceptor H-bond	donor H-bond	acceptor H-bond
nonpolar	9.2 (1.1)	46.8 (2.3)	8.2 (0.8)	46.4 (2.2)	1.0 (0.4)	0.4 (0.3)
polar, neutral	35.6 (2.1)	55.0 (4.5)	9.1 (1.5)	22.3 (2.2)	26.5 (2.5)	32.7 (2.8)
positively charged	98.5 (4.3)	24.6 (2.2)	13.6 (2.1)	24.6 (2.1)	84.9 (3.7)	0.0
negatively charged	5.4 (1.3)	110.0 (7.4)	4.1 (0.8)	20.8 (2.2)	1.3 (0.4)	89.2 (6.5)

3.3.1 Water-Density Distribution at the Protein Surface

To investigate the influence of the protein surface on the structure of water the density profile of water, the water-water hydrogen bond distribution and coordination number of water as a function of distance from the protein atoms, were calculated. For precise evaluation of the water-density distribution, the volume of shells with thickness 0.1 Å as a function of distance from nearest heavy atoms (N, O, C and S) of the protein was calculated, and these data were determined for every tenth saved set of coordinates of the simulation run. Figure 3.6 displays some of such profiles obtained from different sets of coordinates during the simulation run. The shape of the curves reflects the existence of cavities and clefts at the protein surface. For calculation of the water density profile, the number of water molecules (the position of the water molecule is defined as the center of the oxygen atom) within the same shell was counted and divided by the corresponding volume of their shell. Two

pronounced density maxima close to the protein can be identified, at distances of 2.85 and 3.65 Å with peak heights of 0.0713 and 0.0426 Å⁻³, respectively (Figure 3.7a). These results are in qualitative agreement with previous molecular dynamics simulations of protein solutions [78,130]; however, owing to the short step size (0.1 Å) used here for the calculation of the water density profile close to the protein surface, two distinct peaks in the density profile were observed.

For further evaluation of the water density profile around different types of residues, the water molecules were labelled in the following way: If the nearest atom of the protein to a given water molecule is a polar atom (N, O or S), it is called class I water molecules, if it is a carbon atom, it is denoted class II water molecules. The average density profiles of class I and II water molecules are depicted in Figure 3.7b. The number density of water at distances > 7.5 Å equals the number density of bulk TIP3P water at ambient conditions (0.0329 Å⁻³). These data show that, concerning structural properties, about 2-3 water layers are perturbed by the protein surface, only. To define the hydration water, a simple distance criterion was used: water molecules belong to the hydration shell when they are closer to any heavy atom of the protein than some cut-off distance. For estimation of the average density of the protein hydration shell, one has to calculate the number of water molecules within the hydration shell, and the volume accessible to these water molecules is defined by the difference between the volume within the shell of a given cut-off distance from the centres of any heavy atom and the corresponding solvent-excluded volume, V_{SE} [57]. Due to hydrogen bond formation between the hydration water and the protein surface, it is difficult to define a clear-cut interface between the protein and the hydration water and to separate which volume corresponds to the protein and to the hydration water, respectively. To clarify this problem, the solvent-excluded volume of the protein with and without hydrogen atoms of the protein was calculated. It was obtained a 0.3 % density increase over the bulk in a shell of 4.5 Å and of 0.6 % in a shell of 4.0 Å, when only the heavy atoms of SNase for calculating V_{SE} were used. One may use these values as a lower limit for the density of the hydration water. In the case of implicit hydrogens of the protein it was obtained a 6.7 and 5.5 % density increase over the bulk in shells of 4.0 and 4.5 Å respectively. One may use these values as an upper limit for the density of the hydration water. These data are in quantitative agreement with recently reported data obtained using small-angle X-ray and neutron scattering techniques [131]. Furthermore, the data clearly show that the first peak of the water density profile arises from water molecules which are bonded to polar atoms (N, O and S) at the surface of SNase, whereas the second peak clearly corresponds to the localization of water molecules near apolar atoms (C) (Figure

3.7b). Also, to define the hydration shell one may also use different cut-off distances for different types of the protein atoms (hydrophobic and hydrophilic). In fact, the minima of the water local density profile near the protein surface are located at distances of 3.5 and 4.5 Å for hydrophilic (N, O and S) and hydrophobic (C) atoms of the protein, respectively. This yields an about 11 % density increase over the bulk in a hydration shell with these mixed cut-off distances, which is in good agreement with other simulations. [78, 130]

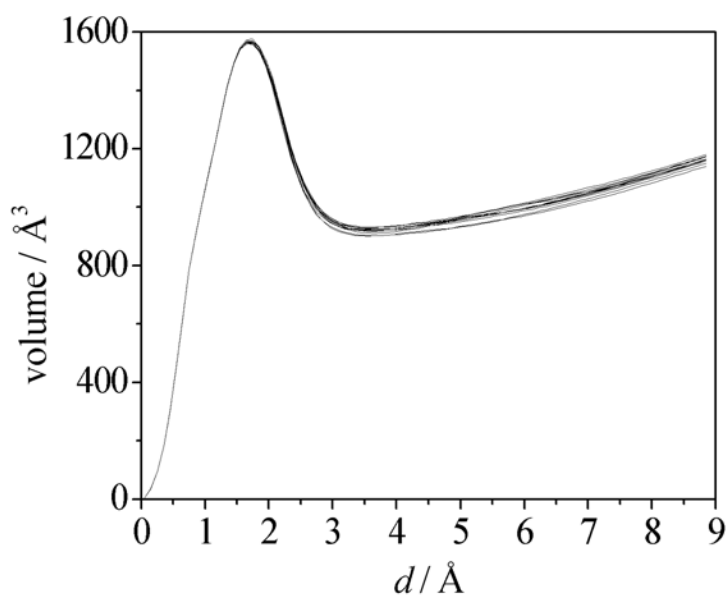


Figure 3.6: Volume of shells as a function of distance (d) from nearest protein heavy atoms. The figure shows profiles obtained from the coordinates at different stages of the simulation run.

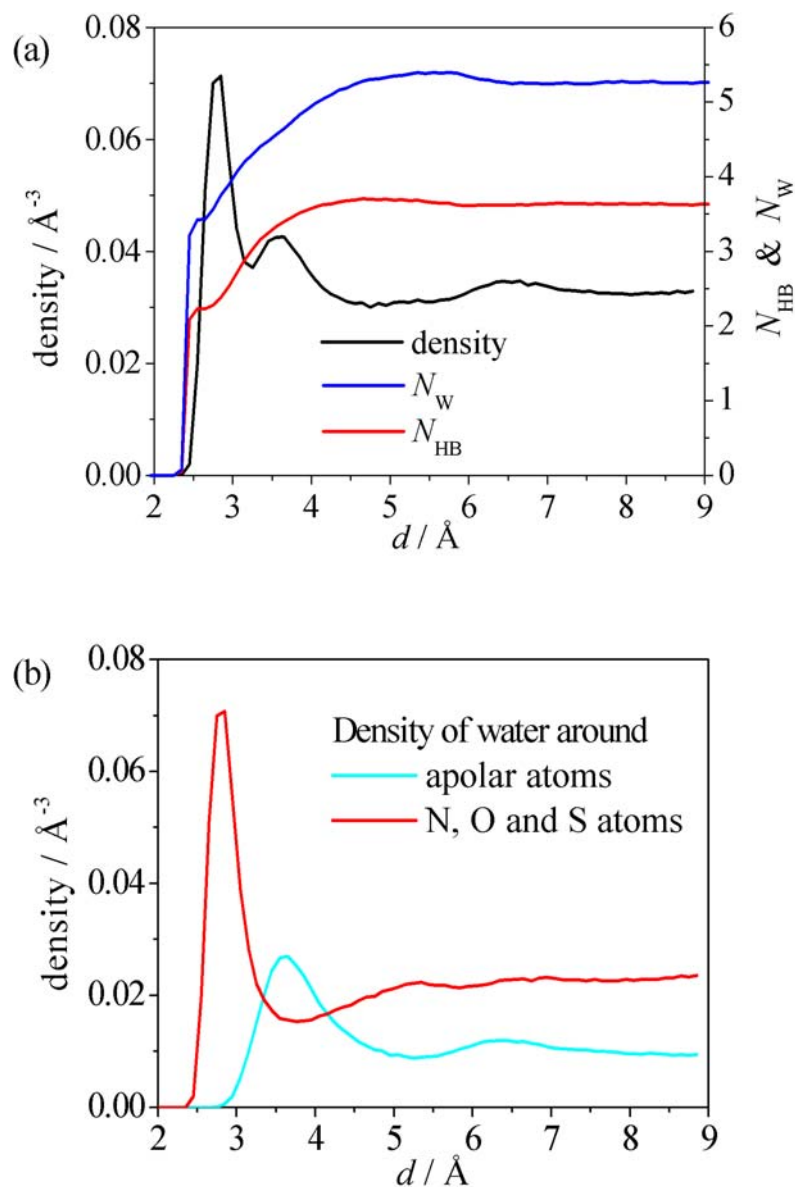


Figure 3.7: (a) Density profile of water (full black line), average number of water-water hydrogen bonds per water molecule (N_{HB}), average coordination number of water molecules (N_w) as a function of distance (d) from nearest protein heavy atoms. (b) Density profile for water as a function of distance from different types of nearest protein atoms.

3.3.2 Water-Water H-Bonds Near the Protein Surface

The average number of water-water H-bonds per water molecule (N_{HB}) and the average coordination number of water (N_w) as a function of distance from nearest heavy atoms of the protein are also shown in Figure 3.7a. In the water bulk region, an average number of

hydrogen bonds of 3.64 per water molecule and an average coordination number of 4.80 are found. Both values are similar to the values found in a control MD simulation of bulk TIP3P water. The numbers of water-water H-bonds, N_{HB} , and the coordination number of water molecules, N_{w} , decrease with decreasing distance from protein atoms, which is due to consecutive substitution of water molecules by surface protein atoms. These profiles exhibit a broad maximum in the region of the second minimum of the water density profile (Figure 3.6a). Interestingly, a similar behavior of H-bond profiles has been observed by Werder et al. [86] in their MD simulation of a water droplet on graphite. At this maximum, the average water-water H-bond number per water molecule is 3.71 (1.9 % above the bulk value), and the average coordination number of water is 5.40 (3.8 % above the bulk value).

3.3.3 Water-Polygons at the Protein Interface

The density profiles of water polygons as defined in paragraph 2.2 of the Method's section are given in Figure 3.8a. The distributions of pentagons, hexagons and heptagons exhibit a maximum which is located in the region of the second minimum of the water density profile (at 4-6 Å). Interestingly, a similar behavior for pentagons and hexagons was observed by Brovchenko et al. [88] in Gibbs ensemble simulations of water in spherical pores with slightly hydrophilic surfaces. In that paper, the authors suggested from a simple geometrical analysis that such a maximum simply reflects water-rings which include water molecules from the second water layer. For small water rings (3- and 4-rings), this maximum is shifted closer to the protein surface (4-4.5 Å). For 8- and 9-sided polygons, the peak height decreases; the peak broadens and is shifted towards larger distances (~6 Å). On the surface of the protein, polar, charged and apolar atoms and residues are located. Hence, an analysis of the distribution of non-short-circuited hydrogen bonded water rings was done in a way similar to that presented above for the water density profiles (Figure 3.8a), labelling the polygons according to the chemical nature of the protein heavy atom closest to the centre of the rings. If the centre of the ring is close to a carbon atom of the protein, we name that ring class II ring, and class I ring otherwise.

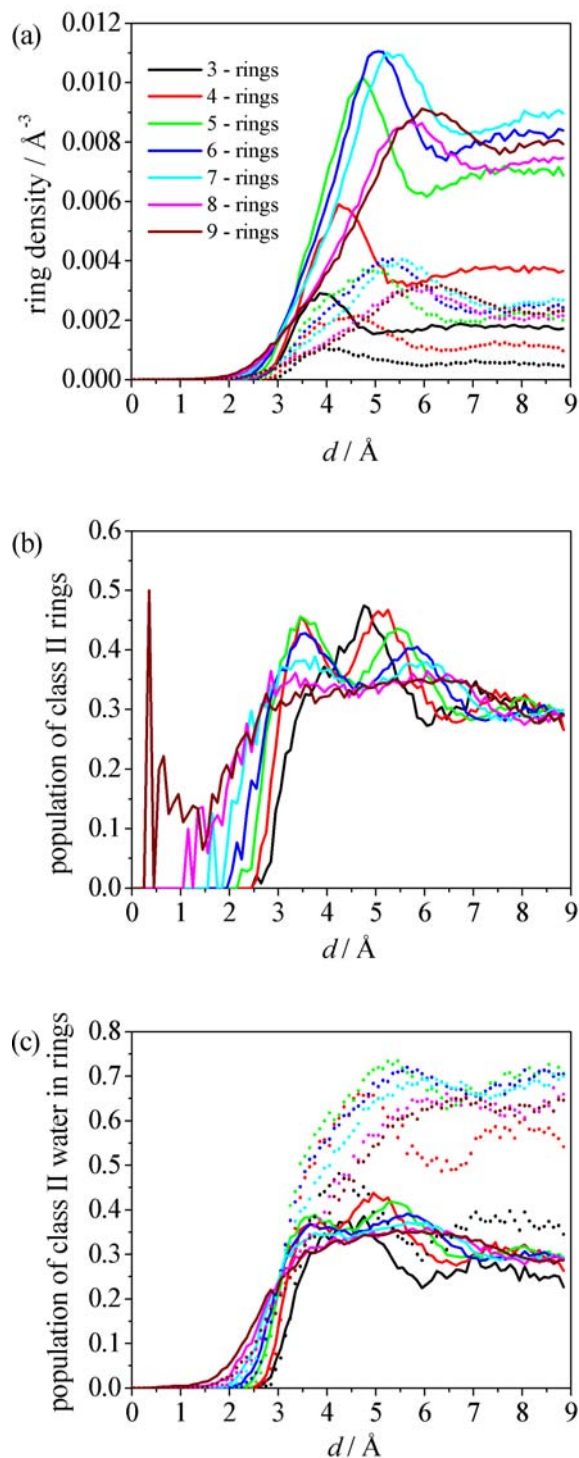


Figure 3.8: (a) Total number density of "non-short-circuited water rings" (solid lines) and number density of class II water rings (dotted lines) as a function of distance from nearest protein heavy atoms (class II water rings are denoted those water rings, whose centre of mass is closer to a carbon atoms than to a polar atom of the protein). (b) Population of class II water rings as a function of distance from nearest protein heavy atoms. (c) The average population of class II water molecules in all water rings (solid lines) and in class II water rings (dotted lines) (class II water molecule denotes water being located closer to a carbon atom than to a polar atom of the protein).

The plot of the number density of class II rings of sizes 4-7 exhibits a prepeak around 3-4 Å (Figure 3.8a, dotted lines). For a better understanding of the nature of this prepeak, the population of the class II rings as a function of distance from nearest protein heavy atoms was calculated (Figure 3.8b) by dividing the number of class II rings by the total number of rings in the corresponding shell at a given distance from the protein atoms. It is noticeable that a more or less pronounced first peak appears at a distance of ~ 3.45 Å for all polygons, except for 3-rings, which exhibit still a shoulder in the distribution function at that distance, however. This peak corresponds to the prepeak in the density profile of class II rings (Figure 3.8a). The second peak shifts to larger distances with increasing ring size; it appears at 4.75 Å for 3-rings and at 6.05 Å for 7-rings. For 8- and 9-rings, a broad distribution is observed only. The first pronounced peak corresponds to class II rings with centres of mass located in the first hydration shell of carbon atoms (these types of rings, evidently, are hydrogen bonded to polar side chain and/or backbone atoms), while the second peak relates to cases when some water molecules in rings are shifted to the second hydration shell of the carbon atoms. The minimum observed around 4.5 Å between the two density maxima of class II rings is due to the fact that at this location most of the water-rings are class I rings and arranged in such a way that they are bonded to oxygen or nitrogen atoms of the protein by one or more hydrogen bonds, and most of their water molecules are located in the second hydration shell of the polar atoms with orientations of these rings being more perpendicular to the protein surface; the population of class II rings at this distance is lower. Interestingly, Figure 3.8b also reveals that a significant population ($\sim 80\%$) of larger polygons have centers located at rather short distances to protein atoms, indicating that these large class I rings are sitting close to the surface of the protein, surrounding some of the polar a. a. side chains (but the concentration of such clusters is very low as seen in Figure 3.8a).

Figure 3.8c depicts the average population of class II water molecules in the total number of water polygons (solid lines) and in class II rings (dotted lines) as a function of distance from the protein nearest atoms. The maxima observed for the average population of class II water molecules in water polygons seem to be related to the maxima in the profiles for the population of class II rings (Figure 3.8b). An average population of class II water molecules at the position of the first peak in Figure 3.7b with values higher than 0.50 indicates that most class II rings are directly hydrogen bonded to polar protein atoms. The population of class II water in class II rings in the region of the second maximum is still higher and equals to 0.65 for 4-rings and to ~ 0.70 for 5-7-rings, i.e., most water molecules in the class II rings are located above carbon atoms here and only very few in the second

hydration shell are close to nitrogen or oxygen atoms at the protein surface. In the case of 3-rings, most of them are oriented such that only one water molecule in these rings is located above a carbon atom.

The distribution of water rings around different types of a. a. residues near the protein surface was analysed in great detail. A water ring was attributed to a particular type of a. a. residue if its center has the closed approach to that particular residue. The population of water-rings near different types of a. a. residues at distances far from the protein surface ($> 6.5 \text{ \AA}$) does not change anymore and simply reflects the SASA of that particular type of residue. Upon approaching the protein surface, however, this population changes owing to local topographic properties, density changes and specific interactions with the various a. a. residues. In particular, with approaching the protein surface, the fraction of rings near non-polar and positively charged a. a. residues decreases while that near polar neutral and negatively charged a. a. residue increases. One may suspect that since essentially polar atoms of the negatively charged a. a. are accessible to the solvent (see above) and near such groups the density is higher, the possibility of forming water-rings increases. In support of that it was found that in the region between 3.0 and 5.0 \AA from the atoms of positively charged a. a., about half of the ring atoms are located near side chain polar atoms and another half of the ring covers the nonpolar side chain atoms. On the contrary, around negatively charged a. a. residues, water-rings cover essentially the polar side chain atoms. This may be due to a simple size effect of the negatively charged residues having shorter side chains compared to the positively charged ones.

Figures 3.9a and 3.9b display the pentagon-pentagon radial distribution function, $g_{55}(r)$, and the pentagon-hexagon cross-correlation function, $g_{56}(r)$, for the control bulk water simulation and for the protein-water interfacial region (only polygons within a shell of radius 6.5 \AA around protein atoms were considered). Owing to the excluded volume effect by the protein, these correlation functions were normalised to the calculated density of the water-rings in the volume between 2.0 and 6.5 \AA from the centers of the protein heavy atoms.

In Figure 3.9a, the pentagon-pentagon radial distribution function $g_{55}(r)$ is shown. Three pronounced maxima appear, which are located at 0.75, 1.65 and 3.2 \AA . The first and second peaks correspond to compact arrangements of pentagons (two or three common sites), while the peak at 3.2 \AA corresponds to a more open, clathrate-like arrangement [81, 83] when pentagons share only one site. Two further broad peaks appear at 4.7 \AA and between 6 and 7 \AA , respectively. In reference [81], the corresponding configurations are given.

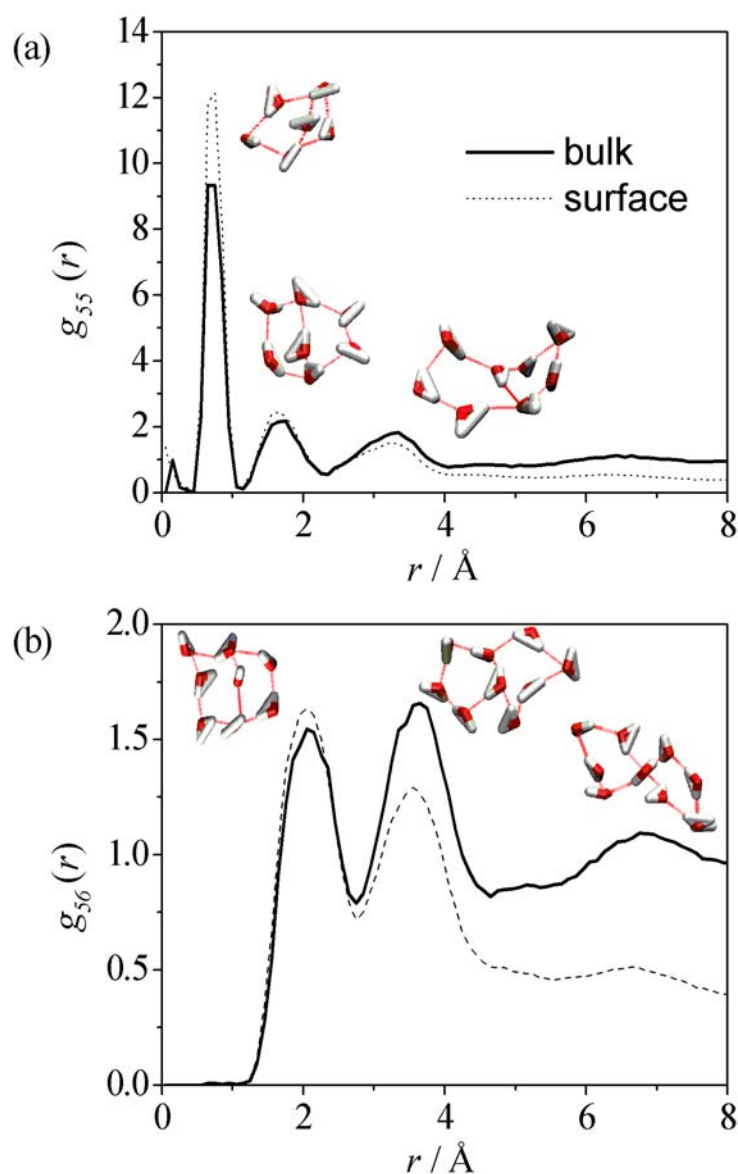


Figure 3.9: (a) The pentagon-pentagon and (b) the pentagon-hexagon radial distribution functions for bulk water (solid lines) and for water rings within 6.5 Å from protein heavy atoms (dashed lines). The figures visualize the arrangement of water rings corresponding to the positions of the respective correlation function.

Figure 3.9b shows the existence of three peaks for the pentagon-hexagon radial distribution function $g_{56}(r)$. Maxima of $g_{56}(r)$ at 2 and 3.6 Å correspond to cases, where the pentagon and hexagon share two sites or only one site, respectively. The broad peak at 6.7 Å corresponds to configurations with common water molecules in hexagon-pentagon arrangements. Some representative arrangements are depicted at the corresponding peak positions of Figure 3.9b.

Despite the excluded volume effect, both correlation functions for the protein-water system exhibit distinct maxima which have a similar position and shape as the bulk water system. Some of their maxima seem to be even slightly higher. Due to the existence of distinct correlations between pentagon-pentagon and pentagon-hexagon units and the fact that the number density of water rings is rather high at distances 4-5 Å, one may infer that a significant part of these water rings are oriented essentially parallel to the protein surface.

3.4 Temperature effect

In pressure perturbation calorimetry (PPC) [132], a relatively new and efficient technique, to study the solvation and volumetric properties of proteins in their native and unfolded state, the coefficient of thermal expansion of the partial volume of the protein is deduced from the heat consumed or produced after small isothermal pressure jumps, which strongly depends on the interaction of the protein with the solvent or co-solvent at the protein-solvent interface.

Proteins contain both hydrophilic and hydrophobic surface groups, also in the native state. Vicinal water molecules close to charged residues are expected to be oriented to the surface charges, leading to a more or less layered structure [133]. In fact, this layering, if substantial, should be reflected in thermodynamic properties of the protein, such as the partial molar volume and its apparent coefficient of thermal expansion, α . Just how many water layers with properties different from bulk solvent may form at the protein surface, depends principally on two factors: the charge density and its spatial distribution and the strength of the thermal forces that tend to disrupt the induced organization. Increase of temperature is expected to lead to a disruption of the solvent layer, which, in turn, should be reflected in the temperature dependence of α .

To yield a molecular interpretation of the different terms contributing to the partial protein volume and its temperature dependence and hence a better understanding of the PPC data, MD computer simulations on SNase were also carried out and compared with the experimental data. Figure 3.10 shows the local density profiles of water from the simulation runs at ambient pressure and different temperatures, normalized by the density of bulk water at the same conditions. With increasing temperature, the height of the maximum of the water density profile, which corresponds to the hydration water ("bound water"), decreases and broadens, indicating a temperature-induced weakening of the water-protein interactions and smearing out of the structure of the hydration shell.

In this study, one cut-off distance for all heavy atoms of the protein to define the hydration shell of SNase, and all atoms were used for calculating the physical surface properties of the protein molecule. Figure 3.11 shows the average density of the hydration shell of SNase as a function of the cut-off distance d at different temperatures. A pronounced average density maximum close to the protein interface can be identified in all simulations, and appears at a similar distance of about 3.8 Å. As expected, this density maximum broadens with increasing temperature. About two water layers are perturbed at the protein surface and have different densities compared to bulk water.

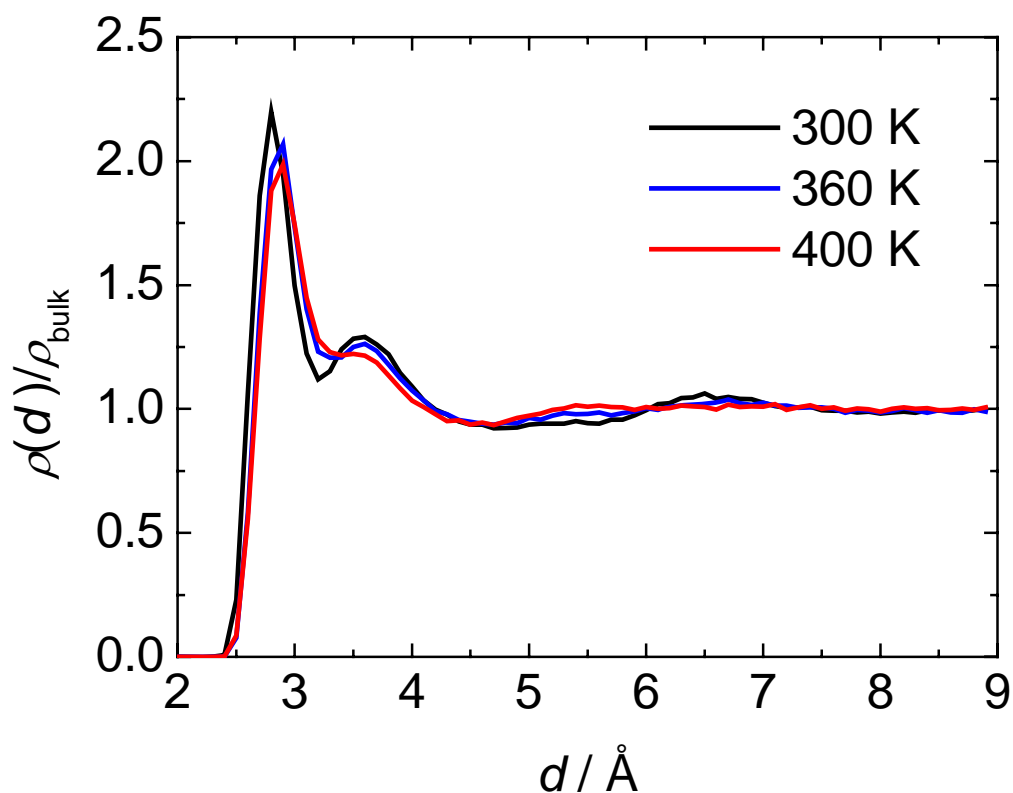


Figure 3.10: MD results for the local density profile of water, $\rho(d)$, near the surface of SNase as a function of distance (d) from nearest protein heavy atoms, normalized by the density of bulk water at the same conditions. At $p = 1$ bar and $T = 300, 360$ and 400 K.

The thermal expansion coefficient, $\alpha = -\rho^{-1}(\partial\rho/\partial T)_p$ of the hydration shell has been determined from simulations at two (ρ, T) data points using $\alpha = -(2/(\bar{\rho}_1 + \bar{\rho}_2))((\bar{\rho}_2 - \bar{\rho}_1)/(T_2 - T_1))$, where $\bar{\rho}_1, \bar{\rho}_2$ are the mean values of the density at

temperatures T_1 and T_2 , respectively. Figure 3.12 exhibits the thermal expansion coefficient of the water within of the hydration shell of SNase as a function of the cut-off distance, d . It can be clearly seen that α is much larger at low temperatures close to the protein interface, and drastically decreases at the higher temperature. To be able to compare the properties of the hydration and bulk water, α for the bulk TIP3P water by simulation of pure water under the same conditions was determined (Table 3.4). The results are in agreement with previous simulations [134]. TIP3P bulk water-like behaviour is observed above ~ 5.5 Å, i.e., beyond the second hydration layer.

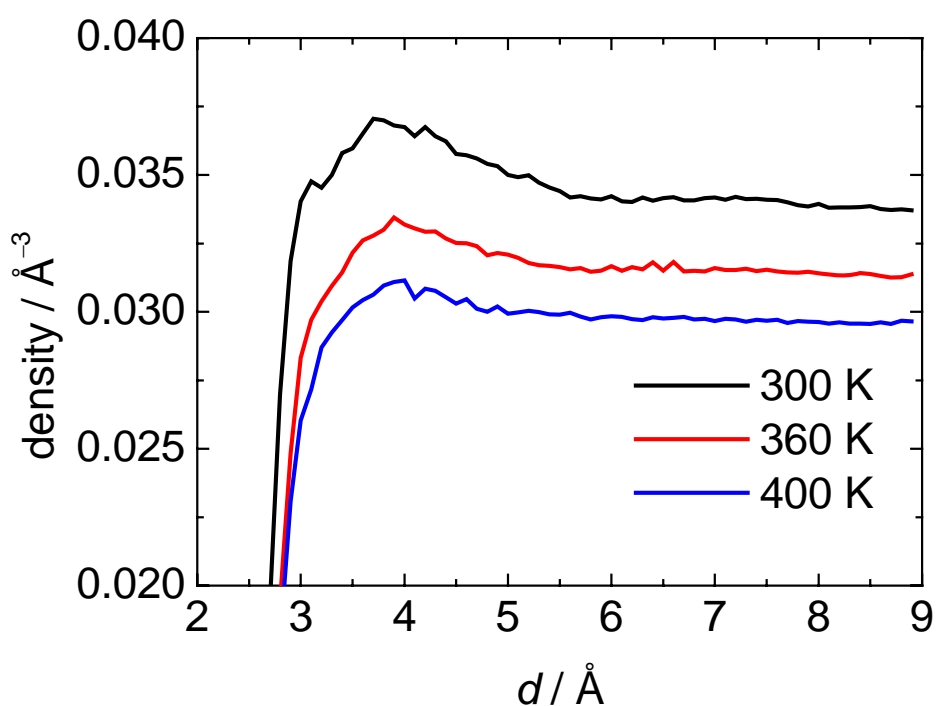


Figure 3.11: The average number density of the hydration water shell near the surface of SNase as a function of the cut-off distance (d) from nearest protein heavy atoms. At $p = 1$ bar and $T = 300, 360$ and 400 K.

In the discussion below, only the solvent-accessible and the molecular surface was considered. The molecular and the solvent-accessible surface areas of SNase at different temperatures are shown in Table 3.5. In addition, the ASA was divided into different types of residues by their physical-chemical properties: nonpolar, polar neutral, positively or negatively charged. The molecular and solvent-accessible surface areas of SNase slightly increase (~ 0.8 and 0.3% , respectively) with increasing temperature to 360 K, but slightly decrease again at 400 K compared to ambient temperature conditions. The reason is a drastic

surface area increase of polar neutral and nonpolar groups of about 8-10%, and a concomitant, largely compensating decrease of surface areas of positively and negatively charged groups. Hence, a significant redistribution of surface residues is revealed leading to a higher population of nonpolar and polar neutral residues. Upon further temperature increase, a reduction in the number of nonpolar surface area essentially occurs, only. These findings probably depend on the specific surface topography of the protein under consideration.

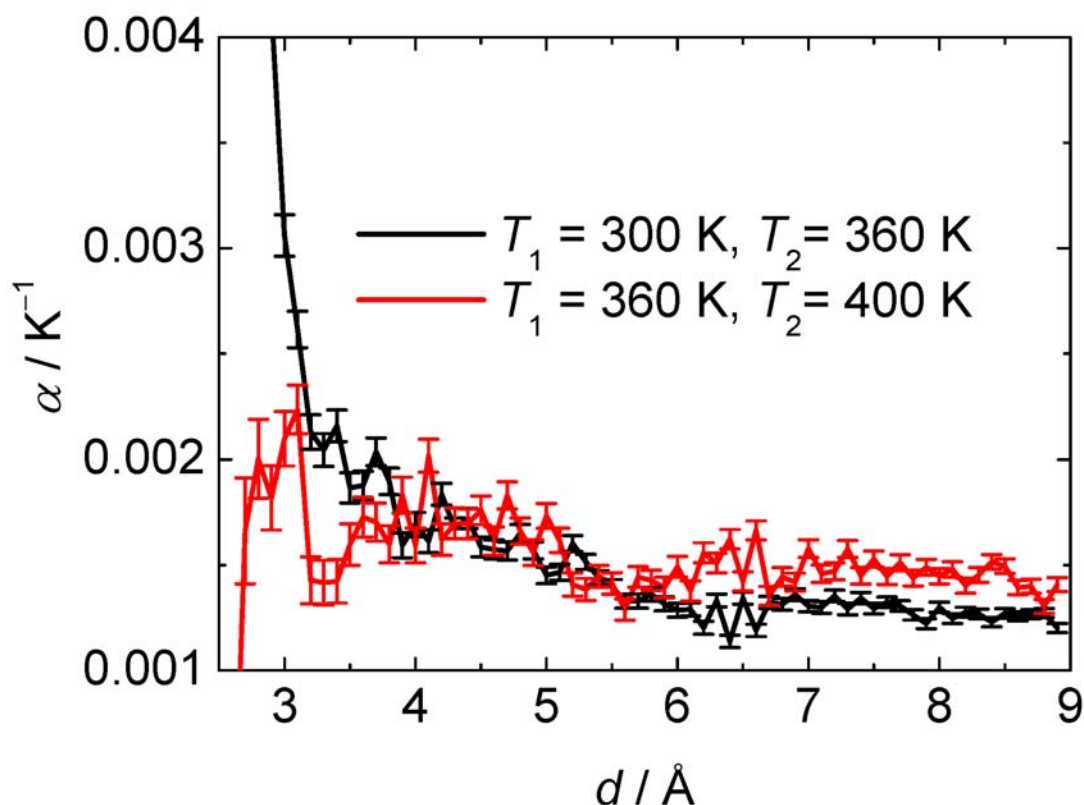


Figure 3.12: Thermal expansion coefficient α of the hydration shell of SNase as a function of the cut-off distance d from nearest protein heavy atoms.

The partial volume, V_p , of a protein in solution is defined as the volume of the solution minus the volume of the solvent in the absence of solute. The MD partial specific volume, $V_S = V_p/m_p$, of SNase slightly changes from $0.680 \text{ cm}^3/\text{g}$ at ambient temperature ($T = 300 \text{ K}$) to $0.684 \text{ cm}^3/\text{g}$ at $T = 400 \text{ K}$ (still far from MD unfolding conditions), respectively. Experimental data on SNase [107] show a similar increase of $V_p(T)$ in the native state, though the absolute values are slightly larger ($\sim 4\%$). At low temperatures, the temperature dependence of V_S is larger than at higher temperatures just before the unfolding temperature. Such behaviour is in agreement with our simulations: With increasing temperature from 300

to 360 K, an about 0.6% increase was obtained in the V_S value. Further increase of temperature does not change the V_S value significantly.

In Table 3.6, the various protein volumes of SNase obtained for the different temperatures are shown. In general, V_P can be represented by the sum of the intrinsic geometrical volume, $V_{\text{int}} = V_{\text{vdW}} + V_{\text{void}}$, occupied by the protein molecule itself (which can be approximated by the solvent excluded volume V_{SE} of the protein molecule) and by the changes in the solvent volume, ΔV_h , resulting from its interaction with accessible a. a. side chains, i.e., $V_P = V_{\text{vdW}} + V_{\text{void}} + \Delta V_h$ [135, 136]. The hydration contribution, ΔV_h , to the partial volume of the protein molecule reflects the interaction volume associated with the hydration of solvent accessible protein residues. $\Delta N = \Delta V_h \rho_{\text{bulk}}$ is defined here as the number of water molecules that have to be added or released to the hydration shell in order to yield the bulk density value (ρ_{bulk} is the number density of water under the same temperature conditions).

Table 3.4: The thermal expansion coefficient of the bulk TIP3P and hydration water of SNase at different temperature conditions. Errors are standard error estimates.

	$\alpha / 10^{-3} \text{ K}^{-1}$	
	Bulk water	Hydration water
$T_1 = 300 \text{ K}, T_2 = 360 \text{ K}$	1.06 ± 0.02	1.90 ± 0.06
$T_1 = 360 \text{ K}, T_2 = 400 \text{ K}$	1.38 ± 0.03	1.60 ± 0.09

Table 3.5: The molecular surface and solvent-accessible surface areas of various types of a.a. residues of SNase from MD simulations at different temperatures. Errors are standard error estimates. Standard deviations are given in parentheses.

System	Area / \AA^2					
	Molecular surface	Solvent-accessible surface				
		all residues	nonpolar	polar, neutral	positively, charged	negatively charged
SNase, 300 K, 1 bar	7359 ± 12 (107)	9417 ± 18 (160)	1871 ± 7 (60)	2271 ± 8 (70)	3792 ± 8 (74)	1483 ± 6 (50)
SNase, 360 K, 1 bar	7418 ± 14 (131)	9446 ± 24 (216)	2054 ± 7 (66)	2455 ± 11 (95)	3615 ± 10 (90)	1322 ± 9 (80)
SNase, 400 K, 1 bar	7334 ± 14 (124)	9395 ± 20 (180)	1984 ± 8 (77)	2468 ± 12 (110)	3607 ± 11 (96)	1336 ± 8 (72)

Hence, at least for a semiquantitative interpretation of the partial volume of the protein in terms of hydration [137], one may use the following relationship

$$V_p = V_{SE} + n_h(V_h - V_0). \quad (3.1)$$

V_0 and V_h are the partial volumes of water in the bulk and in the hydration shell of the solute, respectively, and n_h is the “hydration number”, that is, the number of water molecules in the hydration shell of the solute. In case of protein hydration, one has to consider V_h as an average effective partial volume of water in the hydration shell due to the presence of different types of the amino acid residues at the protein surface.

The partial volume in the simulation is simply the difference between the volumes of the system before and after adding the protein to the solution. The partial volume of SNase from the simulations was determined, using:

$$V_p = V_{BOX} - m_w / \rho_{bulk} - V_{Cl}, \quad (3.2)$$

where V_{BOX} is the volume of the simulation box, m_w is the mass of water in the simulation box, ρ_{bulk} is the density of bulk water at same conditions, and V_{Cl} is the partial volume of the counter ions, which was determined by considering the ions as spheres with radius $r = r_{vdW} + 0.5 \text{ \AA}$, where r_{vdW} is their van der Waals radius and 0.5 \AA is the thickness of the empty layer surrounding the counter ion [138].

It is well known that for polar solutes, ΔV_h is negative and for nonpolar solutes positive [139]. Around a hydrophilic solute, the water molecules are more tightly packed, while in presence of the hydrophobic solutes the water structure is less dense [139]. For proteins, ΔV_h is generally found to be negative. Remarkably, with increasing temperature, the absolute value of ΔV_h increases (for nonpolar and polar solutes), as the decrease of the bulk water density is faster than the decrease of the hydration water density with increasing temperature. As expected, V_{vdW} is more or less independent of temperature. Due to a marked increase of V_{void} (~4%) up to 360 K, V_{SE} of SNase increases (~1%) with increasing temperature. The van der Waals and void volumes make up 86 and 14% of the solvent excluded volume, respectively at $T=300$ K. The void volume of the protein increases markedly with increasing temperature (~4%), from 3010 \AA^3 at 300 K to 3127 \AA^3 at 360 K. For these two temperatures, the negative values of the hydration contribution to the partial volume are essentially the same, which is probably due to some surface rearrangement of the protein, leading to a decrease of the polar and concomitant increase of nonpolar and polar

neutral protein surface area as discussed above (Table 3.5). This leads to a decreasing absolute value of ΔV_h . At the higher temperature of 360 K, ΔV_h increases in fact, as the decay of the bulk water density with increasing temperature is larger than the decay of the water density of the hydration water.

Table 3.6: The solvent-excluded, van der Waals, void and partial volumes of the SNase at different temperatures ($p = 1$ bar). Also, the hydration contribution to the partial volume and $\Delta N = \Delta V_h \rho_{\text{bulk}}$ are given. Errors are standard error estimates. Standard deviations are given in parentheses.

System	Volume (\AA^3)					$\Delta N = \Delta V_h \rho_{\text{bulk}}$
	Solvent-excluded volume, V_{SE}	Van der Waals volume, V_{vdW}	Void volume, V_{void}	Partial protein volume, V_p	Hydration contribution to the partial volume	
SNase, 300 K, 1 bar	21482 ± 10 (86)	18472 ± 2 (21)	3010 ± 8 (73)	18987 ± 43 (393)	-2495 ± 6 (52)	-82
SNase, 360 K, 1 bar	21604 ± 14 (109)	18476 ± 2 (21)	3127 ± 11 (100)	19113 ± 49 (446)	-2491 ± 6 (58)	-77
SNase, 400 K, 1 bar	21665 ± 15 (111)	18480 ± 2 (22)	3186 ± 11 (99)	19059 ± 56 (512)	-2606 ± 8 (70)	-76

With further increasing the temperature from 360 to 400 K, a further but less pronounced increase of the void volume is visible. Obviously, the temperature dependence of the void volume is largest at lower temperatures and reaches a limiting value at high temperatures, probably as a further expansion without unfolding of the protein is not feasible anymore. As the contribution of the polar and nonpolar amino acid residues at the protein surface remain the same at 360 and 400 K, an increase of the absolute value of ΔV_h was observed due to the temperature increase in this case.

Additionally, the thermal expansion coefficient of the different protein volumes was determined (Table 3.7a), using the same dissection method as described above. In a first approximation, the thermal expansion coefficient of the protein partial volume can be expressed as:

$$\alpha_{V_p} = \varphi_{V_{\text{vdW}}} \alpha_{V_{\text{vdW}}} + \varphi_{V_{\text{void}}} \alpha_{V_{\text{void}}} + \varphi_{\Delta V_h} \alpha_{\Delta V_h}, \quad (3.3)$$

where φ_i is the volume fraction of the corresponding protein volume, $\varphi_i = V_i/V_p$ (see Table 3.7b).

Table 3.7: The thermal expansion coefficients α_i (a) and relative contributions $\varphi_i\alpha_i$ (b) of the different contributions to the partial protein volume of SNase at different temperatures. Errors are standard error estimates.

a)

SNase	$\alpha / 10^{-3} \text{ K}^{-1}$				
	Solvent excluded volume	Van der Waals volume	Void volume	Hydration contribution to the partial volume	Protein partial volume
$T_1 = 300 \text{ K},$ $T_2 = 360 \text{ K}$	0.094 ± 0.009	0.004 ± 0.003	0.63 ± 0.08	≈ 0	0.10 ± 0.01
$T_1 = 360 \text{ K},$ $T_2 = 400 \text{ K}$	0.07 ± 0.02	0.005 ± 0.005	0.33 ± 0.13	1.13 ± 0.41	-0.06 ± 0.03

b)

SNase	$\varphi_i\alpha_i / 10^{-3} \text{ K}^{-1}$			
	Van der Waals volume	Void volume	Hydration contribution to the partial volume	Protein partial volume
$T_1 = 300 \text{ K},$ $T_2 = 360 \text{ K}$	0.00389	0.102	≈ 0	0.10 ± 0.01
$T_1 = 360 \text{ K},$ $T_2 = 400 \text{ K}$	0.00484	0.0546	-0.151	-0.06 ± 0.03

As can be seen from Table 3.7b, for the temperature interval from 300 to 360 K, the main contribution to the expansion coefficient of the protein partial volume is due to the expansion of the void volume of the protein ($0.13 \% \text{ K}^{-1}$). At higher temperatures (360 to 400 K), a competition between a decreasing but still positive contribution of the void volume ($0.05 \% \text{ K}^{-1}$) to the expansion coefficient of the protein and an increasing negative hydration contribution to the partial protein volume ($-0.12 \% \text{ K}^{-1}$) was observed. The drastic decrease of the expansion coefficient of the protein partial volume with increasing temperature, which has also been observed experimentally using PPC, can thus be attributed to a decrease of the expansibility of voids with increasing temperature and a decrease of the expansibility of the hydration layer.

The MD computer simulations revealed that the thermal expansion coefficient of the native protein's partial volume is largely determined by the expansibility of its internal voids and by a significant hydration contribution, and both contributions decrease with increasing temperature.

3.5 Pressure Effect

The mechanisms of protein folding and unfolding have been the subject of intensive experimental and theoretical studies for several decades now, and a detailed understanding of the structural, dynamic and thermodynamic properties of the unfolded state of proteins has been a major topic of research in the field of protein chemistry and biophysics in recent years. While the study of protein folding has been undertaken primarily using perturbation by temperature or chemical denaturants, a complete thermodynamic description of the folding and unfolding reaction of proteins requires the characterization of the response of the protein structure also to pressure [140]. Pressure- as well as temperature-induced effects in proteins are closely connected with the presence of water. Changes in the interaction with the solvent (hydration) and the imperfect packing effects in the interior (void volume and cavities) contribute to the structure and dynamics of proteins in aqueous solutions. For example, a negative activation volume at high pressure suggests a mechanism of water penetration and binding before the unfolding and denaturation of proteins [140]. The water in the hydration shell of proteins has been proposed to play a crucial role in high-pressure induced protein unfolding, and it was proposed that the penetration of water molecules into the hydrophobic core might induce unfolding of the system [141-144]. The water penetration model is also supported by high pressure molecular dynamics simulations [145]. It is also well known that the thermal expansion and the compressibility of the protein contain contributions from the hydration as well as from the cavities. The role of cavities on volume effects of proteins was already emphasized by Silva and Weber [5]. Pressure-induced conformational changes of proteins and the role of water in this process deserve still further investigations, however, as the experimental and theoretical data are still scarce.

The volumetric properties of globular proteins were extensively studied by means of various experimental techniques [134, 135, 146-154] and by theoretical approaches [55, 56, 155-163]. It is well known, that the hydration water on the protein surface strongly influences the structural and dynamical properties of the proteins and enables their function [10, 11, 24, 33, 164]. The behaviour of the hydration water at the protein surface in different environments has also been the subject of a number of computer simulation studies [14, 20, 36, 38, 67, 68, 71, 73, 130, 165, 166], and there are excellent review articles on high pressure simulation of proteins [163] as well as the discussion of volume effects of proteins [167].

Some structural, thermodynamic, and dynamic properties of native, mutated, truncated, and denatured SNase were already reported in experimental work [96-107], theoretical and MD simulation studies [108-114].

In principle, computer simulation methods can offer a structural view of a protein during the process of unfolding. MD simulation combines detailed information at the atomic level with high resolution in time [168]. This is in contrast to experiments, which provide mostly only little structural information about the conformations adopted by a protein in the process of unfolding. Hence, MD simulations can help to furnish a more detailed view both in time and in atomic detail [120,169]. In common, results of unfolding simulations are quite consistent with available experimental results [170-173], and, in some cases, MD helps with the interpretation of experimental results [174]. Unfolding of proteins at high pressure by molecular dynamics simulation is difficult, however. Hünenberger and co-authors showed that a pressure of 10 kbar significantly slows down the motions of protein and water, and they did not observe unfolding of lysozyme within 210 ps [175]. Molecular dynamics simulations at high pressure have also been performed on bovine pancreatic trypsin inhibitor, but no unfolding has been observed up to 10 kbar [143,176]. Wroblowski and co-workers managed to unfold bovine pancreatic trypsin inhibitor at a higher pressure between 10 and 15 kbar [141]. Recently, the small-angle neutron scattering data of the reversible folding/unfolding transition of SNase at ambient temperature and 1-2 kbar pressures were complemented by molecular dynamics simulations [142]. Here, the unfolding of SNase in molecular dynamics simulations was initiated by water insertion into the protein interior, however, i.e., the putative initial step when water starts to penetrate into the protein hydrophobic core of the protein is missing. In order to understand the initial stages of pressure-induced protein unfolding, such as changes in hydration properties and reorganization of surface groups, the properties of the hydration water at the protein surface at different pressures were studied in detail in this work.

The presence of a solute generally affects the properties of water. Besides, density oscillations in the close neighbourhood of the solute surface are observed which are due to packing effects. To investigate the influence of the protein surface on the structure of water, the local density profile of water as a function of distance from the protein atoms was calculated. Figure 3.13 shows the local density profiles of water from the simulation runs at ambient temperature and various pressures (from 1 bar to 5 kbar), normalized by the densities of the bulk TIP3P water at the same conditions. The height of the second peak increases and shift markedly toward the protein surface with increasing pressure, indicating that the

hydration of the interfacial hydrophobic part of the protein increases, as it has been shown that this second peak corresponds to water molecules that are located near hydrophobic atoms [20].

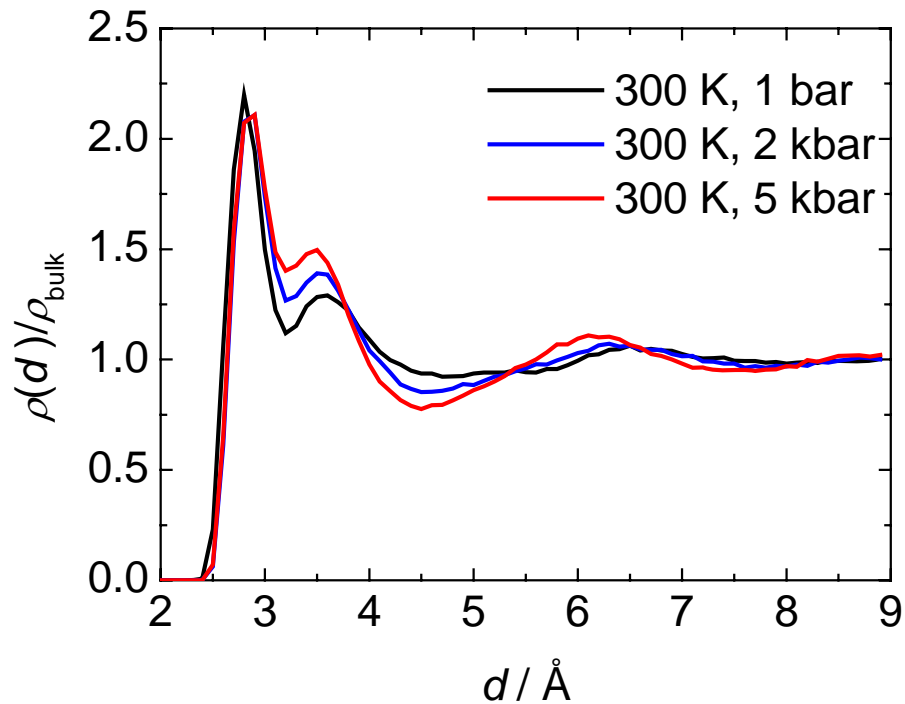


Figure 3.13: Local density profiles of water, $\rho(d)$, near the surface of SNase as a function of distance (d) from nearest protein heavy atoms, normalized by the densities of the bulk water at the same conditions. At $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar.

Figure 3.14 shows the average density of the hydration shell of the SNase as a function of the cut-off distance, d , at different pressure conditions. A pronounced broad maximum of the average density close to the protein can be identified in all simulations, appearing at a similar distance of about 3.8 Å.

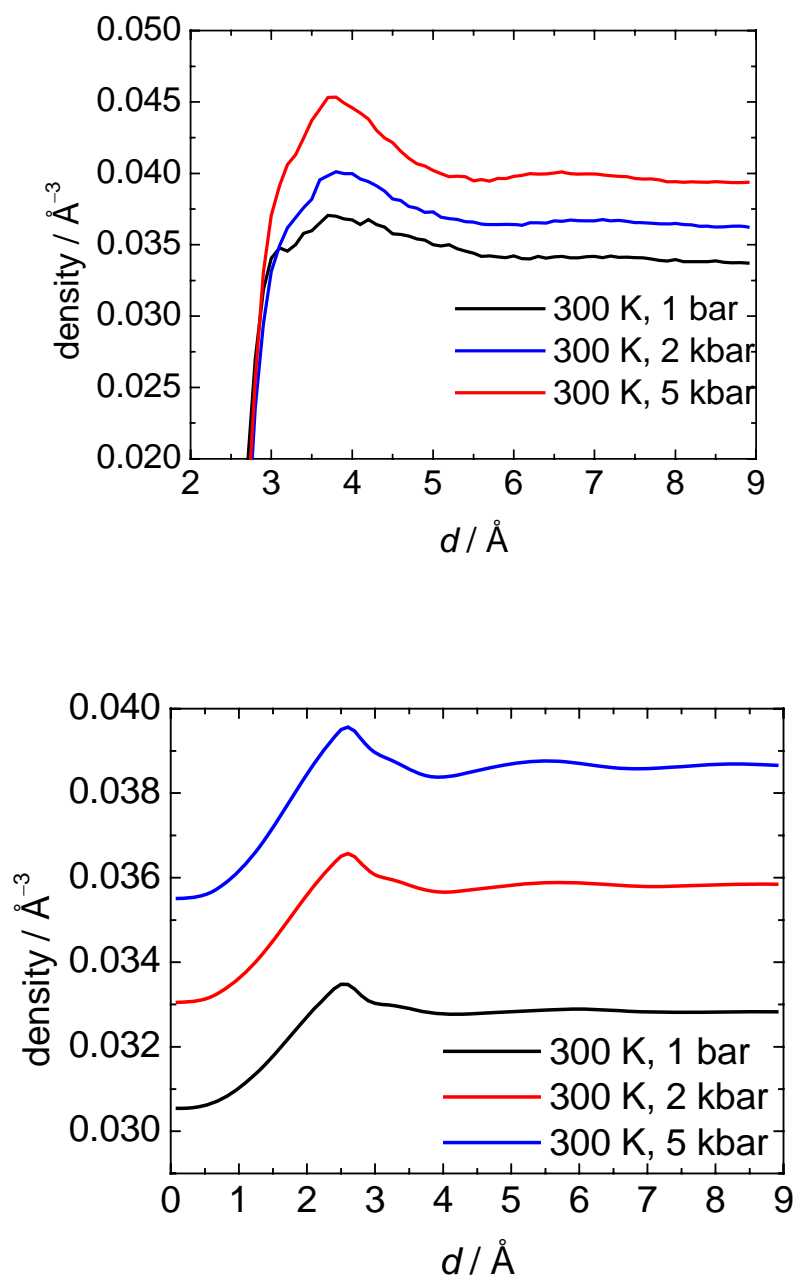


Figure 3.14: The average number density of the hydration water shell (upper panel) near the surface of SNase and the average number density of the water outside the hydration shell (lower panel) as a function of the cutoff distance (d) from nearest protein heavy atoms. At $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar.

The lower panel of Figure 3.14 shows the average density of the water outside the hydration shell of the protein as a function of the cut-off distance at different pressure conditions. The location of its maximum corresponds to the minimal distance between water molecules and protein atoms. When the cut-off, which defines the width of the hydration

shell, was moved far from the protein surface, water molecules that are localised near the protein surface contribute less and less to the average water density outside the hydration shell, finally reaching values typical for the bulk water density at the given pressure conditions. The average water density out of the hydration shell at cut-off distances > 3.8 - 4.0 Å is close to the number density of bulk TIP3P water under the same conditions, indicating that the location of the average density maximum of the hydration water around 3.8 Å is a good measure of the water shell that is mainly disturbed by the protein surface.

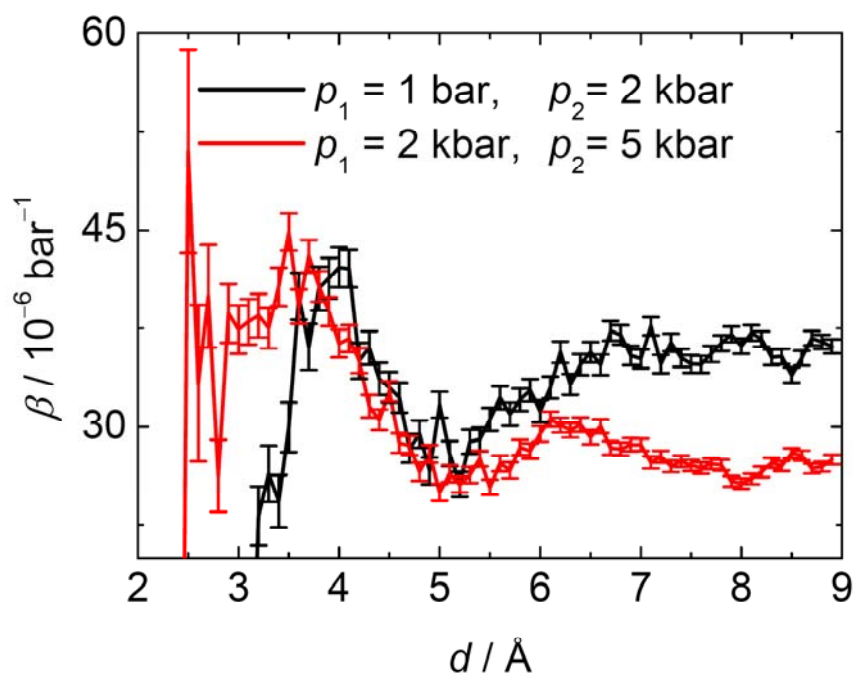


Figure 3.15: Coefficient of isothermal compressibility β of the hydration shell of SNase as a function of the cutoff distance d from nearest protein heavy atoms for two pressure ranges.

The coefficient of isothermal compressibility, $\beta = \rho^{-1}(\partial\rho/\partial p)_T$, of the hydration shell has been determined from simulations at two (ρ, p) data points using $\beta = (2/(\bar{\rho}_1 + \bar{\rho}_2))((\bar{\rho}_2 - \bar{\rho}_1)/(p_2 - p_1))$, where $\bar{\rho}_1, \bar{\rho}_2$ are the mean values of the density at pressures p_1 and p_2 , respectively. Figure 3.15 shows the isothermal compressibility data of the water within the hydration shell of SNase as a function of the cut-off distance, d . It can be clearly seen that β of the water close to the protein surface is significantly larger at higher pressures. To be able to compare the properties of the hydration water with bulk water, the β value for the bulk TIP3P water by simulation at the same p, T conditions was determined. The value of the isothermal compressibility of bulk water decreases from $(42.4 \pm 0.03) 10^{-6}$ to $(25.78 \pm 0.03) 10^{-6} \text{ bar}^{-1}$ in the same pressure range. The results are in good agreement with

previous simulations [134]. It is noticeable that the volumetric properties of the hydration water strongly depend on the polarity of the solute [177]. Hence, the isothermal compressibility of the hydration water is very different from its bulk value and strongly depends on the pressure.

Several structural properties of the SNase as a function of pressure are shown in Table 3.8. The root mean square deviation (RMSd) of all atoms from the starting native conformation increases only slightly with increasing pressure, indicating that the compact conformation is largely preserved within the simulation time of 14 ns. The root mean square fluctuations (RMSf) decrease slightly with increasing pressure. A slight decrease (of $\sim 0.2 \text{ \AA}$), i.e. compaction, is also indicated for the radius of gyration, R_g , of the protein upon initial pressurization. The change in R_g is within the error bar of the simulation, however.

Table 3.8: Changes in molecular structural properties of SNase at 300 K with increasing pressure. Standard deviations are given in parentheses.

	1 bar	2000 bar	5000 bar
RMSd / \AA	2.2 (0.3)	2.6 (0.3)	3.2 (0.3)
RMSf / \AA	1.07 (0.74)	0.91 (0.6)	0.88 (0.71)
R_g / \AA	15.7 (0.1)	15.5 (0.1)	15.5 (0.1)

Table 3.9: The molecular surface and solvent-accessible surface area of SNase for simulations at different pressures for $T = 300 \text{ K}$. Errors are standard error estimates. Standard deviations are given in parentheses.

System	Area / \AA^2					
	Molecular surface	Solvent-accessible surface				
		all residues	nonpolar	polar neutral	positively charged	negatively charged
SNase, 300 K, 1 bar	7359 ± 12 (107)	9417 ± 18 (160)	1871 ± 7 (60)	2271 ± 8 (70)	3792 ± 8 (74)	1483 ± 6 (50)
SNase, 300 K, 2 kbar	7309 ± 11 (104)	9313 ± 18 (162)	1955 ± 4 (39)	2208 ± 8 (70)	3827 ± 8 (74)	1323 ± 7 (60)
SNase, 300 K, 5 kbar	7288 ± 9 (81)	9437 ± 15 (136)	1908 ± 5 (42)	2400 ± 5 (49)	3831 ± 8 (76)	1298 ± 5 (44)

The solvent-accessible and the molecular surface areas of SNase at different pressures are shown in Table 3.9. The solvent-accessible surface area data was also divided into the contributions from different types of amino acid residues by their physical-chemical

properties: polar neutral, positively or negatively charged, nonpolar. The molecular surface area of SNase decreases ($\sim 1\%$) with increasing pressure up to 5 kbar. The solvent-accessible surface area decreases slightly with increasing pressure to 2 kbar, but then slightly increases again up to 5 kbar. The reason is a surface area increase of polar neutral amino acid residues and a largely compensating decrease of surface area of negatively charged and nonpolar groups. Pressure increase up to 2 kbar leads to an $\sim 1\%$ reduction (compaction) of the surface area, which is largely due to a significant decrease of surface area from negatively (exposed loop regions) and polar neutral groups, with a partially compensating increase of non-polar surface area. Upon further pressure increase up to 5 kbar, essentially a slight increase of polar neutral surface area takes place, leading to the above-mentioned increase of the overall solvent-accessible surface area. The increase of the solvent accessible surface area with pressure seems to be due to a pressure-induced alteration of the protein surface structure, which is also reflected in an increase of the surface roughness.

Table 3.10: The solvent-excluded, van der Waals, void and partial volumes of SNase at $T = 300$ K and different pressures. Also, the hydration contribution to the partial volume is given. Errors are standard error estimates. Standard deviations are given in parentheses.

System	Volume / \AA^3				
	Solvent-excluded volume	Van der Waals volume	Void volume	Partial protein volume	Hydration contribution to the partial volume
SNase, 300 K, 1 bar	21482 ± 10 (86)	18472 ± 2 (21)	3010 ± 8 (73)	18987 ± 43 (393)	-2495 ± 6 (52)
SNase, 300 K, 2 kbar	21099 ± 9 (79)	18403 ± 2 (19)	2696 ± 8 (71)	18874 ± 33 (301)	-2225 ± 4 (35)
SNase, 300 K, 5 kbar	20950 ± 9 (73)	18350 ± 2 (20)	2601 ± 7 (63)	18754 ± 26 (238)	-2196 ± 3 (28)

In Table 3.10, the various protein volumes determined for SNase at the different pressures are shown. The partial and solvent accessible protein volumes decrease slightly upon pressurization (about 1-2%), the van der Waals volume, as expected, much less. Hence, the observed changes in V_p and V_{SE} are mainly due to the decrease in void volume (overall $\sim 15\%$) and hydration changes (overall $\sim 12\%$).

The onset of the pressure-induced unfolding may be indicated by the occurrence of cavities and fissures on the protein surface containing more or less water molecules. Such surface defects would be reflected in an increase of the molecular roughness. One can

determine the molecular surface roughness, which can be calculated as $\eta = R_s/R_v$, where R_s and R_v are the radii of a sphere whose surface and volume, respectively, equal those of the protein (the roughness of an ideal sphere is 1). The molecular roughness increases slightly with pressure, only: from 1.402 at 1 bar to 1.408 at 5 kbar.

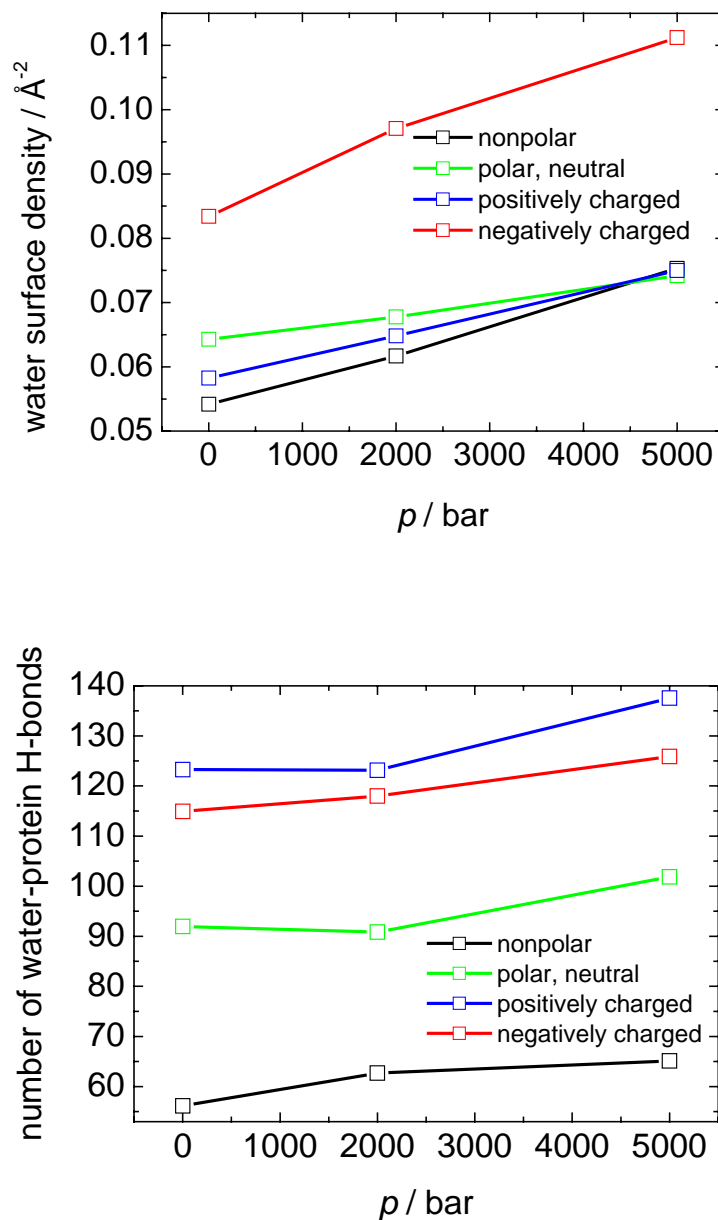


Figure 3.16: Upper panel: Pressure dependence of the surface water density (number of water molecules in the hydration shell divided by the surface area) around SNase. Lower panel: Pressure dependence of the number of water-protein H-bonds ($T = 300$ K). The colors of the various types of amino acid residues are given in the legend of the figure.

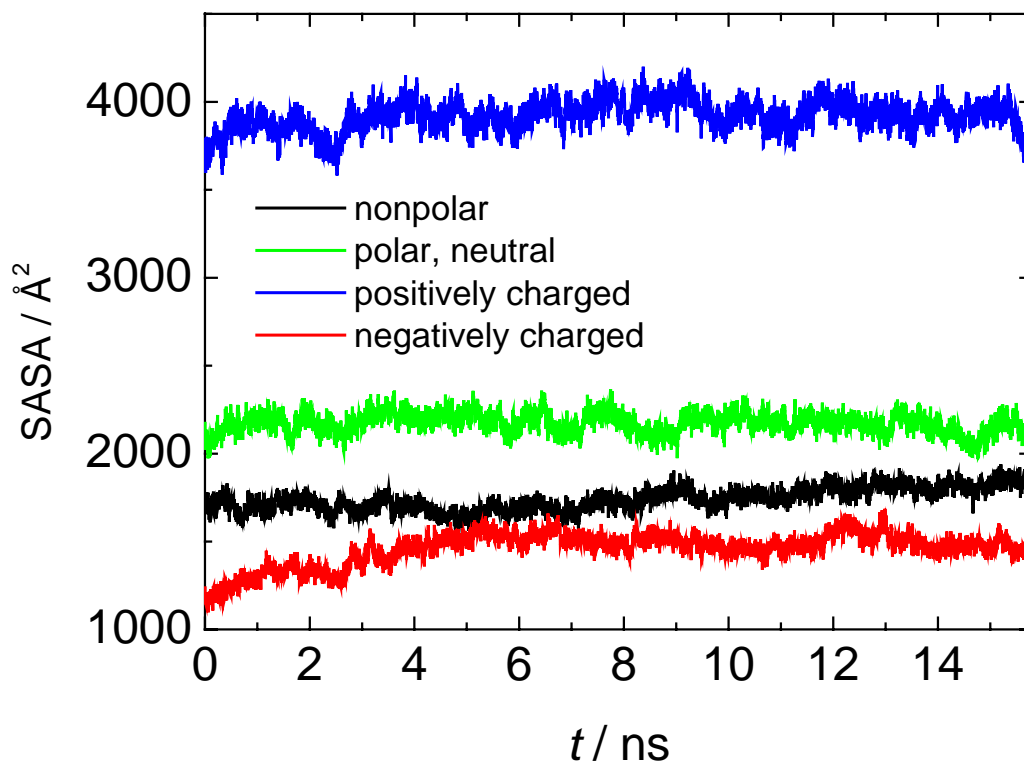


Figure 3.17: Time evolution of the solvent-accessible surface area of the different types of amino acid residues of SNase at 5 kbar ($T = 300$ K). The colors of the various types of amino acid residues are given in the legend of the figure.

The surface water density (number of water molecules in the hydration shell divided by the surface area) increases with increasing pressure, in particular around nonpolar and negatively charged residues at the protein surface (Figure 3.16, upper panel). The lower panel of Figure 3.16 exhibits the pressure dependence of the number of water-protein H-bonds. Up to ~ 2 kbar, very little changes of the H-bond network are observed, probably due to the rearrangement of a. a. residues at the protein surface. Above 2 kbar, the number of water-protein hydrogen bonds increases for all types of residues. A similar trend for the behavior of water-protein hydrogen bonds upon pressurisation of proteins has been observed in simulations before and is discussed in a review by Paci [163]. The increase of the number of water-protein H-bonds is probably not only due to the appearance of new possibilities to form hydrogen bonds between hydration water and protein surface atoms with increasing pressure, but also due to a preference of water molecules to make hydrogen bonds with the protein at the expense of water-water H-bonds [163]. Additionally, Figure 3.17 shows the time evolution of the solvent-accessible surface area of the different types of amino acid residues at

5 kbar. The solvent-accessible surface area of the negatively charged groups increases up to 5 ns and then remains essentially constant. Beyond 5 ns and until the end of the simulation run, a slight and monotonous increase of the SASA of non-polar residues was observed.

Table 3.11 shows the different contributions to the apparent coefficient of isothermal compressibility of SNase at 300 K, at medium and high pressures, respectively. The compressibility of the protein's partial volume, which consists of the van der Waals and void volume as well as a hydration contribution [161], decreases by about 30% upon compression. The corresponding solvent excluded volume, which consists of the van der Waals and void volume compression, decreases even more drastically. The coefficient of isothermal compressibility of the protein partial volume can be expressed as $\beta_p = \varphi_{\text{vdW}} \beta_{\text{vdW}} + \varphi_{\text{void}} \beta_{\text{void}} + \varphi_h \beta_h$, where φ_i is the volume fraction of the corresponding protein volume, $\varphi_i = V_i/V_p$ (Table 3.11, lower panel). As can be clearly seen, the major contribution to β_p , particularly in the lower pressure regime, stems from the void volume compression and from hydration changes.

Table 3.11: The coefficients of isothermal compressibility of the various protein volumes (a) and the relative contributions $\varphi_i \beta_i$ to the partial protein volume (b) of SNase at 300 K and different pressures. Errors are standard error estimates.

a)

SNase	$\beta / 10^{-6} \text{ bar}^{-1}$				
	Protein partial volume	Solvent excluded volume	Van der Waals volume	Void volume	Hydration contribution to the partial volume
$p_1 = 1 \text{ bar},$ $p_2 = 2 \text{ kbar}$	3.0 ± 0.5	9.0 ± 0.4	1.87 ± 0.09	55.0 ± 2.0	57.2 ± 5.0
$p_1 = 2 \text{ kbar},$ $p_2 = 5 \text{ kbar}$	2.1 ± 0.3	2.4 ± 0.3	0.96 ± 0.05	12.0 ± 1.2	4.4 ± 0.5

b)

SNase	$\varphi_i \beta_i / 10^{-6} \text{ bar}^{-1}$			
	Van der Waals volume	Void volume	Hydration contribution to the partial volume	Protein partial volume
$p_1 = 1 \text{ bar},$ $p_2 = 2 \text{ kbar}$	1.8 ± 0.1	8.3 ± 0.4	-7.1 ± 0.5	3.0 ± 0.5
$p_1 = 2 \text{ kbar},$ $p_2 = 5 \text{ kbar}$	0.9 ± 0.1	1.7 ± 0.3	-0.5 ± 0.1	2.1 ± 0.3

3.6 The Temperature Induced Unfolding Pathway of SNase

The temperature-induced unfolding of SNase was also studied by MD simulation. Figures 3.18 and 3.19 show the time evolution of the RMSd, R_g , V_{vdw} , V_{SE} and SASA at 500 K, i.e., at temperature-induced unfolding MD conditions of SNase. The R_g and backbone atom's RMSd show two distinct plateaus, from 1 to 3 ns and from about 5 to 7 ns, respectively, followed, after a more and less pronounced decrease, by a steep increase of RMSd and R_g up to the end of the simulation run. R_g fluctuates significantly and reaches values of about 23 Å after 12 ns (Figure 3.18, upper panel). The van der Waals volume, V_{vdw} , of SNase increases slightly and essentially continuously with time from the starting value of 18470 to 18650 Å³ (Figure 3.18, lower panel). Upon the temperature-induced unfolding, the void volume of the protein is expected to change drastically, which should be reflected in the time evolution of the solvent excluded volume, V_{SE} . V_{SE} increases slightly up to 4.7 ns. Then, in fact, a sudden decrease occurs, which is due to the disappearance of cavities inside the protein. Above ~8 ns, after adjustment of the simulation box, a further marked decrease of the solvent excluded volume occurs (Figure 3.18, lower panel). As depicted in Fig. 3.19, the increase of the solvent accessible surface area upon temperature-induced unfolding is mostly due to the exposition of the nonpolar amino acid residues.

The time evolution of the secondary structure elements provides a more detailed insight into the mechanism of unfolding (Figure 3.20). An overview of protein snapshots during the unfolding transition is displayed in Figure 3.21. At the beginning of the unfolding process, the protein passes through a slightly expanded, native-like state before the β -domain unfolds (Figure 3.20 and Figure 3.21, 0.0-4.0 ns). The β -domain then collapses, which is followed by various intermediate states distinguished by fluctuating packing arrangements of α -helices (Figure 3.21, 4.0-6.0 ns). Finally, a conformation of the protein is observed around 7 ns, which consists of a highly disordered tertiary structure and some remaining fluctuating α -helical structures (Figure 3.21, 6.0-7.0 ns). Obviously, the most significant changes in the protein structure already occur in the first 7 ns. Longer simulations reveal a fluctuating extended random coil structure (data not shown). Hence, one may divide the temperature-induced unfolding trajectory into four main stages, namely: (i) from 0.0 to 3.0 ns, the native structure is perturbed, forming a molten globule kind of state; (ii) from 3.0 to 4.0 ns, the β -barrel disappears; (iii) from 4.0 to 7.0 ns, disruption of α -helical native packing occurs; (iv)

after ~ 7.0 ns further unfolding and formation of a more extended random coil structure with some fluctuating α -helical conformations is observed.

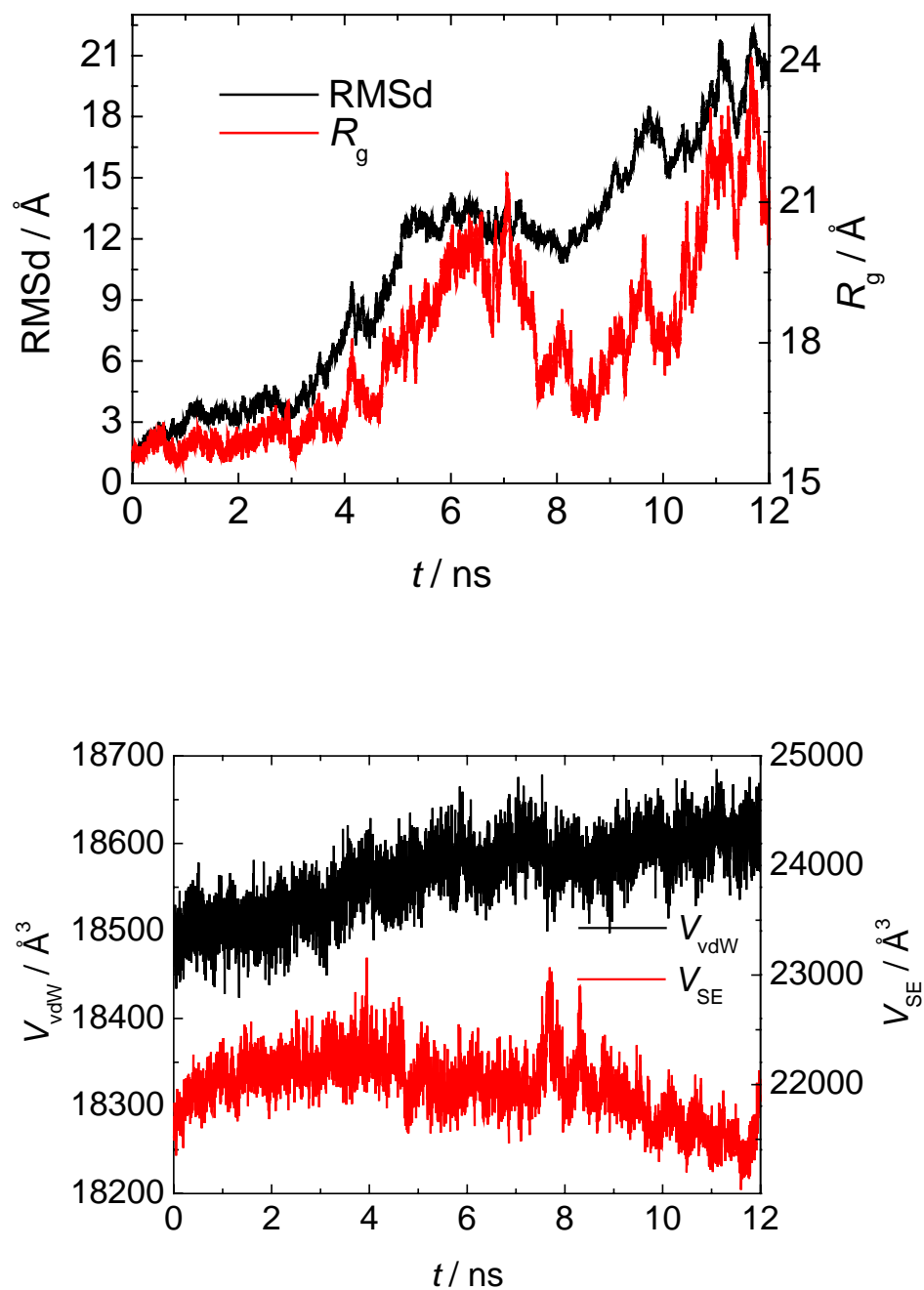


Figure 3.18: The backbone atoms RMS deviations from the energy-minimized crystal structure and the radius of gyration of SNase at 500 K as a function of the simulation time (upper panel). The solvent excluded and van der Waals volumes of SNase as a function of simulation time (lower panel).

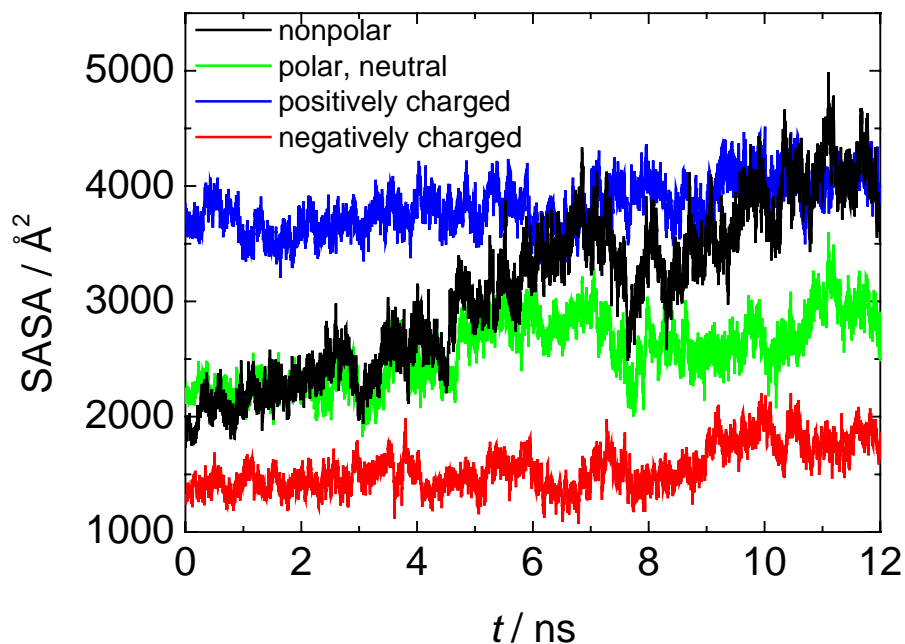


Figure 3.19: The solvent accessible surface area (SASA) of the various types of amino acid residue at $T = 500$ K as a function of simulation time. The colors of the various types of amino acid residues are given in the legend of the figure.

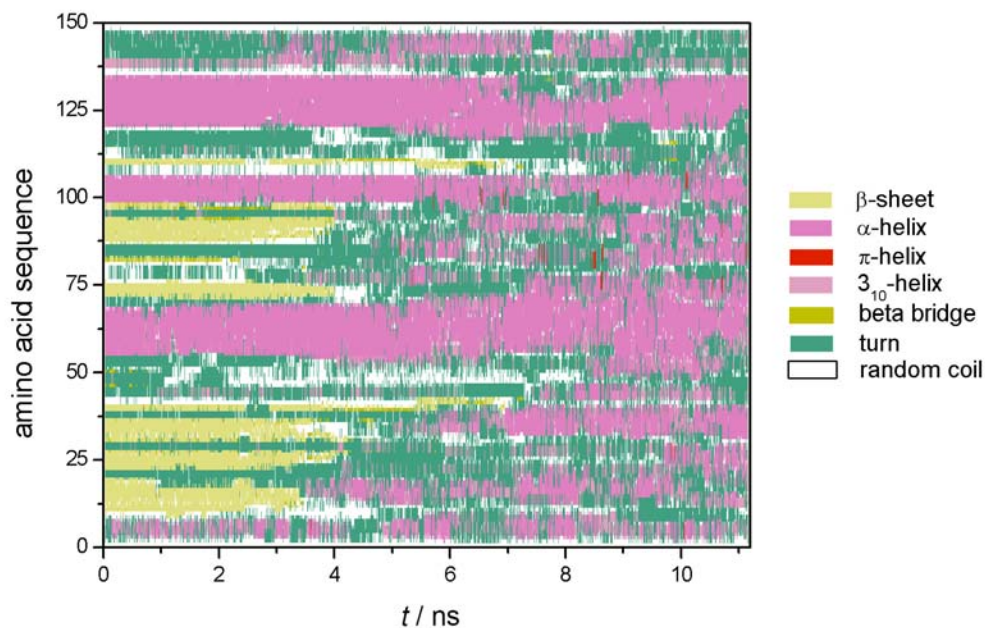


Figure 3.20: Secondary structure elements of SNase at 500 K as a function of simulation time (DSSP classification).

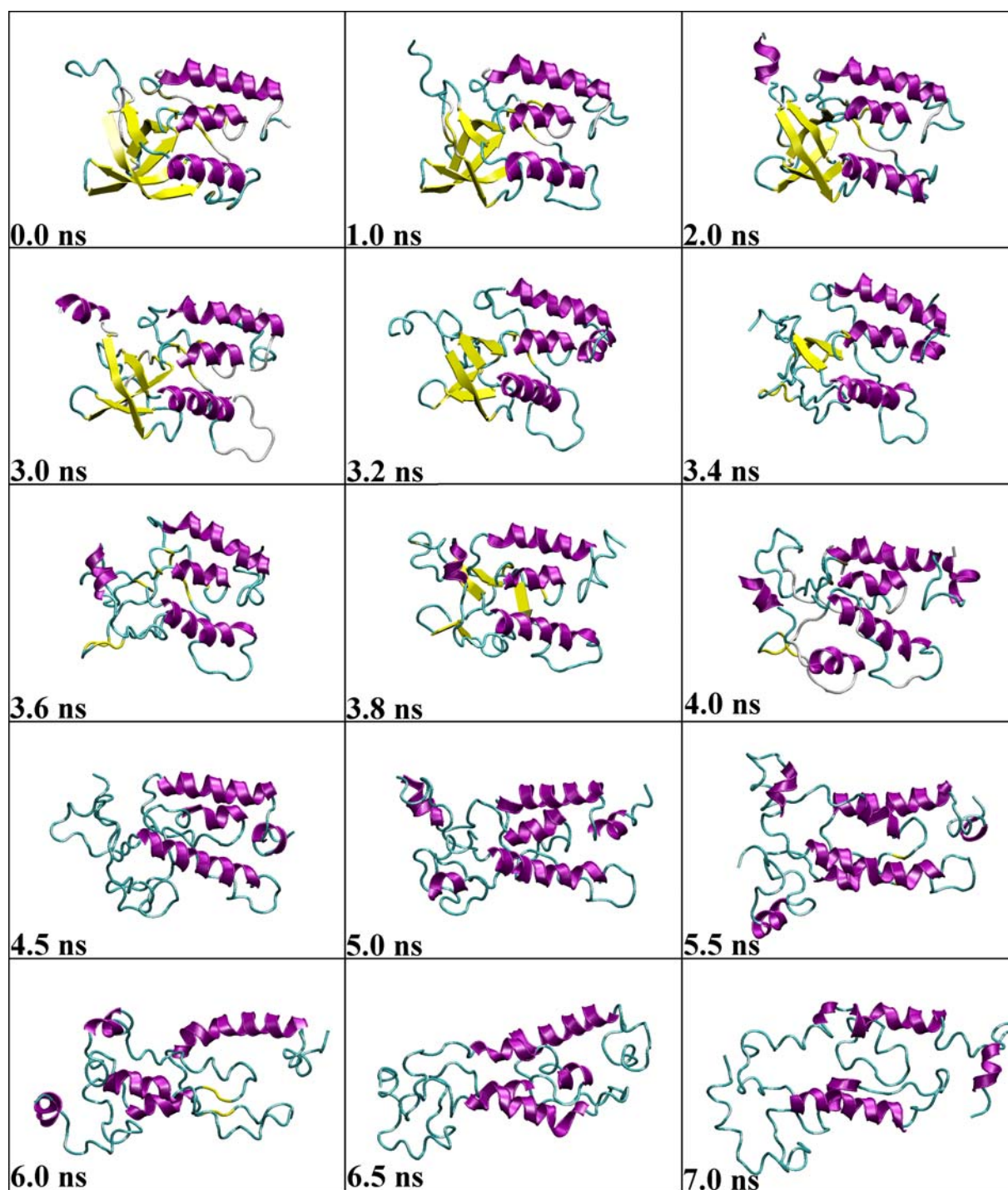


Figure 3.21: Snapshot pictures during the temperature-induced unfolding of SNase at 500 K.

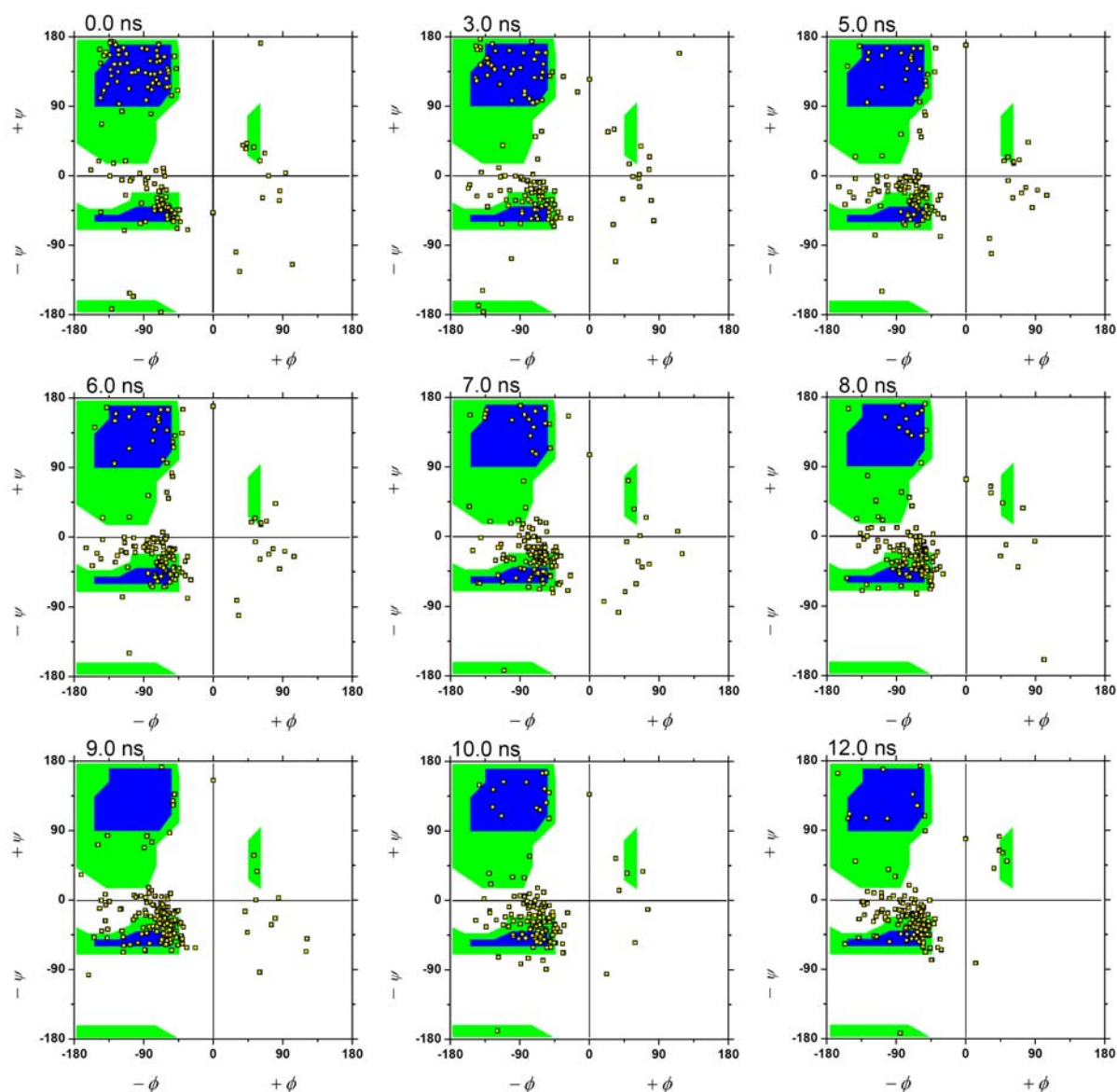


Figure 3.22: Ramachandran plots of the backbone torsion angles during the temperature-induced unfolding of SNase at 500 K.

The secondary structure may also be defined by specific backbone torsion angles and a specific main chain hydrogen bond pairing pattern. Figure 3.22 and 3.23 exhibit the conformational changes of SNase during the temperature-induced unfolding scenario as seen in the corresponding Ramachandran and contact plots. The Ramachandran plots shown in Figure 3.22 show the phi and psi values from -180 to 180 degrees, running along the horizontal and vertical axis, respectively. The freely available conformational space is shaded in blue (over 80% of the a. a. residues of proteins are found in this region of the Ramachandran plot).

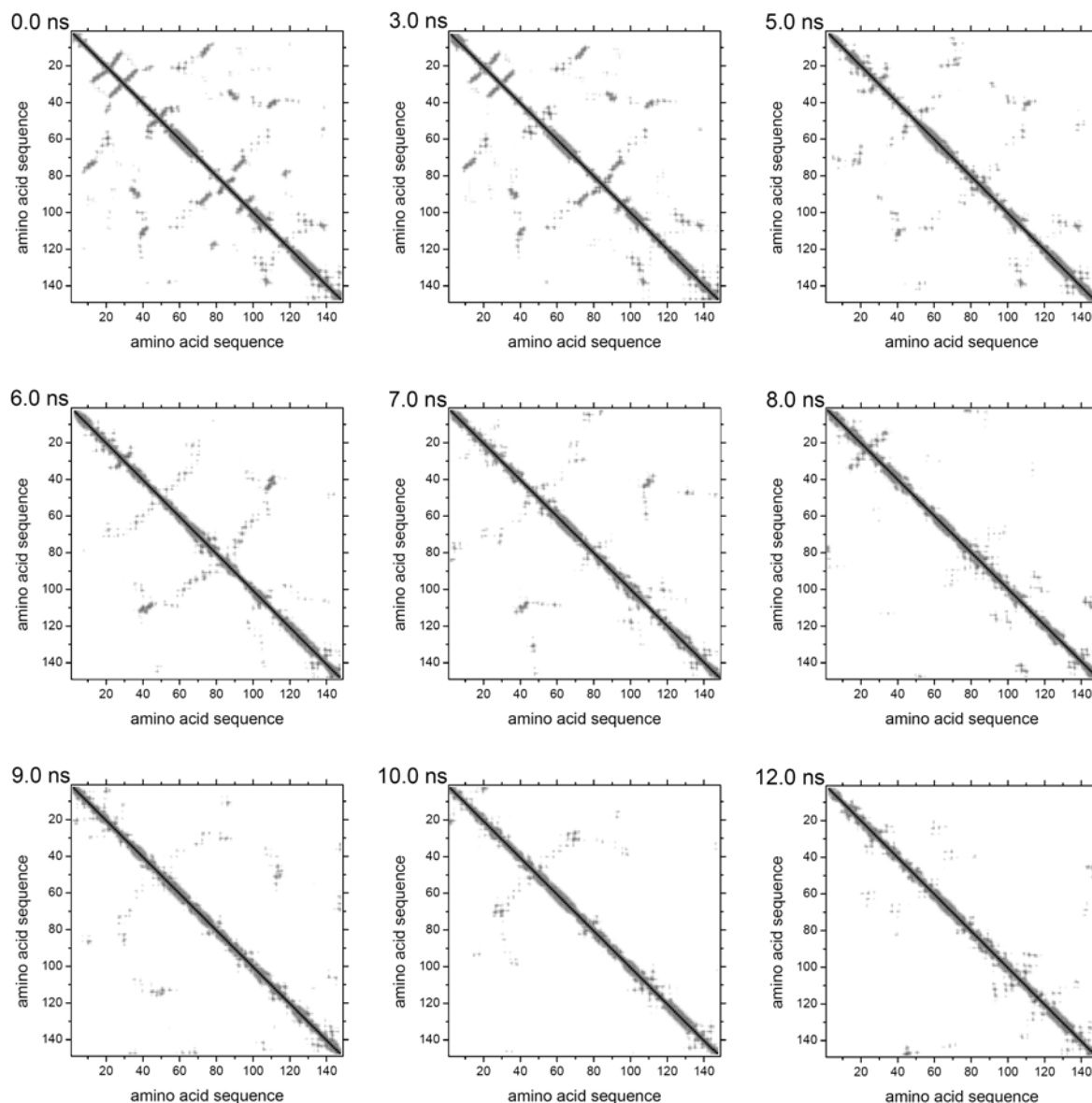


Figure 3.23: Contact maps during the temperature-induced unfolding of SNase at 500 K. The contact distances are displayed as a colour-coded matrix where dark areas indicate residues which are close to each other and lighter colours indicate residue pairs which are distant from each other (dark black at a distance of 0.0 Å, followed by a linear grey scale up to white at 10.0 Å).

Less favorable regions of the Ramachandran space are colored in green. All other regions are effectively disallowed with the minor exception of a small region representing left-handed helical structures. Right-handed α -helices fall at $-57, -47$ (ϕ, ψ)-values (in degrees) (Figure 3.22, lower-right quarter), left-hand α -helices appear at about $+60, +60$ (Figure 3.22, upper-right quarter), and 3_{10} helices at $-49, -26$. Parallel and antiparallel β -strands exhibit dihedral angles at $-119, +113$ and $-139, +135$, respectively (Figure 3.22, upper-left quarter). The corresponding contact maps indicate amino acid residue contacts and

provide useful information about the protein's secondary structure. Moreover, they also capture non-local interactions, giving clues to tertiary structural changes upon unfolding. Clusters of contacts represent particular secondary structures: α -helices appear as bands along the main diagonal since they involve contacts between one amino acid and its four successors; β -sheets appear as thick bands, parallel or antiparallel to the main diagonal. The Ramachandran plot as well as the contact map clearly show that the β -sheet structures are destabilized and disappear at about 5 ns, and a large amount of sterically hindered α -helical conformations, including 3_{10} helices, appear. Beyond a simulation time of ~ 7 ns, the α -helical structures start to fluctuate largely on the ns time-scale, and finally disappear in longer simulation runs.

3.7 Structural and Dynamic Properties of Hydration Water

Figures 3.24 and 3.25 compares the relative values of the average number of water-water H-bonds per water molecule and the average coordination number of water (number of water molecules in the first shell of a given water molecule) as a function of the distance from nearest heavy atoms of the protein, normalized by the corresponding values for the bulk water at the same conditions. The height of the peak of the relative values of the average number of water-water H-bond seems to increase slightly with increasing temperature and decrease with increasing pressure (Figure 3.24). The relative value of the average water coordination number with increasing temperature does not differ from the behaviour at ambient conditions ($p = 1$ bar, $T = 300$ K). With increasing pressure, the coordination number of water slightly decreases. The change in the height of the peaks is within the error bar of the simulation, however (Figure 3.25).

Table 3.12 shows the average number of water-water hydrogen bonds per water molecule in the bulk, within the hydration shell of SNase, and near the non-polar atoms of SNase non-polar atoms as a function of the shell thickness at different p , T conditions. The average number of hydrogen bonds in bulk water decreases with increasing temperature (from 3.64 at 300 K to 3.16 at 400 K). A pressure increase up to 5 kbar leads to a $\sim 9\%$ increase of the average number of hydrogen bonds in bulk water. A similar trend was observed for the hydration water (Table 3.12). The average number of water-water hydrogen bonds per water molecule within the hydration shell of 4.5 \AA decreases with increasing temperature from 3.00 at $T = 300$ K to 2.56 at 400 K, but increases with increasing pressure to 3.30 at $p = 5$ kbar.

Moreover, the average numbers of water-water hydrogen bonds for the water molecules that are located near the hydrophobic atoms at different temperatures and pressures were obtained.

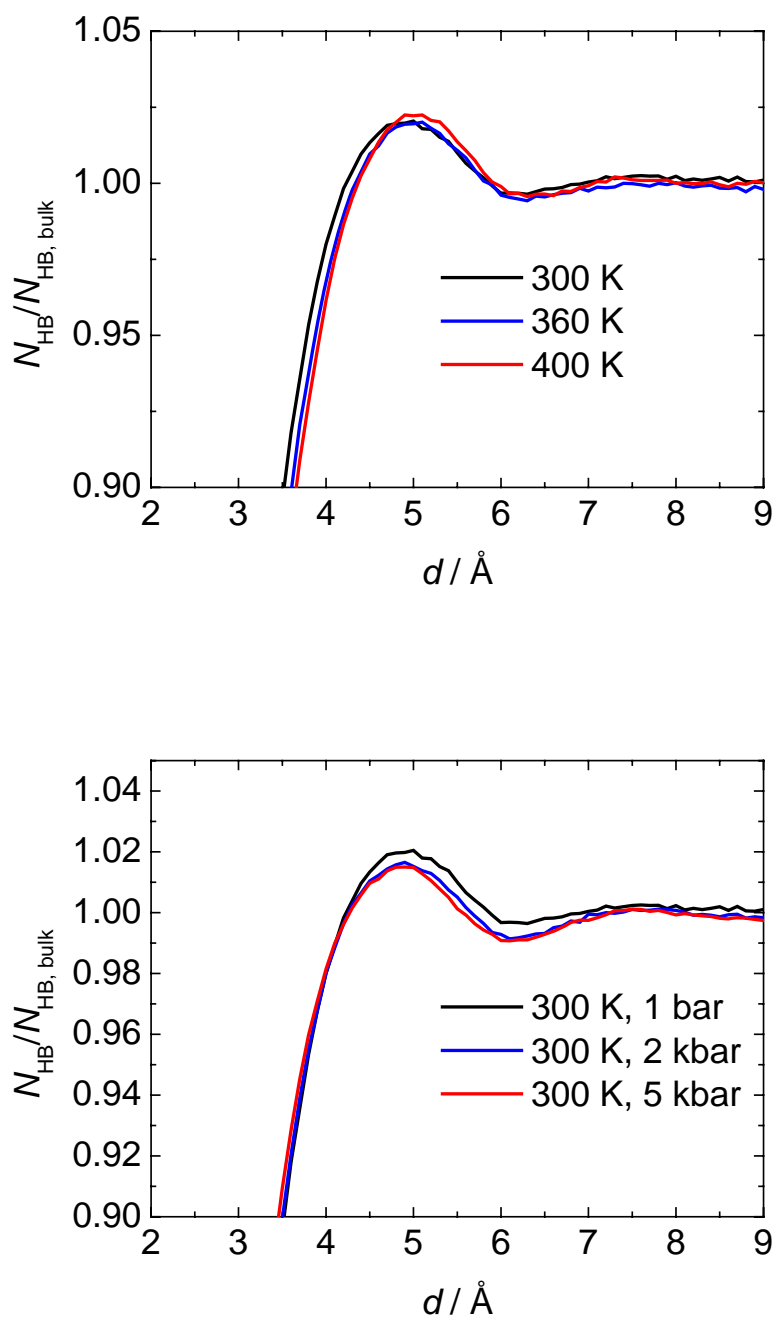


Figure 3.24: Temperature and pressure dependencies of the relative average number of water – water hydrogen bonds per water molecule (N_{HB}) as a function of the distance from nearest protein heavy atoms, normalized by the average number of water – water hydrogen bonds per water molecule in bulk water at the same conditions. Upper panel: at $p = 1$ bar and $T = 300, 360$ and 400 K. Lower panel: at $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar.

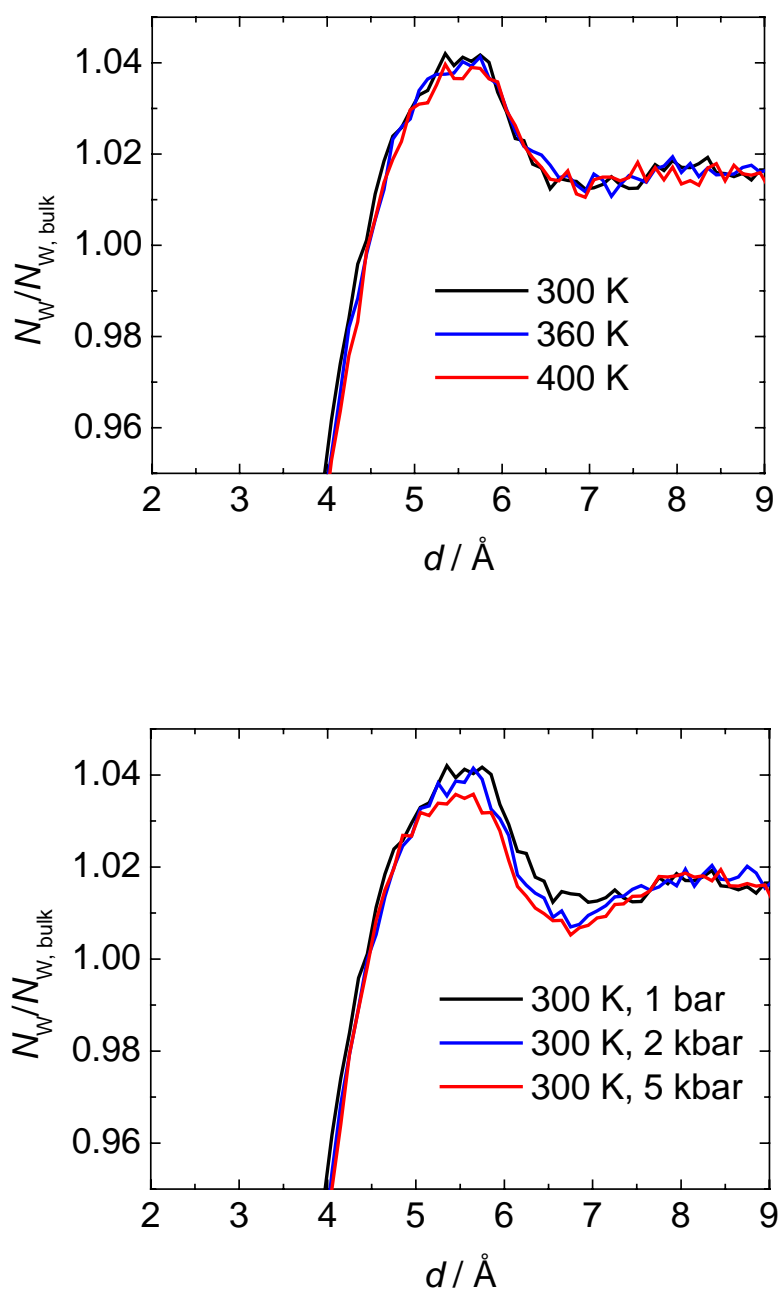


Figure 3.25: Temperature and pressure dependencies of the relative average coordination number of water molecules as a function of the distance from nearest protein heavy atoms, normalized by the average coordination number of water molecules in bulk water at the same conditions. Upper panel: at $p = 1$ bar and $T = 300, 360$ and 400 K. Lower panel: at $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar.

The average coordination numbers of water in the bulk, within the hydration shell of SNase, and near the non-polar atoms of SNase as a function of shell thickness at different p , T conditions are shown in Table 3.13. A temperature increase from 300 K to 400 K leads to a $\sim 12\%$ decrease of the average coordination number of water within the hydration shell of 4.5 Å. The average coordination number of water in the bulk decreases with increasing temperature from 5.18 at $T = 300$ K to 4.70 at 400 K, but increases with increasing pressure to 6.56 at $p = 5$ kbar.

Table 3.12: Average number of water-water hydrogen bonds per water molecule in the bulk, within a hydration shell of different thickness and near protein non-polar atoms at different temperatures and pressures.

	bulk	Hydration shell			Non-polar atoms		
	N_{HB}	N_{HB} (3.5 Å)	N_{HB} (4.0 Å)	N_{HB} (4.5 Å)	N_{HB} (3.5 Å)	N_{HB} (4.0 Å)	N_{HB} (4.5 Å)
300 K, 1 bar	3.64	2.61	2.85	3.00	3.11	3.30	3.38
360 K, 1 bar	3.37	2.39	2.61	2.75	2.81	2.99	3.08
400 K, 1 bar	3.16	2.24	2.43	2.56	2.57	2.75	2.84
300 K, 2 kbar	3.79	2.78	3.01	3.15	3.26	3.44	3.52
300 K, 5 kbar	3.95	2.95	3.17	3.30	3.44	3.60	3.67

Figure 3.26 compares the orientational probability distribution $P(\cos\theta)$ of the water dipole moment at different distances from nearest protein heavy atoms at ambient conditions. θ is the angle between the water dipole vector and the surface normal that points to the water phase. The two peaks correspond to water molecules which bind to the protein surface. The left peak at $\theta = 130^\circ$ and the right peak at about 40° correspond to the protein-water hydrogen bonds when atoms from the protein act as acceptor or as donor of hydrogen bonds, respectively. The distribution becomes broader at distances which correspond to the location of water near hydrophobic protein atoms. Figure 3.27 shows the average orientational probability distribution of the water dipole moment within the protein hydration shell as a function of the shell thickness. It is found that when the thickness of the hydration shell is small (3.5 Å), two pronounced peaks of the distribution are observed. With increasing shell thickness (4.5 Å), the average orientational probability distribution becomes broader due to an increasing number of water molecules, which locate near the protein's non-polar atoms.

The average orientational probability distributions of the water dipole moment within the hydration shell of 4.5 Å near SNase at different p , T - conditions are shown in Figure 3.28. The distribution becomes slightly broader and both peaks decrease with increasing temperature, indicating a temperature-induced weakening of the water-protein interactions and smearing out of the orientational structure of the hydration shell. With increasing pressure the left peak at $\theta = 130^\circ$ slightly decreases. The right peak increases slightly and shifts to θ -values ($\theta \approx 75^\circ$) which correspond to an orientation of water molecules almost parallel to the protein surface. This could be due to a pressure-induced increase of the relative population of water molecules near hydrophobic atoms of the protein.

Table 3.13: Average coordination number of water in the bulk, within hydration shells of different thickness and near the protein non-polar atoms at different conditions.

	Bulk	Hydration shell			Non-polar atoms		
	N_w	N_w (3.5 Å)	N_w (4.0 Å)	N_w (4.5 Å)	N_w (3.5 Å)	N_w (4.0 Å)	N_w (4.5 Å)
300 K, 1 bar	5.18	3.91	4.14	4.33	4.22	4.44	4.57
360 K, 1 bar	4.93	3.67	3.89	4.07	3.94	4.16	4.30
400 K, 1 bar	4.70	3.50	3.70	3.87	3.69	3.91	4.05
300 K, 2 kbar	5.81	4.37	4.61	4.79	4.63	4.87	5.02
300 K, 5 kbar	6.56	4.90	5.16	5.36	5.17	5.45	5.60

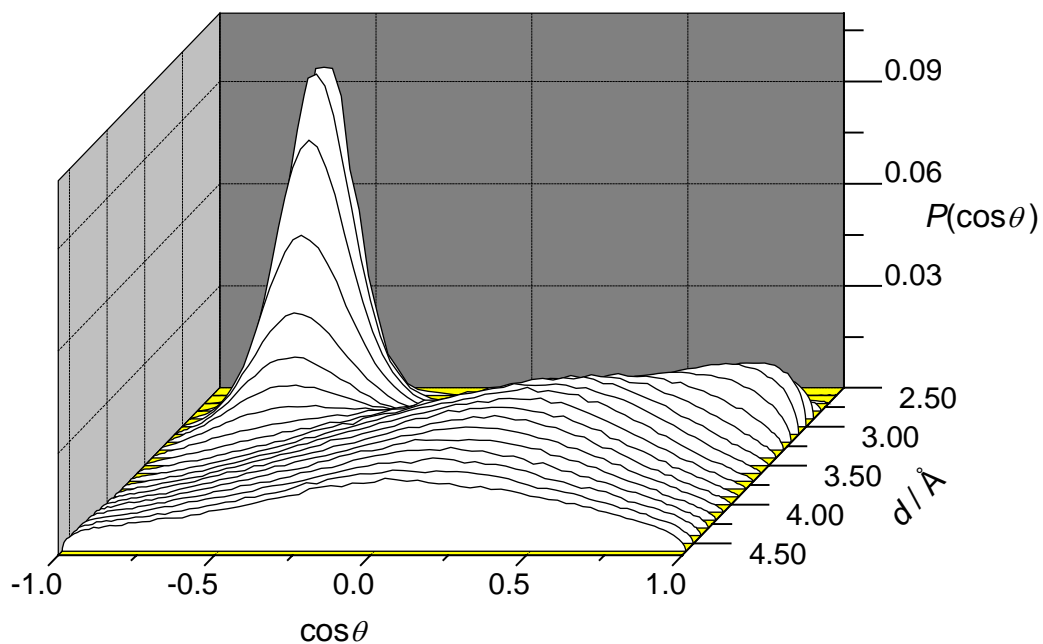


Figure 3.26: The orientational probability distribution $P(\cos\theta)$ of the water dipole moment as a function of distance (d) from nearest protein heavy atoms. θ is the angle between the water dipole vector and the surface normal that points to the water phase ($T = 300\text{ K}$, $p = 1\text{ bar}$).

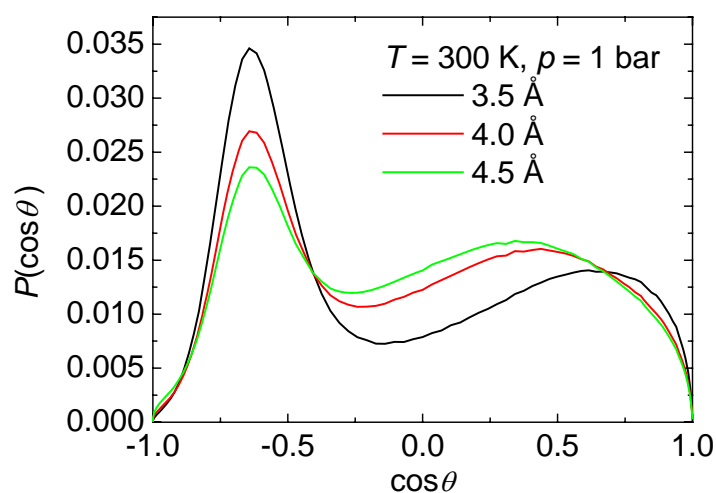


Figure 3.27: The orientational probability distribution $P(\cos\theta)$ of the water dipole moment as a function of the thickness of the hydration shell.

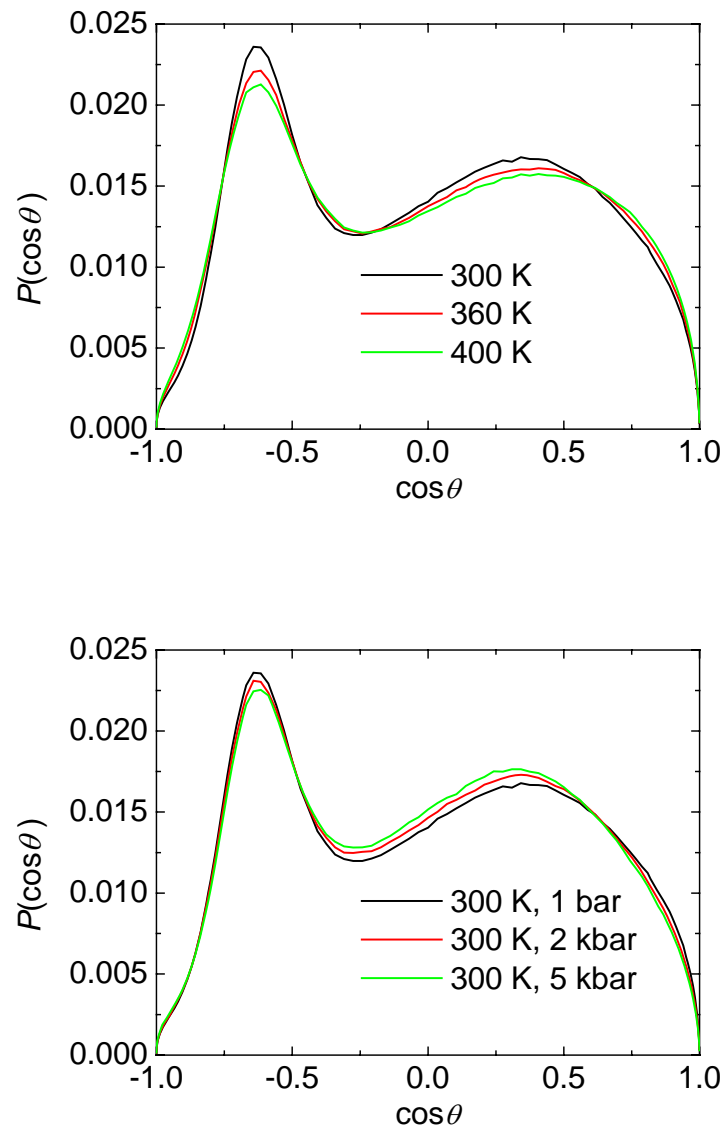


Figure 3.28: The orientational probability distribution $P(\cos\theta)$ of the water dipole moment within the hydration shell of 4.5 Å thickness. Upper panel: at $p = 1$ bar and $T = 300, 360$ and 400 K. Lower panel: at $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar.

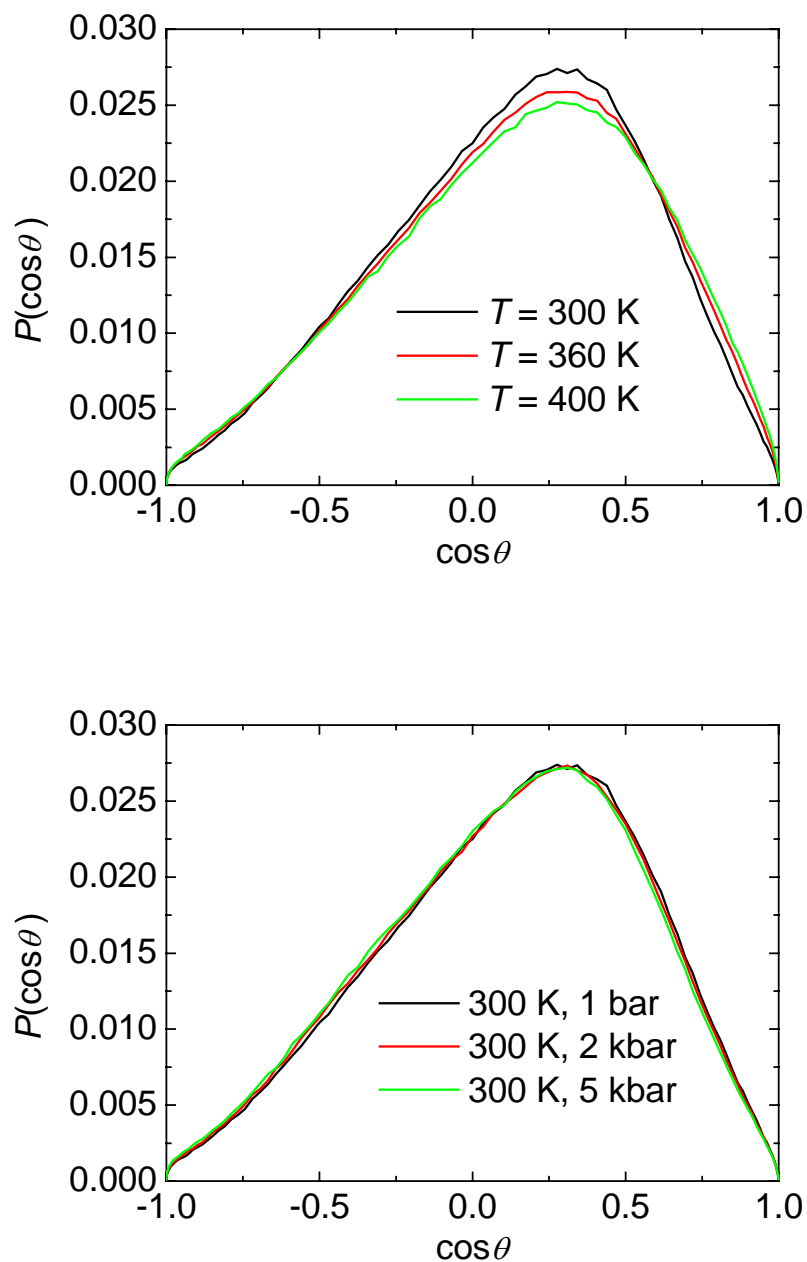


Figure 3.29: The orientational probability distribution $P(\cos\theta)$ of the water dipole moment in the hydration shell of 4.5 Å thickness near nonpolar protein atoms. Upper panel: at $p = 1$ bar and $T = 300, 360$ and 400 K. Lower panel: at $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar.

Figure 3.29 depicts the orientation of the average dipole moment within the hydration shell of SNase near the nonpolar protein atoms as a function of temperature and pressure, respectively. The peak of the distribution function corresponds to those water molecules with almost parallel orientations very close to the protein surface ($\theta \approx 75^\circ$). With increasing temperature, the height of the maximum of the distribution decreases and broadens, demonstrating a weakening of the hydration of the hydrophobic atoms at the protein surface. No changes are observed in the orientational distribution of the water molecules near the hydrophobic protein atoms upon pressurisation up to 5 kbar. Figure 3.28 (lower panel) shows an increasing population of the water molecules that are located near nonpolar protein atoms oriented parallel to the surface. Namely, the number of water near the hydrophobic atoms of protein increases upon pressurisation.

The average number of protein-water hydrogen bonds and the average number of water molecules in the hydration shell (4.5 Å) of all amino acid residues in the protein sequence were calculated for different pressures and temperatures (Figures 3.30 and 3.31). The mean number of hydrogen bonds and the average number of water molecules in the hydration shell of amino acids, on the average, decrease with increasing temperature and increase with increasing pressure. These data are in agreement with previous molecular dynamics simulation results [32, 175].

The average lifetimes of the protein-water hydrogen bonds, τ_{HB} , were calculated as a function of the residue sequence number (Figure 3.32). The relatively high values of τ_{HB} at ambient conditions are obtained for residues located at the edges of the β -sheets (residues 10, 11, 19, 21, 36, 41, 43, 75, 89, 94, 95), near the α -helices (residues 66, 122, 135, 140), and in loops (residues 52, 57, 142, 143, 146). With increasing temperature, the τ_{HB} values decrease along the amino acid sequence. For almost all residues, the average number of SNase-water hydrogen bonds increases with increasing pressure, but the average lifetime decreases, and a similar behaviour is also observed for the water-water hydrogen bonds (Figure 3.32 and Table 3.14).

The average values of the residence times of the water molecules in the hydration shell were also calculated for each amino acid residue at different p , T - conditions (Figure 3.33). In general, the results indicate decreasing water residence times with increasing temperature and increasing water residence time with increasing pressure. In Table 3.14, the average residence times of the water molecules in the hydration shell of SNase at different p , T - conditions are given. The average residence time increases from 2.67 ps at $p = 1$ bar and $T = 300$ K to 3.49 ps at $p = 5$ kbar and $T = 300$ K and decreases to 2.07 ps at $p = 1$ bar and $T = 400$ K.

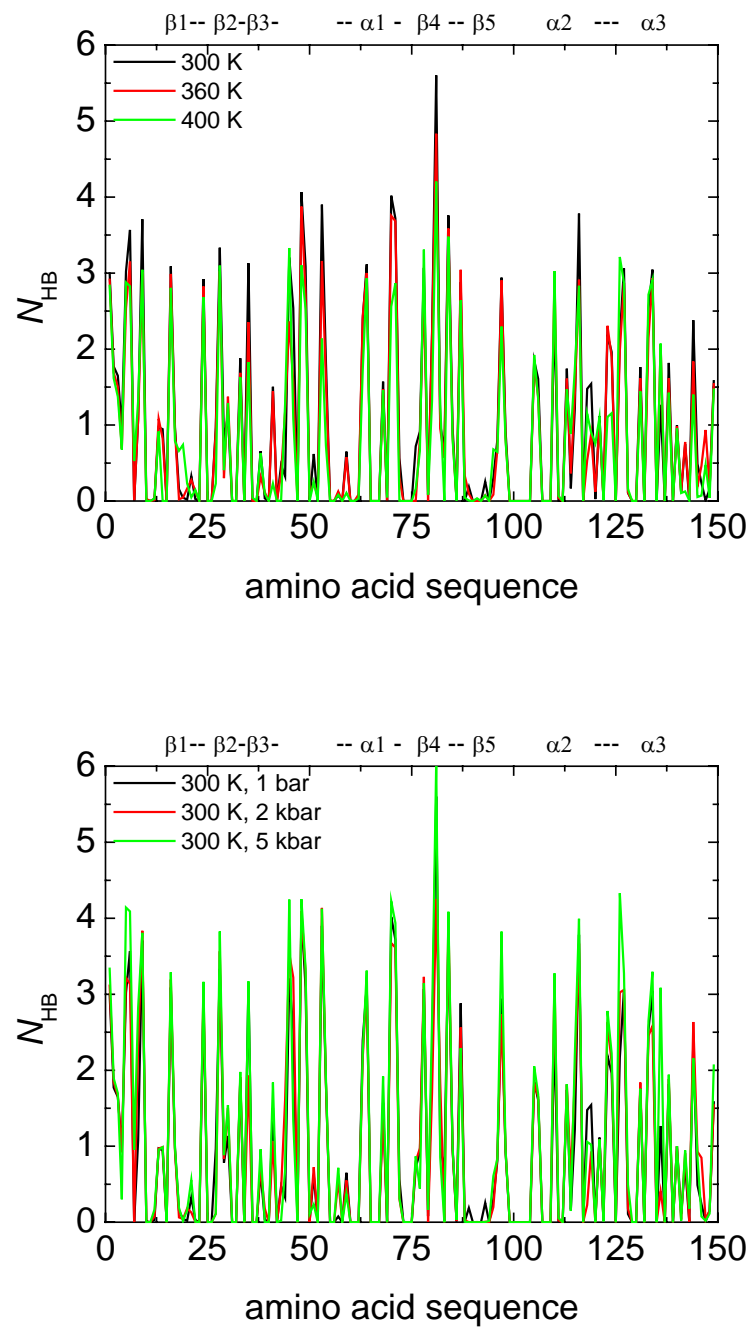


Figure 3.30: The average number of water-protein hydrogen bonds in the hydration shell of SNase as a function of the amino acid residue sequence number. At $T = 300, 360$ and 400 K and $p = 1$ bar (upper panel). At $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar (lower panel).

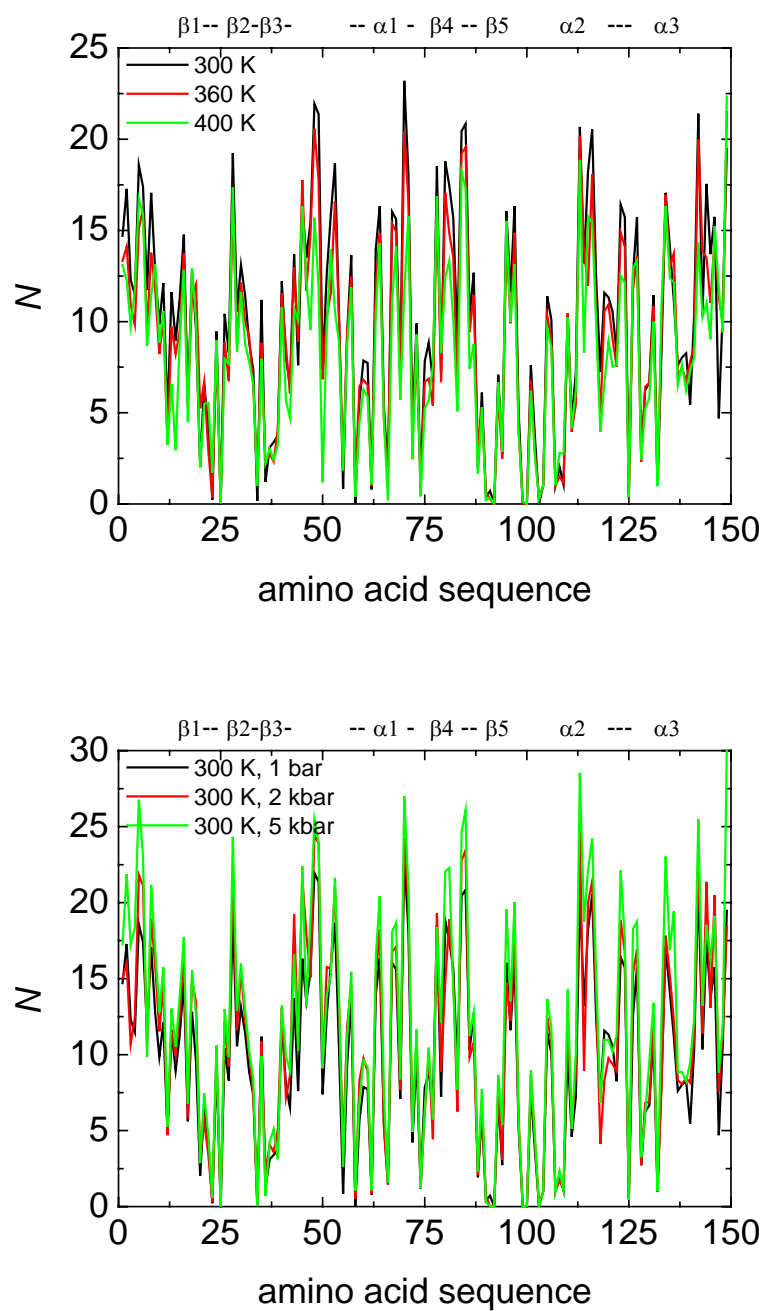


Figure 3.31: The average number of water molecules in the hydration shell of SNase (up to $d = 4.5 \text{ \AA}$) as a function of the amino acid residue sequence number. At $T = 300, 360$ and 400 K and $p = 1 \text{ bar}$ (upper panel). At $T = 300 \text{ K}$ and $p = 1 \text{ bar}, 2 \text{ kbar}$ and 5 kbar (lower panel).

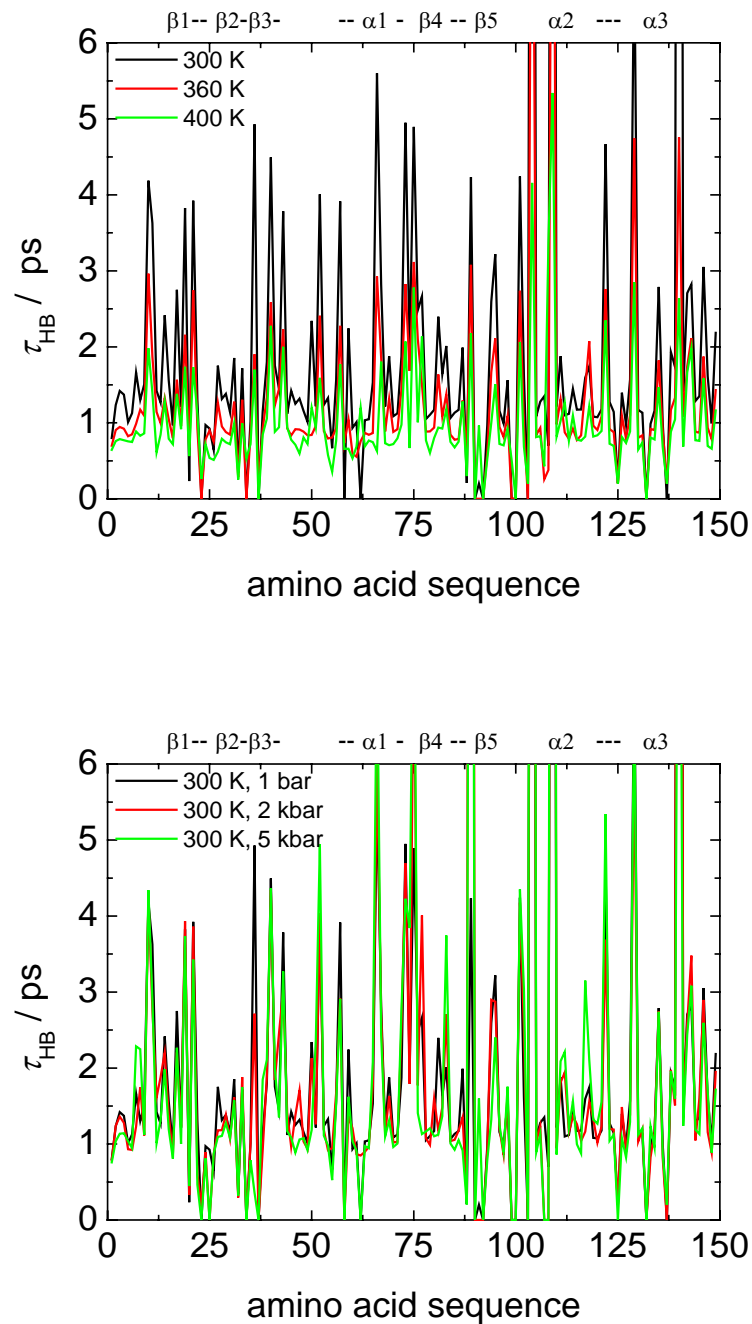


Figure 3.32: The average life time of water-protein H-bonds at the surface of SNase as a function of the amino acid residue sequence number. At $T = 300, 360$ and 400 K and $p = 1$ bar (upper panel). At $T = 300$ K and $p = 1$ bar, 2 kbar and 5 kbar (lower panel).

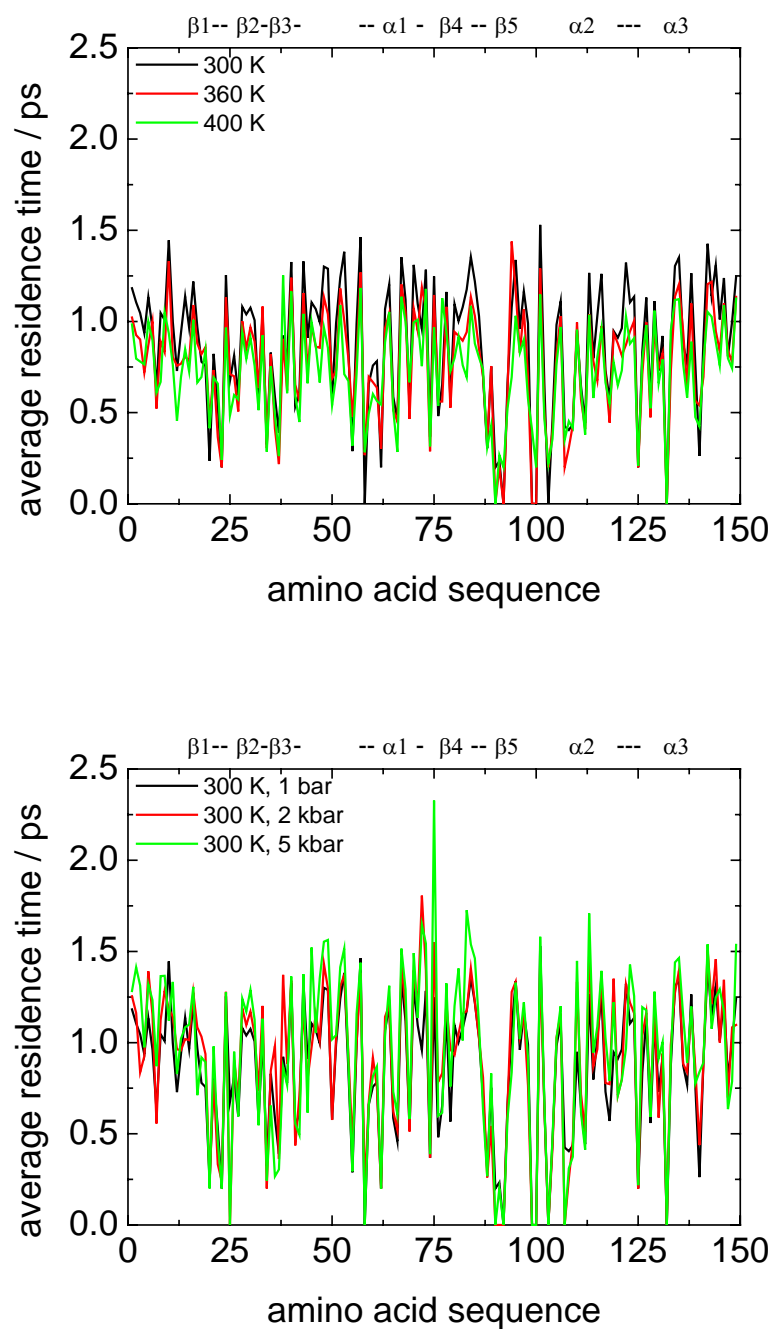


Figure 3.33: The average residence time of the water molecules in the hydration shell (up to $d = 4.5 \text{ \AA}$) of SNase as a function of the amino acid residue sequence number. At $T = 300, 360$ and 400 K and $p = 1 \text{ bar}$ (upper panel). At $T = 300 \text{ K}$ and $p = 1 \text{ bar}, 2 \text{ kbar}$ and 5 kbar (lower panel).

Table 3.14: Residence time of water molecules in the hydration shell of 4.5 Å thickness of SNase. Average lifetime of the hydrogen bonds of water-protein and water-water in the hydration shell, of water-water in the bulk, and of hydration water - bulk water at different p and T conditions.

	Average life-time of H-bonds / ps				Average residence time / ps
	water-water			protein-water	
	<i>bulk-bulk</i>	<i>shell-shell</i>	<i>shell-bulk</i>		
300 K, 1 bar	0.88	0.72	0.47	1.89	2.67
360 K, 1 bar	0.64	0.55	0.39	1.39	2.28
400 K, 1 bar	0.54	0.47	0.36	1.16	2.07
300 K, 2 kbar	0.82	0.74	0.47	1.74	3.06
300 K, 5 kbar	0.76	0.75	0.47	1.67	3.49

In order to obtain information about the dynamic properties of hydration water, the average lifetimes of water-water hydrogen bonds were calculated as well (Table 3.14). An analysis of different types of water-water hydrogen bonds was performed. The average lifetimes of the hydrogen bonds between two water molecules were calculated (i) when both molecules are located within the protein hydration shell and (ii) when one water molecule is within the hydration shell and another one located out of the hydration shell. Additionally, the average lifetime of the water-water hydrogen bonds in bulk water was calculated for the same conditions.

The average lifetime of water-water hydrogen bonds in the bulk region decreases with increasing pressure or temperature. Notably, a similar trend for the lifetime of the water-water hydrogen bonds was observed by Stanley and co-workers in liquid water using molecular dynamics simulations [210]. It is interesting that the average lifetime of the water-water hydrogen bonds, when both water molecules are located within the hydration shell of the SNase, slightly increases upon pressurization and decreases with increasing temperature.

The average residence time of water in the hydration shell of SNase of 4.5 Å decreases with increasing temperature from 2.67 ps at $T = 300$ K to 2.07 ps at 400 K, but increases with increasing pressure to 3.49 ps at $p = 5$ kbar.

3.8 Co-solvent effects

Proteins are seldom solvated by pure water. Intracellular solutions are crowded with many types of proteins, nucleic acids, metabolites, osmolytes, and other molecules. Co-solvents in aqueous solution can have profound effects on protein stability, structure, and function. The use of these solutions to stabilize or destabilize proteins, depending on the type of co-solvents, is commonplace. The presence of the co-solvents generally alters protein equilibria and reaction kinetics by perturbing the chemical potential of the protein system. Co-solvents perturb the chemical potential of the protein system by associating either more strongly or more weakly with the protein than water. Furthermore, irrespective of whether co-solvents directly bind to or are rejected by the protein surface, they are expected to induce significant changes in the quantity of “bound” water and its associated physical properties. In general, protein hydration, as well as its packing and dynamics, also changes in the course of protein un- and refolding, aggregation and binding events. The use of co-solvents such as glycerol has gained much importance, primarily because of their ability to stabilize the folded protein through a mechanism that may not involve direct contact, but rather alter the hydration shell around the protein. It has been proposed that the driving force for stabilizing the protein conformation is a non-specific solvation effect in which the preferential exclusion of solvents from the protein surface arises from enhanced solvent ordering (structure makers) [2, 178]. In contrast, when denaturing co-solvents bind to proteins, water-protein and water-co-solvent interactions are replaced by relatively stronger co-solvent-protein interactions with a concomitant release of water molecules into the bulk phase. Furthermore, these co-solvents (e.g. urea) tend to reduce the solvent ordering [179].

To study the changes of protein hydration by adding co-solvents to protein solutions, simulations of SNase in urea and glycerol solutions were performed. Details of each simulation are given in Table 3.15, namely the number of water and co-solvents molecules present. Four urea simulations were performed, UR1, UR2, UR3 and UR4, with urea mole fractions of 0.029, 0.066, 0.115 and 0.184, respectively. Three glycerol simulations were performed, GL1, GL2 and GL3, with glycerol mole fractions of 0.013, 0.043 and 0.061, respectively.

The density profiles of water, urea and glycerol were computed for different co-concentrations and are shown in Figures 3.34 and 3.35. For computing the density profiles, each co-solvent molecule was treated as a point at its center of mass.

Figure 3.34 shows the local density profiles of water near the surface of SNase in urea and glycerol solutions. The presence of urea changes significantly the shape of the water density profile in the region of the second peak, which corresponds to water which is located near hydrophobic atoms of the SNase. The perturbation of the water density profile extends to about 6 Å. In the case of the glycerol solution, due to the larger size of the glycerol molecule, such perturbation is observed even up to 7 Å.

Beyond the regions of total exclusion from the protein surface, well pronounced first-coordination shells of co-solvents can be clearly seen (Figure 3.35). The peaks of the first-coordination shells become broader and distant from the atoms of the protein as the size of the molecules they correspond to increases. The coordination peaks of co-solvents are located at distances of about 3.8 Å and 4.5 Å in urea and glycerol solutions, respectively. The sizes of the glycerol and urea molecule are about 6.5 Å and 4.5 Å, respectively. Significantly smaller second-coordination shell peaks are also visible for the urea solvating SNase. At distances greater than 7 Å from the protein, the solvation shells of the co-solvents cannot be discerned, and the local density of water, urea, and glycerol reach their bulk values. The results are in good agreement with previous molecular dynamics simulations of Rnase A in urea and glycerol solutions [180]. The authors showed a similar density distribution of urea and glycerol molecules near the Rnase A surface.

Table 3.15: Details of the molecular dynamics simulations on the SNase / co-solvent systems performed.

number of molecules	SNase in urea solutions				SNase in glycerol solutions		
	mole fraction of co-solvents						
	0.029	0.066	0.115	0.184	0.013	0.043	0.061
water	7981	7284	6312	4460	8076	7037	6617
co-solvent	243	524	840	1041	106	323	441

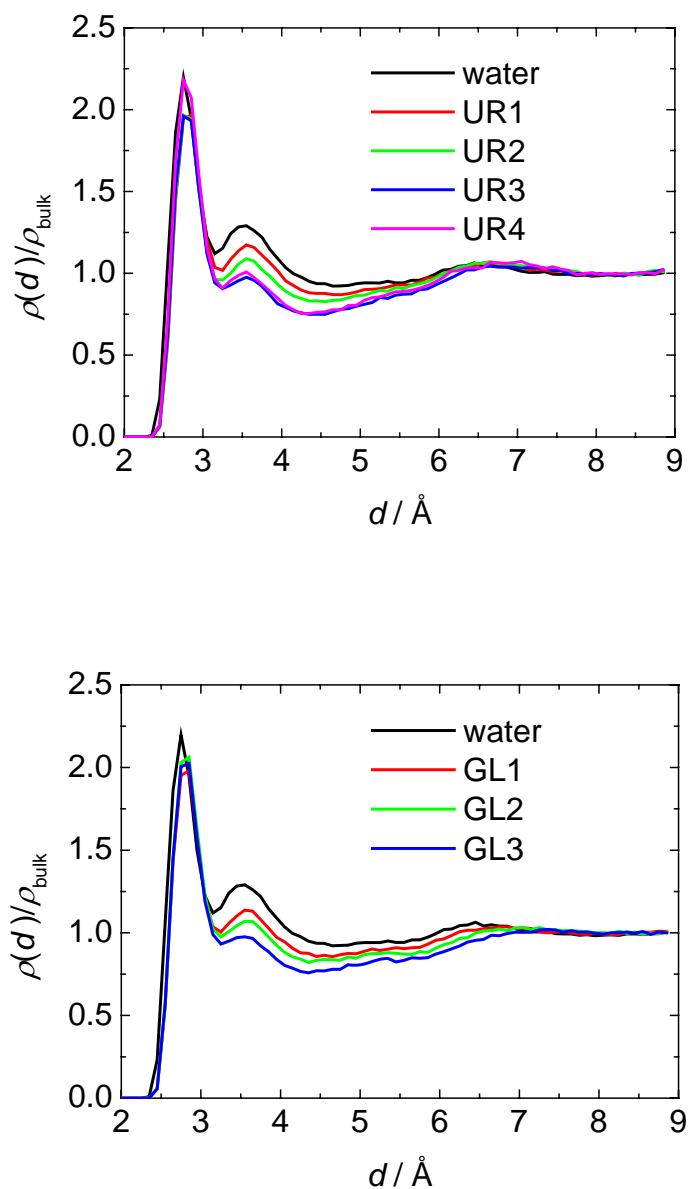


Figure 3.34: MD results of the local density profile of water, $\rho(d)$, near the surface of SNase in urea (upper panel) and glycerol (lower panel) solutions as a function of distance (d) from nearest protein heavy atoms, normalized by the density of water in the bulk region.

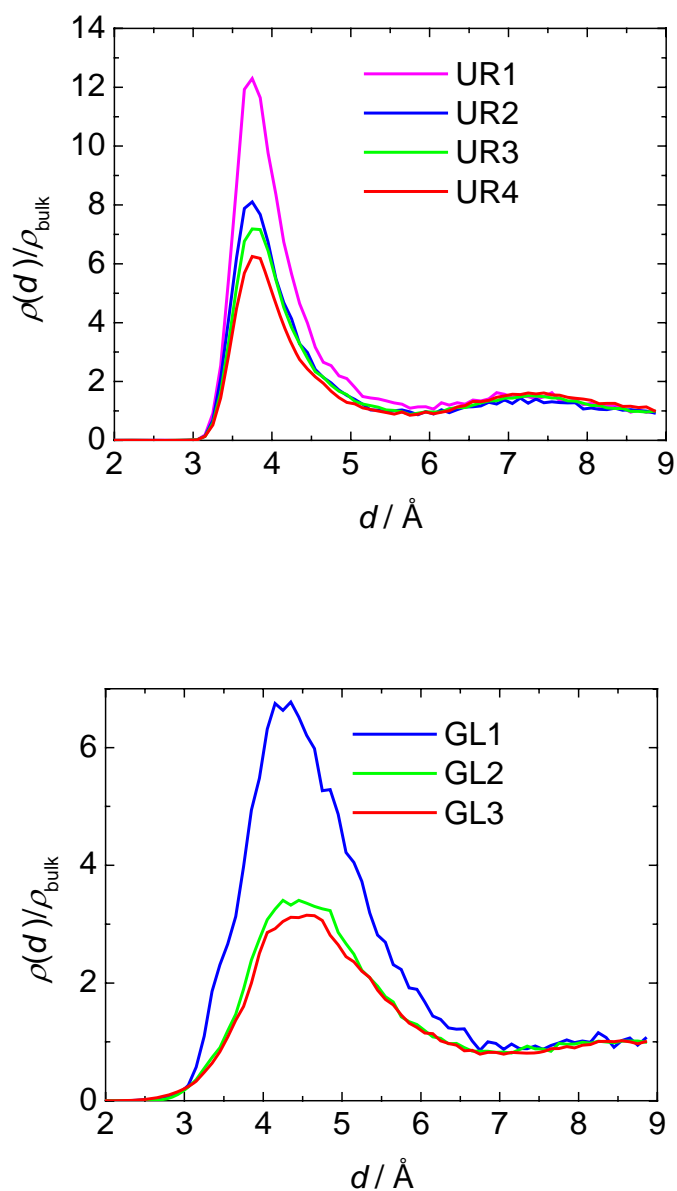
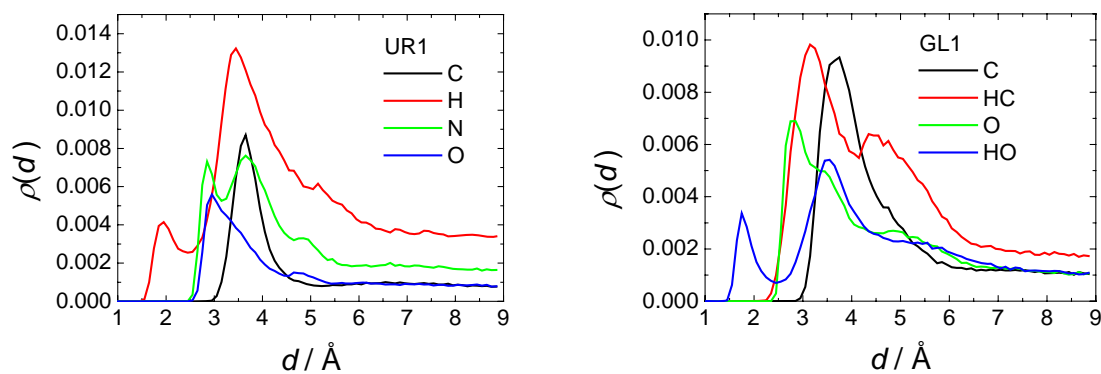
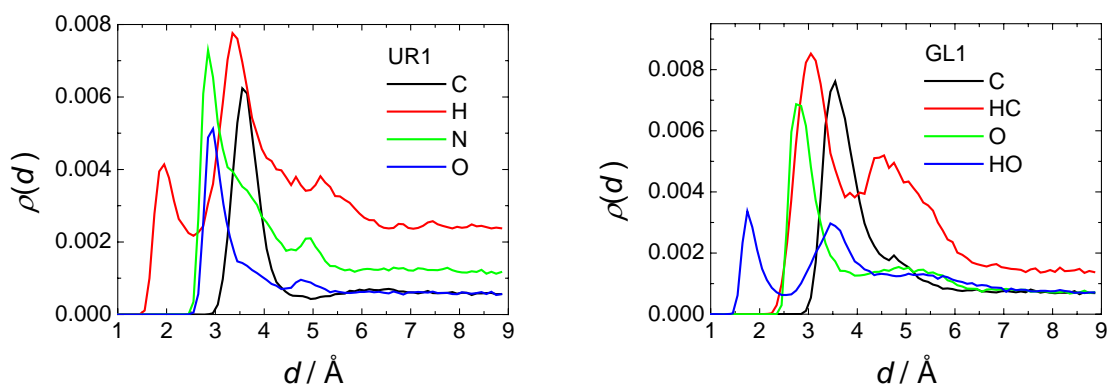


Figure 3.35: Local density profile, $\rho(d)$, of urea and glycerol near the surface of SNase in urea (upper panel) and glycerol (lower panel) solutions as a function of the distance (d) from nearest protein heavy atoms, normalized by the density of the co-solvent in the bulk region.

a)



b)



c)

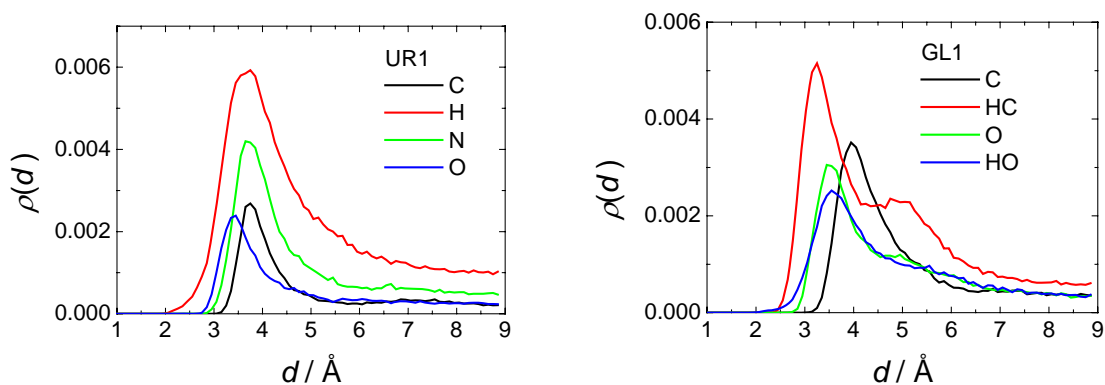


Figure 3.36: Local density profile, $\rho(d)$, of atoms of urea and glycerol near the surface of SNase in UR1 (left panels) and GL1 (right panels) solutions as a function of distance (d) from nearest protein heavy atoms. a) The total density profiles of the co-solvent atoms. b) The density profile of co-solvent atoms near the polar and c) near the nonpolar atoms of the SNase.

In order to understand the properties of the co-solvents near the SNase surface in more details, the local density profile of urea and glycerol atoms near the protein surface were calculated. Figure 3.36 shows such profiles for UR1 (left panels) and GL1 (right panels) solutions. The total density profiles of the co-solvent atoms are shown in Figure 3.36a. For further evaluation of the urea and glycerol density profiles around different types of protein atoms, the density profiles around polar and nonpolar atoms of SNase were calculated (Figure 3.36b and 3.36c, respectively). The left panel of Figure 3.36a shows the pronounced peaks of the hydrogen and oxygen atoms of urea near the surface. Those peaks correspond to the urea molecules which form H-bonds with the polar atoms of the SNase (Figure 3.36b, left panel). A similar results for the glycerol solution was observed (Figure 3.36a and 3.36b, left panels). Besides, orientations of the urea and glycerol molecules near the nonpolar atoms of SNase almost parallel to the surface of SNase were found.

Variations in the local density profiles with distance for each co-solvent are obvious up to about 6-7 Å, or two hydration shells of water, away from the protein. Glycerol is not excluded from close contact with the protein. These results are in agreement with previous molecular dynamics simulation of the protein in glycerol solutions [180].

To examine the nature of the co-solvent solvation, the hydrogen bonds between the protein and the co-solvent molecules were analysed (Tables 3.16 and 3.18). Table 3.16 shows that at low concentrations of urea (UR1), the urea hydrogen binds relatively uniformly to the residues at the protein surface, namely with protein-water/protein-urea H-bonds ratio from 0.15 to 0.32. With increasing urea concentration, the urea molecules replace more and more water molecules from the nonpolar and charged residues of SNase. It was found for the UR4 system that the average H-bonds ratio (protein-water/protein-urea) for positively and negatively charged amino acids are 0.68 and 0.58, respectively, while for neutral and nonpolar groups they are 1.08 and 0.72, respectively (Table 3.16). Table 3.18 shows that the glycerol molecules are preferentially hydrogen bonded to the charged amino acid residues of the protein.

Tables 3.17 and 3.19 lists the number of urea, glycerol and water molecules in the hydration shell of SNase during the production run of each molecular dynamics simulation. The hydration shell contains much more urea molecules than in the bulk solvent region, although the initial distribution was essentially uniform. The average ratio of the number of water to number of urea molecules within a shell of 4.5 Å increases from 0.08 to 0.46 in the UR1 and UR4 systems, respectively.

Table 3.16: Hydrogen bond statistics of SNase in urea solutions.

number of urea-SNase / water-SNase hydrogen bonds / ratio				
types of residues	UR1	UR2	UR3	UR4
all residues	63 / 318 / 0.20	100 / 291 / 0.34	125 / 279 / 0.45	169 / 235 / 0.72
nonpolar	14 / 44 / 0.32	19 / 39 / 0.49	21 / 43 / 0.49	22 / 36 / 0.61
polar, neutral	17 / 74 / 0.23	28 / 62 / 0.45	42 / 54 / 0.78	52 / 48 / 1.08
positively charged	15 / 101 / 0.15	30 / 95 / 0.32	31 / 92 / 0.34	50 / 74 / 0.68
negatively charged	17 / 99 / 0.17	23 / 95 / 0.24	31 / 90 / 0.34	45 / 77 / 0.58

Table 3.17: Statistics of the hydration shell of SNase in urea solutions. A urea or water molecule belongs to the hydration shell of SNase if the center of mass is within 4.5 Å of any atom of SNase.

number of urea / water molecules within a hydration shell of 4.5 Å / ratio				
types of residues	UR1	UR2	UR3	UR4
all residues	55 / 649 / 0.08	104 / 571 / 0.18	141 / 507 / 0.28	188 / 412 / 0.46
nonpolar	23 / 214 / 0.11	41 / 190 / 0.22	48 / 177 / 0.27	65 / 139 / 0.47
polar, neutral	24 / 248 / 0.10	46 / 198 / 0.23	66 / 171 / 0.39	77 / 161 / 0.48
positively charged	41 / 363 / 0.11	68 / 324 / 0.21	77 / 311 / 0.25	113 / 236 / 0.48
negatively charged	15 / 190 / 0.08	26 / 177 / 0.15	33 / 155 / 0.21	48 / 122 / 0.39

Table 3.18: Hydrogen bond statistics of SNase in glycerol solutions.

number of glycerol-SNase / water-SNase hydrogen bonds / ratio			
types of residues	GL1	GL2	GL3
all residues	37 / 336 / 0.11	65 / 305 / 0.21	104 / 282 / 0.37
nonpolar	7 / 47 / 0.15	9 / 45 / 0.20	14 / 42 / 0.33
polar, neutral	7 / 83 / 0.08	14 / 73 / 0.19	22 / 70 / 0.31
positively charged	11 / 107 / 0.10	19 / 101 / 0.19	27 / 94 / 0.29
negatively charged	12 / 99 / 0.12	24 / 87 / 0.28	41 / 76 / 0.54

Table 3.19: Statistics of the hydration shell of SNase in glycerol solutions. A glycerol or water molecule belongs to the hydration shell of SNase if the center of mass is within 4.5 Å of any atom of SNase.

number of glycerol / water molecules within a hydration shell of 4.5 Å / ratio			
types of residues	GL1	GL2	GL3
all residues	38 / 655 / 0.06	70 / 568 / 0.12	101 / 521 / 0.19
nonpolar	19 / 216 / 0.09	29 / 192 / 0.15	47 / 173 / 0.27
polar, neutral	16 / 254 / 0.06	31 / 207 / 0.15	43 / 200 / 0.22
positively charged	27 / 380 / 0.07	46 / 341 / 0.13	61 / 318 / 0.19
negatively charged	10 / 183 / 0.05	20 / 159 / 0.13	30 / 136 / 0.22

3.9 Discussion and Conclusions

In this study, a 7 ns constant pressure, constant temperature MD simulation of fully hydrated SNase was carried out. The trajectories were analysed with respect to changes in protein global structural properties: RMSd, RMSf, SASA, volume, R_g and secondary structure elements. The data for the RMSd and RMSf are in good agreement with previous molecular dynamics simulation of native, liganded and unliganded SNase [113,114]. The radius of gyration is very close to the experimental value (15.7 Å) [103], and the structure of the protein is stable under the simulation conditions. Most of the SASA covered by polar a. a. residues (in particular neutral and positively charged groups). The fractional solvent accessibilities for these residues are higher than those for non-polar amino acid residues. Mainly those water molecules located in the first hydration shell of SNase around exposed side chain atoms of positively and negatively charged polar residues are orientationally ordered. This is in agreement with data obtained by other simulations of fully hydrated proteins [73, 75].

The electrostatic properties of the a. a. residues play an important role in the distribution of the protein-water H-bonds at the protein surface. Most of the water molecules are hydrogen bonded to positively and negatively polar residues as an acceptor or donor of hydrogen atom, respectively. A similar behavior has been reported in molecular dynamics simulation of copper plastocyanin [32].

Our investigation provides a detailed picture of the water structure at the protein surface. Water density oscillations near the protein surface due to packing effects are observed. The average water density of the first hydration shell of SNase is ~0.3-0.6% larger than that of the bulk solvent. This tendency is in qualitative agreement with recently reported [131] data obtained by using small-angle X-ray and neutron scattering techniques. Our calculation of the density increase is lower than that predicted from molecular dynamics simulations of fully hydrated lysozyme [78]. They revealed that the first hydration layer is 15 % denser than bulk water, but about two-thirds of that value is due to geometrical effects caused by the definition of the protein surface and only a 5% density increase originates from the perturbation of the average water structure from the bulk. Levitt and Sharon [130] calculated the density of the water at the surface of BPTI using Voronoi polyhedra. A small increase in density for water molecules ~2.5 Å from the protein atoms and a small decrease in

density for water 3-4.5 Å from the surface was observed. It is noticeable that an increase or decrease of the density of water in the first hydration shell of the protein depends on the chemical characteristics of the protein surface, and that the calculation of the density of the hydration layer is very sensitive to the definition of the volume corresponding to protein and hydrating water, respectively. The broad maximum of the average number of water-water H-bonds in the range 4-6 Å must be due to an increase of the average coordination number of water molecules, which is a consequence of the increased water density in the neighboring coordination shells.

In this work, the TIP3P H-bond network was also analysed in terms of rings of hydrogen bonded water molecules. A similar analysis was performed for bulk ST2 water (at ambient conditions [79, 81], for supercooled and stretched water [83]), for the ST4 water model (for solutions of methane [93]), for SPC water (solution of hydrophobic and hydrophilic amino acids [94, 181]), and for the TIP4P water model (bulk [82] and confined water [88, 89]). In particular, the distribution of water-rings in the different hydration layers of protein was determined. Due to the density oscillation of the water near the protein and due to protein-water H-bond formation, a preferential orientation of the water-rings near the protein surface (essentially parallel near non-polar atoms and perpendicular near polar atoms) was observed when approaching the protein surface. In the first hydration shell of SNase there exist water-rings which are located in the hydration shell of side chain carbon atoms and bonded to polar side chain and/or backbone atoms by one or two hydrogen bonds. Also, we observed class I water rings, water rings near polar atoms at the protein surface, at distances 4-6 Å of the protein which are bonded to oxygen or nitrogen atoms by hydrogen bonds and oriented essentially perpendicular to the protein surface. Interestingly, interfacial water structures in a solution of Phe(amide)₂, which exhibits a large hydrophobicity, were investigated by the energy minimization method [182], and the authors found four-, five- and six-membered cyclic water structures and showed that six-membered cyclic structures of water are predominant and mainly distributed at the outside of the first hydration shell. Results of this work are similar to our data.

So far, no analysis has been reported on the structural properties of water in terms of rings of hydrogen-bonded water molecules for a protein-water system. In fact, the analysis of non-short-circuited water rings provides additional information on the local order of water molecules at the protein surface, and, in general, their distribution may serve as valuable order parameter for describing structural properties of water at biomolecular interfaces. Results of this work show that the distribution of water rings sensitively depends on the nature of the

nearest protein atom. It is noticeable that nearby carbon atoms of the protein invoke a distribution of water rings similar to that shown in simulations of water in spherical pores with slightly hydrophilic substrates [88]. The pronounced pentagon-pentagon and the pentagon-hexagon radial distribution functions for the water rings at the protein surface as well as the increased number of class II rings in the first hydration shell indicate a type of the water molecules arrangement which is similar to the structure of clathrate-like cages which are formed around apolar solute molecules in water [90, 183]. Interestingly, also a strengthening of water-pore interactions and a decreasing pore size of water confined in porous systems has been shown to increase the clathrate-like character of the water structure in the first hydration layer [89]. The existence of relatively small (4-, 5- and 6-rings) and large (7-, 8- and 9-rings) water-rings at the protein surface might have the same physical origin as the existence of polygons and hydrogen bonded chains observed in the first outer layer in hydrophilic pores [184]. As revealed by the organization of water rings around the various protein atoms, it is noticeable that both, hydration of the hydrophobic protein atoms (clathrate-like arrangement of the water molecules) and interaction with hydrophilic atoms (protein-water hydrogen bonding), are important features of the native, folded protein's surface. A rather complex hydration scenario of the protein surface is observed, which strongly depends on the local surface topography and contains competing contributions of both types of hydrating water molecules.

To conclude, the molecular dynamics simulation runs on fully hydrated SNase and the detailed analysis of the structural properties of water indicate that the structural behaviour of the water and H-bond network at the protein surface is very complex and depends on the chemical and topographical properties of the protein surface. It is noticeable that in spite of (i) the structural non-uniformity of the protein surface, (ii) the presence of different types of atoms (polar and non-polar) on protein surface and (iii) strong protein-water interactions by hydrogen bond formation, the general behaviour of the population of water-rings and water density oscillation is very similar to that known for smooth surfaces [88, 184]. This work supposes that the H-bond network structure near the protein surface is not only relevant for the structure and dynamics, but also for the biological function of the biomolecule. Hence, it is important to study such H-bond networks in terms of their primary units, such as water-rings. Such analysis can also be helpful to study the perturbation of the water structure near protein surfaces, for example, in the case of cold denaturation and upon addition of chaotropic and kosmotropic co-solvents. These results demonstrate the importance of detailed investigations of the H-bond network at protein surfaces such as that shown in this study.

In order to understand the initial stages of the pressure-induced protein unfolding, a series of MD computer simulations on SNase at 300 K up to pressures of 5 kbar were performed. The perturbation of the water hydrogen-bonded network extends roughly 1-2 water layers from the protein surface at ambient conditions, slightly further at high pressures (Fig. 3.13). This work obtains a $\sim 9\%$ density increase over the bulk in shells of 4.5 Å at 5 kbar. Interestingly, that at high pressure the compressibility of the water close to the protein surface is significantly larger than in the bulk. An oscillation of the isothermal compressibility β of the water within the hydration shell of SNase as a function of the cut-off distance was observed, which reflects the corresponding changes of density, surface structure and hydration. Notably, similar oscillations have been observed for the excess compressibility of water at hydrophobic solutes [185]. The increase of β in the first hydration shell at high pressure is – at least in part – due to an increase of hydrophobic hydration accompanying the restructuring of the protein surface, and might be a hallmark of the onset of the unfolding transition. The values for the different volume contributions of the protein to the isothermal compressibility of the protein, which were observed in our simulations, are in agreement with results obtained in previous molecular dynamics simulations [162]. The isothermal compressibility of the protein itself is essentially due to the void volume compression (Table 3.8, upper panel). In the lower pressure regime, the hydration contribution is significant as well.

During the 15 ns simulation run at 5 kbar, no significant unfolding of the protein is detected. However, the hydration of the hydrophobic protein part at the protein surface increases upon pressurisation, and a slight increase of the molecular surface roughness upon pressurization was observed. Our results indicate that the initial stages of the pressure-induced unfolding are the following steps: increasing roughness of the protein surface due to void volume compression and increasing protein-water interactions (hydration of the protein's nonpolar groups and H-bonding between water and polar atoms of the nonpolar amino acid residues). Further pressurisation may then lead to the penetration of the hydration water into the protein interior and finally into its hydrophobic core, followed by partial unfolding of the protein, as observed experimentally. Also, it was observed that the motions of SNase slow down markedly with increasing pressure, which is in agreement with previous high-pressure molecular dynamics simulations [175]. Certainly, high pressure-induced unfolding of proteins in simulations is a rather slow process, so that also the data obtained here from 15 ns molecular dynamics simulations have to be treated with care. For studies of full unfolding, one would probably have to reside to replica exchange molecular dynamics methods [186].

A high-temperature molecular dynamics simulation of SNase was also performed to reveal its temperature-induced unfolding pathways. Temperature-induced unfolding is much faster and is readily observed in this 12 ns MD simulation study. As depicted in Figure 8, the increase of the solvent accessible surface area upon temperature-induced unfolding is mostly due to the exposition of nonpolar amino acid residues. The protein undergoes essentially a two-step unfolding process. At the beginning of the temperature-induced unfolding process, the protein passes through a slightly expanded, native-like state before the β -domain unfolds at about 4.0 ns. The following nanoseconds are characterized by different and strongly fluctuating α -helical conformations. Finally, conformations of the protein are observed above 7 ns that consist of a highly disordered tertiary structure and some fluctuating α -helical structures, which also disappear in longer simulation runs, leaving a fluctuating extended random coil structure, only.

Recently, it was shown that the transformation of the hydration water shell of a protein from an “ordered” to a “disordered” state might have a direct relation to temperature-induced conformational transitions of proteins, such as protein unfolding [187]. This transition occurs via a quasi-2D percolation transition of the hydration water. Comparable studies of the effect of pressure on the percolation transition of hydration water are necessary to yield a more complete picture of the role of water at the protein surface also under high pressure conditions.

A series of simulations of SNase in the presence of co-solvents (urea and glycerol) have been performed to elucidate the interactions of co-solvents with protein. At the same time information was obtained on how the presence of co-solvents affects the water-protein interactions. Both of these are important for understanding the mechanisms of protein denaturation and stabilisation by co-solvents. Also the interactions of co-solvents with the protein via hydrogen bonding were studied. It was found that glycerol molecules are not excluded from the hydration shell of SNase and preferentially hydrogen bonded to the charged amino acid residues of SNase.

4 Percolation of Hydration Water in Biosystems

4.1 Introduction and System Setup

Water adjacent to the protein surface (so-called “hydration”, “bound” or “biological” water) strongly influences the structural and dynamical properties of proteins and enables their function [10, 11, 27]. With increasing hydration level below monolayer coverage a hydrogen-bonded water network, which fully spans the protein surfaces, appears [10, 34-38, 188-192]. Experimental studies on protein powders indicate that with the appearance of a system-spanning water network via a 2D percolation transition, the onset of some biological function of proteins is observed [34-38]. The relation between these two phenomena is not clear yet. Computer simulations can help to clarify this problem by studying the behaviour of various molecular properties at the percolation threshold.

One may assume that the appearance of protein function is enabled by a 2D water network, spanning over a single protein molecule. In this case, the protein function could be related to some specific changes of protein structure and dynamics or to the charge transfer between its various functional groups. Such an approach is widely used in theoretical and computer simulation studies of hydrated proteins (for example, the study of protonic conductivity at protein surfaces [193] and the glass transition of hydrated proteins [39]). Many properties of single hydrated protein molecules have been studied in detail by computer simulations. However, relating the obtained results to the properties of real protein systems is not trivial. On the other hand, in powder protein molecules form complex extended protein surfaces, and a water network, which spans such a “collective” surface, may cover each protein molecule only partially. Simulation studies of hydrated protein powders could be more directly related to the available experiments [10, 34-38, 188-192]. However, studies of the structural and dynamic properties of such systems are hampered by several problems. The structure of a protein powder as well as its changes with the hydration level is unknown. Additionally, the conformation of the protein molecules could change upon hydration. Finally, any realistic simulation of a protein powder needs an essential number of protein molecules and a variation of their arrangement. With these ramifications in mind, we start approaching this problem by considering a single rigid protein molecule and two simple molecular arrangements at various hydration levels.

In order to understand the onset of protein function, the structural and dynamical properties of hydrated proteins should be studied below and above the percolation threshold of hydration water. Our simulations furnish information about the location of the percolation threshold, the water distribution at protein surfaces close to the percolation transition and particular properties of spanning and non-spanning water networks.

Although some properties of bound water at the surface of single protein molecules at low hydration levels were studied by computer simulations (see, for example, Refs. 30, 32, 194), the formation of a spanning water network via a percolation transition was studied neither for single proteins nor for protein powders. In this thesis, the first computer simulation study of this phenomenon is presented, using a single rigid lysozyme molecule and two model powders of lysozyme. The simplified powder models do not take into account possible changes of the powder structure and lysozyme conformation with the hydration level, but allow to explore the hydration in a wide range, including the percolation threshold. The clustering of water molecules on the protein surface is analysed at various hydration levels and two temperatures (300 and 400 K). The higher temperature was considered to explore the temperature effect on water clustering and to avoid a possible 2D condensation of water (layering transition), which in general could be expected at hydrophilic surfaces at lower temperatures [195]. The percolation transitions in the lysozyme powders were located by analysing the water clustering at various hydration levels, using the conventional methods, which are applicable to infinite systems [59]. In order to develop an appropriate method to study the formation of a spanning network in finite, closed systems, such as the surface of a single protein, the clustering of water on the surface of hydrophilic spheres of several sizes was additionally studied. This method was then applied to locate the percolation transition of water at the surface of a single lysozyme molecule.

Hen egg white lysozyme [196] is a small globular protein with 129 amino acid residues that contains α -helices and a triple-stranded β -sheet in two structural domains. The lysozyme molecule (molecular mass of about 14.5 kDa) was modelled, using the crystallographic heavy atom coordinates from the Protein Data Bank [115] (entry 2LYM [197]) and the AMBER force-field from Ref. 51, which treats all atoms, including hydrogens, explicitly. The TIP4P model [119] was used for water.

Molecular dynamics (MD) simulations of the lysozyme + water systems were done in the NVT ensemble. The temperature was kept constant by a Berendsen thermostat [53] with a coupling time of 0.5 ps. A spherical cut-off at 9 Å was used for the van der Waals interactions and the particle mesh Ewald [118] summation method was used for the calculation of the

electrostatic interactions. Integration time steps of 2 fs and 1 fs were used at $T = 300$ K and 400 K, respectively. For the residues we chose the charged states corresponding to pH 7. The total charge of $+8e$ on the protein surface was then neutralized by a uniform distribution of the opposite charge between all protein atoms (adding a charge of $-8e/1960 \approx -0.004e$ to each atom of the lysozyme molecule). To remove bad contacts and to adopt the chosen system to the force field, the energy minimization was carried out with Steepest Descent and Conjugate Gradient methods.

The rigid model lysozyme molecule was fixed in the centre of a cubic box (edge length 60 Å), and periodic boundary conditions were applied. N_w water molecules were then randomly placed in the free space of the simulation box to provide the chosen hydration level. The water molecules were equilibrated at constant temperature during 1.5 to 3.0 ns in the field of the protein atoms. The number of water molecules varied from $N_w = 200$ to 600 at $T = 300$ K and from 400 to 800 at $T = 400$ K. Subsequently, the water clustering was analysed every hundredth integration step (every 0.2 and 0.1 ps at $T = 300$ and 400 K, respectively), and trajectories from 8 to 15 ns were used for these purposes. The surface coverage of the single lysozyme molecule was estimated as $C = N_w/SASA$, where SASA is the solvent accessible surface area, found to be about 6900 Å².

The structure of an amorphous lysozyme powder, used for the experimental studies, is not available. In low humidity tetragonal crystals [198] with a partial density of lysozyme of about $\rho \approx 0.80$ g cm⁻³, approximately 120 water molecules are in the first hydration shell of lysozyme molecules. In order to explore a wide range of hydration levels up to monolayer coverage (about 300 [192] to 420 [199] water molecules), the partial density of lysozyme in the powder should be < 0.80 g cm⁻³. Therefore, an amorphous model protein powder with a dry protein density of $\rho = 0.66$ g cm⁻³ (densely packed model powder) was studied. It was prepared by placing six rigid lysozyme molecules randomly in a cubic box (edge length 70 Å) without close intermolecular contacts (less than 3.5 Å) between any atoms. Then, during a low temperature simulation in the NPT ensemble ($T = 10$ K, $p = 1$ bar) the length of the simulation box shrank to about 60 Å, and the above given density of the dry powder was obtained.

In order to explore the effect of protein packing on the water percolation transition, an artificial loosely packed powder with a density of the dry protein of $\rho = 0.44$ g cm⁻³ was also simulated. This model powder contained four lysozyme molecules randomly arranged in a cubic box (edge length 60 Å) without close contacts, but each lysozyme molecule had at least

one medium ranged contact with another neighbour in the box (shortest intermolecular distance between 3.5 and 5 Å).

The resulting configurations of the model lysozyme powders, with various numbers N_w of water molecules added, were used as starting points for subsequent MD runs with periodic boundary conditions in all directions. In the densely packed powder, N_w varied from 500 to 1200 at $T = 300$ K and from 700 to 1200 at $T = 400$ K. In the loosely packed powder, N_w varied from 800 to 1400 at $T = 300$ K. Integration steps of 2 fs were used for all simulations of the powder. After an initial equilibration for 1 to 3 ns, the water clustering was analysed every hundredth integration step during runs of 12 to 14 ns.

The water clustering and the percolation on the surfaces of structureless hydrophilic spheres with radii $R_{sp} = 15, 30$ and 50 Å were also investigated by Monte Carlo (MC) simulations in the NVT ensemble. The water-surface interaction was described in this case by a (9-3) Lennard-Jones potential between the water oxygen and the surface with $\sigma = 2.5$ Å and a well-depth $U_0 = -4.62$ kcal/mol. To avoid 2D condensation, the simulations were performed at $T = 425$ K, exceeding the critical temperature of the layering transition ($T \approx 400$ K) for this water-surface interaction [195]. To explore the temperature effect, the clustering at the surface of the smallest sphere was also studied at $T = 475$ K. Various hydration levels of the hydrophilic spheres were obtained by placing different numbers N_w of water molecules in the simulation box. N_w varies from 150 to 450 for the sphere of radius $R_{sp} = 15$ Å, from 900 to 1300 for $R_{sp} = 30$ Å, and from 2500 to 3400 for $R_{sp} = 50$ Å. Configurations after every 1000 MC steps were analysed. That provides up to $5 \cdot 10^5$ analysed configurations for each hydration level.

The number of water molecules in the first hydration shell of the lysozyme molecule was calculated by counting up to the minima in the pair correlation functions between the water oxygen and the heavy atoms of the protein. A water molecule belongs to the first hydration shell when the shortest distance between its oxygen and at least heavy atom of the protein is < 3.5 Å for N, O or S atoms or < 4.5 Å for C atoms [20]. Similarly, in the case of the hydrophilic spheres, a water molecule was attributed to the first hydration shell when the distance from its oxygen to the surface is less than 4.5 Å [194].

Two water molecules (TIP4P) are considered as hydrogen bonded when the distance between the oxygen atoms is < 3.5 Å and the water-water interaction energy is < -2.4 kcal/mol.

The hydration level of the powders can be described not only by the number of inserted water molecules N_w , but also by C , the water mass fraction, and by the ratio h of the water

mass to the mass of the dry protein, which is widely used in studies of protein hydration. In the case of the single protein or the hydrophilic sphere, the hydration level was characterized also by the surface coverage C^* . For a spherical surface, C^* equals $N_w/4\pi(R_{sp} + r_0)^2$, where r_0 is the distance from the surface to the first maximum of the oxygen local density (r_0 is about 3 Å for the considered hydrophilic surface [195]). To estimate the surface coverage of the single lysozyme, its solvent accessible surface area (SASA) was used, i.e. $C^* = N_w/\text{SASA}$. A probe sphere of a radius of 1.4 Å yields an estimation of the SASA for our model lysozyme of about 6900 Å².

A conventional percolation analysis does not give definite numerical criteria to estimate the accuracy of the probability distributions, such as R , n_s , $m(r)$, etc. Normally, visual inspection of the obtained probability distributions and the observation that longer simulations do not change them noticeably are considered to be sufficient to make reliable conclusions. In the present studies, the running average value y_i^{av} of a given probability function y_i at the i -th simulation step was additionally calculated by averaging over all previous configurations: $y_i^{\text{av}} = \frac{y_{i-1}^{\text{av}} \cdot (i-1) + y_i}{i}$. In all samples studied, the number of the analysed configurations provided sufficient sampling: the running average values y_i^{av} did not show any monotonous trend and varied within 1% during the last 10⁴ simulated configurations, only.

4.2 Formation of Spanning Water Networks on Protein Surfaces

4.2.1 2D Percolation Transition of Water in Lysozyme Powders

Some properties of the water clusters in our model of densely packed lysozyme powder at $T = 300$ and 400 K are shown in Figure 4.1 as a function of the water mass fraction C . In infinite systems, R changes sharply from 0 to 1 at the percolation threshold. In finite systems, the percolation threshold is smeared out, and a gradual increase of R is observed in some range of hydration (Figure 4.1, upper panel). Taking into account studies of the percolation in various lattice and continuous models [62-65], the water concentration $C_1(R)$ was defined, where the probability R reaches about 50%, as a lowest boundary of the percolation threshold. We obtained the values $C_1(R) = 0.122 \pm 0.002$ at $T = 300$ K and $C_1(R) = 0.149 \pm 0.002$ at $T = 400$ K, shown in Figure 1 by blue vertical lines. The mean cluster size S_{mean} shows a broad maximum close to the hydration level $C_1(R)$ for both temperatures (Figure 4.1, middle

panel). The mean cluster size S_{mean} in finite systems passes a maximum below the percolation threshold [59]. This supports the attribution of the concentration $C_1(R)$ to the lowest limit for the percolation threshold.

The existence of a permanent spanning cluster corresponds to $R = 1$. If one attributes for the finite powder systems the appearance of a permanent spanning cluster to $R = 0.99$, this corresponds to concentrations $C_2(R) = 0.159$ at $T = 300$ K and to $C_2(R) = 0.181$ at $T = 400$ K (green vertical lines in Figure 4.1). These concentrations could serve as estimates of the upper boundary for the percolation threshold, as the true percolation threshold is reached at $R < 0.99$, independent from the system size and dimensionality [62-65].

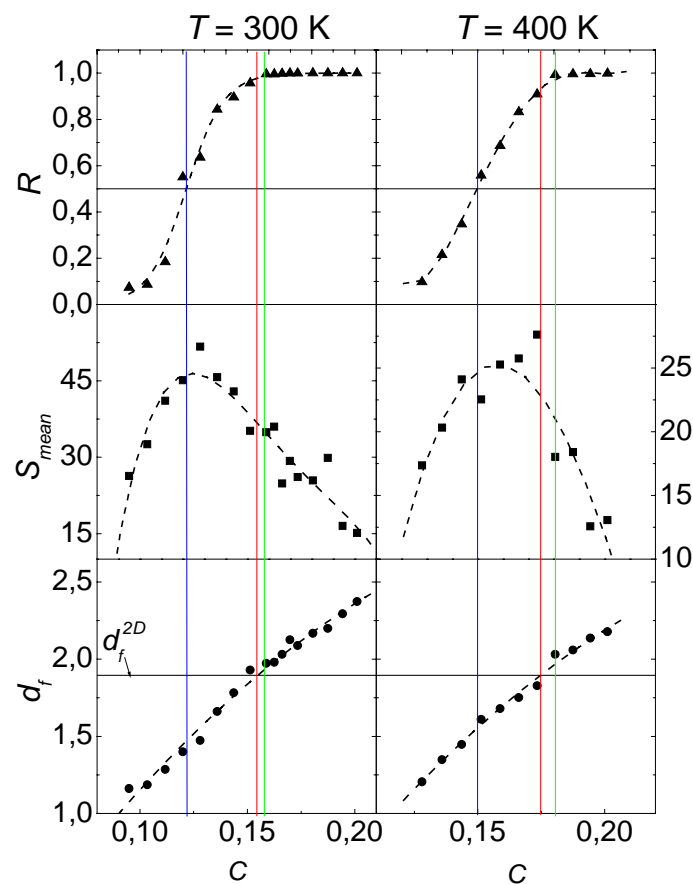


Figure 4.1: 2D percolation transition of water in the hydrated densely packed powder of lysozyme at $T = 300$ K (left panel) and 400 K (right panel). The spanning probability R (triangles, upper panel), mean cluster size S_{mean} (squares, middle panel) and fractal dimension of the largest cluster d_f (circles, lower panel) are shown as a function of the mass fraction C of water. The dashed lines are guides for the eyes only. Vertical lines indicate the water fractions $C_1(R)$ (blue), $C_2(R)$ (green) and $C_2(d_f)$ (red).

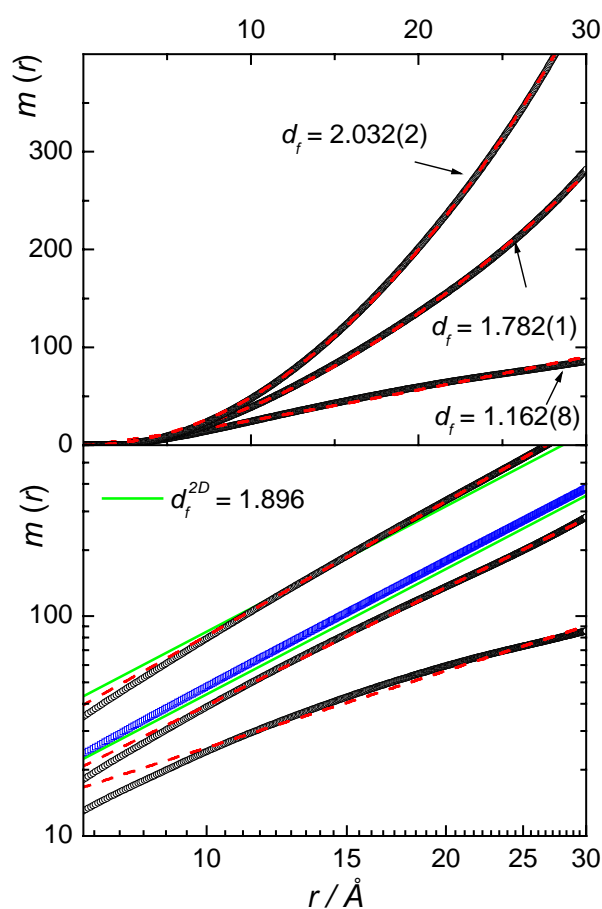


Figure 4.2: Determination of the fractal dimension d_f of the largest water cluster at $T = 300$ K in the densely packed powder. $m(r)$ vs. r (black circles) and their fits to Equation 2.26 (red dashed lines) are shown in a linear scale (upper panel) and in a double logarithmic scale (lower panel). The hydration level increases from bottom to top: $N_w = 500, 800$ and 950 . $m(r)$ for the largest water cluster at an unstructured planar hydrophilic surface at the percolation threshold is shown by blue squares. This data set and the data for $N_w = 950$ are shifted vertically in the lower panel.

To determine the dimensionality of the observed percolation transition, the fractal dimension of the largest cluster was examined. For this purpose, the function $m(r)$, which describes the mass distribution within the largest cluster, averaged over all molecules of this cluster as reference site, was calculated. Typical distributions $m(r)$ at different hydration levels of densely packed lysozyme powder are shown in Figure 4.2 (upper panel). Fits of $m(r)$ to the power law (Equation 2.26) are shown as dashed red lines. The effective values of d_f , found from the fits, are shown in Figure 4.2 (uncertainties in brackets correspond to the confidence level 95%).

Below the percolation threshold, the largest cluster is not a fractal object, and $m(r)$ does not follow a power law behaviour, which can be clearly seen in a double logarithmic scale (see Figure 4.2, bottom-most curve). The obtained values of d_f are essentially effective in this case. Approaching the percolation threshold, the largest cluster evolves to a fractal object. Indeed, a behaviour of $m(r)$ at $T = 300$ K close to a power law is observed in the range of $r > 10$ Å (see middle and upper curves in Figure 4.2). At the hydration level $C = 0.144$ ($N_w = 800$, Figure 4.2, middle curve), the effective value $d_f = 1.782$ is slightly lower than $d_f^{2D} = 1.896$ (green lines), which is observed at the 2D percolation threshold of water on an unstructured planar hydrophilic surface with periodic boundaries [200] (blue squares). The deviation of $m(r)$ downwards from the power law behaviour at $r < 10$ Å, as observed at various hydration levels (Figure 4.2, lower panel), could reflect the structured character of the lysozyme surface.

The obtained effective fractal dimension d_f gradually increases with the hydration level (see Figure 4.1, lower panel). In the hydration range where the true percolation threshold is expected (between $C_1(R)$ to $C_2(R)$), d_f varies from about 1.5 to about 1.95 for both temperatures. This evidences the 2D character of the observed percolation transition of water in densely packed lysozyme powder. The fractal dimension of the largest cluster approaches the value d_f^{2D} at water mass fractions $C_2(d_f) = 0.155 \pm 0.002$ and 0.175 ± 0.002 at $T = 300$ and 400 K, respectively, i.e. very close to the corresponding values of $C_2(R)$. Note, that with a further increase of the hydration level, the fractal dimension d_f increases towards the value of $d_f^{3D} \approx 2.53$ of 3D percolation [61], in agreement with experimental observations [35]. The difference between the mass fractions of water, corresponding to the estimated lower and upper boundaries for the true percolation threshold in densely packed powder, is only about 0.03 and slightly decreases with increasing temperature.

Table 4.1: Location of the percolation transition of water in lysozyme powders, based on the fractal dimension of the largest cluster. The threshold hydration level, when d_f approaches the 2D value $d_f^{2D} = 1.896$, is given in various terms: water mass fraction $C_2(d_f)$, level of hydration h in grams of water per gram of dry protein, and number of water molecules per lysozyme N_w/N_p . n_H is the average number of water-water H-bonds at the threshold.

system	T (K)	$C_2(d_f)$	h	N_w/N_p	n_H
densely packed	300	0.155 ± 0.003	0.183 ± 0.003	146 ± 3	2.30
densely packed	400	0.175 ± 0.002	0.212 ± 0.003	169 ± 2	2.02
loosely packed	300	0.262 ± 0.002	0.355 ± 0.003	282 ± 2	1.94

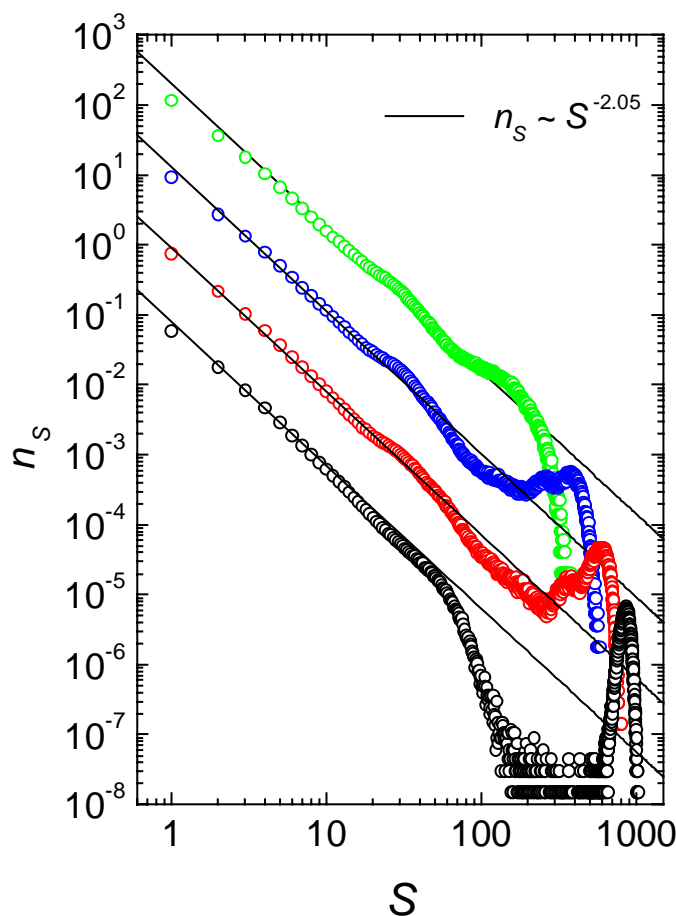


Figure 4.3: Probability distributions n_S of clusters with S water molecules in the densely packed powder at $T = 400$ K. The hydration level increases from $N_w = 700$ (top) to 1200 (bottom). Blue circles represent n_S at $N_w = 850$, when a spanning cluster exists with a probability of about 50%, while the red circles correspond to $N_w = 1000$, when the fractal dimension of the largest cluster is close to the 2D percolation threshold value. The distributions are shifted consecutively by one order of magnitude each, starting from the bottom.

A similar relation between R , S_{mean} and d_f was observed in simulation studies of the water percolation on a planar hydrophilic surface [200]. It was shown that the difference between $C_1(R)$ and $C_2(R)$, caused by the finite size of the simulated system, vanishes with increasing system size in that the value $C_1(R)$ approaches $C_2(R)$ [200]. The hydration level $C_2(d_f)$ of water in finite systems was found not to be very sensitive to the system size. Therefore, the values of $C_2(d_f)$, given in Table 4.1, could be reasonable estimates of the true 2D percolation threshold in the considered densely packed lysozyme powder. Note also, that the average number n_H of water-water H-bonds, which are formed by each water molecule, is

about 2.32 and 2.04 at the mass fraction $C_2(d_f)$ (Table 4.1) at $T=300$ and 400 K, respectively. These values are close to the percolation threshold values in 2D lattices [59]. In the particular case of a square lattice (the most relevant case for a dense water monolayer [184]), the threshold numbers of bonds are 2.00 and 2.37 for bond and site percolation, respectively [59]. The decrease of the threshold value n_H with the temperature could be attributed to the trend toward three-dimensionality due to weaker localization of the water molecules at the surface at higher temperatures (The threshold number of bonds in 3D systems is always lower than in 2D systems [59]). This corroborates with the threshold value of n_H (≈ 1.8) obtained for the quasi-2D water percolation in an aqueous solution of tetrahydrofuran [201].

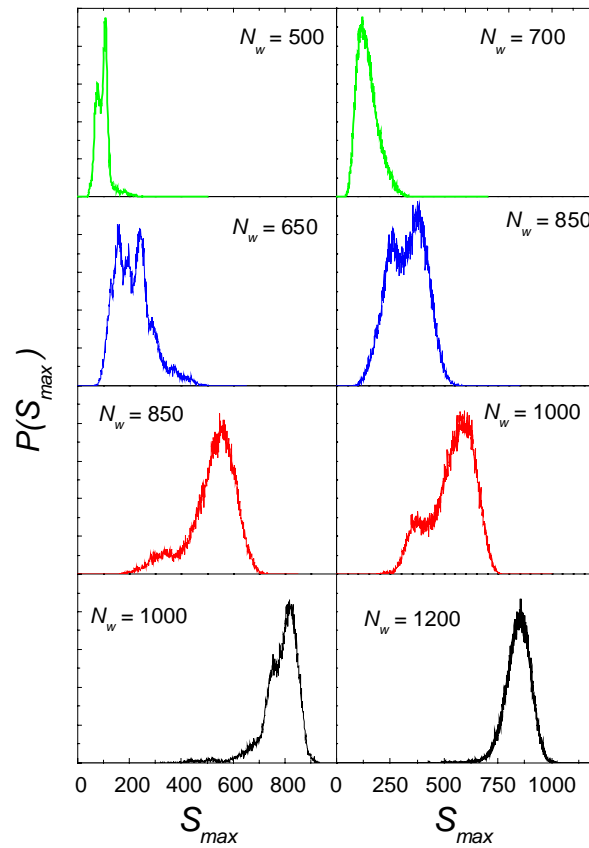


Figure 4.4: Probability distribution $P(S_{\max})$ for the largest cluster size S_{\max} of the water molecules in densely packed lysozyme powder at $T=300$ K (left panel) and at $T=400$ K (right panel) at various hydration levels. Green lines: below the percolation threshold; blue lines: R is about 50% ($C_1(R)$); red lines: 2D percolation threshold ($C_2(d_f)$); black lines: above the percolation threshold.

The cluster size distributions n_S of water in densely packed lysozyme powder at $T = 400$ K at different hydration levels are shown in Figure 4.3. The large clusters are truncated by the finite size of the simulated system, resulting in the appearance of a hump of n_S at large S . Right at the percolation threshold, the cluster size distribution n_S follows the power law $n_S \sim S^{-\tau}$ in a widest range of cluster sizes in various aqueous systems [200, 201]. In the densely packed lysozyme powder, n_S follows a power law in the range of cluster sizes up to 200 molecules, when $N_w = 1000$ (see red circles in Figure 4.3). Oscillatory deviations from the power law behaviour could not be eliminated by improving the statistics and should be attributed to the peculiar arrangement of the lysozyme molecules in the model powder. For large S to the left to the hump position, the distribution n_S deviates from the power law upwards below the percolation threshold and downwards above the percolation threshold [200, 201]. Hence, the cluster size distributions indicate a percolation threshold at $N_w \approx 1000$, corresponding to a water mass fraction $C \approx 0.17$, that is quite close to the threshold value $C_2(d_f) = 0.175$ estimated from the behaviour of the fractal dimension d_f of the largest cluster (see Figure 4.1).

The formation of an infinite water network can also be explored by studying the evolution of the largest cluster from a non-spanning to a spanning structure. In addition to the fractal dimension, the size distribution of the largest cluster $P(S_{\max})$ at various hydration levels was studied. Usually, $P(S_{\max})$ appears as an asymmetric curve with a maximum, which becomes sharper with increasing system size. The distributions $P(S_{\max})$ for the water in the densely packed protein powder at different hydration levels are shown in Figure 4. Close to the percolation threshold, $P(S_{\max})$ obviously shows a two-peak structure at $T = 400$ K. A similar evolution of $P(S_{\max})$ was observed for 2D lattices [202, 203] and for water at a planar hydrophilic surface [201]. It was found that the left peak represents finite largest clusters, while the right peak is due to infinite (spanning) clusters. When the spanning probability R is close to 50%, the two maxima in $P(S_{\max})$ have roughly the same height. The two-peak structure of $P(S_{\max})$ is pronounced in small systems and vanishes with increasing system size.

At lower temperature ($T = 300$ K), the two-peak structure of $P(S_{\max})$ is not so clear (Figure 4.4, left panel) due to the additional structure of the left peak. The splitting of the left peak of $P(S_{\max})$ could reflect the binding of the largest water clusters to some particular hydrophilic sites of lysozyme powder at low hydration levels.

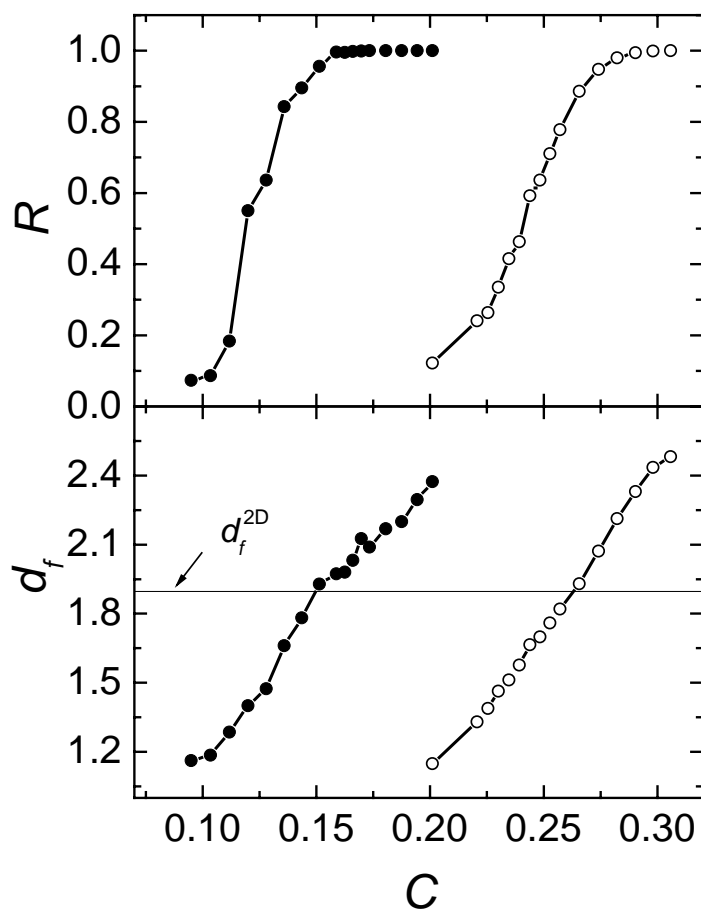


Figure 4.5: Spanning probability R (top) and fractal dimension d_f of the largest water cluster (bottom) as a function of the water mass fraction C in densely packed (solid circles) and loosely packed (open circles) powders at $T = 300$ K.

The same cluster properties were studied for water in the loosely packed lysozyme powder model, as described above. The course of the spanning probability R and the fractal dimension d_f of the largest cluster in loosely packed and densely packed powders with increasing hydration level are compared in Figure 4.5 at $T = 300$ K. In the loosely packed powder, the percolation transition of water is noticeably shifted to higher hydration levels. In particular, the threshold value $C_1(R) = 0.235 \pm 0.002$ in the loosely packed powder is almost twice the corresponding value $C_1(R) = 0.122 \pm 0.002$ in the densely packed powder. Similarly, the threshold value $C_2(d_f)$ is also essentially higher in the loosely packed powder (Figure 4.5, Table 4.1). The percolation threshold value, estimated from the cluster size distributions n_s in the loosely packed powder (not shown here), noticeably exceeds this value of $C_2(d_f)$. These observations, as well as the lowering of n_H at this threshold in comparison with the

densely packed powder (Table 4.1), could mean that the spanning water network at the percolation threshold in loosely packed powder is not two-dimensional. Visual inspection of the simulation systems evidences that the spanning water network consists of 2D sheets at the protein surface as well as of 3D water domains, formed due to capillary condensation of water in hydrophilic cavities. The latter effect causes the essential distortion of various distribution functions of water clusters in the loosely packed powder.

4.2.2 Hydrophilic Spherical Surfaces

Infinite water clusters can not appear on the surface of a finite object, such as a sphere or a single protein molecule. Strictly speaking, the probability R to find an infinite cluster is equal to zero in such cases. Nevertheless, a transition from an ensemble of finite water clusters to a water network, which spans over the whole object, should be expected. To locate and characterize this specific percolation transition of water at the surface of hydrophilic spheres and single molecules, the cluster size distribution n_S , the mean cluster size S_{mean} , the fractal dimension d_f of the largest cluster and its size distribution $P(S_{\text{max}})$ were analysed at various levels of surface coverage C^* .

To detect and characterize the formation of a spanning network on the surface of a finite object, the threshold values of the surface coverage were determined from various properties: the mean cluster size S_{mean} approaches a maximum at $C^*_1(S_{\text{mean}})$, the size distribution of the largest cluster $P(S_{\text{max}})$ shows two peaks of comparable height at $C^*_1(S_{\text{max}})$, the fractal dimension d_f of the largest cluster approaches d_f^{2D} at $C^*_2(d_f)$. The obtained threshold values are shown in Table 4.2. The mean cluster size S_{mean} passes through the maximum practically at the same surface coverage, when the two peaks in $P(S_{\text{max}})$ have similar heights, i.e. $C^*_1(S_{\text{mean}}) \approx C^*_1(S_{\text{max}})$ for the three studied spheres and two temperatures (Table 4.2). They both approach the threshold value $C^*_2(d_f)$ with increasing sphere radius. The average number of H-bonds n_H at this surface coverage (Table 4.2) is close to the threshold values in 2D-lattices [59] and the lysozyme powder (Table 4.1).

The size distributions n_S of water clusters at the surface of a hydrophilic sphere for several hydration levels are shown in Figure 4.6. n_S shows the power law behaviour in the widest range of cluster sizes at the surface coverage $C^*_2(d_f)$ (see Table 4.2). The two maxima of n_S , clearly seen at large S , are directly connected to the two-peak structure of the distribution of the largest cluster $P(S_{\text{max}})$.

Table 4.2: Water coverage C^* (in \AA^{-2}) on the surface of a single lysozyme and on hydrophilic spheres, when: the mean cluster size S_{mean} passes through the maximum, $C^*_1(S_{\text{mean}})$, two peaks in $P(S_{\text{max}})$ have the same height, $C^*_1(S_{\text{max}})$ and d_f approach the 2D threshold value $d_f^{2D} = 1.896$, $C^*_2(d_f)$. System properties at the surface coverage $C^*_2(d_f)$: $n_H(C^*_2)$, average number of H-bonds formed by each water molecule; $N_w(C^*_2)$, total number of water molecules; $N_w^1(C^*_2)$, number of water molecules in the first hydration shell.

system	$C^*_1(S_{\text{max}})$	$C^*_1(S_{\text{mean}})$	$C^*_2(d_f)$	$n_H(C^*_2)$	$N_w(C^*_2)$	$N_w^1(C^*_2)$
lysozyme, 300 K	0.058	0.056	0.065	2.26	450	336
lysozyme, 400 K	0.091	0.089	0.100	2.04	690	354
sphere, $R_{\text{sp}} = 15 \text{\AA}$, 425 K	0.086	0.082	0.096	2.11	390	350
sphere, $R_{\text{sp}} = 30 \text{\AA}$, 425 K	0.080	0.080	0.094	2.14	1270	1143
sphere, $R_{\text{sp}} = 50 \text{\AA}$, 425 K	0.087	0.087	0.092	2.15	3250	2925
sphere, $R_{\text{sp}} = 15 \text{\AA}$, 475 K	0.111	0.110	0.122	1.95	490	390

In general, the growth of the water clusters with increasing hydration level is very similar at the surface of a finite object and in an infinite system such as the lysozyme powder, considered above, or periodic planar hydrophilic surfaces [200]. This similarity allows one to distinguish spanning and non-spanning water clusters on the surface of a finite object and to determine the minimum surface coverage, providing the permanent presence of a spanning network. Namely, a spanning cluster should belong to the right (large size) peak of the distribution $P(S_{\text{max}})$ of the largest cluster, and therefore could be easily detected in the case of a pronounced two-peak structure of $P(S_{\text{max}})$. A spanning network starts to be permanently present when the fractal dimension of the largest cluster approximately approaches the value 1.896 expected at the 2D percolation threshold.

The changes of $P(S_{\text{max}})$ on the surfaces of two hydrophilic spheres with increasing water coverage is shown in Figure 4.7. The two-peak structure of $P(S_{\text{max}})$ is clearly seen in the considered range of surface coverage. It is evident that the two maxima in $P(S_{\text{max}})$ have roughly the same height when the probability to find a spanning cluster is close to 50%. At this hydration level, the two peaks are rather narrow at $T = 425 \text{ K}$ (blue lines in Figure 4.7). They become broader, but still distinct at $T = 475 \text{ K}$. Surprisingly, the two-peak structure of $P(S_{\text{max}})$ seems to be only weakly sensitive to the size of the sphere. Since the surface of a sphere should approach an infinite planar surface with increasing radius, one may expect a smearing of the two-peak structure of $P(S_{\text{max}})$ with increasing sphere radius. To test this expectation, simulations of the water clustering at the surface of a very large sphere (radius

$R_{\text{sp}} = 50 \text{ \AA}$) with a surface area of more than $35\,000 \text{ \AA}^2$ were performed. The evolution of $P(S_{\text{max}})$ with increasing surface coverage at this spherical surface (Figure 4.8) is similar to the ones observed at the smaller spherical surfaces (Figure 4.7). Surprisingly, the two-peak structure of $P(S_{\text{max}})$ is still pronounced at the surface coverage $C^*_1(S_{\text{max}})$ (see Figure 4.9), which is close to $C^*_1(S_{\text{mean}})$, where the mean cluster size S_{mean} passes through the maximum (see Table 4.2).

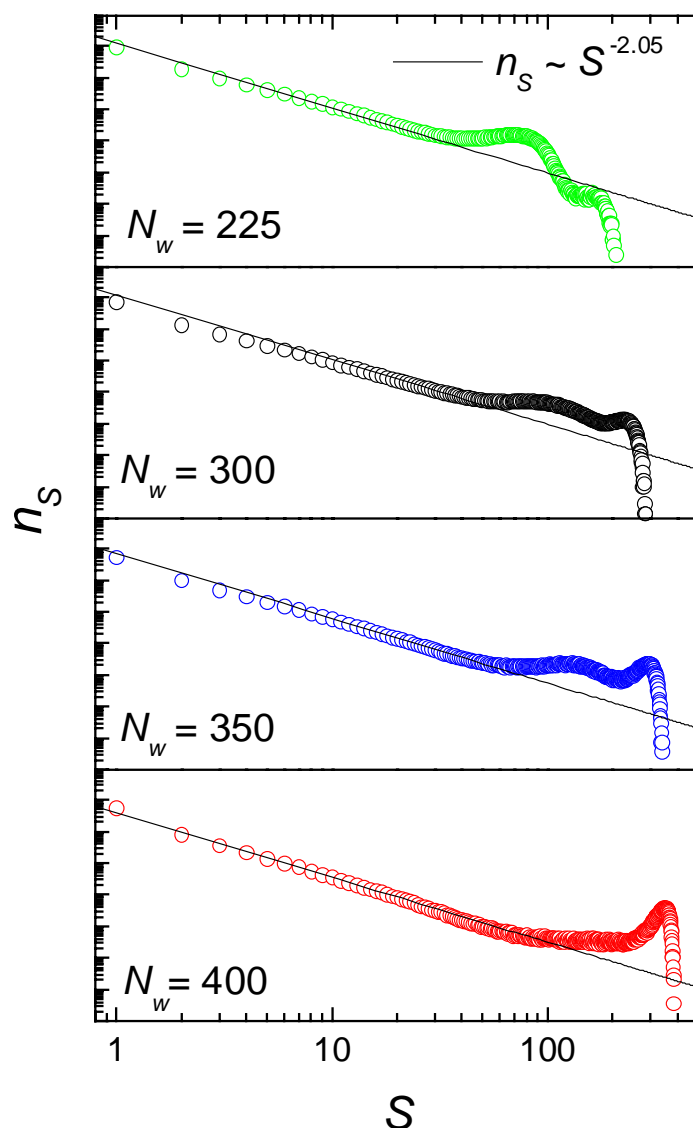


Figure 4.6: Probability distributions n_S of clusters with S water molecules at the surface of a hydrophilic sphere of radius $R_{\text{sp}} = 15 \text{ \AA}$ at $T = 425 \text{ K}$. The surface coverage increases from top to bottom. At $N_w = 350$, a spanning cluster exists with a probability of about 50%. At $N_w = 400$, the fractal dimension of the largest cluster is close to the 2D percolation threshold value.

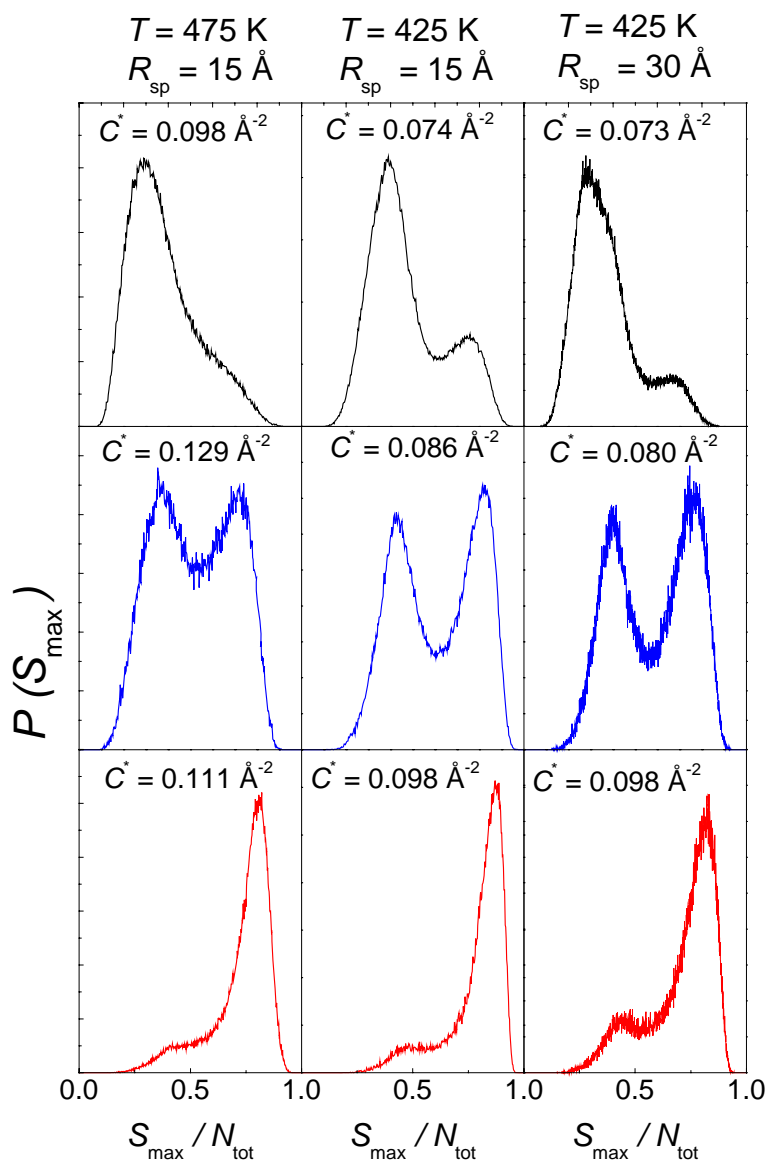


Figure 4.7: Probability distribution $P(S_{\max})$ of the size of the largest water cluster S_{\max} at the surfaces of hydrophilic spheres with the radii $R_{\text{sp}} = 15$ and 30 \AA at various hydration levels and temperatures. Upper panel: below the percolation threshold. Middle panel: the probability of a spanning cluster is about 50%. Lower panel: 2D percolation threshold.

The positions of the two peaks in the distribution $P(S_{\max})$ give the most probable size of the largest non-spanning and spanning clusters. These positions could be estimated visually, or determined from a fit of $P(S_{\max})$. For example, $P(S_{\max})$ could be fitted by a sum of Gaussians:

$$P(S_{\max}) = \sum \frac{a_i}{w_i} \exp\left(-\frac{2(S_{\max} - S_{\max,i})^2}{w_i^2}\right), \quad (4.1)$$

where $S_{\max,i}$, w_i , a_i , are the position, the half-width and the amplitude of the i -th peak, respectively. It is not possible to fit the obtained distributions $P(S_{\max})$ perfectly by two Gaussians, because the right side of the second peak decreases sharply due to the truncation of the spanning cluster on the finite (spherical) surface. Neglecting the steep decay of $P(S_{\max})$ to the right of the second peak and using two Gaussians in Equation 4.1, we found that $S_{\max,2}$ is about twice $S_{\max,1}$ for all spheres studied. In Figure 4.9, we show such a fit of the distribution $P(S_{\max})$ for the largest sphere, imposing $S_{\max,2} = 2S_{\max,1}$ (see Figure 4.9, dot-dashed line). The restriction $S_{\max,2} = 2S_{\max,1}$ also allows to fit the complete distribution $P(S_{\max})$ perfectly by three Gaussians (see Figure 4.9, red line). Summarizing, the most probable size of a spanning water cluster on a spherical surface is about twice the largest non-spanning cluster at any hydration level where both are observable. This observation has not yet been discussed in any percolation theory, but can be seen also in some previously published studies (Refs. 202, 203).

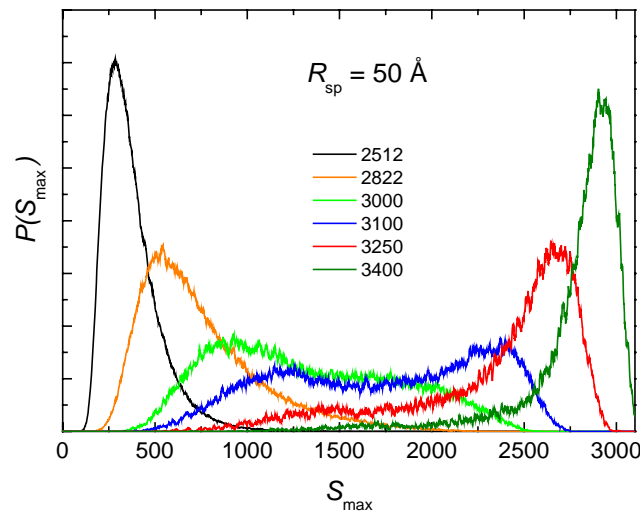


Figure 4.8: Probability distribution $P(S_{\max})$ of the largest water cluster at the surface of a hydrophilic sphere with radius $R_{\text{sp}} = 50 \text{ \AA}$ at $T = 425 \text{ K}$. The numbers of water molecules N_w are given in the legend. The blue line: probability of a spanning water cluster of about 50%. Red line: 2D percolation threshold.

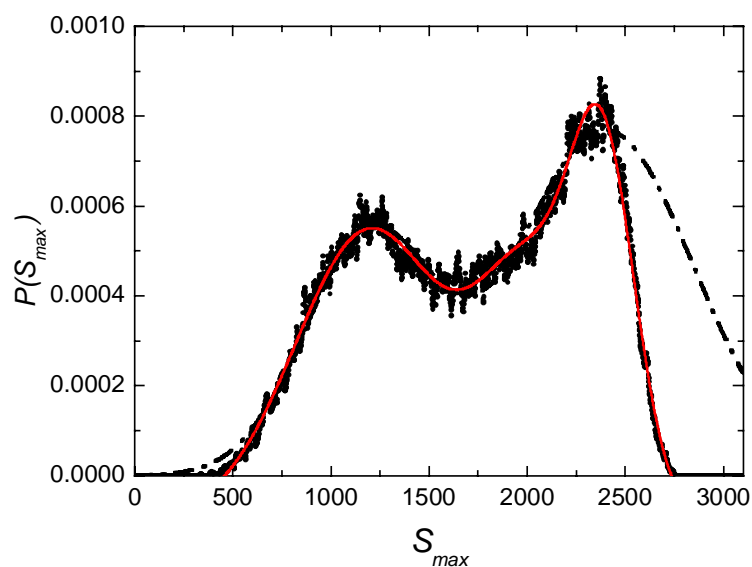


Figure 4.9: Probability distribution $P(S_{\max})$ of the largest water cluster at the surface of the hydrophilic sphere with radius $R_{\text{sp}} = 50 \text{ \AA}$ at $T = 425 \text{ K}$ and hydration level $N_w = 3100$ (symbols). The lines show fits to Equation 2 (see text for the details).

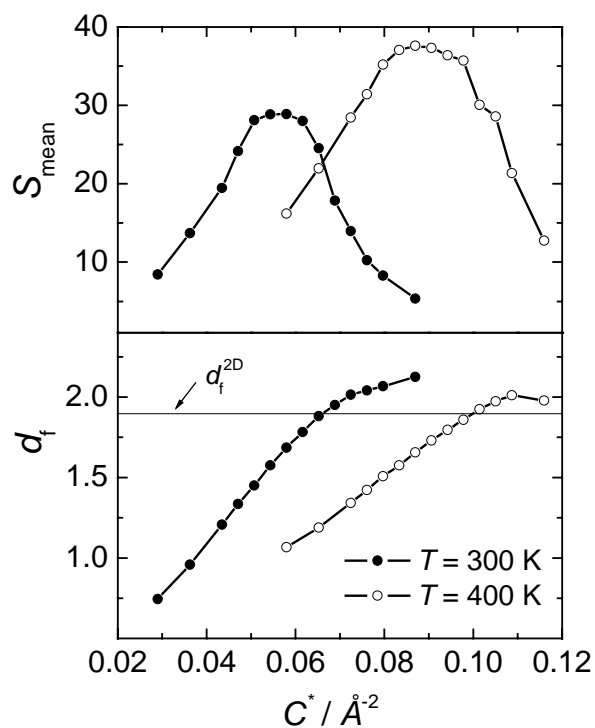


Figure 4.10: Mean cluster size S_{mean} (upper panel) and fractal dimension d_f of the largest water cluster (lower panel) on the surface of a single lysozyme molecule as a function of the surface coverage.

4.2.3 Single Lysozyme Molecule

The clustering of water on the surface of a single lysozyme molecule was analyzed along the lines employed to hydrophilic spherical surfaces. The fractal dimension d_f of the largest cluster approaches the value d_f^{2D} of the 2D percolation threshold approximately at the same surface coverage (Figure 4.10), when the cluster size distribution n_s obeys the power law in the largest range of cluster sizes (Figure 4.11, red circles). This gives the following estimations of the minimum water coverage, which enables the persistent existence of a spanning water network at the surface of the lysozyme molecule: $N_w = 450$ ($C_2^*(d_f) = 0.065 \text{ \AA}^{-2}$) at $T = 300 \text{ K}$ and $N_w = 690$ ($C_2^*(d_f) = 0.100 \text{ \AA}^{-2}$) at $T = 400 \text{ K}$ (Table 4.2).

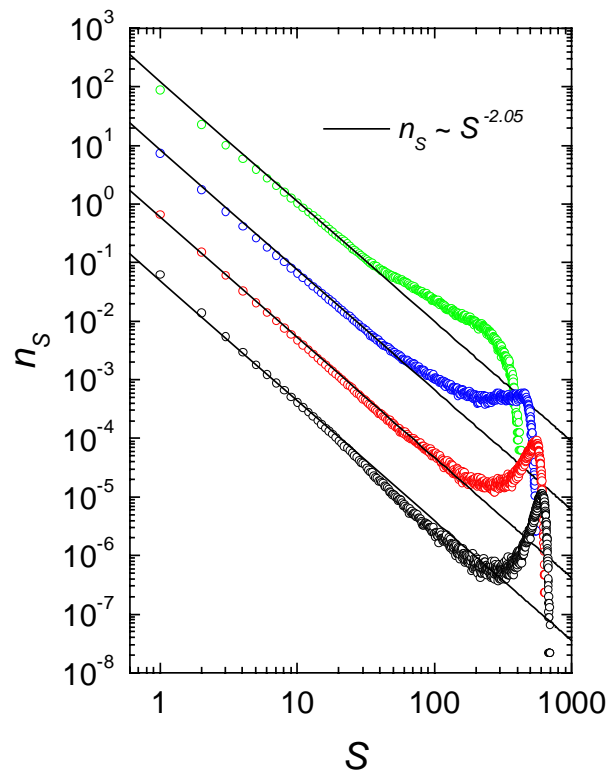


Figure 4.11: Distributions n_s of clusters with S water molecules at the surface of a single lysozyme molecule at $T = 400 \text{ K}$. The surface coverage increases from $N_w = 525$ (top) to 750 (bottom). Blue circles: $N_w = 625$, a spanning network exists with a probability of about 50%. Red circles: $N_w = 700$, the fractal dimension d_f of the largest cluster is close to the 2D percolation threshold value. The distributions are shifted consecutively by one order of magnitude each, starting from the bottom.

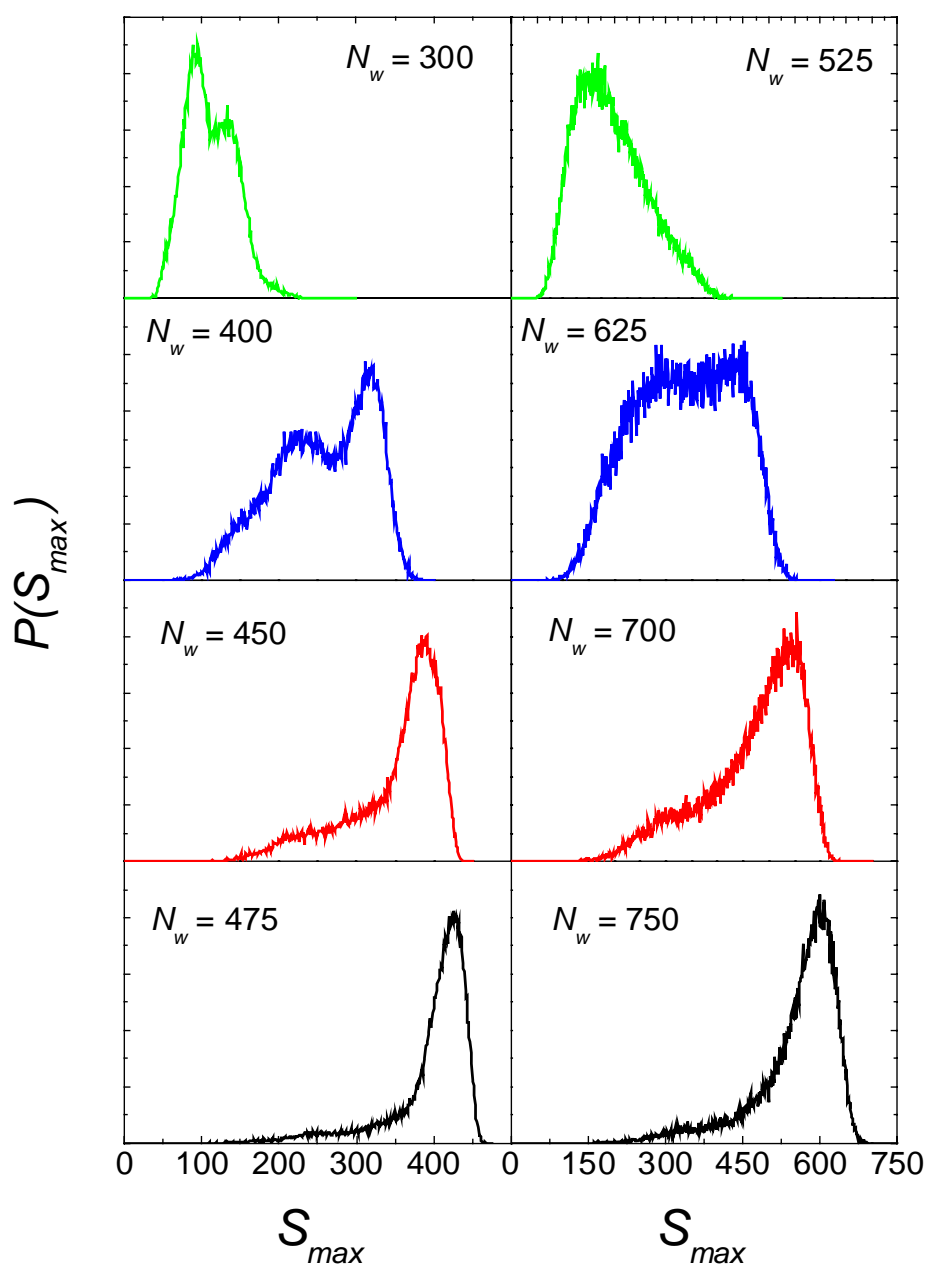


Figure 4.12: Probability distribution $P(S_{\max})$ of the largest water cluster S_{\max} on the surface of a single lysozyme molecule at $T = 300$ K (left panel) and 400 K (right panel) and various hydration levels. Green lines: below the percolation threshold; blue lines: probability of a spanning cluster is about 50%; red lines: 2D percolation threshold; black lines: above the percolation threshold.

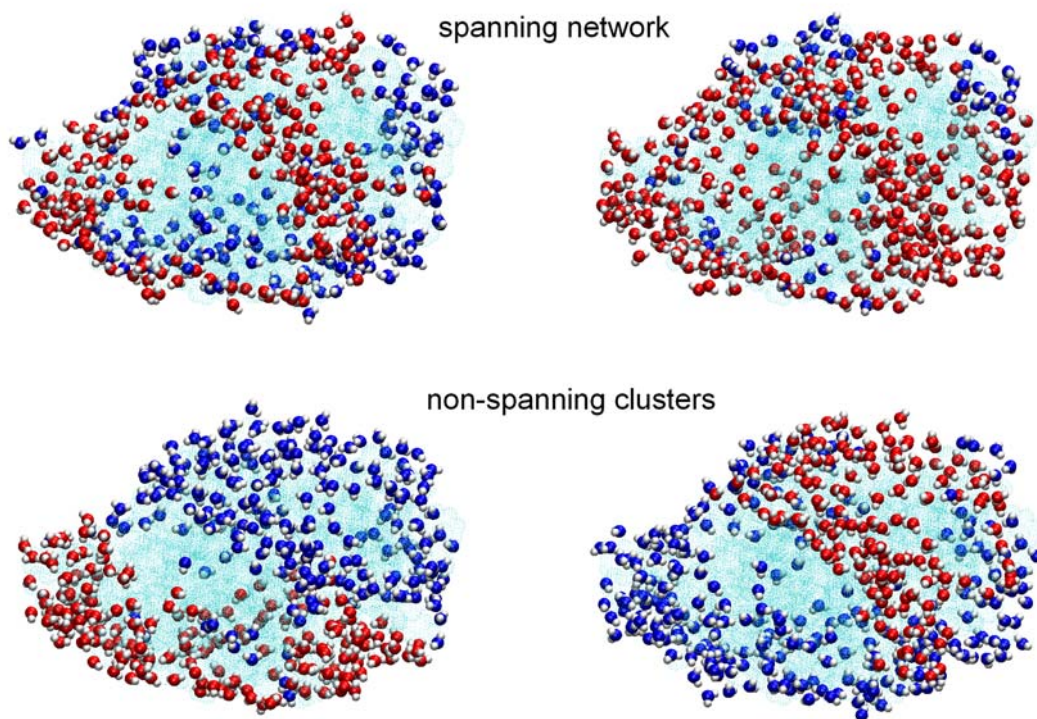


Figure 4.13: Arrangement of the water molecules on the surface of a lysozyme molecule at the hydration level where the probability to find a spanning water network is about 50% ($T = 300$ K, $N_w = 400$). The oxygen atoms of the water molecules, which belong to the largest cluster, are colored in red, those of all other water molecules in blue. Examples of spanning and non-spanning water clusters are shown in the upper and lower panel, respectively.

The mean cluster size S_{mean} passes through a maximum at slightly lower hydration levels: $C_1^*(S_{\text{mean}}) = 0.056$ and 0.089 \AA^{-2} at $T = 300$ and 400 K, respectively (Figure 4.10, upper panel). The evolution of the distribution $P(S_{\text{max}})$ with the hydration level evidences a two-peak structure, which appears as two distinct maxima at $T = 300$ K, and which is still visible at $T = 400$ K (Figure 4.12). The behaviour of $P(S_{\text{max}})$ indicates that the probability for the existence of a spanning water network is about 50% when $N_w \approx 400$ ($C_1^*(S_{\text{max}}) = 0.058 \text{ \AA}^{-2}$) at $T = 300$ K and $N_w \approx 625$ ($C_1^*(S_{\text{max}}) = 0.091 \text{ \AA}^{-2}$) at $T = 400$ K. The ratio of the most probable sizes of spanning $S_{\text{max},2}$ and non-spanning $S_{\text{max},1}$ largest clusters, estimated from the positions of the two peaks of $P(S_{\text{max}})$, is about 1.6 for both temperatures. This value is smaller than the value ~ 2 obtained for spherical hydrophilic surfaces, probably due to the non-spherical shape of lysozyme and/or the inhomogeneous distribution of the hydrophilic sites on the protein surface.

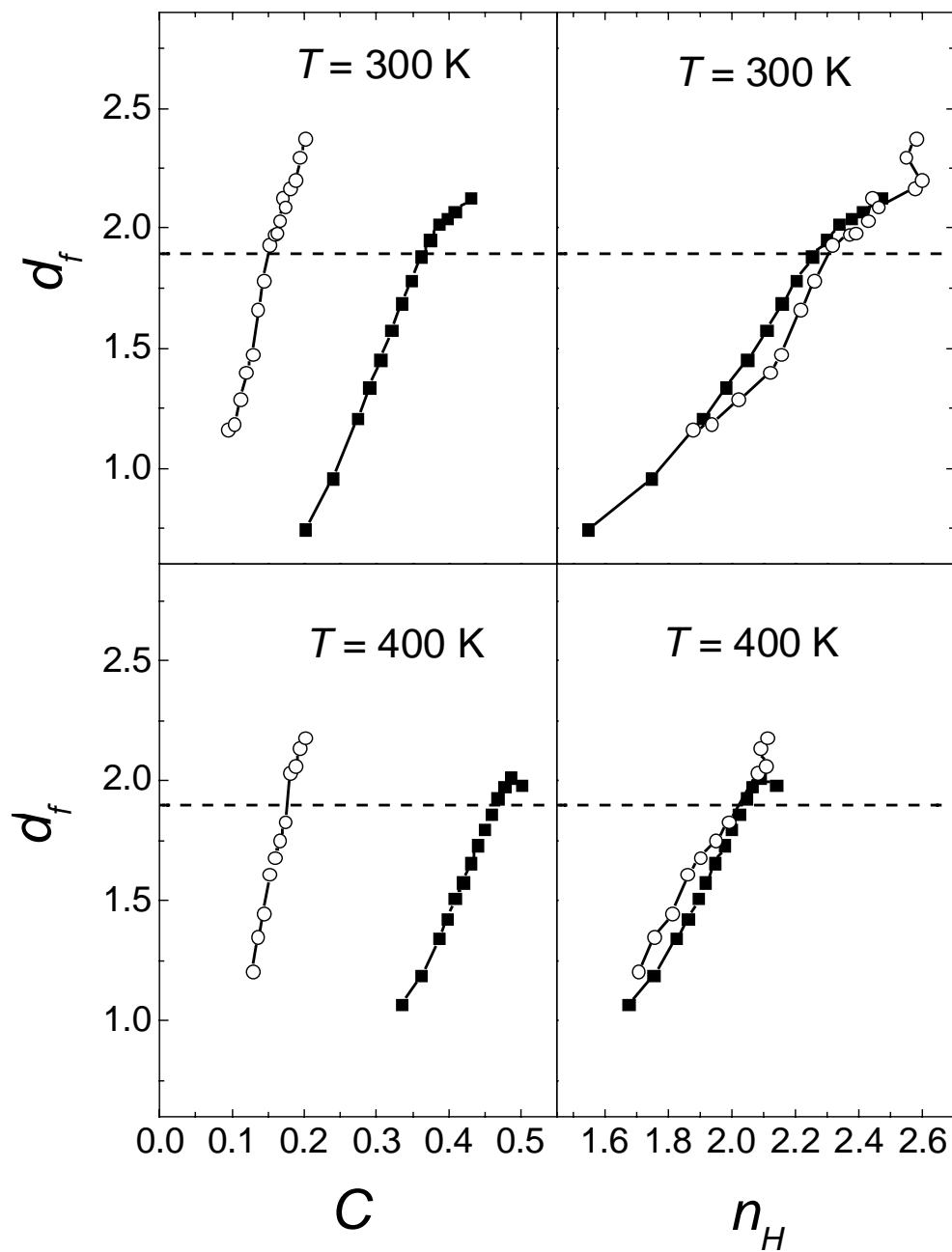


Figure 4.14: Fractal dimension of the largest water clusters as a function of the water mass fraction (left panel) and as a function of the average number of H-bonds n_H (right panel): single lysozyme (solid squares) and densely packed powder (open circles) at $T = 300$ and 400 K. Dashed lines: $d_f^{2D} = 1.896$.

When the two peaks of $P(S_{\max})$ have comparable heights, the surface water exists with equal probability in two quite different states: with a spanning network and without it. Examples of these two kinds of water ordering scenarios are depicted in Figure 4.13. In the case of a spanning water network (Figure 4.13, upper part), the largest cluster envelopes the whole lysozyme molecule. The non-spanning largest water cluster is usually attached to some strongly hydrophilic part of the lysozyme (for example, to the left-bottom or to the cleft in the middle part of the molecule) (Figure 4.13, lower panel). Because of the large difference between $S_{\max,2}$ and $S_{\max,1}$, strong fluctuations of the water network occur at this surface coverage. Spanning and non-spanning structures of the surface water replace each other frequently. The lifetime of the distinct types of configurations is comparable to the life-time of single water-water hydrogen bonds (~ 1.0 ps [204]).

It is interesting to compare the hydration levels, which provide the formation of a spanning water network at the surface of a single lysozyme molecule and in the model lysozyme powder. As an example, the course of the fractal dimension of the largest cluster d_f in the cases of the single lysozyme molecule and the lysozyme powder are compared in the Figure 4.14. The threshold hydration level, expressed as the mass fraction of water, C , is essentially higher in the case of a single lysozyme molecule (Figure 4.14, left panel), because in the powder the accessible surface of the lysozyme molecules decreases due to their contacts and, additionally, the water molecules could simultaneously belong to the hydration shells of two or more lysozyme molecules. Taking into account that the number of bonds per particle at the percolation threshold is rather universal and depends mainly on the system's dimensionality [59], it is reasonable to consider the threshold hydration level in terms of the average number of water-water hydrogen bonds n_H of the water molecule. The dependence of the fractal dimension of the largest cluster d_f on n_H , shown in Figure 4.14 (right panel), indicates a close coincidence of the 2D percolation thresholds in the lysozyme powder and on the surface of a single lysozyme molecule. In particular, the spanning 2D water network appears in the two systems approximately at the same values of n_H at both studied temperatures (Tables 4.1 and 4.2). So, n_H could serve as highly universal occupancy variable for such complex systems as hydrated proteins.

4.3 Properties of spanning water networks at protein surfaces

It was found that the formation of a *spanning* water network at the surface of a single protein molecule occurs in a similar way as in a protein powder. A collective infinite water network in protein powder appears at a low hydration level and covers less than half of the hydrophilic surface of each protein molecule. This means that the first appearance of biological activity may not ultimately require the existence of a *spanning* network around a single protein molecule, but rather the existence of a water network connecting some particular sites of one or several proteins. A comparison of the simulated hydration process [165] with experimental observations [10] indicates that the formation of an individual fractal-like percolating water network, which envelops each protein molecule, can be identified with the first appearance of a water “monolayer”, which restores the full internal dynamics of the protein.

Simulations of hydrated protein *powders* are possible nowadays within crude models only, making it impossible to study dynamical properties as well. On the other hand, the structure and dynamics of a *single* hydrated protein molecule and its hydration water can be studied by modern computer simulation methods, both at low hydration levels and in aqueous solution. This should be done with and without the presence of a *spanning* network of hydration water, i.e., under conditions above and below the percolation threshold. Hence, the location of the *spanning* network of hydration water is a necessary prerequisite for such studies.

It is not obvious how to distinguish *spanning* and *non-spanning* networks at the surface of a finite object. In the chapter 4.2, it was shown that conventional methods for the analysis of clustering and percolation can be used to locate a percolation transition of water at the surface of a hydrophilic sphere or protein (lysozyme) molecule. In the present chapter various topological and dynamic properties of *spanning* and *non-spanning* water networks in such systems was studied as a function of the level of hydration, temperature and object size. Moreover, several criteria how to detect the existence of a *spanning* network at the surface of a finite object in simulations are proposed.

Several properties of the largest clusters S_{\max} were studied at various hydration levels. The size distribution of the largest clusters $P(S_{\max})$, studied in the previous chapter, allows to distinguish spanning and non-spanning largest clusters. The linear extension of the largest

cluster is characterized by the maximum distance L_{\max} between two oxygens of water molecules in the largest cluster. The compactness of the largest clusters can be measured by the radius of gyration R_g :

$$R_g = \sqrt{\frac{\sum_{i=1}^N m_i (\vec{r}_i - \vec{r}_0)^2}{\sum_{i=1}^N m_i}}, \quad (4.2)$$

where m_i is the mass of water molecule, \vec{r}_i is a vector that defines the position of the i -th water molecule relative to the center of a sphere or to the geometrical center of lysozyme, whereas $\vec{r}_i - \vec{r}_0$ is its distance to the center of mass of the largest cluster located at $\vec{r}_0 = \frac{\sum_i m_i \vec{r}_i}{\sum_i m_i}$. The spanning properties of the largest cluster can also be described by the position of the center of mass of the largest cluster H_{\max} relative to the center of mass of the sphere or the protein:

$$H_{\max} = \sqrt{\vec{r}_0^2}. \quad (4.3)$$

The structure of the hydration water in a surface layer was described using oxygen-oxygen radial distribution functions $g_{O-O}(r)$ of water as a function of the hydration level. The lifetime of water hydrogen bonds and the spanning water network were also analysed using conventional methods [204].

In a finite system (with periodic or open boundary conditions) and in a closed system without boundaries (such as the surface of a finite object), the percolation transition is smeared out. As a result, contrary to infinite systems, various properties of clusters indicate the percolation threshold at slightly different hydration levels. In particular, close to the percolation threshold, the probability distribution $P(S_{\max})$ of the size S_{\max} of the largest water cluster shows a pronounced two-peak structure. The small-size peak of $P(S_{\max})$ represents non-spanning (finite) largest clusters, while the large-size peak is due to spanning (infinite) clusters. When the two peaks are of comparable height, the probability R to observe a spanning cluster is about 50%. At this hydration level (denoted as C_1), the mean cluster size S_{mean} , calculated excluding the largest cluster, approximately reaches its maximum. The fractal dimension d_f of the largest cluster approaches its 2D threshold value of $d_f^{2D} \approx 1.896$ at the slightly higher hydration level C_2 . Roughly at the same hydration level, the cluster size distribution n_s obeys the power law behaviour $n_s \sim S^{-2.05}$ in the widest range of cluster sizes. The hydration level C_2 corresponds to the minimum water coverage, which enables the

persistent existence of a spanning water network at the surface. More details of the definitions of the hydration levels C_1 and C_2 can be found in our previous chapter.

It is reasonable to perform the comparison of various properties of spanning and non-spanning water networks at the hydration level C_1 , where both kinds of largest clusters exist with equal probability. This threshold hydration level was determined in chapter 4.2: at $T = 425$ K, C_1 is about 0.086 \AA^{-2} ($N_w \approx 350$), 0.088 \AA^{-2} ($N_w \approx 1200$), and 0.087 \AA^{-2} ($N_w \approx 3100$) for the studied hydrophilic spheres of radius 15, 30 and 50 \AA , respectively. The threshold hydration level C_1 of a single lysozyme molecule was found to be about 0.058 \AA^{-2} ($N_w \approx 400$) at $T = 300$ K and 0.091 \AA^{-2} ($N_w \approx 625$) at $T = 400$ K. The probability distributions $P(S_{\max})$ of size S_{\max} of the largest water cluster on the surfaces of three hydrophilic spheres at the hydration level C_1 are compared in Figure 4.15. The two-peak structure of $P(S_{\max})$ is caused by the finite size of the system and is expected to vanish with increasing surface area. However, Figure 4.15 evidences that the two peaks are weakly sensitive to the surface area and remain pronounced and well-separated even at the surface of a very large sphere (radius $R_{\text{sp}} = 50 \text{ \AA}$) with a surface area of more than $35\,000 \text{ \AA}^2$. Taking into account that the two-peak structure of $P(S_{\max})$ disappears at significantly smaller surface areas in the case of a planar surface with periodic boundary conditions [200], we may conclude that the specific closed surface topology of spherical surfaces enhances such a two-peak structure. This effect also appears in the larger distance between the positions of the peaks of $P(S_{\max})$ in the case of spherical surfaces in comparison with planar surfaces. Namely, in the latter case the average spanning cluster is about 1.6 times larger than the average non-spanning largest cluster [200], whereas at the spherical surface this ratio is about 2 [165]. Such peculiarity of a spherical surface allows a clear separation of the spanning and non-spanning networks, which is very useful for their comparative analysis, as will be seen below. Examples of the arrangement of water molecules in the cases of spanning and non-spanning largest water clusters on a spherical hydrophilic surface are shown in Figure 4.16.

The probability distributions of the maximum linear extension L_{\max} of the largest water cluster $P(L_{\max})$ are shown in Figure 4.17 as a function of L_{\max} , normalized to the effective diameter of a hydrophilic sphere, $2(R_{\text{sp}} + 3 \text{ \AA})$, which accounts for the distance of about 3 \AA from the surface to the location of the water oxygens in the first hydration layer. For clusters with all water molecules in the first hydration shell, the value of $L_{\max}/2(R_{\text{sp}} + 3 \text{ \AA})$ does not exceed 1. Figure 4.17 evidences that, even at low hydration levels, the largest clusters extend through the essential part of the spherical surface. The radial distribution function $g_{\text{O-O}}(r)$ of

surface water at smooth hydrophilic surfaces shows a specific maximum at $r \approx 5.4 \text{ \AA}$, indicating the presence of chain-like water structures [184]. Due to such structure formation, the largest cluster is ramified, and even at low surface coverage ($C \sim 0.065 \text{ \AA}^{-2}$) its extension is comparable with the diameter of the sphere. At the surface coverage C_1 , when the probability R to observe a spanning cluster is about 50%, for the vast majority of the largest clusters (both spanning and non-spanning), L_{\max} exceeds $2(R_{\text{sp}} + 3 \text{ \AA})$ (the blue lines in Figure 4.17). At larger surface coverage, L_{\max} noticeably exceeds $2(R_{\text{sp}} + 3 \text{ \AA})$ because the largest cluster includes water molecules, which do not belong to the first hydration shell.

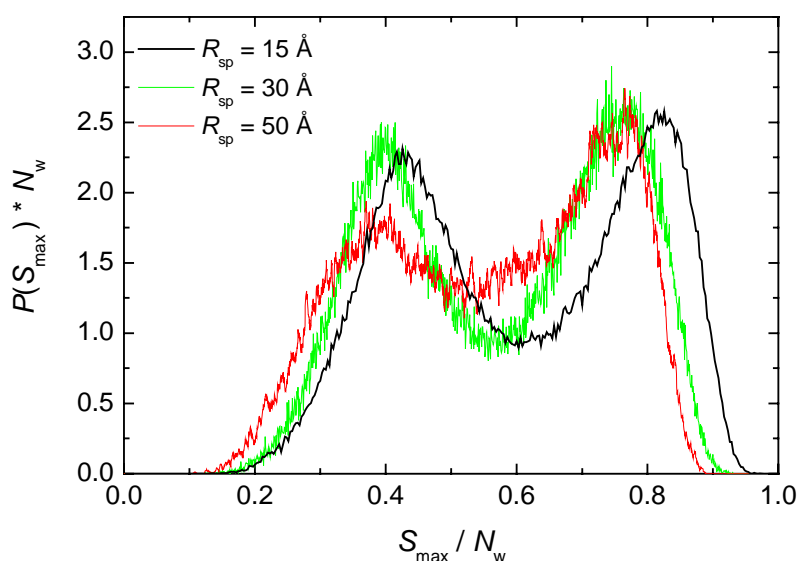


Figure 4.15: Probability distribution $P(S_{\max})$ of the size S_{\max} of the largest water cluster normalized to the total number of water molecules N_w on the surface of hydrophilic spheres with radii $R_{\text{sp}} = 15$, 30 and 50 \AA at the hydration level C_1 , where the probability to find a spanning water network is about 50%.

To analyse the behaviour of L_{\max} separately for spanning and non-spanning largest clusters, one has to classify each largest cluster of the size S_{\max} using the two-peak structure of the distribution $P(S_{\max})$. For this purpose, the joint probability distributions $P(L_{\max}, S_{\max})$ were calculated for various hydration levels. For the spherical surface with $R_{\text{sp}} = 30 \text{ \AA}$, $P(L_{\max}, S_{\max})$, obtained at the threshold hydration level C_1 , is shown in Figure 4.18. The two clearly separated peaks in $P(L_{\max}, S_{\max})$ correspond to the spanning and non-spanning largest clusters. The projection of this probability distribution on the plane $L_{\max} S_{\max}$ is shown in the upper panel of Figure 4.19 (shading is proportional to probability density). The two shadowed areas in Figure 4.19 (upper panel) correspond to the two peaks in Figure 4.18: the left-hand

and right-hand areas represent the largest non-spanning and spanning water clusters, respectively. It is clearly seen from Figs. 4.18 and 4.19 that these two peaks of $P(L_{\max}, S_{\max})$ are hardly resolved, if considering the projection of $P(L_{\max}, S_{\max})$ on L_{\max} axis. Indeed, the non-spanning largest clusters can be characterized by the average value of the maximum linear extension L_{\max} of about 70 Å, whereas for the spanning clusters $L_{\max} \approx 72$ Å (Figure 4.19, upper panel). These values correspond to $L_{\max}/2(R_{\text{sp}} + 3 \text{ Å}) = 1.06$ and 1.09 for the largest non-spanning and spanning cluster, respectively, and they can be hardly distinguished as a faint shoulder in Figure 4.17 (lower panel, blue line). This difference becomes totally indistinguishable for smaller spheres (Figure 4.17, upper panel, blue line).

A similar analysis of the maximum extension L_{\max} of the largest water cluster was performed for the hydrated lysozyme molecule. The evolution of the probability distribution $P(L_{\max})$ with increasing hydration level is shown in Figure 4.20. Contrary to the spherical surface, the two maxima in $P(L_{\max})$ can be seen in a wide range of hydration. With increasing temperature, $P(L_{\max})$ becomes smoother, but the two maxima are still pronounced (Figure 4.20, lower panel). Similarly to the smooth surface, the radial distribution function $g_{\text{O-O}}(r)$ of water near the lysozyme surface shows a specific maximum at $r \approx 5.4$ Å, reflecting its chain-like structure in a wide range of hydration levels.

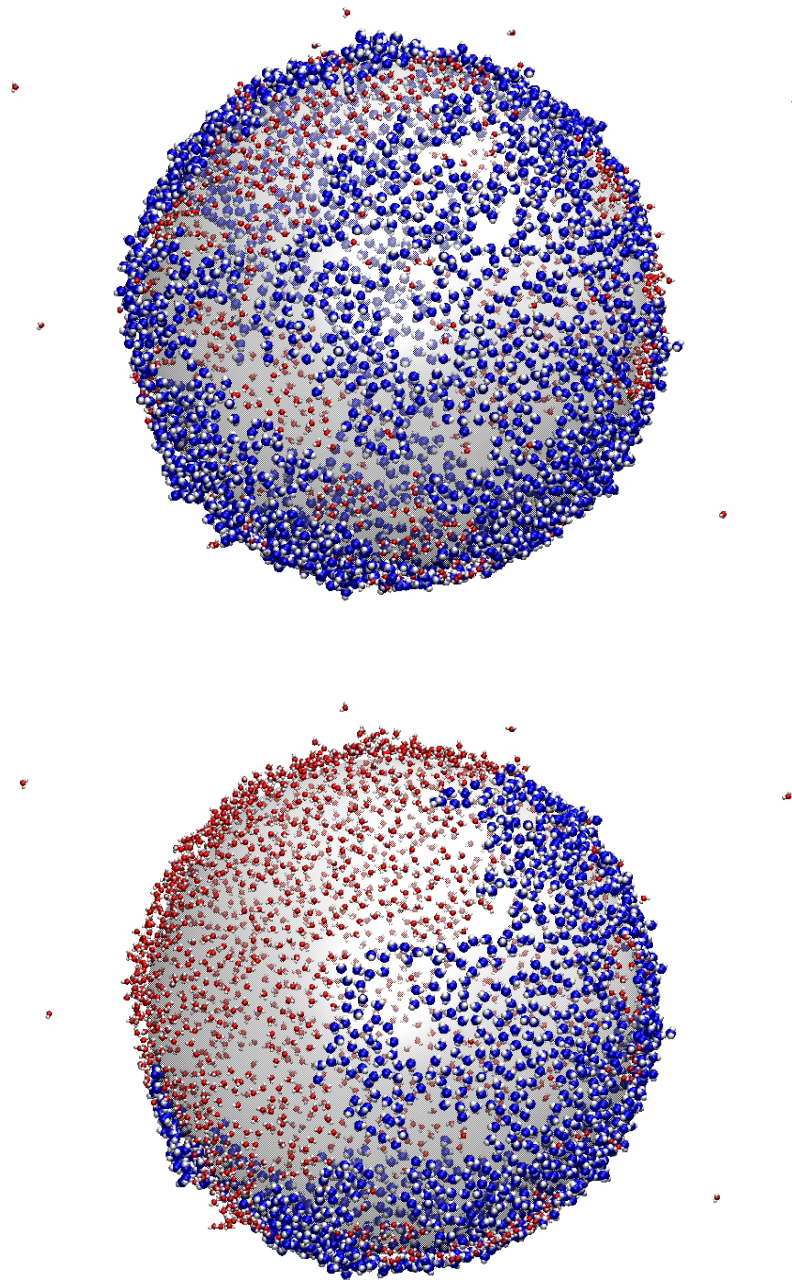


Figure 4.16: Arrangement of water molecules on the surface of a (transparent) hydrophilic sphere of radius $R_{\text{sp}} = 50 \text{ \AA}$ at the hydration level C_l , where the probability to find a spanning water network is about 50% ($T = 425 \text{ K}$, $N_w = 3100$). The oxygen atoms of the water molecules that belong to the largest cluster are colored in blue, those of all other water molecules in red. An example of a spanning and a non-spanning largest water cluster is shown in the upper and lower panel, respectively.

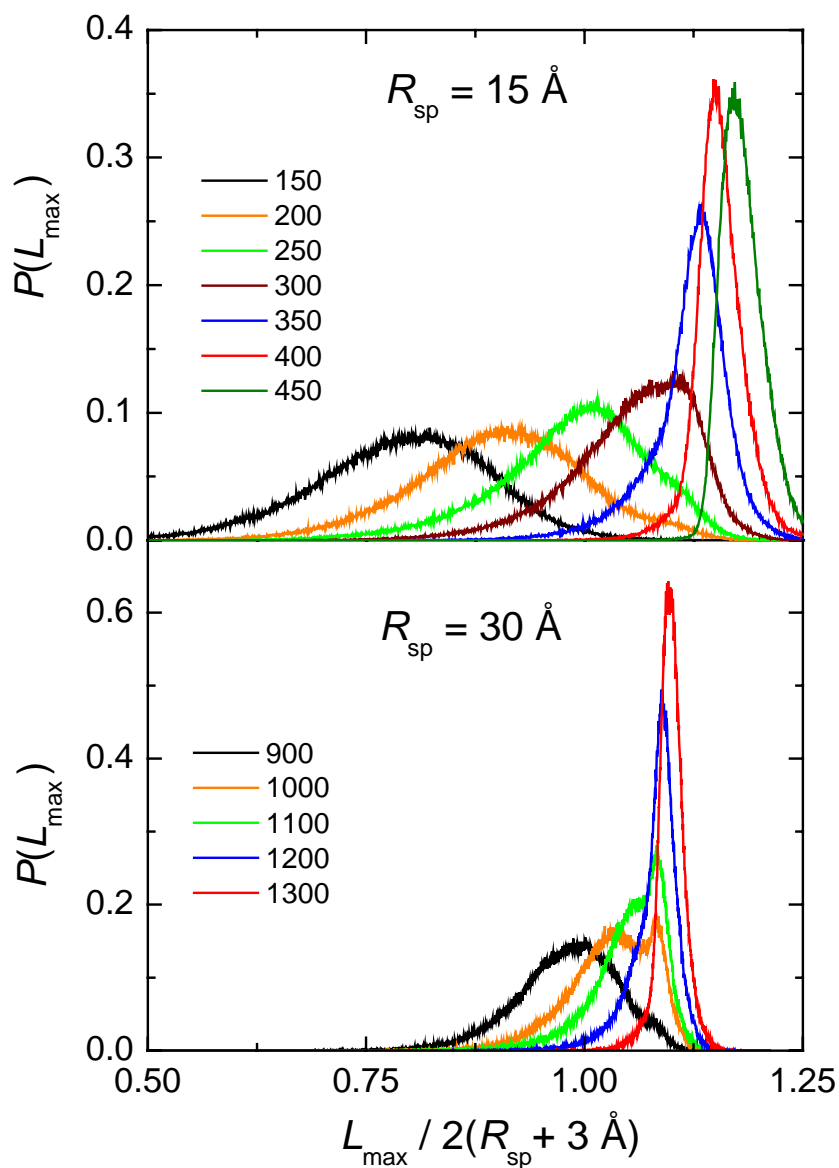


Figure 4.17: Probability distribution $P(L_{\max})$ of the maximum linear extension L_{\max} of the largest water cluster at the surface of two hydrophilic spheres ($R_{sp} = 15$ and 30 \AA) at $T = 425 \text{ K}$ and various hydration levels N_w , given in the inset. The blue and red lines approximately correspond to the surface coverages C_1 and C_2 , respectively.

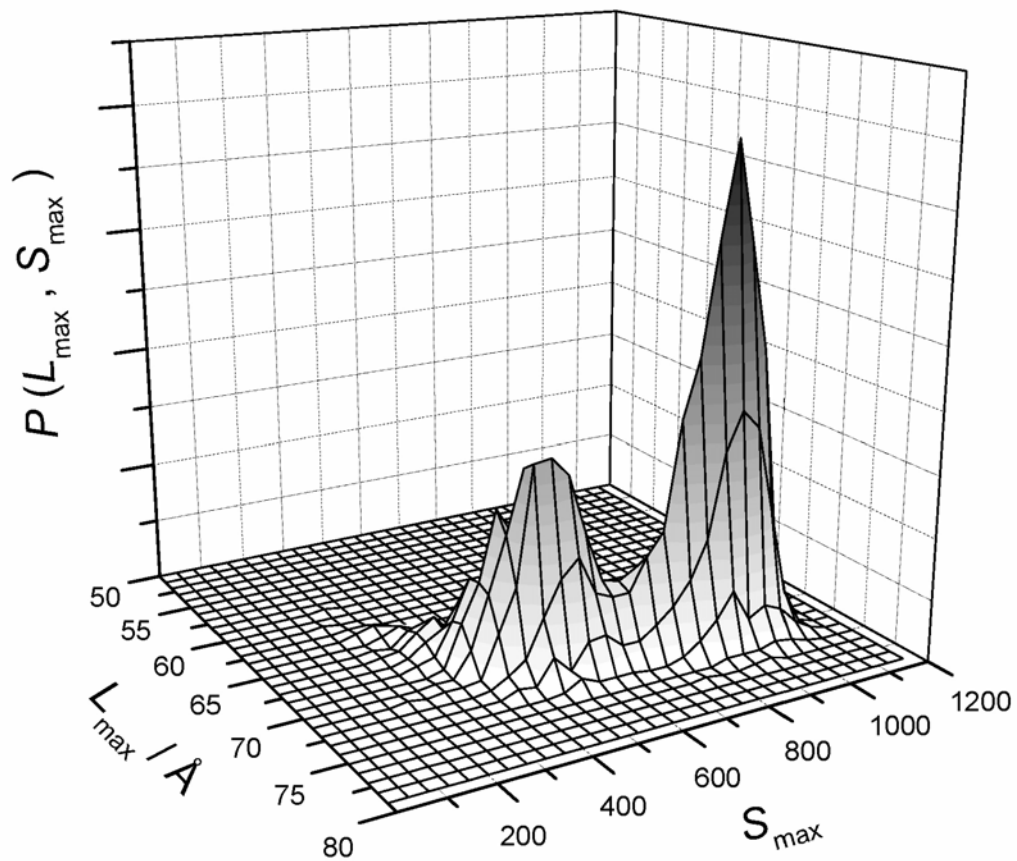


Figure 4.18: Joint probability distribution $P(L_{\max}, S_{\max})$ of the maximum linear extension L_{\max} and size S_{\max} of the largest water cluster at the spherical surface of radius $R_{\text{sp}} = 30 \text{ \AA}$ at $T = 425 \text{ K}$ and hydration level $C_1 \approx 0.088 \text{ \AA}^{-2}$ ($N_w = 1200$).

The projection of the joint probability distribution $P(L_{\max}, S_{\max})$, calculated at the threshold hydration level C_1 for two temperatures, is shown in Figure 4.19. The two dark areas in both the middle and lower panels in Figure 4.19 can be used to distinguish spanning and non-spanning largest clusters at each temperature. Figures 4.19 and 4.20 indicate that the average maximum extension L_{\max} of non-spanning clusters at the lysozyme surface is about 40 \AA , whereas the average value of $L_{\max} \approx 55 \text{ \AA}$ can be attributed to spanning clusters.

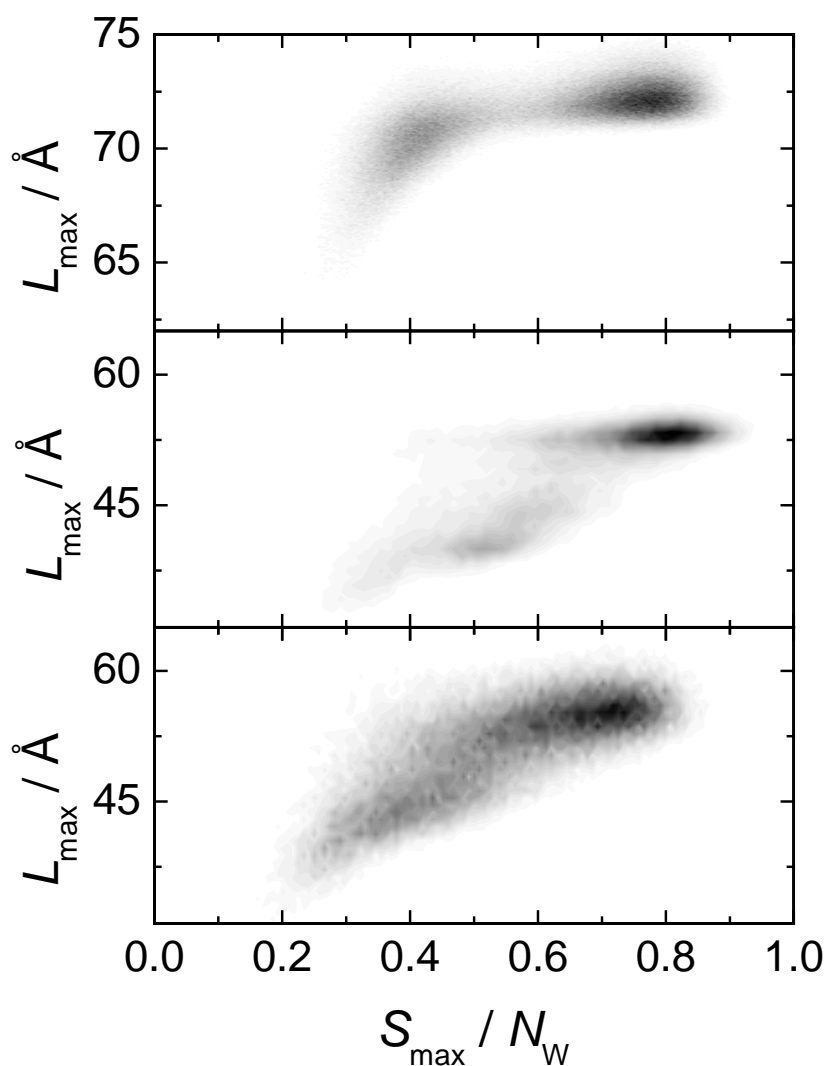


Figure 4.19: Projection of the joint probability distribution $P(L_{\max}, S_{\max})$ of the linear extension L_{\max} and size S_{\max} of the largest water cluster at the surface of a hydrophilic sphere of $R_{\text{sp}} = 30 \text{ \AA}$ ($N_w = 1200$) at $T = 425$ K and on the surface of lysozyme at $T = 300$ K (middle panel, $N_w = 400$) and $T = 400$ K (bottom panel, $N_w = 625$). The hydration level corresponds to the threshold hydration level C_1 in all cases. Shading is proportional to the probability density. Each panel has its own (proper) probability scale.

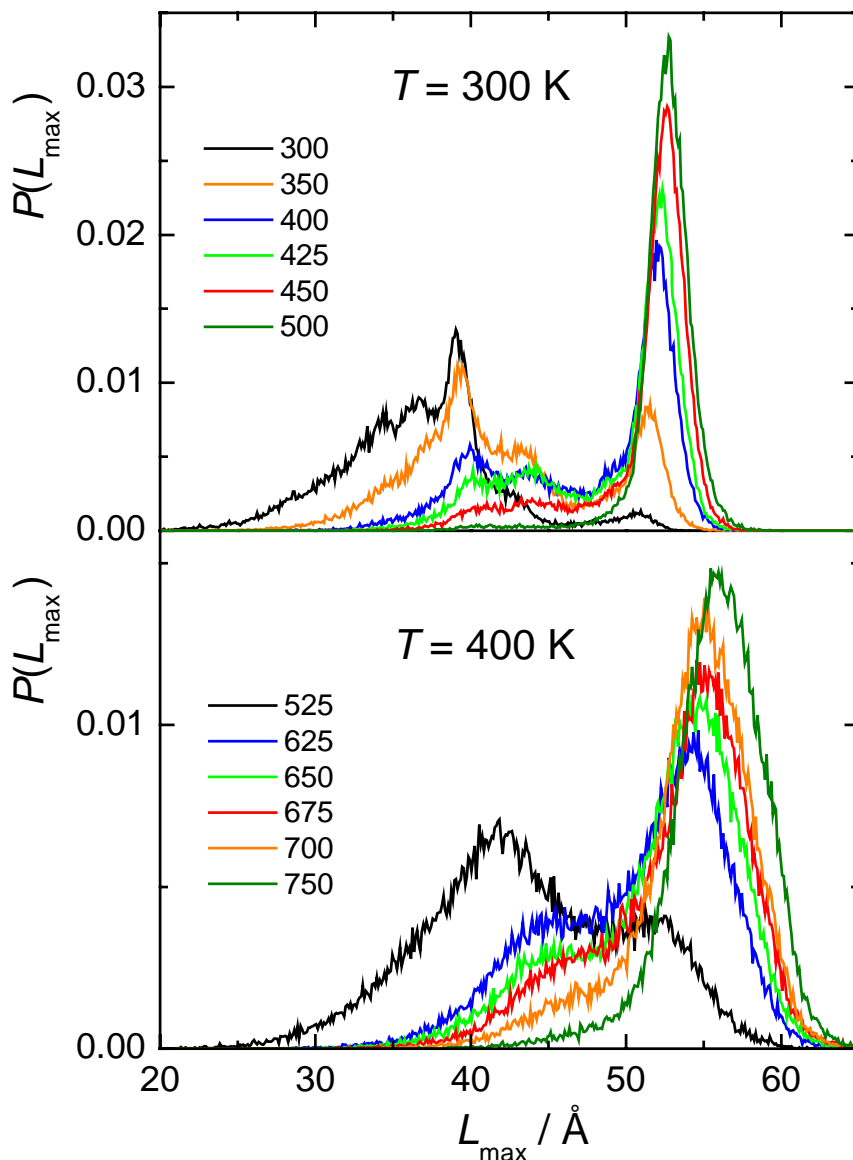


Figure 4.20: Probability distribution of the maximum linear extension L_{\max} of the largest water cluster at the surface of lysozyme at two temperatures and various hydration levels (N_w is given in the inset). The blue and red lines approximately correspond to the threshold surface coverages C_1 and C_2 , respectively.

The noticeable difference between the average values of L_{\max} for spanning and non-spanning clusters on the lysozyme surface is obviously determined by the topology of the protein molecule. It could be due to the non-spherical (ellipsoid-like) shape of the globular protein or due to the specific arrangement of hydrophilic sites on the lysozyme surface. If the first factor is dominant, the difference in L_{\max} could be used to separate spanning and non-

spanning water clusters in computer simulations of hydrated globular proteins. The effect of a particular shape of the protein molecule on the distribution $P(L_{\max})$ deserves further studies.

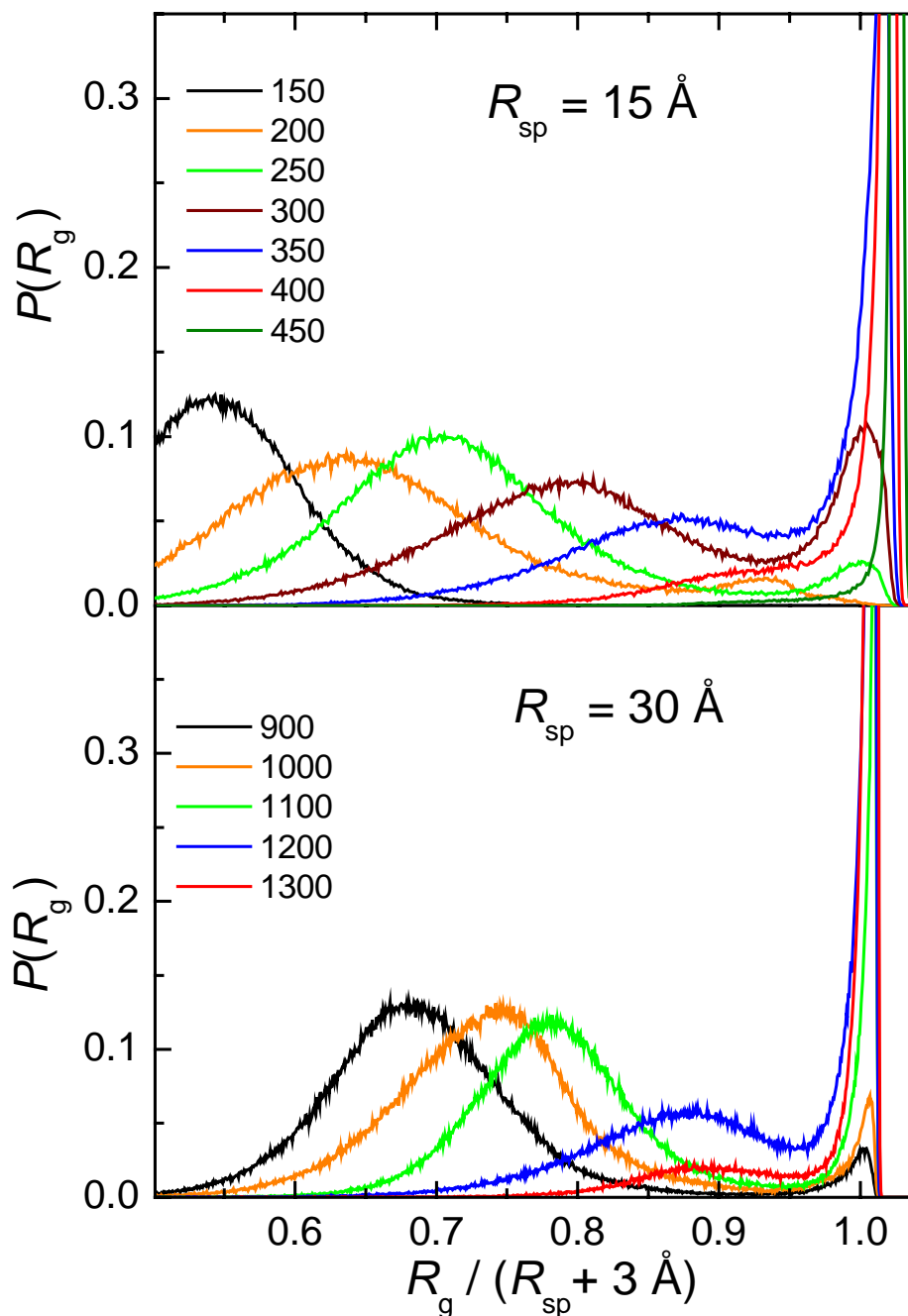


Figure 4.21: Probability distributions $P(R_g)$ of the radius of gyration R_g of the largest water cluster at the surfaces of two hydrophilic spheres of $R_{sp} = 15, 30 \text{ \AA}$ at $T = 425 \text{ K}$ and hydration levels N_w given in the inset. Blue and red lines approximately correspond to the surface coverage C_1 and C_2 , respectively.

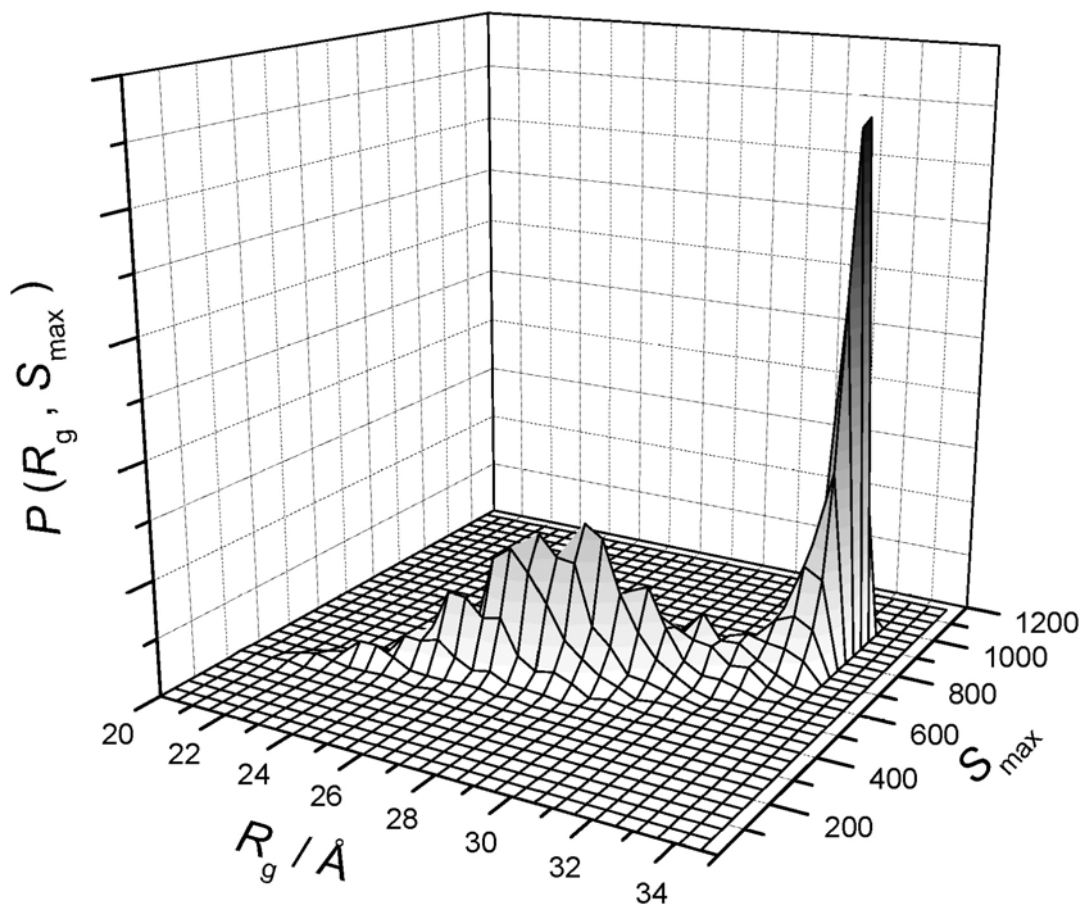


Figure 4.22: Joint probability distributions $P(R_g, S_{\max})$ of the radius of gyration R_g and size S_{\max} of the largest water cluster at the spherical surface of radius $R_{\text{sp}} = 30 \text{\AA}$ at $T = 425 \text{ K}$ and hydration level $C_1 \approx 0.088 \text{\AA}^{-2}$ ($N_w = 1200$).

The radius of gyration R_g of the largest water cluster was analyzed in a similar way as its maximum extension L_{\max} . Figure 4.21 evidences a pronounced two-peak structure of the probability distribution $P(R_g)$ of the radius of gyration in a wide range of hydration levels. This could mean that spanning and non-spanning networks are characterized by very different values of R_g . To check this possibility, the joint probability distribution $P(R_g, S_{\max})$ of the radius of gyration R_g and the size of the largest cluster S_{\max} were calculated, which is shown in Figure 8 for a sphere of radius $R_{\text{sp}} = 30 \text{\AA}$ and the threshold hydration level C_1 . Indeed, the sharp peak of $P(R_g, S_{\max})$ at large S_{\max} values evidences that the radius of gyration of a spanning cluster is close to the effective radius of a sphere ($R_{\text{sp}} + 3 \text{\AA}$). The low and wide peak, positioned at smaller S_{\max} in Figure 4.22, corresponds to non-spanning clusters. Figure 4.21 indicates that R_g of the spanning clusters is not sensitive to the hydration level, whereas R_g of the non-spanning largest clusters continuously increases with increasing hydration level.

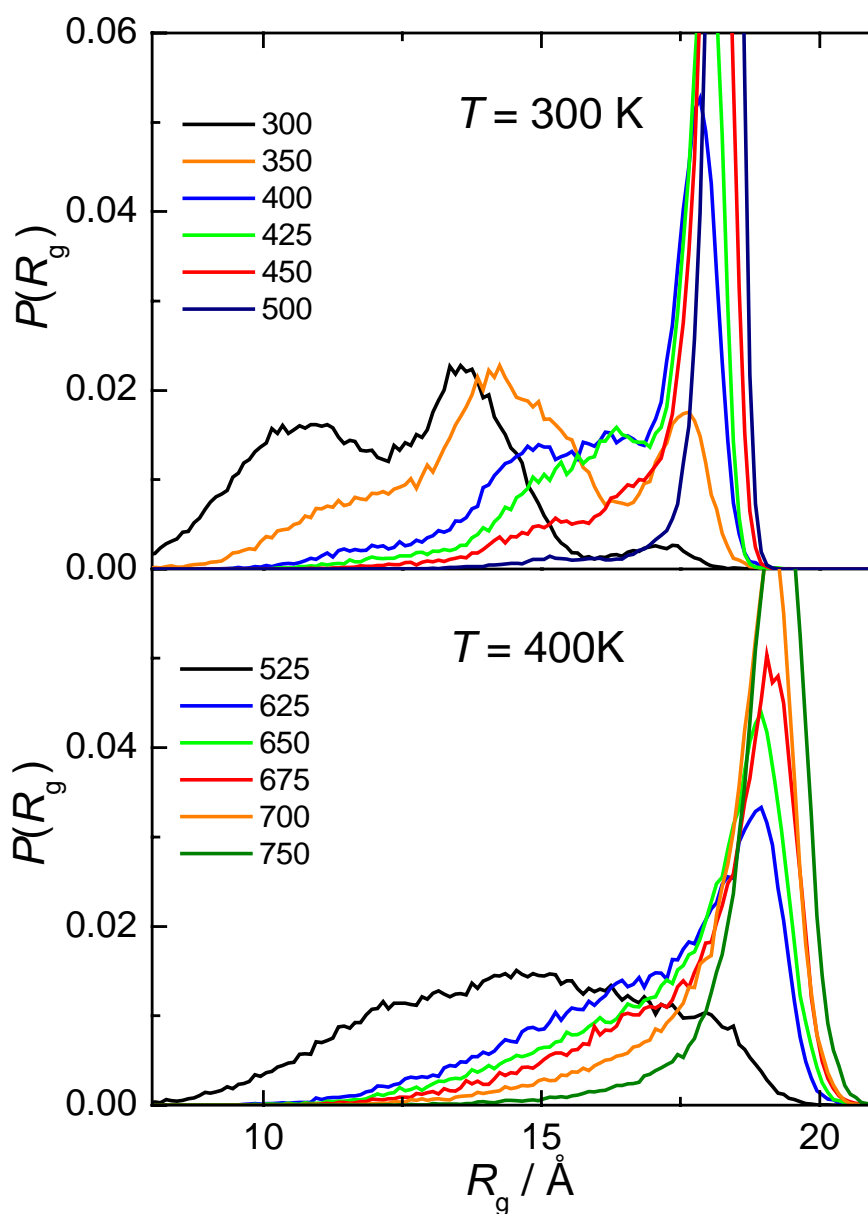


Figure 4.23: Probability distributions $P(R_g)$ of the radius of gyration R_g of the largest water cluster on the surface of lysozyme at two temperatures and various hydration levels (N_w is given in the inset). The blue and red lines approximately correspond to the surface coverage C_1 and C_2 , respectively.

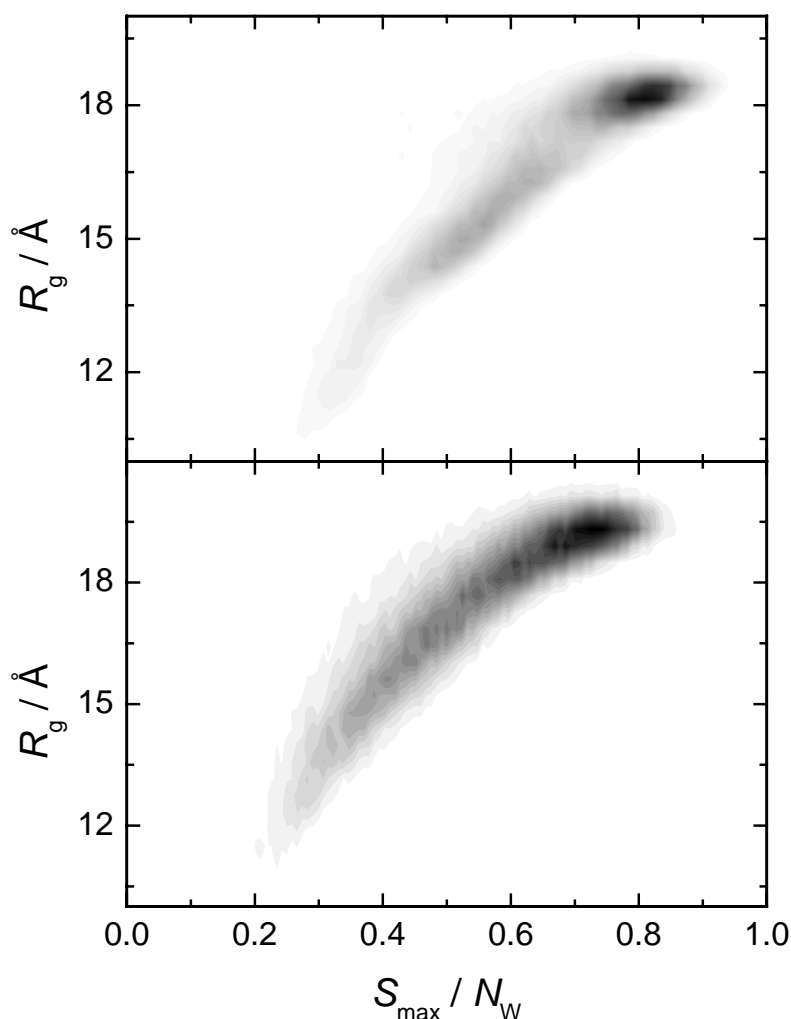


Figure 4.24: Projection of the joint probability distribution $P(R_g, S_{\max})$ of the radius of gyration R_g and size the S_{\max} of the largest water cluster at the surfaces of lysozyme at the threshold surface coverage C_1 at $T = 300$ K (upper panel $N_w = 400$) and $T = 400$ K (bottom panel $N_w = 625$). Shading is proportional to probability density. Each panel has its own (proper) probability scale.

The behavior of the radius of gyration R_g of the largest water cluster on the surface of lysozyme is quite similar to that observed at smooth hydrophilic spheres. The change of the probability distribution $P(R_g)$ of the radius of gyration of the largest cluster on the surface of lysozyme with increasing hydration level is shown in Figure 4.23. The correlation between the radius of gyration R_g and the size of the largest cluster S_{\max} could be analysed using the joint probability distribution $P(R_g, S_{\max})$ (Figure 4.24). This correlation allows one to consider separately the radius of gyration for spanning and non-spanning clusters. The average value of R_g of the non-spanning largest cluster continuously increases with the hydration level (Figure

4.23). R_g of the spanning water cluster is practically independent of the hydration level and close to $R_g \approx 18 \text{ \AA}$ at $T = 300 \text{ K}$ and $R_g \approx 19 \text{ \AA}$ at $T = 400 \text{ K}$ (Figures 4.23 and 4.24). Such universality makes the radius of gyration the appropriate indicator of the spanning character of the largest water cluster.

The radius of gyration of the simulated model lysozyme molecule is about 14 \AA . For a homogeneous sphere of radius R , the radius of gyration is given by $R_g = \sqrt{3/5} \cdot R$. Thus, for a spherical lysozyme molecule with a uniformly distributed mass, the radius should be about 18 \AA . The system of homogeneously distributed water molecules on the smooth surface of a sphere with such a radius should have a radius of gyration of $R_g \approx 18 \text{ \AA} + 3 \text{ \AA} = 21 \text{ \AA}$ (see Figure 4.21 and its discussion above). The obtained lower values of R_g for the spanning water network at the lysozyme surface are obviously due to the non-spherical shape of the lysozyme molecule and the non-homogeneous distribution of water at the lysozyme surface. The latter factor originates from a non-regular distribution of hydrophilic residues, which form the preferential sites for adsorption of water molecules at the surface of proteins.

At low hydration levels, we observe a splitting of the first peak of $P(R_g)$, corresponding to non-spanning water networks (Figure 4.23, $T = 300 \text{ K}$). This splitting correlates with a similar behavior of the size distribution of the largest cluster $P(S_{\max})$ at low hydration levels (Figure 4.12) and obviously reflects a preferential hydration of two comparatively large hydrophilic parts of the lysozyme surface.² The importance of the particular structure of the protein surface for various properties of the largest water cluster vanishes with increasing temperature (Figure 4.23, $T = 400 \text{ K}$).

The probability distribution of the distance H_{\max} from the center of mass of the largest water cluster to the center of a sphere also shows a two-peak structure in a wide range of hydration level (see Figure 4.25). Water clusters that cover the spherical surface completely or homogeneously are represented by the sharp peak of $P(H_{\max})$ at small values of $H_{\max}/(R_{\text{sp}} + 3 \text{ \AA})$ (left peaks in Figure 4.25). The second peak, positioned at higher values of $H_{\max}/(R_{\text{sp}} + 3 \text{ \AA})$, obviously represents the non-spanning largest clusters. With decreasing water coverage, the right peak moves towards $H_{\max}/(R_{\text{sp}} + 3 \text{ \AA}) = 1$, that reflects increasing localization of the largest cluster within a small surface area.

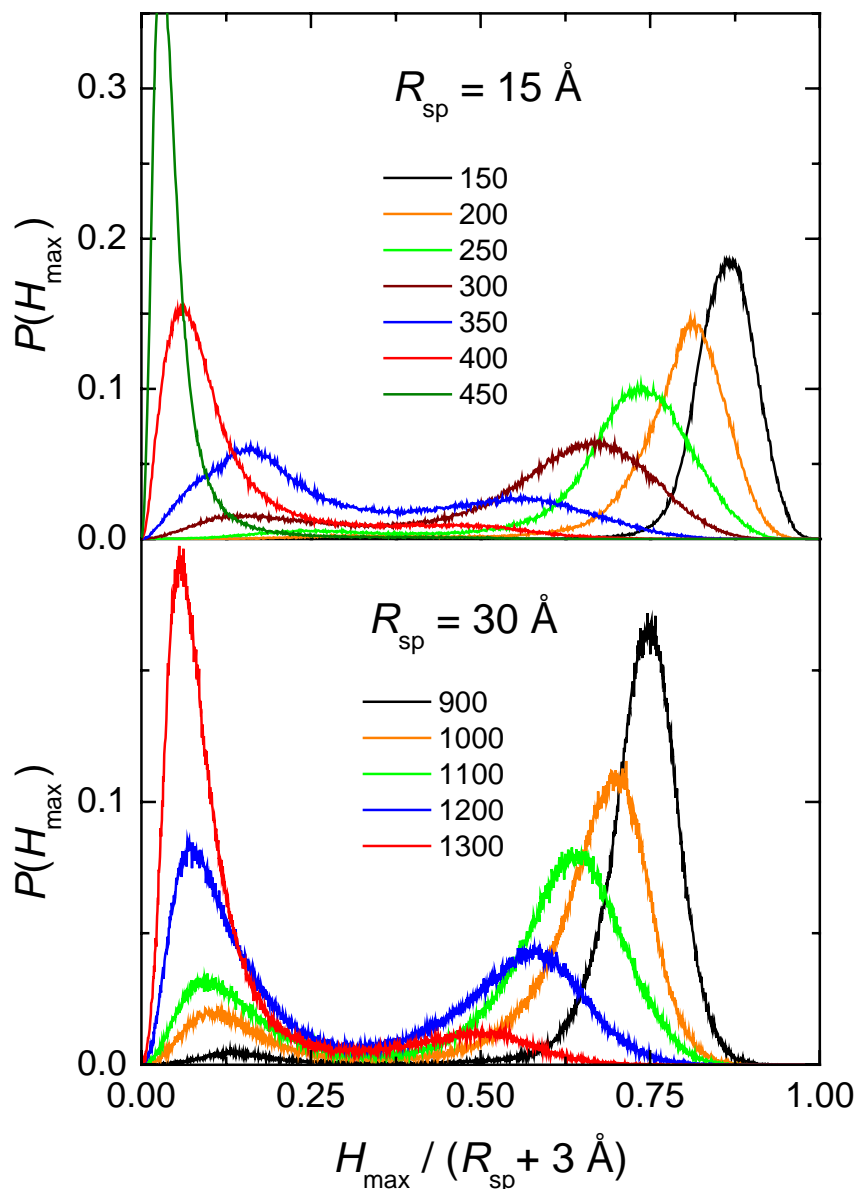


Figure 4.25: Probability distribution $P(H_{max})$ of the distance H_{max} between the center of mass of the largest cluster and the center of the sphere of radius $R_{sp} = 15$ and 30 \AA at $T = 425 \text{ K}$ and various hydration levels (N_w is given in the inset). The blue and red lines approximately correspond to the threshold surface coverages C_1 and C_2 , respectively.

For all spheres studied, a clear minimum of $P(H_{max})$ is observed near $H_{max} \approx 0.3 (R_{sp} + 3 \text{ \AA})$. Note, that the center of mass of a surface spherical segment of height h is located at the distance $R_{sp} - h/2$ from the center. Therefore, the center of mass of an infinitely thin empty hemisphere is at the distance $H_{max} \approx R_{sp}/2$. If water covers a hemisphere

homogeneously, the centre of mass of the water molecules should be at $(R_{\text{sp}} + 3 \text{ \AA}) / 2$ from the center of the sphere. So, the distributions $P(H_{\text{max}})$ at all surface coverages (Figure 4.25) evidence a low probability of the largest clusters to cover the hemisphere and slightly larger areas homogeneously. This fact could be treated as an instability of the largest clusters of such sizes. On the spherical surface of a radius $R_{\text{sp}} = 30 \text{ \AA}$, a small peak in the probability distribution $P(H_{\text{max}})$, located at $H_{\text{max}} \leq 0.1 (R_{\text{sp}} + 3 \text{ \AA})$, appears already at the surface coverage $C \approx 0.066 \text{ \AA}^{-2}$ (the black line in Figure 4.25 lower panel), which is noticeably below the percolation threshold $C_1 = 0.088 \text{ \AA}^{-2}$ (which corresponds roughly to the blue line). This observation evidences that any spanning cluster spans essentially more than half of the spherical surface. Obviously, the small largest clusters must be strongly ramified to span such a large area.

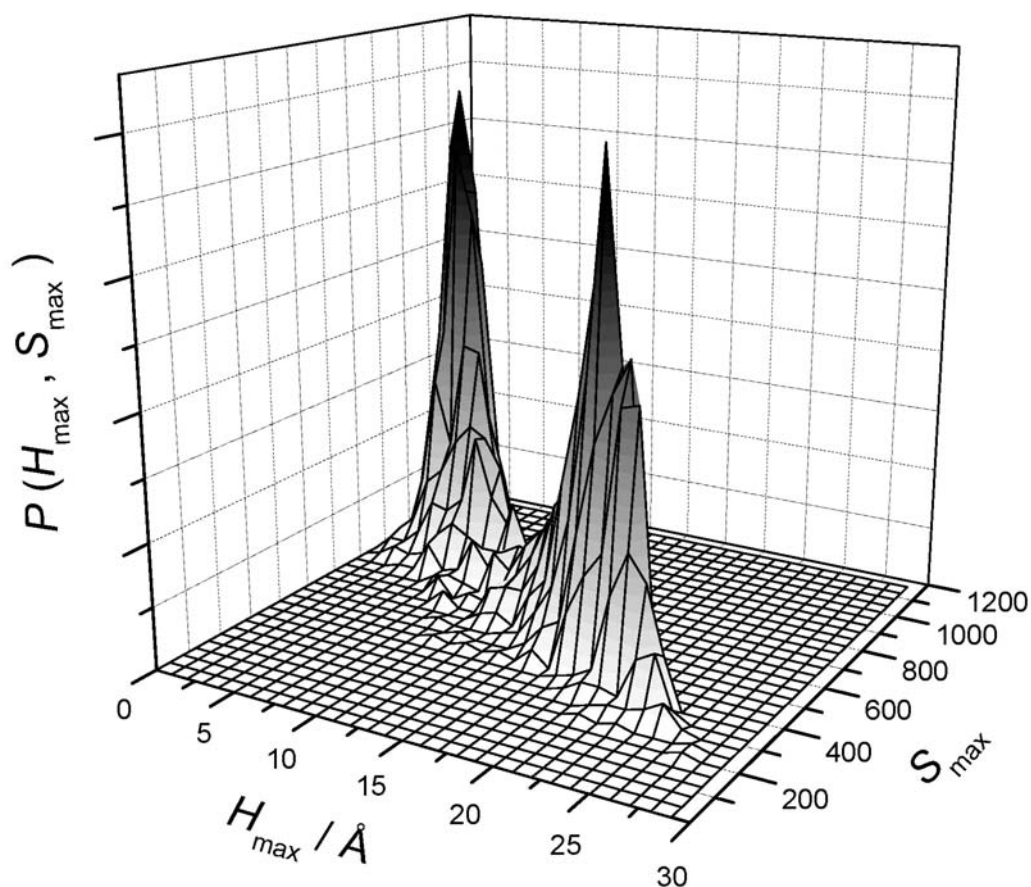


Figure 4.26: Joint probability distributions $P(H_{\text{max}}, S_{\text{max}})$ of the distance H_{max} and size S_{max} of the largest cluster at a spherical surface of radius $R_{\text{sp}} = 30 \text{ \AA}$ at $T = 425 \text{ K}$ and hydration level $C_1 \approx 0.088 \text{ \AA}^{-2}$ ($N_w = 1200$).

The joint probability distribution $P(H_{\text{max}}, S_{\text{max}})$ of the distance H_{max} and the size of the largest cluster S_{max} at the threshold hydration level C_1 is shown in Figure 4.26 for a sphere of

radius $R_{\text{sp}} = 30 \text{ \AA}$. Two sharp peaks, separated by a deep well, evidence that the calculation of the distance H_{max} is an appropriate parameter to be used for the detection of spanning clusters on the surface of a finite object.

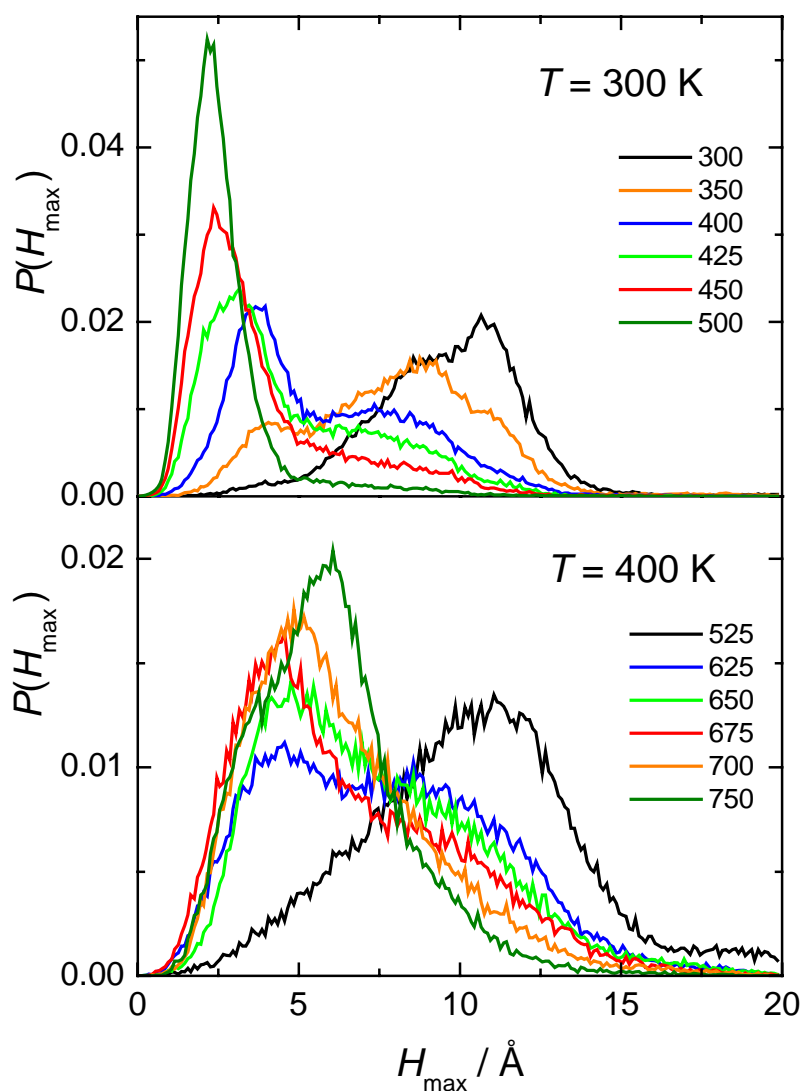


Figure 4.27: Probability distributions $P(H_{\text{max}})$ of the distance H_{max} between the center of mass of lysozyme and the center of mass of the largest water cluster at the surface of lysozyme at two temperatures and various hydration levels (N_w is given in the inset). The blue and red lines approximately correspond to the threshold surface coverages C_1 and C_2 , respectively.

The probability distributions $P(H_{\text{max}})$, calculated for the largest water cluster on the surface of a lysozyme molecule at various hydration levels, are shown in Figure 4.27. The evolution of $P(H_{\text{max}})$ with increasing hydration level reflects the existence of two states of the largest clusters. The shallow minimum of $P(H_{\text{max}})$ between two states at about 6 \AA at 300 K

transforms into an isosbestic-like point at 400 K (Figure 4.27). This behavior differs from that observed for smooth spheres, where two peaks are separated much more clearly (Figure 4.25). Obviously, the more complex shape of the lysozyme molecule and the chemical heterogeneity of its surface smears out the two-peak structure of $P(H_{\max})$. The most probable value of H_{\max} for spanning clusters at the lysozyme surface is about 3 Å at $T = 300$ K and about 5 Å at $T = 400$ K. The minimum of $P(H_{\max})$ corresponds to the most unstable largest clusters. The value of H_{\max} at the minimum could serve as a boundary between spanning and non-spanning clusters, which is about 6 and 7.5 Å at $T = 300$ and 400 K, respectively. These values are in good agreement with the results for clusters at spherical surfaces. Indeed, if the lysozyme molecule is replaced by an effective sphere of radius 18 Å (see above), one may expect the minimum of $P(H_{\max})$ to occur at $H_{\max} \approx 0.3(18 + 3) \text{ Å} \approx 6.3 \text{ Å}$.

Finally, the correlations between the two most significant properties, R_g and H_{\max} of the largest water cluster were analysed, which could be used as sensitive indicators of the presence of a spanning water network. Correlations between H_{\max} and R_g obtained from the calculation of the joint probability distribution $P(H_{\max}, R_g)$ for the three studied spherical surfaces are shown in Figure 4.28. Little scatter of the data points around a parabolic-like dependence R_g on H_{\max} indicates a strong correlation of these two parameters. R_g and H_{\max} , defined by Equations 4.2 and 4.3, respectively, are related to each other by the following equation:

$$R_g^2 = \frac{\sum \vec{r}_i^2}{N_w} - H_{\max}^2 \quad (4.4)$$

If all N_w molecules are at the same distance $R_{sp} + 3 \text{ Å}$ from the center of a sphere, $|\vec{r}_i^2|$ is simply equal to $N_w(R_{sp} + 3 \text{ Å})^2$, i.e.

$$R_g = \sqrt{(R_{sp} + 3 \text{ Å})^2 - H_{\max}^2} \quad (4.5)$$

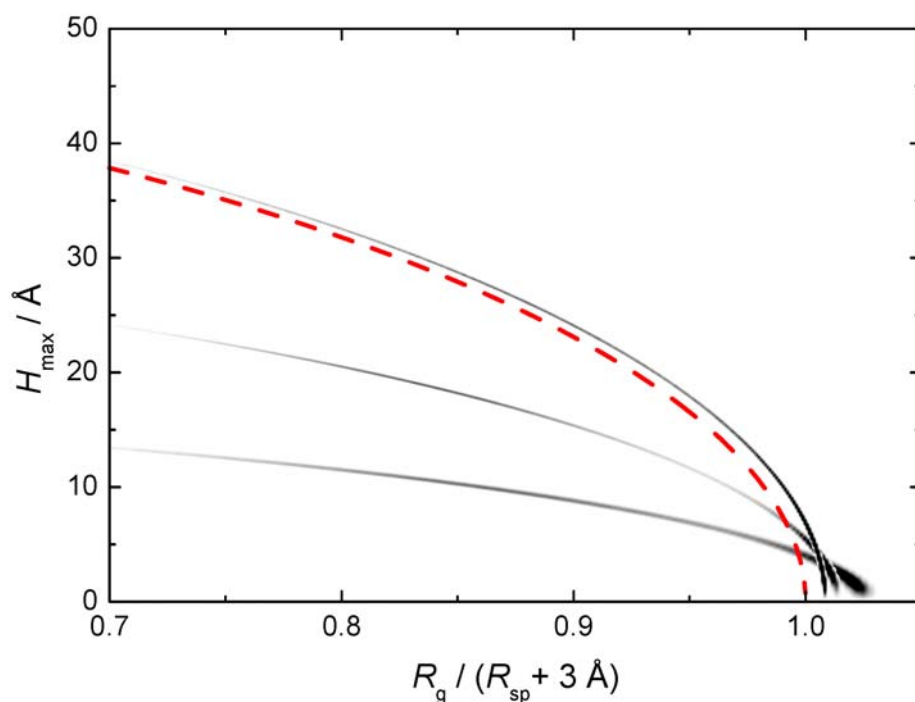


Figure 4.28: Correlation between the distance H_{\max} and the radius of gyration R_g of the largest water cluster at spherical surfaces obtained from the joint probability distribution $P(H_{\max}, R_g)$. The radius of the spherical surface, $R_{\text{sp}} = 15, 30$ and 50 \AA , increases from the bottom to the top. The red dashed line shows the dependence $R_g(H_{\max})$ as obtained from Equation 4.5 for a sphere of $R_{\text{sp}} = 50 \text{ \AA}$. The hydration levels approximately correspond to the threshold surface coverages C_1 in each system studied. Shading is proportional to the probability density.

In the case of small values of H_{\max} , corresponding to spanning clusters, the dependence $R_g(H_{\max})$, described by Equation 4.4, is close to parabolic. For large H_{\max} (non-spanning largest clusters) it approaches a linear dependence. The Equation 4.5 for a spherical surface of radius $R_{\text{sp}} = 50 \text{ \AA}$ is shown by a dashed line in Figure 4.28. Deviations of the calculated correlation between R_g and H_{\max} from Equation 4.5 can be attributed to the fact that some water molecules in the largest cluster are out of the first water monolayer. The latter effect is especially pronounced in the case of a large spanning cluster (H_{\max} close to zero).

The correlation between H_{\max} and R_g of the largest water cluster at the surface of the lysozyme molecule, obtained from the calculation of the joint probability distribution $P(H_{\max}, R_g)$, is shown in Figure 4.29. In general, this correlation looks qualitatively similar to the case of water at the spherical surface: H_{\max} decreases, when the radius of gyration increases. The spanning and non-spanning largest clusters can be clearly separated at both temperatures. However, the spanning and non-spanning clusters obviously show different correlations between H_{\max} and R_g . This reflects the non-spherical shape of the lysozyme

molecule and the non-homogeneity of its surface. Hence, contrary to an ideal spherical surface, H_{\max} and R_g are not as closely related in the case of protein molecules.

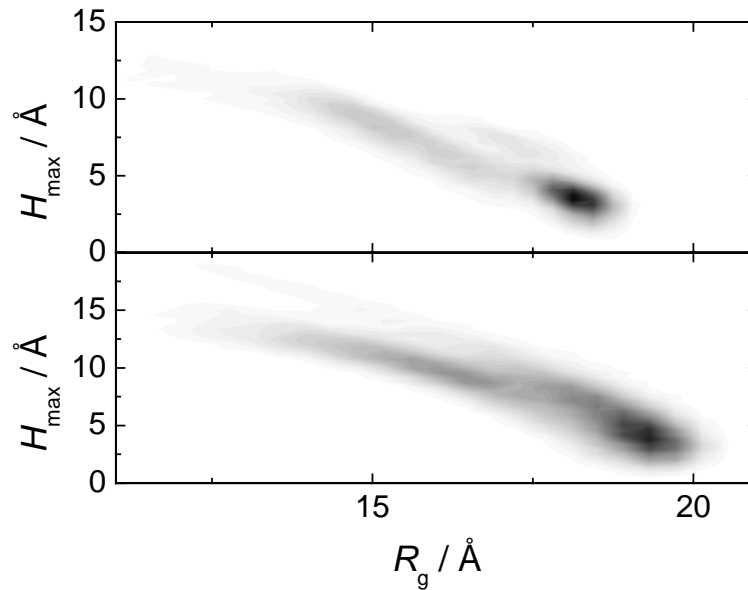


Figure 4.29: Joint probability distribution $P(H_{\max}, R_g)$ of the distance H_{\max} and radius of gyration R_g of the largest water cluster at the surface of lysozyme at the threshold surface coverage C_1 at $T = 300$ K (upper panel, $N_w = 400$) and $T = 400$ K (lower panel, $N_w = 625$). Shading is proportional to the probability density. Each panel has its own (proper) probability scale.

Additionally, the lifetime of the spanning water networks at the surface of lysozyme was analysed at various hydration levels. The two-peak probability distribution of the largest cluster size was used to distinguish between spanning and non-spanning clusters (Figure 4.15 and Figure 4.12). In particular, at $T = 300$ K the largest cluster, which includes more than 270 molecules, was considered as a spanning one. Every hundredth MD step (0.2 ps) we checked for the presence of a spanning cluster. The time t between the appearance and disappearance of a spanning cluster was defined as the product of the time-step 0.2 ps and the number of consecutive configurations when the cluster remains spanning. The numbers $h(t)$ of spanning clusters, which have a life-time t , allow to calculate the number $N(t)$ of spanning networks living for a time t or longer [204]:

$$N(t) = \sum_{t'=t}^{t'=\infty} h(t'). \quad (4.6)$$

The value $N(0)$ of the decay function $N(t)$ corresponds to the total number of spanning networks observed in the simulation run, i.e., the number of spanning networks which are present at least during one analysed configuration. The simulated function $N(t)/N(0)$ of spanning water networks at the surface of lysozyme is shown in Figure 4.30 for various hydration levels.

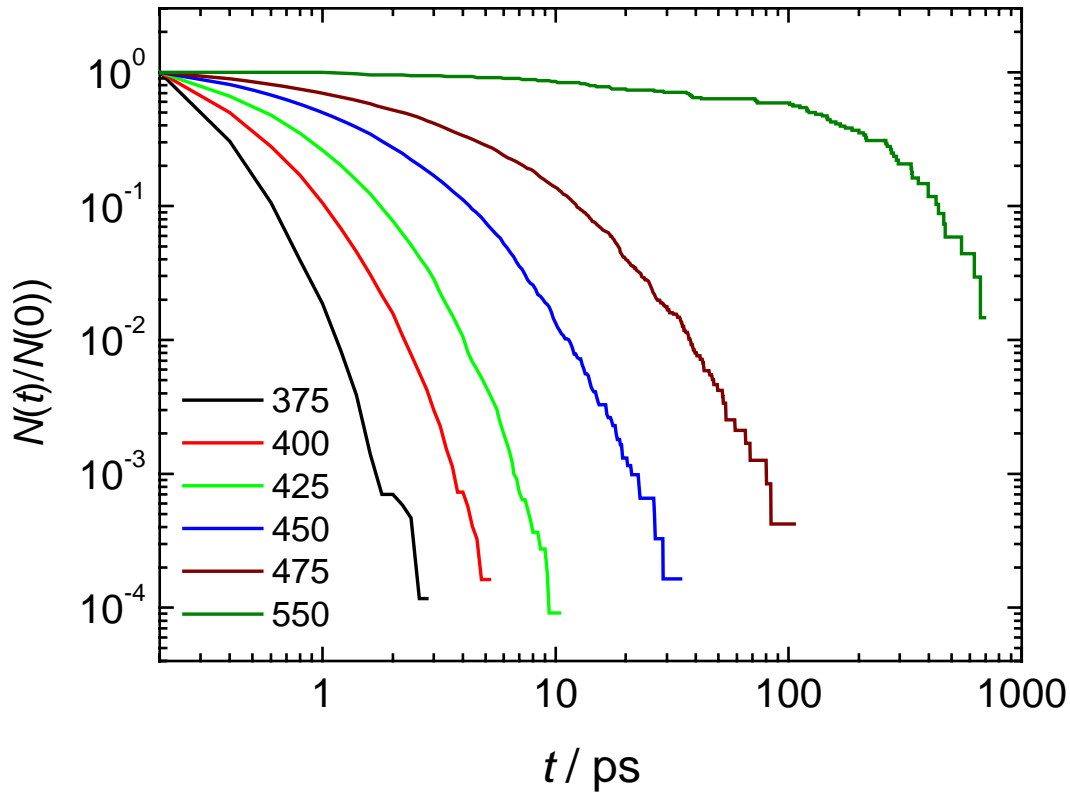


Figure 4.30: Decay function $N(t)/N(0)$ of the lifetime of a spanning water network at the surface of lysozyme at $T = 300$ K and various hydration levels (N_w is given in the inset). The blue and red lines approximately correspond to the threshold surface coverages C_1 and C_2 , respectively.

At $t \geq 1$ ps, the function $N(t)/N(0)$ can be well fitted to a stretched exponential function:

$$N(t)/N(0) = \exp\left(-\left(\frac{t}{\tau_{st}}\right)^\alpha\right), \quad (4.7)$$

where τ_{st} is a characteristic time and α is a stretching parameter. The deviation of the stretching parameter α from 1 is a signature of the nonexponential behaviour of $N(t)/N(0)$. In

order to take all data points equally into account at all time scales, we fitted the data to Equation 4.7 in a logarithmic form. The obtained fitting parameters τ_{st} and α are listed in Table 4.3. The average lifetimes $\langle \tau_{st} \rangle$, calculated at various hydration levels by the equation:

$$\langle \tau_{st} \rangle = \frac{\tau_{st}}{\alpha} \Gamma\left(\frac{1}{\alpha}\right), \quad (4.8)$$

with Γ as the gamma function, are given in Table 1.

The average lifetime $\langle \tau \rangle$ of a spanning water network can also be calculated directly, i.e. without any assumptions concerning the behaviour of $N(t)/N(0)$:

$$\langle \tau \rangle = N(0)^{-1} \int_0^{\infty} N(t') dt' . \quad (4.9)$$

There is a good agreement between the average lifetimes of spanning water networks, obtained in these two different ways (compare $\langle \tau \rangle$ and $\langle \tau_{st} \rangle$ in Table 4.3). This evidences a stretched exponential decay of the spanning water clusters with time at all hydration levels. The deviation from the exponential decay increases with the hydration level and seems to approach a maximum ($\alpha \approx 0.6$) at the percolation threshold, indicating the existence of a broad range of time scales. Far above the percolation threshold, the decay function $N(t)$ of the spanning water network could not be fitted satisfactorily to Equation 4.7. As a result, the quality of the fit for $N_w = 550$ is worse compared to the quality of the fits at the other hydration levels given in Table 4.3.

It is reasonable to compare the lifetime of a spanning network with the average lifetime $\langle \tau_{HB} \rangle$ of a single water-water H-bond. Various procedures can be used to calculate $\langle \tau_{HB} \rangle$ from MD simulations. In the present thesis, a procedure similar to the one used to study the lifetime of a spanning water network is adopted. Namely, the number $N(t)$ of water-water H-bonds which occur and remain continuously unbroken during the time interval t or longer are calculated. The distribution $N(t)$ and the average lifetime $\langle \tau_{HB} \rangle$ are found to be rather sensitive to the sampling frequency. In fact, the use of a long interval between sampled configurations assumes that breaking and restoring of some H-bonds at shorter time intervals are ignored. The time interval between the analysed configurations was chosen to be equal to 0.2 ps, i.e., exactly the same as used in the analysis of the lifetime of a spanning water network. The average lifetimes $\langle \tau_{HB} \rangle$ of water-water H-bonds near the lysozyme surface at

various hydration levels are shown in Table 4.3. Note, that $\langle \tau_{\text{HB}} \rangle$ calculated for hydration water at the lysozyme surface is slightly higher than the value $\langle \tau_{\text{HB}} \rangle \approx 0.9$ ps obtained by applying the same procedure for the bulk liquid water at ambient conditions. The latter value is in good agreement with the average water-water hydrogen bond lifetime, obtained in bulk water using the same sampling frequency 1.0 ps.⁴²

Table 4.3: Parameters τ_{st} and α for lysozyme at 300 K, computed by fitting of $N(t)/N(0)$ with a stretched exponential function (Equation 4.7). The lifetime of a single H-bond $\langle \tau_{\text{HB}} \rangle$, calculated by Equation 4.9 and the lifetimes of a spanning water network $\langle \tau_{\text{st}} \rangle$ and $\langle \tau \rangle$, calculated by Equations 4.8 and 4.9, respectively, are also given.

N_w	α	$\tau_{\text{st}} / \text{ps}$	$\langle \tau_{\text{st}} \rangle / \text{ps}$	$\langle \tau \rangle / \text{ps}$	$\langle \tau_{\text{HB}} \rangle / \text{ps}$
375	0.857	0.202	0.219	0.197	1.088
400 ($\approx C_1$)	0.835	0.371	0.408	0.355	1.084
425	0.800	0.613	0.695	0.653	1.077
450 ($\approx C_2$)	0.633	1.038	1.462	1.164	1.075
475	0.573	2.648	4.524	4.727	1.072
550	(1)	184	184	176	1.059

4.4 Discussion and Conclusions

An analysis of the water clustering in hydrated model lysozyme powder and on the surface of a single lysozyme molecule reveals the formation of a spanning water network with increasing hydration level via a percolation transition of the hydration water adsorbed on the protein surface, in agreement with experimental studies of protein powders [10, 34-38, 188-192]. The fractal dimension of the largest water cluster d_f and the average number of H-bonds n_H at the percolation threshold (Figures 4.1, 4.5, 4.10, 4.14) evidence the 2D character of the percolation transition, which was in fact established experimentally by determining the critical exponent of the protonic conductivity near the percolation threshold of water in powders of lysozyme [35], purple membrane [36], maize seeds [37] and yeast [38].

The percolation threshold in a model of densely packed lysozyme powder at $T = 300$ K is in good agreement with the experimental studies. The obtained mass fraction of water at the percolation transition, $C_2(d_f) = 0.155$ (hydration level $h = 0.183$), is close to the experimental

values for hydrated lysozyme powders, $C = 0.132$ ($h = 0.152 \pm 0.016$) [34]. So, the rather crude model of densely packed lysozyme powder satisfactorily reproduces the clustering of water in real systems near the percolation threshold. Taking into account that in the artificial loosely packed model powder ($\rho = 0.44 \text{ g cm}^{-3}$), the percolation transition of water is distorted and strongly shifted toward higher water content, an improved model powder should have a density of the dry protein slightly above 0.66 g cm^{-3} . Further improvement of the model should account for changes of the powder structure and density with the level of hydration. Finally, the powder should consist of flexible protein molecules. However, the current levels of simulation technique and computer facilities make the application of the latter two improvements a major effort.

The current simulations allow a detailed exploration of the arrangement of the water molecules near hydrated proteins and its changes during the percolation transition. Experimental studies [34] provide only the average number of water molecules per one protein ($N_w/N_p \approx 120$) at the percolation threshold. However, this number may not correspond to the average number of water molecules N_w^1 in the first hydration shell of each protein in the powder. In the densely packed powder used in this work, $N_w/N_p \approx 146$ and $N_w^1 \approx 149$ at the percolation threshold at 300 K, indicating that most of the water molecules belong to a hydration shell which is not shared by several protein molecules. These numbers are significantly smaller than $N_w = 450$ and $N_w^1 \approx 336$ for the percolation threshold at the surface of a single lysozyme molecule at the same temperature (Table 4.2). Such a strong difference can be attributed to a significant decrease of the accessible surface area of proteins in the powder due to close contacts. So, the 2D percolation transition of water in protein powder appears as the formation of a water network, which spans the extended “collective” surface created by closely packed protein molecules, covering each protein molecule only partially.

The crude model of a protein powder does not change its structure with the level of hydration, and therefore is not reliable at high hydration levels. Nevertheless, one can make some predictions, based on the simulations of a single hydrated protein molecule. In particular, at the percolation threshold in the powder ($h = 0.183$), about 190 molecules per protein molecule are missing to form a spanning network around each protein. This could be achieved at $h \sim 0.30$, if all excess water molecules above the threshold hydration are shared by two proteins. Finally, the formation of separate spanning networks could be expected at $h \approx 0.42$.

The analysis of the experimental data [10, 192] indicated three important hydration levels of lysozyme: the 2D percolation threshold is observed at $h \approx 0.152$; the water motion increases strongly at $h \approx 0.25$; and the full internal motions of the protein recover at $h \approx 0.38$. Dividing the characteristic values of h estimated above from the simulations by a factor of about 1.15, we obtain a surprising coincidence with these experimentally observed specific hydration levels. The factor 1.15 is equal to the ratio of SASA of our model lysozyme (6900 \AA^2) and SASA calculated from the crystal structure (6000 \AA^2) [10]. This seems to be the main origin of the differences between the hydration levels estimated from experiment and computer simulations. The close similarity of the protein hydration in experiments and simulations shows that an individual (non-shared) spanning water network around the protein molecule is a necessary condition for its full internal motions. The formation of a fractal-like spanning network on a protein molecule via a 2D percolation transition reflects the first appearance of an individual hydration shell.

It is interesting to estimate which part of the lysozyme surface is covered by water at the percolation threshold. A water monolayer with bulk-like structure (corresponding to a density of about 0.033 \AA^{-3} at ambient conditions) gives a surface coverage of about 0.1 \AA^{-2} . This coverage does not change strongly due to packing effects near planar smooth surfaces [195] or near model protein surfaces [20, 78] and, therefore, 10 \AA^2 could be used as an average area occupied by a water molecule at the surface. At the percolation threshold on the surface of a single lysozyme at $T = 300 \text{ K}$, water covers about 50% of the total lysozyme surface, or about 66% of the hydrophilic part of the lysozyme surface. (About 74% of the lysozyme surface is estimated to be hydrophilic, considering polar and charged residues.) The value of 66% is close to the site occupancy at the percolation threshold of square and honeycomb 2D lattices, which consists of sites with 4 and 3 nearest neighbours, respectively. The above mentioned surface occupancy should not be mixed up with the so-called space occupation probability, which is 0.45 ± 0.03 at the 2D site percolation threshold regardless of the lattice structure [205]. The latter parameter was used in the analysis of water percolation at protein surfaces [10, 34-37, 192]. Note, that the space occupation probability uses a specific normalization to the whole surface area instead of the conventional normalization to the complete coverage, defined as the maximum packing of particles. Neglecting this difference gives diverse non-universal critical values for the space occupation probability [10, 34, 36, 37].

There are several experimental estimates of the water coverage at the percolation threshold in protein systems. They were determined as the ratio of the hydration level h_c at the percolation threshold and at some higher hydration level h_m , which was attributed to the

monolayer coverage. The value h_c/h_m as well as h_c and h_m vary strongly from system to system. In lysozyme powder, h_c/h_m is 0.40 [34] (compared to the value 0.42 from the present simulations, see above), in purple membrane, $h_c/h_m = 0.18$ [36], in embryo and endosperm of maize seeds h_c/h_m is 0.36 and 0.70, respectively [37]. The present simulations indicate that h_m is close to the threshold coverage of a single molecule and, therefore, should depend mainly on the protein's hydrophilicity. The value h_c depends on the hydrophilicity of the protein molecules, topological features of its surface and packing of molecules in powder. Therefore, the experimentally observed strong variations of the ratio h_c/h_m reflect mainly the peculiarity of protein packing.

At $T = 300$ K, about 75% of the water molecules belong to the first hydration shell of a single lysozyme molecule at the percolation threshold, and this fraction noticeably decreases with a further increase of the hydration level. For comparison, in the case of hydrophilic spheres, this value is about 90 % at $T = 425$ K. This evidences that the lysozyme surface is noticeably less hydrophilic than the studied smooth hydrophilic surface, which approximately corresponds to the minimum hydrophilicity necessary for the occurrence of a layering transition (2D condensation) [195]. So, a layering transition can not be expected on a lysozyme surface.

The formation of a quasi-two-dimensional spanning water network at the surface of a single protein molecule can be identified with the first appearance of a water "monolayer", which is a necessary condition for completely restoring the internal dynamics of proteins, i.e. the dynamics observed in the case of full hydration. Besides, even in the latter case (protein in aqueous solution), the spanning network of surface water can break with temperature and pressure or due to the addition of co-solvents. This may have important consequences for the protein structure, dynamics and hence its function. Therefore, the properties of spanning networks of hydration water on the surface of proteins should be studied in detailed, in particular by computer simulations.

Percolation theory does not predict how a percolating (spanning) network should appear and how it can be detected and characterized in such closed systems as the surface of a finite object. Computer simulation studies of various properties of spanning and non-spanning water networks at the simplest closed surface, namely, smooth hydrophilic spherical surfaces were performed, and this analysis was expected to the surface of a real protein, a lysozyme molecule.

The two-peak structure of the probability distribution $P(S_{\max})$ of the largest cluster size reflects the two different contributions from largest non-spanning and spanning clusters,

respectively. This property allows to identify the spanning character of the largest cluster, based on its size. To explore the specificity of various topological properties of a spanning water network, their mutual correlations and correlations with the size S_{\max} of the largest cluster were studied.

It was found that, in the case of a sphere, the maximum linear extension L_{\max} of the largest cluster approaches values exceeding the diameter of the sphere, even for comparatively small non-spanning clusters. The non-spherical shape of the lysozyme molecule seems to increase the difference between the values of L_{\max} for spanning and non-spanning clusters. However, we cannot exclude the existence of two kinds of spanning clusters, which envelop the ellipsoid-like lysozyme molecule in two different ways: along the major or along the minor axis of the ellipsoid.

The distance H_{\max} from the centre of mass of the largest water cluster to the centre of the sphere and the radius of gyration of the largest cluster R_g are the topological parameters of the network, which directly reflect its spanning or non-spanning character. H_{\max} is close to zero for all spanning clusters, whereas for the majority of non-spanning clusters, $H_{\max} > 0.5R_{\text{sp}}$. A deep minimum in the probability distribution $P(H_{\max})$ indicates an instability of the largest clusters which span roughly half of the spherical surface. In the case of a sphere, all spanning clusters have a radius of gyration R_g close to the radius of a sphere R_{sp} (more correctly, $R_g \approx R_{\text{sp}} + 3 \text{ \AA}$ for water near smooth hydrophilic spheres).

The behavior of H_{\max} and R_g of the largest water cluster on the surface of the lysozyme molecule does not allow for such a clear geometrical interpretation as in the case of an ideal sphere with a smooth (structureless) surface. Due to the complex shape and heterogeneous chemical structure of a protein surface, water molecules are distributed non-homogeneously in the hydration shell. Besides, the properties of water molecules noticeably depend on their location in the particular place of the protein surface. Obviously, these factors essentially affect the structural properties of the non-spanning (local) water networks at low hydration levels and low temperatures. The properties of the spanning water network seem to be much more universal. In particular, the radius of gyration of any spanning water cluster should exceed at least the radius of gyration of the lysozyme molecule and approach the effective radius of the completed first hydration shell. This allows one to use both H_{\max} and R_g to distinguish spanning and non-spanning clusters of water at the surface of the protein molecule.

The average lifetime of a spanning water network strongly depends on the hydration level, whereas the lifetime of a single H-bond is about 1 ps in a wide range of hydration levels

(see Table 4.3). At the percolation threshold, the lifetime $\langle\tau\rangle$ of a spanning network is comparable with the life-time of a single water-water H-bond. This indicates that each water molecule of a largest cluster breaks and creates on the average one H-bond before this cluster changes its character from spanning to non-spanning or vice versa. It is noticeable that the lifetime $\langle\tau\rangle$ of a spanning water network near the percolation threshold is significantly smaller than the residence time of water molecules at the lysozyme surface in solution [77].

Some internal protein movements could be affected (or even activated) by the spanning water network when their characteristic time-scales become comparable to the lifetime of the spanning water network. This represents an example for the "slaving" effect mentioned in the Introduction. For example, the collective dynamics of polypeptide side chains of lysozyme is described by characteristic times of ~ 4 ps [206]. Such motions could be influenced/activated by a spanning water network just above the percolation threshold (see Table 4.3). Indeed, such protein dynamics are not observed in dry proteins [207].

Hence, the presence of a spanning network of water molecules, connected via relatively strong (stable) and flexible H-bonds, could have specific effects on the dynamical properties of the protein+water system. Besides the appearance of the eigen modes of such networks [208], the "rubber-like" character of a spanning water network may not only facilitate the dynamics of various movements of the protein molecule, but may also lead to their coupling.

Finally, the perspectives of computer simulations of water percolation in hydrated protein systems will be discussed. More realistic powder models with flexible molecules should be used to clarify the role of protonic conductivity and protein dynamics at the onset of protein function. It seems useful to explore first the dynamic properties of a single protein as well as its structure at hydration levels on both sides of the percolation threshold of the surface water. Besides, the influence of the large fluctuations of the water network close to the percolation threshold on the properties of the hydrated protein should be clarified. In addition, studying the spanning water network around protein molecules in aqueous solutions and its breakage by temperature or co-solvents could also be useful for the understanding of the protein behaviour in complex solutions. Results [187] show that the clustering of the water molecules near the flexible protein and peptide molecules in infinitely diluted aqueous solution is similar to that presented above for a low-hydrated single lysozyme molecule.

Percolation transitions in closed systems, such as the surface of a finite object, have physical significance not only for the processes in the hydration shells of proteins, but also for relaxation processes in glasses [209]. This percolation problem did not attract much attention

till now, but definitely deserves further studies in the framework of a percolation analysis. In particular, various properties of spanning and non-spanning networks as well as appropriate criteria to detect spanning networks in simulations should be studied systematically.

Finally, studies of various dynamical properties of hydrated proteins with and without spanning networks of hydration water should be carried out to elucidate their role in protein dynamics and function.

5 Summary

In this work, the various properties of SNase and hydration water under ambient conditions, high hydrostatic pressure, high temperature, and cosolvent perturbation, were investigated in atomic details by computer simulation techniques.

This study reveals a detailed picture of the water properties at the protein surface at different conditions. It has been shown that the initial steps of the pressure-induced unfolding are increasing roughness of the protein surface and increasing protein-water interactions. The compressibility of the water close to the protein surface at high pressure is significantly larger than in the bulk. The computer simulations have also shown that the thermal expansion coefficient of the native protein's partial volume is largely determined by the expansibility of its internal voids and by a significant hydration contribution, and both contributions decrease with increasing temperature. The volumetric and structural properties of the SNase were extensively studied at various temperature and pressure conditions. The temperature-induced unfolding simulation shows that the SNase undergoes essentially a two-step unfolding process and its completed after 10 ns. The results of the molecular dynamics simulations of SNase in the presence of cosolvents (urea and glycerol) furnish detailed information about structural properties of the cosolvents at the protein surface.

To analyse the structure of the hydration water, the percolation transition threshold of the hydration water, the first computer simulation study of the percolation transition of water in a model protein powder and on the surface of a single protein molecule was performed. Strong fluctuations of the surface water network are observed close to the percolation threshold. The simulations also furnish a microscopic picture for understanding the specific values of the experimentally observed hydration levels, where different steps of increasing mobility in the hydrated powder are observed. An appropriate method to study the formation of a spanning network in finite, closed systems, such as the surface of a single protein, was developed. This method was then applied to locate the percolation transition of water at the surface of a single lysozyme molecule. The radius of gyration of the spanning cluster always exceeds the effective radius of the protein. Any spanning cluster envelops essentially more than half of the surface area. This approach allows studying the structural and dynamical properties of hydrated biosystems below and above the percolation threshold of hydration water. Such simulations are necessary for understanding the onset of protein function.

6 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurden verschiedene Eigenschaften des Proteins SNase und dessen Hydratwasser unter Normalbedingungen sowie unter dem Einfluss von hohem hydrostatischem Druck, bei hohen Temperaturen und bei Variation der Lösungsumgebung untersucht. Mit Hilfe von Computersimulationstechniken wurden dazu Erkenntnisse in atomarer Auflösung erhalten.

Diese Untersuchung trägt zu einem detaillierten Verständnis der Eigenschaften von Wasser an der Proteinoberfläche unter verschiedenen Bedingungen bei. Dabei wurde aufgezeigt, dass die druckinduzierte Entfaltung anfangs durch eine erhöhte Rauigkeit der Proteinoberfläche und vermehrte Protein-Wasser-Wechselwirkungen gekennzeichnet ist. Die Kompressibilität von Wasser direkt an der Proteinoberfläche ist unter Einwirkung von Hochdruck größer als die des Bulk-Wassers. Die molekulardynamischen Simulationen haben außerdem ergeben, dass der thermische Ausdehnungskoeffizient des partiellen Volumens des nativen Proteins weitgehend durch die Fähigkeit zur Ausdehnung von proteininternen Kavitäten sowie einem signifikanten Hydratationsbeitrag bestimmt wird, wobei beide Beiträge mit steigender Temperatur abnehmen. Die volumetrischen und strukturellen Eigenschaften von SNase wurden ausgiebig unter verschiedenen Temperatur- und Druckbedingungen untersucht. Die Simulation der temperatur-induzierten Entfaltung von SNase zeigte, dass das Protein innerhalb von ca. 10 ns in einem – im Wesentlichen - zweistufigen Prozess vollständig entfaltet. Die Ergebnisse der molekulardynamischen Simulation des Proteins in Gegenwart von Co-Solventien (Harnstoff und Glycerol) lieferten detaillierte Informationen über strukturelle Eigenschaften des Co-Solvens an der Proteinoberfläche.

Zur Analyse der Struktur des Hydratwassers und dessen Perkulationsübergangsschwellenwert wurde die erste Computersimulation des Perkulationsübergangs von Wasser an einem Modellprotein-Pulver sowie auf der Oberfläche eines einzelnen Proteinmoleküls durchgeführt. In der Nähe des Perkulationsübergangs wurden starke Fluktuationen des Netzwerks von Wasser an der Proteinoberfläche beobachtet. Die Simulationen trugen des Weiteren auf mikroskopischer Ebene zum Verständnis der experimentell beobachteten Hydratationsgrade bei. Hierbei wurden verschiedene distinkte Mobilitätszunahmen im hydratisierten Pulver beobachtet. Um die Bildung eines umspannenden Netzwerk in finiten,

geschlossenen Systemen, wie z.B. der Oberfläche eines Proteins, zu untersuchen, wurde eine geeignete Methode entwickelt. Unter Anwendung dieser Methode wurde der Perkulationsübergang von Wasser auf der Oberfläche eines Lysozym-Monomers charakterisiert. Der Gyrationradius des umspannenden Clusters übertrifft dabei immer den effektiven Radius des Proteins. Jeder umspannende Cluster bedeckt mindestens die Hälfte der Proteinoberfläche. Auf diese Weise können strukturelle und dynamische Eigenschaften von hydratisierten Biosystemen in der Nähe des Perkulationsgrenzwertes von Hydratwasser untersucht werden. Die durch diese Simulationen gewonnenen Erkenntnisse sind unerlässlich zum Verständnis der biologischen Aktivität von Proteinen in Abhängigkeit des Hydratationsgrades.

References

- 1) Stryer, L. *Biochemistry*; 3rd Ed., W. H. Freeman and co.: New York, 1988.
- 2) Timasheff, S. N. *Ann. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67.
- 3) Dunbar, J.; Yennawar, H. P.; Banerjee, S.; Luo, J.; Farber, G. K. *Protein Sci.* **1997**, *6*, 1727.
- 4) Vidugiris, G. J. A.; Markley, J. L.; Royer, C. A. *Biochemistry* **1995**, *34*, 4909.
- 5) Silva, J. L.; Weber, G. *Annu. Rev. Phys. Chem.* **1993**, *44*, 89.
- 6) Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859. Schellman, J. A. *Annu. Rev. Biophys. Chem.* **1987**, *16*, 115.
- 7) McMurry, J.; Fay, R. C. *Chemistry*; 4th Ed., Prentice Hall: 2003.
- 8) Horton, H. R.; Moran, L. A.; Ochs, R. S.; Rawn, D. J.; Scrimgeour, K. G. *Principles of Biochemistry*; 3rd Ed., Prentice-Hall: 2001.
- 9) Timasheff, S. N. *Biochemistry* **2002**, *41*, 13473.
- 10) Rupley, J. A.; Careri, G. *Adv. Protein Chem.* **1991**, *41*, 37.
- 11) Kuntz, I. D. Jr.; Kauzmann, W. *Adv. Protein Chem.* **1974**, *28*, 239.
- 12) Teeter, M. M. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 577.
- 13) Teeter, M. M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014.
- 14) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **2000**, *79*, 2966.
- 15) Teeter, M. M.; Yamano, A.; Stec, B.; Mohanty, U. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11242.
- 16) Dorbez-Sridi, R.; Cortez, R.; Mayer, E.; Pin, S. *J. Chem. Phys.* **2002**, *116*, 7269.
- 17) Nakasako, M. *J. Biol. Phys.* **2002**, *28*, 129.
- 18) Head-Gordon, T.; Sorenson, J. M.; Pertsemlidis, A.; Glaeser, R. M. *Biophys. J.* **1977**, *73*, 2106.
- 19) Cheng, Y.-K.; Rosky, P. *Nature* **1998**, *392*, 696.
- 20) Smolin, N.; Winter, R. *J. Phys. Chem. B* **2004**, *108*, 15928.
- 21) Karvounis, G.; Nerukh, D.; Glen, R.C. *J. Chem. Phys.* **2004**, *121*, 4925.
- 22) Tarek, M.; Tobias, D. J. *Biophys. J.* **2000**, *79*, 3244.
- 23) Dellerue, S.; Bellissent-Funel, M.-C. *Chem. Phys.* **2000**, *258*, 315.
- 24) Bizzarri, A.R.; Cannistraro, S. *J. Phys. Chem. B* **2002**, *106*, 6617.
- 25) Russo, D.; Hura, G.; Head-Gordon, T. *Biophys. J.* **2004**, *86*, 1852.

- 26) Halle, B. *Phil. Trans. R. Soc. Lond. B* **2004**, 359, 1207.
- 27) Pal, S. K.; Zewail, A. H. *Chem. Rev.* **2004**, 104, 2099.
- 28) Brooks III, C. L.; Karplus, M. *J. Mol. Biol.* **1989**, 208, 159.
- 29) Steinbach, P. J.; Loncharich, R. J.; Brooks, B. R. *Chem. Phys.* **1991**, 158, 383.
- 30) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, 90, 9135.
- 31) Wang, C. X.; Bizzarri, A. R.; Xu, Y. W.; Cannistraro, S. *Chem. Phys.* **1994**, 183, 155.
- 32) Bizzarri, A. R.; Wang, C. X.; Chen, W. Z.; Cannistraro, S. *Chem. Phys.* **1995**, 201, 463.
- 33) Phillips, G. N. Jr.; Pettitt, B. M. *Protein Science* **1995**, 4, 149.
- 34) Careri, G.; Giansanti, A.; Rupley, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 6810.
- 35) Careri, G.; Giansanti, A.; Rupley, J. A. *Phys. Rev. A* **1988**, 37, 2703.
- 36) Rupley, J. A.; Siemankowski, L.; Careri, G.; Bruni, F. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, 85, 9022.
- 37) Bruni, F.; Careri, G.; Leopold, A. C. *Phys. Rev. A* **1989**, 40, 2803.
- 38) Sokolowska, D.; Krol-Otwinowska, A.; Moscicki, J. K. *Phys. Rev. E* **2004**, 70, 052901.
- 39) Smith, J. C.; Merzel, F.; Bondar, A.-N.; Tournier, A.; Fischer, S. *Phil. Trans. R. Soc. Lond. B*, **2004**, 359, 1181.
- 40) Daniel, R. M.; Finney, J. L.; Re'at, V.; Dunn, R.; Ferrand, M.; Smith, J. C. *Biophys. J.* **1999**, 77, 2184.
- 41) Turnier, A. L.; Xu, J.; Smith, J. C. *Biophys. J.* **2003**, 85, 1871.
- 42) Fenimore, P. W.; Frauenfelder, H.; McMahan, B. H.; Parak, F. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 16047.
- 43) Allen, M. P.; Tildesly, D. J. *Computer simulation of liquids*; Clarendon Press: Oxford, 1987.
- 44) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, 106, 765.
- 45) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, 4, 187.
- 46) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, 23, 1513.
- 47) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, 110, 1657.

- 48) Rapaport, D. C. *The art of molecular dynamics simulation*; Cambridge University Press: Cambridge, 1995.
- 49) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Academic Press: San Diego, 1996.
- 50) Sadus, R. J. *Molecular simulation of fluids: theory, algorithms and object-orientation*; Elsevier Science B.V.: Amsterdam, 1999.
- 51) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- 52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- 53) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- 54) Adams, D. J.; Adams, E. M.; Hills, G. J. *Mol. Phys.* **1979**, *38*, 387.
- 55) Lee, B.K.; Richards, F.M. *J. Mol. Biol.* **1971**, *55*, 379.
- 56) Richards, F.M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
- 57) Connolly, M.L. *J. Am. Chem. Soc.* **1985**, *107*, 1118.
- 58) Ramachandran, G. N.; Sasiskharan, V. *Adv. Protein Chem.* **1968**, *23*, 283.
- 59) Stauffer, D. *Introduction to Percolation Theory*; Taylor & Francis: London and Philadelphia, 1985.
- 60) Sator, N. *Phys. Rep.* **2003**, *376*, 1.
- 61) Jan, N. *Physica A* **1999**, *266*, 72.
- 62) Hovi, J.-P.; Aharony, A. *Phys. Rev. E* **1996**, *53*, 235.
- 63) Newman, M.E.J.; Ziff, R.M. *Phys. Rev. Lett.* **2000**, *85*, 4104.
- 64) Martins, P.H.L.; Plascak, J.A. *Phys. Rev. E* **2003**, *67*, 046119.
- 65) de Oliveira, P.M.C.; Nobrega, R.A.; Stauffer, D. J. *Phys. A: Math. Gen.* **2004**, *37*, 3743.
- 66) Ahlström, P.; Teleman, O.; Jönsson, B. *J. Am. Chem. Soc.* **1988**, *110*, 4198.
- 67) Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 2773.
- 68) Brunne, R. M.; Liepinsh, E.; Otting, G.; Wüthrich, K.; van Gunsteren, W. F. *J. Mol. Biol.* **1993**, *231*, 1040.
- 69) Garcia, A. E.; Stiller, L. *J. Comp. Chem.* **1993**, *14*, 1396.
- 70) Lounnas, V.; Pettitt, B. M.; Phillips, G. N. Jr. *Biophys. J.* **1994**, *66*, 601.
- 71) Rocchi, C.; Bizzarri, A. R.; Cannistraro, S. *Phys. Rev. E* **1998**, *57*, 3315.
- 72) Makarov, V. A.; Feig, M.; Andrews, B. K.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 150.

- 73) Garcia, A. E.; Hummer, G. *Proteins Struc. Func. Gen.* **1999**, *36*, 175.
- 74) Garcia, A. E.; Hummer, G. *Proteins Struc. Func. Gen.* **2000**, *38*, 261.
- 75) Luise, A.; Falconi, M.; Desideri, A. *Proteins Struc. Func. Gen.* **2000**, *39*, 56.
- 76) Sterpone, F.; Ceccarelli, M.; Marchi, M. *J. Mol. Biol.* **2001**, *311*, 409.
- 77) Marchi, M.; Sterpone, F.; Ceccarelli, M. *J. Am. Chem. Soc.* **2002**, *124*, 6787.
- 78) Merzel, F.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378.
- 79) Rahman, A.; Stillinger, F. H. *J. Am. Chem. Soc.* **1973**, *95*, 7943.
- 80) Speedy, R. J. *J. Phys. Chem.* **1984**, *88*, 3364.
- 81) Speedy, R. J.; Mezei, M. *J. Phys. Chem.* **1985**, *89*, 171.
- 82) Speedy, R. J.; Madura, J. D.; Jorgensen, W. L. *J. Phys. Chem.* **1987**, *91*, 909.
- 83) Mausbach, P.; Schnitker, J.; Geiger, A. *J. Tech. Phys.* **1987**, *28*, 67.
- 84) Lee, C. Y.; McCammon, J. A.; Rossky, P. J. *J. Chem. Phys.* **1984**, *80*, 4448.
- 85) Lee, C. Y.; Rossky, P. J. *J. Chem. Phys.* **1994**, *100*, 3334.
- 86) Werder, T.; Walther, J. H.; Jaffe, R. L.; Halicioglu, T.; Koumoutsakos, P. *J. Phys. Chem. B* **2003**, *107*, 1345.
- 87) Brovchenko, I.; Paschek, D.; Geiger, A. *J. Chem. Phys.* **2000**, *113*, 5026.
- 88) Brovchenko, I.; Geiger, A.; Paschek, D. *Fluid Phase Equilibria* **2001**, *183-184*, 331.
- 89) Brovchenko, I.; Geiger, A. *J. Mol. Liq.* **2002**, *96-97*, 195.
- 90) Geiger, A.; Rahman, A.; Stillinger, F. H. *J. Chem. Phys.* **1979**, *70*, 263.
- 91) Zichi, D. A.; Rossky, P. J. *J. Chem. Phys.* **1985**, *83*, 797.
- 92) Cheng, Y.-K.; Sheu, W.S.; Rossky, P. J. *Biophys. J.* **1999**, *76*, 1734.
- 93) Head-Gordon, T. *Proc. Natl. Acad. Soc. U.S.A.* **1995**, *92*, 8308.
- 94) Pertsemlidis, A.; Saxena, A. M.; Soper, A. K.; Head-Gordon, T.; Glaeser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10769.
- 95) Lipscomb, L. A.; Peek, M. E.; Zhou, F. X.; Bertrand, J. A.; VanDerveer, D.; Williams, L. D. *Biochemistry* **1994**, *33*, 3649.
- 96) Chen, H. M.; Markin, V. S.; Tsong, T. Y. *Biochemistry* **1992**, *31*, 1483.
- 97) Carra, J. H.; Anderson, E. A.; Privalov, P. L. *Biochemistry* **1994**, *33*, 10842.
- 98) Carra, J. H.; Privalov, P. L. *Biochemistry* **1995**, *34*, 2034.
- 99) Vidugiris, G. J. A.; Truckses, D. M.; Markley, J. L.; Royer, C. A. *Biochemistry* **1996**, *35*, 3857.
- 100) Eftink, M. R.; Ionescu, R.; Ramsay, G. D.; Wong, C.-Y.; Wu, J. Q.; Maki, A. H. *Biochemistry* **1996**, *35*, 8084.
- 101) Hinck, A. P.; Truckses, D. M.; Markley, J. L. *Biochemistry* **1996**, *35*, 10328.

- 102) Walkenhorst, W. F.; Green, S. M.; Roder, M. *Biochemistry* **1997**, *36*, 5795.
- 103) Panick, G.; Malessa, R.; Winter, R.; Rapp, G.; Frye, K. J.; Royer, C. A. *J. Mol. Biol.* **1998**, *275*, 389. Panick, G.; Vidugiris, G. J. A.; Malessa, R.; Rapp, G.; Winter, R.; Royer, C. A. *Biochemistry* **1999**, *38*, 4157.
- 104) Kataoka, M.; Ferrand, M.; Goupil-Lammy, A. V.; Kamikubo, H.; Yunoki, J.; Oka, T.; Smith, J. C. *Physica B* **1999**, *266*, 20.
- 105) Ravindra, R.; Royer, C.; Winter, R. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1952.
- 106) Filfil, R.; Chalikian, T. V. *J. Mol. Biol.* **2000**, *299*, 827.
- 107) Seemann, H.; Winter, R.; Royer, C. A. *J. Mol. Biol.* **2001**, *307*, 1091.
- 108) Bahar, I.; Wallqvist, A.; Covell, D. G.; Jernigan, R. L. *Biochemistry* **1998**, *37*, 1067.
- 109) Zhou, H.-X. *Biophys. J.* **2002**, *83*, 2981.
- 110) Yamaotsu, N.; Moriguchi, I.; Kollman, P. A.; Hirono, S. *Biochim. Biophys. Acta* **1993**, *1163*, 81.
- 111) Lamy, A.; Smith, J. C. *J. Am. Chem. Soc.* **1996**, *118*, 7326.
- 112) Ikura, T.; Tsurupa, G. P.; Kuwajima, K. *Biochemistry* **1997**, *36*, 6529.
- 113) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301.
- 114) Wrabl, J. O.; Shortle, D.; Woolf, T. B. *Proteins Struct. Funct. Gen.* **2000**, *38*, 123.
- 115) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. Jr.; Tasumi, M. *J. Mol. Biol.* **1977**, *12*, 535. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.
- 116) Hynes, T. R.; Fox, R. O. *Proteins: Struct. Funct. Genetics* **1991**, *10*, 92.
- 117) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber*, version 6; University of California: San Francisco, CA, 1999.
- 118) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- 119) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- 120) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1993**, *32*, 4175.
- 121) Huyhn, T.; Smith, J.C.; Sanson, A. *Biophys. J.* **2002**, *83*, 681.

- 122) Duffy, E. M.; Kowalczyk, P. J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1993**, *115*, 9271.
- 123) Chelli, R.; Procacci, P.; Cardini, G.; Della Valle, R. G.; Califano, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 871.
- 124) Kabsch, W.; Sander, C. *Biopolymers*, **1983**, *22*, 2577.
- 125) Mol_Volume - a program in the MDTools utility package at the Theoretical Biophysics group, an NIH Resource for Macromolecular Modeling and Bioinformatics, <http://www.ks.uiuc.edu/Development/MDTools>
- 126) Humphrey, W.; Dalke, A.; and Schulten, K. *J. Molec. Graphics* **1996**, *14.1*, 33.
- 127) Flyvberg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461.
- 128) Finchan, D.; Quirke, N.; Tildesley, D. J. *J. Chem. Phys.* **1986**, *84*, 4535.
- 129) Eisenberg, D.; Kauzmann, W. *The Structure and Properties of Water*; Oxford University Press: New York, 1969.
- 130) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557.
- 131) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267.
- 132) Ravindra, R.; Winter, R. *Chem. Phys. Chem.* **2003**, *4*, 359.
- 133) Toney, M. F.; Howard, J. N.; Richer, J.; Borges, G. L.; Gordon, J. G.; Melroy, O. R.; Wiesler, D. G.; Yee, D.; Sorensen, L. B. *Nature*, **1994**, 368, 444.
- 134) Jorgensen W. L.; Jenson, C. *J. Comp. Chem.* **1998**, *19*, 1179.
- 135) Chalikian, T. V.; Breslauer, K. J. *Biopolymers* **1996**, *39*, 619.
- 136) Chalikian, T. V.; Breslauer, K. J. *Curr. Opin. Struct. Biol.* **1998**, *8*, 657.
- 137) Millero, F. J.; Lo Surdo, A.; Shin, C. *J. Phys Chem.* **1978**, *82*, 784.
- 138) Lee, B. *J. Phys. Chem.* **1983**, *87*, 112.
- 139) Gutmann, V. *Pure Appl. Chem.* **1991**, *63*, 1715.
- 140) Heremans, K.; Smeller, L. *Biochim. Biophys. Acta* **1998**, *1386*, 353.
- 141) Wroblowski, B.; Diaz, J. F.; Heremans, K.; Engelborghs, Y. *Proteins* **1996**, *25*, 446.
- 142) Paliwal, A.; Asthagiri, D.; Bossev, D. P.; Paulaitis, M. E. *Biophys. J.* **2004**, *87*, 3479.
- 143) Kitchen, D. B.; Reed, L. H.; Levy, R. M. *Biochemistry* **1992**, *31*, 10083.
- 144) Doster, W.; Gebhardt, R. *Chem. Phys.* **2003**, *292*, 383.
- 145) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1552.
- 146) Gekko, K.; Noguchi, H. *J. Phys. Chem.* **1979**, *83*, 2706.

- 147) Chalikian, T. V.; Totrov, M.; Abagyan, R.; Breslauer, K. J. *J. Mol. Biol.* **1996**, *260*, 588.
- 148) Chalikian, T. V.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1012.
- 149) Taulier, N.; Chalikian, T. V. *Biochim. Biophys. Acta* **2002**, *1595*, 48.
- 150) Lin, L. N.; Brandts, J. F.; Brandts, J. M.; Plotnikov, V. *Anal. Biochem.* **2002**, *302*, 144.
- 151) Loupiac, C.; Bonetti, M.; Pin, S.; Calmettes, P. *Eur. J. Biochem.* **2002**, *269*, 4731.
- 152) Ravindra, R.; Winter, R. *Z. Phys. Chem.* **2003**, *217*, 1221.
- 153) Ravindra, R.; Winter, R. *Chem. Phys. Chem.* **2004**, *5*, 566.
- 154) Rösgen, J.; Hinz, H.-J. *Biophys. Chem.* **2000**, *83*, 61.
- 155) Richards, F. M. *J. Mol. Biol.* **1974**, *82*, 1.
- 156) Lee, B. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 622.
- 157) Fraunfelder, H.; Hartmann, H.; Karplus, M.; Kuntz, I. D.; Kuriyan, J. Jr., Parak, F.; Petsko, G. A.; Ringe, D.; Tilton, R. F. Jr., Connolly, M. L.; Max, N. *Biochemistry* **1987**, *26*, 254.
- 158) Harpaz, Y.; Gerstein, M.; Chotia, C. *Structure* **1994**, *2*, 641.
- 159) Paci, E.; Marchi, M. *J. Phys. Chem.* **1996**, *100*, 4314.
- 160) Gerstein, M.; Chotia, C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10167.
- 161) Paci, E.; Marchi, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11609.
- 162) Dadarlat, V.M.; Post, C.B. *J. Phys. Chem. B* **2001**, *105*, 715.
- 163) Paci, E. *Biochim. Biophys. Acta* **2002**, *1595*, 185.
- 164) Teeter, M. M. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 577.
- 165) Oleinikova, A.; Smolin, N.; Brovchenko, I.; Geiger, A.; Winter, R. *J. Phys. Chem. B* **2005**, *109*, 1988.
- 166) Smolin, N.; Oleinikova, A.; Brovchenko, I.; Geiger, A.; Winter, R. *J. Phys. Chem. B* **2005**, *109*, 10995.
- 167) Paci, E.; Velikson, B. *Biopolymers* **1997**, *41*, 785.
- 168) Mark, A. E.; Van Gunsteren, W. F. *Biochemistry* **1992**, *31*, 7745.
- 169) Bond, C. J.; Wong, K.-B.; Clarke, J.; Fersht, A. R.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13409.
- 170) Li, A.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10430.
- 171) Li, A.; Daggett, V. *J. Mol. Biol.* **1996**, *257*, 412.
- 172) Caflisch, A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1746.
- 173) Mao, Y.; Ratner, M. A.; Jarrold, M. F. *J. Phys. Chem. B* **1999**, *103*, 10017.

- 174) Alonso, D. O. V.; Daggett, V. *J. Mol. Biol.* **1995**, *247*, 501.
- 175) Hünenberger, P. H.; Mark, A. E.; Van Gunsteren, W. F. *Proteins* **1995**, *21*, 196.
- 176) Brunne, R. M.; Van Gunsteren, W. F. *FEBS Lett.* **1993**, *323*, 215.
- 177) Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M. *Folding & Design* **1998**, *3*, 105.
- 178) Brandts J. F.; Hunt, L. *J. Am. Chem. Soc.* **1967**, *89*, 4826.
- 179) Bennion B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142.
- 180) Baynes, B. M.; Trout, B. L. *J. Phys. Chem. B* **2003**, *107*, 14058.
- 181) Head-Gordon, T.; Sorenson, J. M.; Pertsemliadis, A.; Glaeser, R. M. *Biophys. J.* **1997**, *73*, 2106.
- 182) Kim, S.; Yoon, C. N.; Jhon, M. S. *J. Comp. Chem.* **1988**, *9*, 125.
- 183) Guillot, B.; Guissani, Y. *J. Chem. Phys.* **1993**, *99*, 8075.
- 184) Brovchenko, I.; Geiger, A.; Oleinikova, A. In *New Kinds of Phase Transitions: Transformations in Disordered Substances*; Brazshkin, V. V.; Buldyrev, S. V.; Ryshov, V. N.; Stanley, H. E.; Eds.; Kluwer Academic Publishers: Norwell, MA, 2002; p 367.
- 185) Matubayasi, N.; Levy, R. M. *J. Phys. Chem.* **1996**, *100*, 2681.
- 186) Paschek, D.; Garcia, A. E. *Phys. Rev. Lett.* **2004**, *93*, 238105. Paschek, D.; Gnanakaran, S.; Garcia, A.E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6765.
- 187) Brovchenko, I.; Krukau, A.; Smolin, N.; Oleinikova, A.; Geiger, A.; Winter, R. *J. Chem. Phys.* **2005**, *123*, 224905.
- 188) Pissis, P.; Konsta, A. A. *J. Phys. D: Appl. Phys.* **1990**, *23*, 932.
- 189) Klammler, F.; Kimich, R., *Croat. Chem. Acta* **1992**, *65*, 455.
- 190) Konsta, A. A.; Laudat, J.; Pissis, P. *Solid State Ionics* **1997**, *97*, 97.
- 191) Careri, G. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 223.
- 192) Careri, G.; Peyrard, M. *Cell. Mol. Biol.* **2001**, *47*, 745.
- 193) Careri, G.; Milotti, E. *Phys. Rev. E* **2003**, *67*, 051923.
- 194) Tarek, M.; Tobias, D. J. *J. Am. Chem. Soc.* **1999**, *121*, 9740; *Biophys. J.* **2000**, *79*, 3244.
- 195) Brovchenko, I.; Geiger, A.; Oleinikova, A. *J. Chem. Phys.* **2004**, *120*, 1958.
- 196) McKenzie, H. A.; White, F. H. Jr. *Adv. Prot. Chem.* **1991**, *41*, 174.
- 197) Kundrot, C. E.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 157.
- 198) Kodandapan, R; Suresh, C. G.; Vijayan, M. *J. Bio. Chem.* **1990**, *265*, 16126.
- 199) Kocherbirov, V.; Arnebrant, T.; Söderman, O. *J. Phys. Chem. B* **2004**, *108*, 19036.

- 200) Brovchenko, I.; Oleinikova, A. In *Handbook of Theoretical and Computational Nanotechnology*; Rieth, M., Schommers, W., Eds.; American Scientific Publishers: 2005.
- 201) Oleinikova, A.; Brovchenko, I.; Geiger, A.; Guillot, B. *J. Chem. Phys.* **2002**, *117*, 3296.
- 202) Sen, P. *J. Phys. A: Math. Gen.* **2001**, *34*, 8477.
- 203) da Silva, C. R.; Lyra, M. L.; Viswanathan, G. M. *Phys. Rev. E* **2002**, *66*, 056107.
- 204) Geiger, A.; Mausbach, P.; Schnitker, J.; Blumberg, R. L.; Stanley H. E. *J. Phys. (Paris)* **1984**, *45*, C7-13.
- 205) Zallen, R.; Scher, H. *Phys. Rev. B* **1971**, *4*, 4471.
- 206) De Francesco, A.; Marconi, M.; Cinelli, S.; Onori, G.; Paciaroni, A. *Biophys. J.* **2004**, *86*, 480.
- 207) Fitter, J.; Lechner, R. E.; Büldt, G.; Dencher, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7600.
- 208) Pal, S.; Balasubramanian, S.; Bagchi, B. *J. Chem. Phys.* **2004**, *120*, 1912.
- 209) Campbell, I. A.; Flesselles, J.-M.; Jullien, R.; Botet, R. *Phys. Rev. B* **1988**, *37*, 3825.
- 210) Starr, F. W.; Nielsen, J. K.; Stanley, H. E. *Phys. Rev. E* **2000**, *62*, 579.

Lebenslauf

Name	Nikolai Smolin
Geburtsdatum	02.08.1977
Geburtsort	Ortschaft Komsomolskij, Bezirk Sowjetskij, Gebiet Tjumen, Russland

Schulausbildung

September 1984 bis Oktober 1986	Grundschule No. 1 in Ortschaft Sowjetskij, Bezirk Sowjetskij, Gebiet Tjumen, Russland
November 1986 bis Juni 1992	Schule No. 117, Chelyabinsk, Russland
September 1992 bis Juni 1994	Physik & Mathematik, Lyzeum No. 31, Chelyabinsk, Russland

Studium

September 1994 bis Juni 2000	Studium der Angewandten Mathematik und Physik an der Südurale Staatliche Universität, Chelyabinsk, Russland.
September 1999 bis Juni 2000	Diplomarbeit „The computer simulation studies of the atomic structure and the inter-atomic potential of the liquid Ga and the amorphous alloy Ni-Zr based on the diffraction experimental data “
01. Juni 2000	Diplom (mit Auszeichnung)
01. Juni 2001	Beginn der Promotion am Lehrstuhl für Physikalische Chemie I bei Prof. Dr. R. Winter, Universität Dortmund
16. Mai 2006	Disputation

Berufstätigkeit

Mai 1997 bis Februar 2001	Wissenschaftliche Hilfskraft
März 2001 bis Juni 2006	Wissenschaftlicher Angestellter