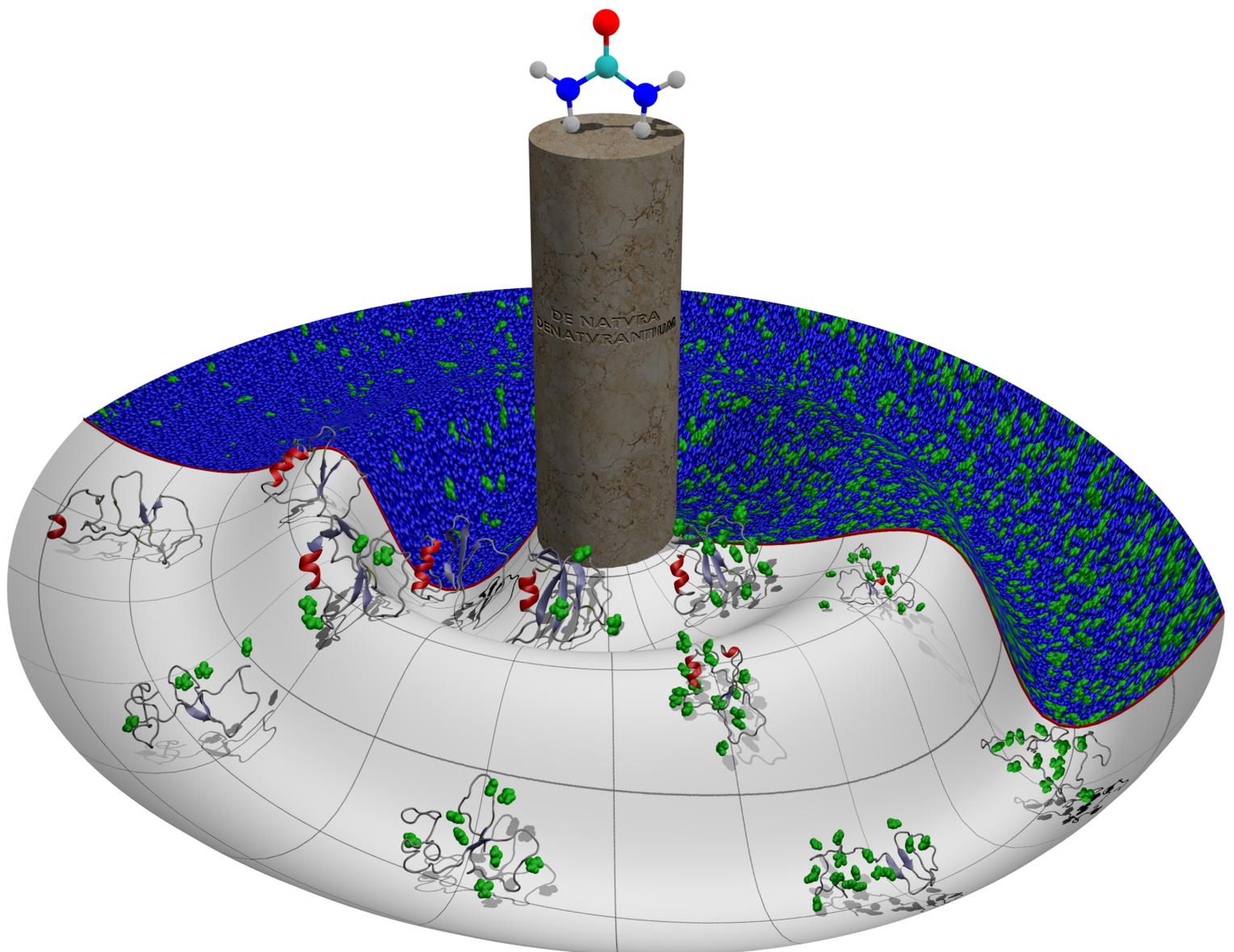


De natura denaturantium

On the molecular basis of
urea-induced protein denaturation



De natura denurantium

—

**On the molecular basis of urea-induced protein
denaturation**

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

Göttingen 2007

vorgelegt von

Martin Stumpe

aus Bad Pyrmont

D 7

Referent: Prof. Dr. Tim Salditt

Koreferent: Prof. Dr. Helmut Grubmüller

Tag der mündlichen Prüfung (Disputation): 29.10.2007

Contents

1	Introduction	11
2	Protein Structure, Folding, and Unfolding — A brief overview	15
2.1	Protein structure	16
2.2	Protein folding and unfolding — a brief overview	18
3	Molecular Dynamics Simulations	21
3.1	Principle	22
3.2	Simulation Details	25
3.3	Relevant Observables — Trajectory Analysis	29
4	Aqueous urea solutions	35
4.1	Introduction	36
4.2	Simulation setup and additional methods	37
4.2.1	Simulation setup	37
4.2.2	Calculation of spatial density distributions	37
4.2.3	Quantification of urea aggregation	38
4.2.4	Comparison to experiments, controls	39
4.3	Results	40
4.3.1	Spatial density distributions	40
4.3.2	Hydrogen bonds	44
4.3.3	Urea aggregation	46
4.3.4	Comparisons to experiments, controls	49
4.4	Summary and Conclusion	50
4.5	Appendix: Quantification of aggregation under periodic boundary conditions	51

5	Direct Interactions of Urea with Amino Acids	53
5.1	Introduction	54
5.2	Methods	54
5.2.1	Simulation setup	54
5.2.2	Force field energies	55
5.3	Results	55
5.3.1	Contact coefficients	55
5.3.2	Hydrogen bonding	57
5.3.3	Driving forces for preferential interactions	59
5.4	Discussion	60
5.5	Conclusions	62
6	Polar or apolar? — The role of polarity for the denaturation power of urea	65
6.1	Introduction	66
6.2	Methods	66
6.2.1	Simulation setup	66
6.3	Results and Discussion	68
6.3.1	Dynamics of the native state in water/urea	68
6.3.2	Protein stability in hypo- and hyperpolar urea	70
6.3.3	Unfolding pathways in hypopolar urea	74
6.3.4	Residual structure in the denatured state	80
6.4	Summary and conclusion	82
7	The effect of urea on partially unfolded proteins	85
7.1	Regular urea — Hydrophobic collapse lost.	86
7.1.1	Introduction	86
7.1.2	Methods	86
7.1.3	Results and discussion	89
	Dynamics of the native state	89
	Dynamics of partially unfolded structures	90

Interactions between urea and the protein	94
Markov model for urea-induced unfolding	96
7.1.4 Summary and conclusion	97
7.2 Hyperpolar urea — Hydrophobic collapse regained	99
7.2.1 Introduction	99
7.2.2 Simulation setup	99
7.2.3 Results	99
7.2.4 Conclusion	103
8 Urea-induced denaturation of the Trp-Cage mini-protein on realistic timescales	105
8.1 Introduction	106
8.2 Simulation setup	106
8.3 Results and Discussion	106
8.3.1 Denaturation of the Trp-Cage	106
8.3.2 Characterization of unfolding processes.	107
8.3.3 Interactions of urea with the Trp-Cage.	108
8.3.4 Observation of a folding intermediate.	110
8.4 Conclusions	112
9 Beyond Urea — A comparison of amino acid interaction preferences for Urea, DMSO, Guanidinium, and TMAO	115
9.1 Introduction	116
9.2 Methods and simulation setup	118
9.3 Results and discussion	119
9.3.1 DMSO	119
9.3.2 Guanidinium	121
9.3.3 TMAO	122
9.4 Conclusions	124
10 Summary, Conclusion, and Outlook	131
Appendix	137
Danksagung	141

1

Introduction

Proteins are the nano-machines of life. They play key roles in every process in living cells, and they are one of the four macromolecular building blocks upon which life is built, together with lipids, polysaccharides, and nucleic acids. The central importance of proteins for living organisms has already been sensed by their name giver, the Swedish chemist Berzelius in 1838 — the word *protein* is derived from Greek $\pi\rho\omega\tau\omicron\sigma$ (“*protos*”), meaning “*the first*” or “*of primary importance*”. The numerous functions of proteins include signal transmission and reception, process regulation and control, enzymatic catalysis, transport and storage, and building the cytoskeleton.

The specific function of a protein is intimately linked to its characteristic three-dimensional structure. The latter is determined by the sequence of amino acids, of which a protein is built as linear heteropolymer.¹ However, the exact relation between the sequence (the chemical structure) and the “fold” (the three-dimensional structure) is unclear. After a protein is synthesized from single amino acids, the disordered protein chain arranges itself to adopt the complex and characteristic three-dimensional structure of the protein, the “fold” or “native state”. This process, the “protein folding”, can be described as a complex sequence of moves on a high-dimensional free energy landscape, in which the native conformation represents a minimum. Due to the high dimensionality of this free energy landscape, protein folding would take an astronomical amount of time, if a protein was sampling conformational space randomly. Yet, proteins fold within seconds or less. This discrepancy, the famous “Levinthal-paradox”,² suggests that the protein does not randomly sample all possible conformations, but instead follows a directed search. While it is now widely accepted that the hydrophobic effect is the main driving force for protein folding,³ the mechanism of protein folding has puzzled scientists for several decades, and has, despite significant progress, just recently been named one of the 125 biggest unsolved problems in science.⁴

A major reason for the complexity of the protein folding problem lies in the intricate interaction of the protein with its solvent environment; the thermodynamic stability of the folded state is a result of the complex interplay of enthalpic and entropic contributions from protein–protein, protein–solvent, and solvent–solvent interactions. In particular, the folded state of a protein is only marginally stable under physiological conditions, and small changes in these conditions can tip the energetic balance and destabilize the folded state. Since protein folding is a reversible process, destabilization of the folded state leads to *unfolding* of the protein. In this process, which is also termed “denaturation”, the protein loses its characteristic three-dimensional structure and adopts conformations of the *denatured state* or *unfolded ensemble*. Unfolding can be induced by different means, e.g., change of temperature or pH, application of high pressure, or the addition of *denaturants* — chemical substances that cause protein denaturation.

Denaturants are widely used to study protein folding and stability, and one of the most commonly used denaturants is urea. Despite its ubiquitous use, however, the molecular mechanism of urea-induced protein denaturation is not yet understood. Understanding the interactions between denaturants and proteins would not only explain chemical denatura-

tion, but would also give invaluable insight into protein stability and solvent interactions in general. Thus, the question how denaturants work at the molecular level, and in particular how urea works, has received considerable attention during the past 40 years. However, despite extensive efforts and literature on the subject, no clear picture has emerged yet, and the situation is still quite controversial. It is the aim of this thesis to advance the understanding of urea-induced protein denaturation.

The central question about urea-induced protein denaturation is “What are the driving forces for denaturation?”, or in other words “*Why* does urea denature proteins?”. Two basic concepts have guided the investigation of the driving forces of urea-induced protein denaturation in the past, and still set the framework for ongoing discussions. According to the first model, urea induces changes in the water structure, which in turn weaken the hydrophobic effect and thus cause protein denaturation.^{5,6} In this model of *indirect* interactions, two alternative views can be distinguished in which urea is regarded either to break,^{5,6} or to enhance⁷ water structure. The second model, in contrast, attributes the denaturing effect of urea to *direct* interactions between urea and the protein.^{8,9,10} Also this model comprises different aspects; either the interaction of urea with polar residues or the peptide backbone, mainly via hydrogen bonding⁹ — or hydrophobic interaction with apolar residues.⁸

All of these possibilities, and also various combinations thereof, have been suggested as the primary driving force of denaturation, and are still controversially discussed. While some recent studies have provided support for the dominance or significant contribution of indirect effects,^{11,12,13,14,15,16} this concept has often been rejected or severely criticized,^{10,17,18} and many recent studies suggest that direct interactions are the cause for denaturation.^{19,20,21,22,23,24,25,26,18,27,28,29} Among these, however, the dominance of either polar interactions^{21,15,26,16,27,29} or apolar interactions^{8,30,19,20,22,23,24,25,28} between urea and the protein is also discussed controversially.

Aside from this central question about the driving forces for denaturation, there are further open questions — for example “*How* does urea denature proteins?”. In particular, does urea actively destabilize the folded state, or rather stabilize the denatured state? Which intermediate states during folding/unfolding are affected most? Calorimetric studies³⁰ point towards a stabilization of the denatured state. But even with this thermodynamic picture in mind, the underlying atomistic processes are still unknown.

Related to this issue is the question of denaturation pathways. Are there specific pathways for unfolding, or is denaturation a more heterogeneous process? Whereas for folding, the existence of specific pathways or mechanisms follows from the Levinthal-paradox, this is not obvious for unfolding, since the denatured state comprises a large collection of conformers rather than one well-defined conformation, and, therefore, is not difficult to “find”.

The structure of denatured proteins is a third topic of particular interest. While it has long been assumed that denatured proteins are random-coils, recent studies strongly suggest residual structure in denatured proteins.^{31,32,33,34,35,36}

Aims and organization of this thesis

This thesis mainly addresses the central question “*What are the driving forces for urea-induced denaturation?*”. In particular, “*What is the relevance of indirect versus direct interactions for denaturation?*”, and “*What is the relevance of polar or hydrophobic interactions between urea and the protein?*”. To this end, molecular dynamics simulations are employed, which shed light on the atomistic details of the interactions and processes involved. Since protein denaturation proceeds on time scales that are several orders of magnitude longer than those accessible in state-of-the-art molecular dynamics simulations, ways beyond the straight-forward simulation of the denaturation process had to be developed to circumvent this problem.

This thesis is organized as follows. After a brief overview of protein structure and folding (Chap. 2) and an introduction into molecular dynamics simulations and relevant observables (Chap. 3), the structure and energetics of aqueous urea solutions are analyzed to address the relevance of indirect effects via urea-induced changes in the water structure (Chap. 4). Subsequently, the specific molecular interactions of urea with all 20 amino acids that build up proteins are investigated in simulations of tripeptides, to address direct interactions and to separate them from effects of sequence of structure (Chap. 5). Motivated by these results, “Gedankenexperiment”-simulations of the CI2 protein solvated in urea with modified polarity explore whether polar or apolar contacts are the driving force for denaturation (Chap. 6). Further, these simulations serve to examine unfolding pathways and possible residual structure of the denatured state. In Chapter 7, the effect of urea on partially unfolded structures of the Cold Shock protein Bc-CsP is investigated. In addition to providing further and independent support for the results of the previous chapter, this Chapter also addresses the question how urea affects different states of the protein. Chapter 8 concludes the studies about urea-induced protein denaturation by testing the ability of molecular dynamics simulations and the employed force fields to describe urea-induced protein denaturation with sufficient accuracy. To this end, microsecond simulations of the Trp-Cage mini-protein are performed which represent the first successful simulations of denaturant-induced protein unfolding on experimental time scales. Finally, Chapter 9 aims to put the results obtained for urea into a more general context. To this end, the molecular interactions of amino acids with urea are compared to those with other denaturing or protective osmolytes, and the transferability of the results for urea is discussed. This Chapter leads naturally to the Conclusions and Outlook Chapter.

2

Protein Structure, Folding, and Unfolding — A brief overview

This chapter introduces some basic concepts of protein structure, folding, and denaturation. Since the main aim here is to provide a background for the study at hand, most topics are addressed only briefly. For further reading, the reader is referred to common textbooks^{37,38} or review articles about protein folding and denaturation.^{39,31,40,41,42,43,44,45,46,47,48,49,50,51,52} A comprehensive account on the history of protein research is given in Ref. 53.

2.1 Protein structure

Proteins can be divided into three main classes. *Globular proteins* are the most numerous of cellular proteins. Since they natively occur in the aqueous intracellular environment or in the plasma, they are generally soluble in water. In terms of function, globular proteins are the most diverse class, and most regulative and enzymatic functions are performed by globular proteins. *Fibrous proteins* usually have a very rigid, often rod-like, structure, and hence their functions are often related to cellular structure. They are generally not soluble in water. *Membrane proteins* occur in lipid membranes, where they typically serve as channels for ions, water, or other molecules, or as receptors. Because they natively exist in the hydrophobic environment of the lipid membrane, they are generally not soluble in water.

Chemically, proteins are linear chains of covalently bonded amino acids. All proteins are built from the same set of 20 naturally occurring amino acids. Each of these amino acids consists of two parts — the backbone, which is identical¹ for all amino acids and links adjacent amino acids via a peptide bond — and the sidechain, which gives each amino acid its characteristic physicochemical properties. The sequence of amino acids is encoded in the genetic information. It unambiguously identifies a protein, and determines its complex three-dimensional structure, and thus also its function.

Protein structure is organized hierarchically, and four different structure levels are distinguished:

The *primary structure* defines the sequence of amino acids. During folding or unfolding, it is not usually not changed. Exceptions are given by posttranslational modifications, like, for instance, backbone-cleavage or formation of disulphide bridges, where covalent bonds are formed or broken after protein synthesis.^{54,40} These processes are usually very system-specific and therefore not addressed in general folding or unfolding studies.

The *secondary structure* level describes local structure motifs. Secondary structure formation is regarded to be driven by two main forces: first, the need for compactness due to the hydrophobic effect,^{55,56,57} and second, satisfaction of the hydrogen bond sites of the backbone.^{58,59,60,61} Conformations that fulfill these requirements very effectively are the common α -helix and β -sheet motifs. Formally, secondary structure elements are defined by

¹with the exception of proline, where the sidechain is not only linked to the C $_{\alpha}$ atom, but also to the N atom

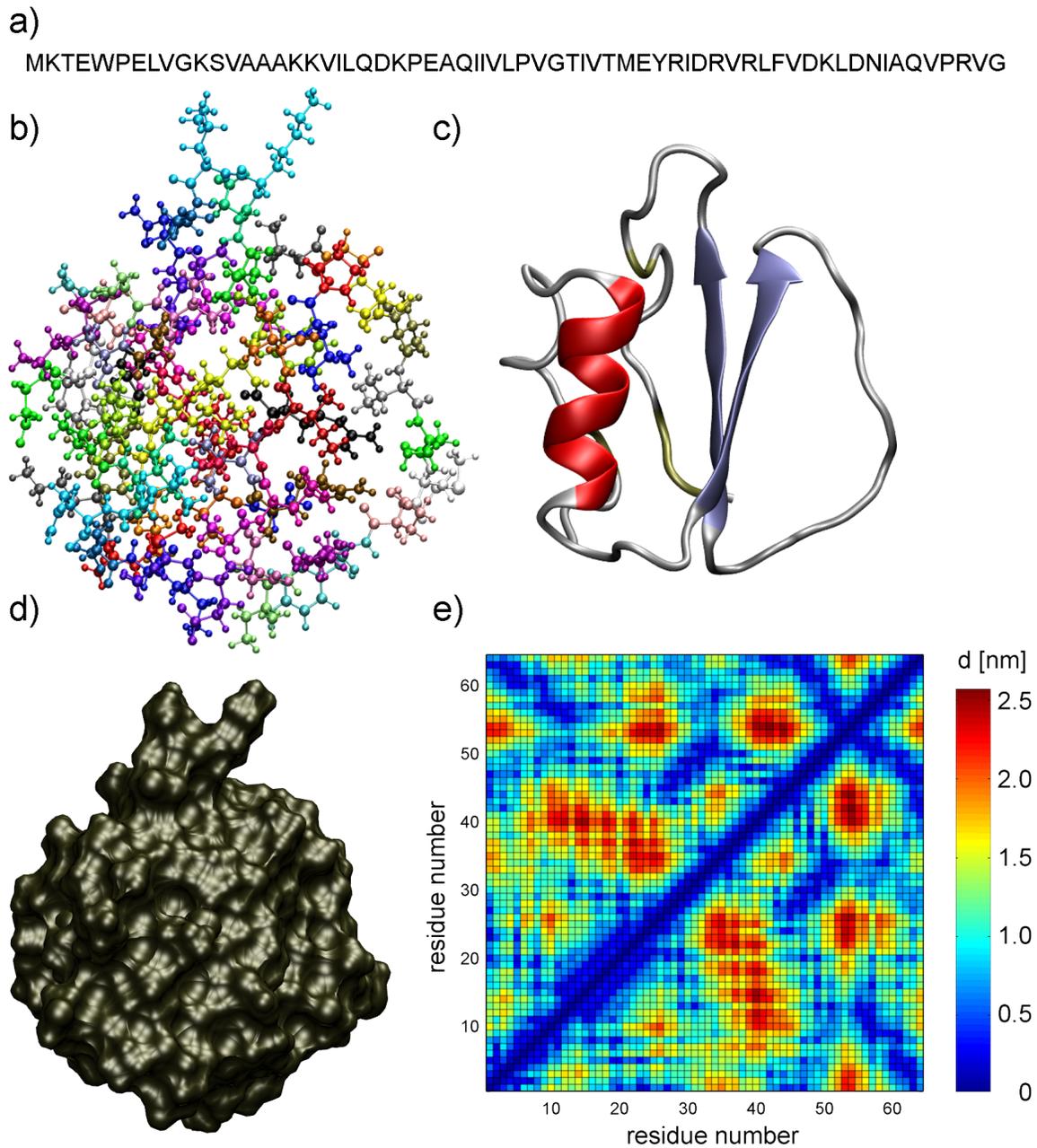


Figure 2.1: Different levels of protein structure, illustrated for the CI2 protein. (a) the letter code denotes the sequence of amino acids, the *primary structure*, (b) all-atom representation of the protein, (c) cartoon representation which highlights *secondary structure* elements (red: α -helix, blue: β -sheet), (d) surface representation to illustrate the compactness, (e) native contact map.

characteristic hydrogen bond patterns and dihedral angle configurations of the backbone. The standard method to assign secondary structure when atomic coordinates are given is the DSSP algorithm by Kabsch and Sander.⁶² The formation or loss of secondary structure elements is one of the main structural changes in folding or unfolding, respectively. Hence, the fraction of native secondary structure content can be used as a reaction coordinate for folding-unfolding transitions.

The *tertiary structure* is a description of the global three-dimensional shape and the way the whole peptide chain is spatially arranged. Forces stabilizing the tertiary structure are the hydrophobic effect, salt-bridges, hydrogen bonds, van-der-Waals dispersion forces, and disulfid bonds. Since the late 1980s, the hydrophobic effect is regarded to be the dominant driving force for tertiary structure formation in globular proteins.^{39,3} Indeed, a common feature of globular proteins is the existence of a hydrophobic core, where less polar residues are buried and excluded from solvent contact. Despite this common feature, thousands of different protein folds and families are known today.^{63,64,65,66,67} One common way to define the tertiary structure of a protein is via the pairwise distances of all its residues, which can be visualized in a contact map (Fig. 2.1e). Since the formation and loss of tertiary structure are the main structural changes during folding and unfolding, respectively, contact maps can be used to define reaction coordinates for folding-unfolding transitions.

The *quaternary structure* applies only to protein complexes which consist of two or more individual peptide chains, which are non-covalently linked to form a complete unit, and describes their spatial arrangement. In folding and unfolding transitions, changes on the quaternary structure level comprise the association and dissociation, respectively, of the individual proteins, and their correct relative placement. However, protein complexes are rather used to investigate protein-protein interactions, instead of general mechanisms of folding and denaturation. In particular, no protein complexes are regarded in this work, and hence, quaternary structure will not be addressed. Yet, since the driving forces for tertiary structure formation are also relevant for protein association, the effects of denaturants on tertiary structure might, to some extent, be transferable to quaternary structure.

2.2 Protein folding and unfolding — a brief overview

After a protein is synthesized, the disordered, linear chain adopts the complex and unique three-dimensional shape which is characteristic for the protein, and essential to its biological activity. The main driving force for the folding of globular proteins is the hydrophobic effect — the “oildrop-effect”: the hydrophobic side chains of less polar amino acids reduce the entropy of solvation water molecules. Hence, water contact of these residues is energetically unfavorable, such that they aggregate to form the *hydrophobic core* of the protein — in essence, hydrophobic residues become buried and excluded from solvent contact, while polar and charged residues are located mainly on the protein surface. A priori, this hy-

drophobic collapse leads to a rather disordered state, called *molten globule*. Subsequently, the protein reorganizes to eventually adopt its native conformation — the folded state.

As Levinthal pointed out, the conformational search for the well-defined folded state would require astronomical time scales if the protein was to randomly sample all possible conformations.² Yet, folding occurs typically within millisecond to seconds. Consequently, the folded process is directed. A common way to illustrate this concept is the “folding funnel” of the free energy landscape: the conformational search of the protein for the minimum of the free energy landscape, the native state, is guided by a gradient of the free energy, which enormously accelerates the process. The mechanism of this efficient search, however, is not yet revealed and has been puzzling scientists for decades.

Aside from this question about the folding pathway and kinetics, another basic question of primary importance addresses the relation between sequence and structure. The structure of the protein is defined by its amino acid sequence,¹ but how exactly does the sequence of amino acids define the characteristic three-dimensional shape, the fold of the protein? Recent studies suggest that the sequence of hydrophobic or hydrophilic amino acids is the main determinant for the fold. In particular, mutation studies where hydrophobic/hydrophilic amino acids were replaced by other hydrophobic/hydrophilic amino acids have shown a significant degree of similarity in the folds of the mutations.³

But folding can also go wrong: If the folding process does not proceed correctly, the protein can adopt a metastable non-native conformation which prevents the protein from correctly fulfilling its biological function. Since this *misfolding* is the molecular origin of numerous serious diseases (which include Alzheimer’s disease, Parkinson’s disease, Creutzfeldt-Jakob, type II diabetes, and various forms of cancer), the topic of protein folding is also highly relevant from a medical perspective.

In vivo, folding is even more complex. Here, interactions with other macromolecules, like other proteins, e.g., can complicate the folding process. In particular, misfolding is more likely to happen under these conditions. To prevent proteins in the cellular environment from misfolding, nature has invented *chaperones* — proteins which have the specific function to assist the correct folding process of other proteins. While the working mechanism of chaperones in general is not yet well understood, some chaperones are known to assist folding by providing an isolated volume for the protein, and to shield it from interactions with other proteins.⁶⁸

The investigation of folding processes poses considerable challenges to experiments and simulations. In particular, lab experiments lack the atomic detail or time resolution to study folding at the molecular level. For computer simulations, on the contrary, the processes are usually still too slow or too rare to be accessible at atomic detail. In the past few years, however, several studies have successfully bridged this gap for *ultrafast folders*, mini proteins like BBA5,⁶⁹ Trp-Cage,^{70,71,72} Trp-Zipper⁷³ or villin headpiece.^{74,75} With these, also the speed limit of folding processes is explored, since folding can be a barrierless downhill process for these small proteins, and folding times below the microsecond are reached.^{76,77}

Since the folded state is only marginally stable under physiological conditions, protein unfolding² can be initiated by even small changes in the conditions. Hence, a variety of ways exist to unfold proteins. A priori, folding and unfolding are reversible processes, but they can become irreversible under some conditions — for instance, when an egg is cooked. Here, crowding effects with other unfolded proteins result in a “sticky mass” of unfolded proteins and render unfolding irreversible. Aside from high temperatures, surprisingly, also low temperature can induce unfolding. While the heat denaturation can be explained relatively easily by the increased entropy gain of the flexible, unfolded protein with respect to the folded protein, the mechanism of cold denaturation is more complex. The current view is that the entropic cost for water to solvate hydrophobic residues, and thus the hydrophobic effect, is reduced at lower temperatures.⁷⁸

The complex interactions of proteins with destabilizing or stabilizing osmolytes, denaturants or counter-denaturants, respectively, are not well understood at all, and are a topic of intensive ongoing research — as in this thesis. Finally, also the interactions of salts with proteins are highly interesting, but not well understood. The effect of ions on protein stability is known to follow a Hofmeister-series,⁷⁹ but the molecular origin of these effects is still not well understood. In particular, guanidinium salts have very heterogeneous effects on protein stability: GdmSCN is a stronger denaturant than GdmCl, and (Gdm)₂SO₄ has no denaturing effect at all.^{80,81}

The interactions of proteins with other denaturants than urea, or with counter-denaturants, will be addressed in Chapter 9.

²The terms *unfolding* and *denaturation* are sometimes used with slightly different meanings. Loss of the native protein structure is termed *denaturation*, while *unfolding* additionally implies a conformational change to a (significantly) less compact structure. This differentiation is not necessary in this work, and the terms are used synonymously.

3

Molecular Dynamics Simulations

All investigations in this thesis were performed using the method of molecular dynamics (MD) simulations. Therefore, this chapter describes the general methodology for MD simulations and trajectory analysis used in this work. Additional methods relevant in single approaches only will be described in the respective chapters.

Molecular dynamics (MD) simulations describe molecular movement on the atomic level. The first MD simulation of a protein (BPTI), published in 1977,⁸² covered a simulation time of 8.8 ps without explicit treatment of the solvent environment. Since then, much progress has been made in terms of computational power and methodology, such that state-of-the-art simulations cover simulation times up to a microsecond for small proteins or describe nanosecond dynamics of huge systems with more than 10^9 interacting atoms. The basic principle of MD simulations, however, is unchanged and will be briefly described below. A detailed description and review of current techniques is given in Refs. [83](#), [47](#), [84](#), [85](#).

3.1 Principle

An accurate treatment of atomic motions would imply solving the time-dependent Schrödinger equation for both, the nuclei and the electrons. However, this is computationally far too expensive even for small systems of more than two or three atoms. The following three fundamental approximations are therefore made to drastically reduce the computational effort:

1. *Separation of electronic and nuclear degrees of freedom*
(Born-Oppenheimer approximation)

The dynamics of any physical system is described by the time-dependent Schrödinger equation,

$$H\psi = i\hbar \frac{d\psi}{dt} \quad , \quad (3.1)$$

where H denotes the Hamilton-operator, ψ the wave-function, and \hbar the reduced Planck constant. Here, the wave-function ψ is a function of the positions of all electrons and nuclei. Due to the much smaller mass and thus higher velocity of the electron with respect to the nucleus, changes in the electronic configuration can be regarded to follow essentially instantaneously the nuclear motion. With this assumption, the Born-Oppenheimer approximation,⁸⁶ the wave-function ψ can be separated into a nuclear part ψ_n and an electronic part ψ_e ,

$$\psi = \psi_n + \psi_e \quad , \quad (3.2)$$

and Equation [3.1](#) separates into two parts. The electronic dynamics are described by a time-independent Schrödinger equation for given positions of the nuclei. The nuclear dynamics are described by a time-dependent Schrödinger equation within the quasi-stationary electronic potential.

2. *Approximation of the electronic potential by a classical force field*

The Born-Oppenheimer approximation allows to describe the electronic configuration of a system with a time-independent Schrödinger equation depending only on the positions of the nuclei. Solving this equation is, however, computationally still not feasible. Therefore, the electronic potential is approximated by a computationally manageable analytic function, the so-called force field $V(r_1 \dots r_N)$, which also depends on the positions of the nuclei. Typical force fields are composed of functionally simple terms describing forces between covalently bonded atoms (bond-stretching and bond-angle vibrations, extraplanar motions, dihedral rotations) as well as longer ranged forces acting between all pairs of atoms (Coulomb interaction, Pauli-repulsion, and van der Waals interactions). The Pauli-repulsion and van der Waals interaction are commonly described by a combined Lennard-Jones potential. Figure 3.1 illustrates these force field terms.

3. *Classical description of nuclear dynamics*

Within the Born-Oppenheimer approximation, the nuclear dynamics are described by a time-dependent Schrödinger equation within the electronic potential. But even with an approximation of the electronic potential by simple force field functions, the computational effort for this calculation is still so large that only few particles can be described. To overcome this limitation, the dynamics of the nuclei are described in classical approximation, i.e., solving the time-dependent Schrödinger equation is replaced by solving Newton's equations of motion,

$$m_i \frac{d^2 r_i(t)}{dt^2} = -\nabla_i V(r_1, \dots, r_N) \quad , \quad (3.3)$$

where m_i and r_i denote the mass and position of the i -th nucleus ($i = 1, \dots, N$), N is the number of atoms, and $V(r_1, \dots, r_N)$ is the electronic ground state energy, approximated by the force field.

These three approximations enable one to simulate the dynamics of large systems such as a solvated protein. The resulting classical description is sufficient to describe the protein dynamics in many kinds of processes; but there are also limitations. In particular, the formation or breaking of covalent bonds can not be described. In protein folding/unfolding transitions, however, no covalent bonds are formed or broken¹, such that the classical description was appropriate for the study at hand. Note that for the description of processes that involve quantum effects, like enzymatic reactions, quantum/classical hybrid simulations ("QM/MM") can be performed, which treat the large part of the system classically, and a small subsystem (e.g., the reactive part of an enzyme) quantum mechanically.⁸⁷

¹as mentioned in the previous chapter, an exception is given by posttranslational modifications, like, for instance, formation of disulphide bridges or backbone-cleavage in some proteins

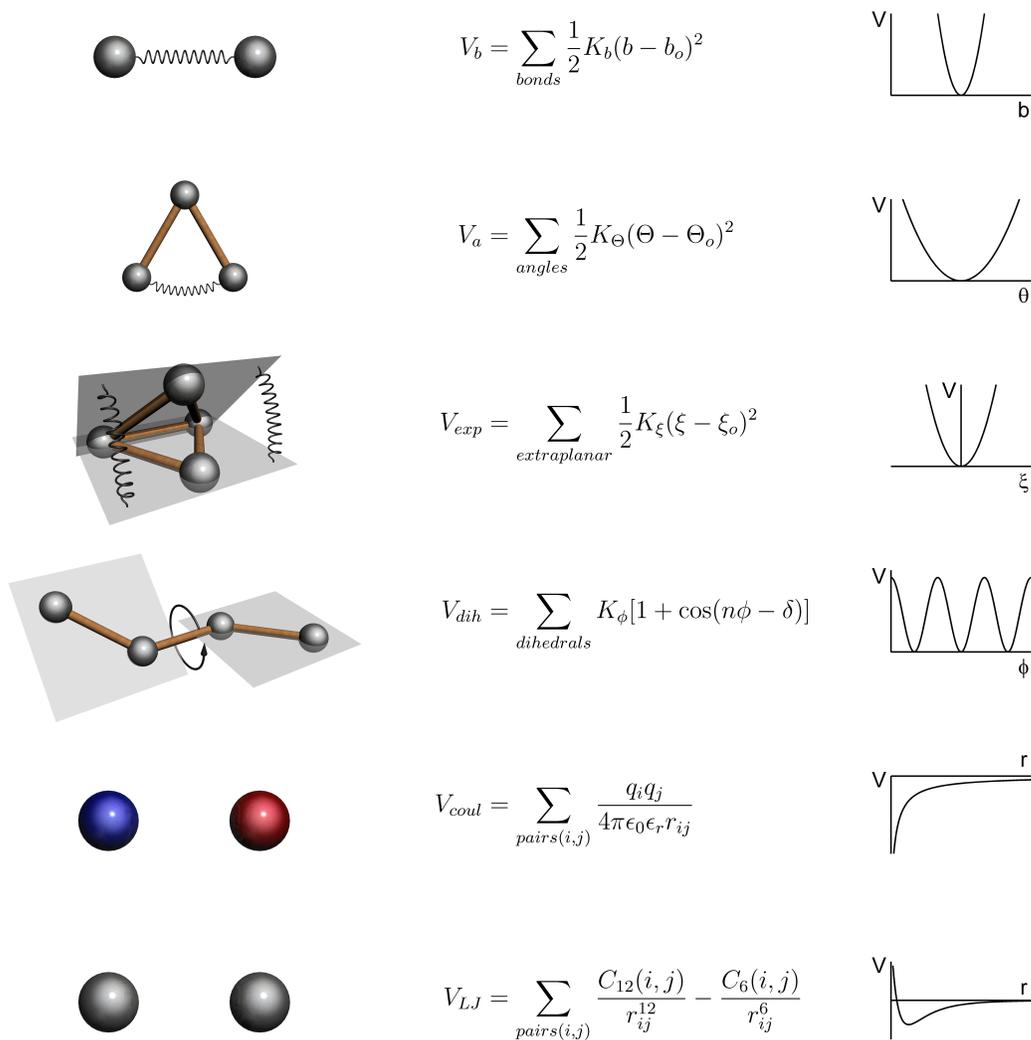


Figure 3.1: Energy terms that constitute a force field for molecular dynamics simulations; from top to bottom: bond-stretching, bond-angle vibrations, extraplanar motions, dihedral angles, Coulomb interactions, and Lennard-Jones interactions. The interactions are illustrated on the left, the corresponding energy terms are shown in the middle, and sketched on the right.

3.2 Simulation Details

The software package *GROMACS*^{88,89,90} was used for all simulations in this work. It contains efficient implementations of methods which will be described briefly.

Integrator. With the approximations described above, the numerical integration of Newton’s equations of motion is the essence of MD simulations. Several algorithms exist, e.g., Runge-Kutta,⁹¹ Leap-Frog,⁹² or Verlet.^{93,94} The latter two have the advantage over the Runge-Kutta algorithm that the forces have to be calculated only once per integration step. In this work, the Leap-Frog algorithm has therefore been used. It iteratively calculates positions x at time t and velocities v at time $t - \frac{\Delta t}{2}$ from the forces $F(t)$:

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m} \Delta t \quad , \quad (3.4)$$

$$x(t + \Delta t) = x(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t \quad . \quad (3.5)$$

An upper limit for the integration time step Δt is imposed by the fastest motions of the system, which typically are the bond-stretching vibrations of hydrogen atoms with periods of 10–20 fs. To avoid numerical instabilities, the time step must be at least one order of magnitude smaller than this period, leading to a time step of 1 fs. However, these high frequency and low amplitude oscillations are almost exclusively in their vibrational ground state, and are therefore better represented by holonomic constraints than by a harmonic potential.⁹⁵ For this purpose, several algorithms exist (e.g., SHAKE,⁹⁶ LINCS⁹⁵) which rescale the bond length to the equilibrium value. LINCS is known to be faster than SHAKE, with the same accuracy,⁹⁵ and therefore was used in this work. With these fastest motions constrained, the next fastest motions are bond-angle vibrations with a typical period of about 20 fs. Hence, the integration time step can be increased to 2 fs, such that the use of constraints leads not only to a more correct description of the dynamics but also to improved efficiency.

Initial conditions. To describe the dynamics of a molecule, proper initial conditions for the positions and velocities of all atoms have to be specified. While the initial velocities can be randomly assigned from a Maxwell-Boltzmann distribution, the initial positions for each atom, i.e., the initial structure of the protein, have to be specified. Protein structures at atomic resolution can be obtained from the Brookhaven Protein Data Bank.⁹⁷ Today (2007), it contains more than 45000 structures, which are all obtained from X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. The structures for the proteins used in this work, the CI2 (PDB-code 1YPC⁹⁸), the Cold Shock Bc-Csp (PDB-code 1C9O,⁹⁹ and the Trp-Cage (PDB-code 1L2Y¹⁰⁰) were also obtained from this data bank.

Solvent environment. The native environment of globular proteins, such as the three used in this work, is water. Because the solvent environment strongly affects the structure, energetics and dynamics of proteins and polymers,^{101,102,103,104} the proteins are simulated here in a physiological solvent consisting of water and sodium chloride ions at a physiological concentration of 150 mM. The main effects of the solvent include the hydrophobic effect, friction, and the dielectric shielding of charges due to the (orientational and electronic) solvent polarizability.

Since the inclusion of a sufficient number of solvent molecules increases the total number of particles typically by a factor of ten, and hence the computational effort significantly, implicit solvent models have been developed to mimic the different effects of the solvent, e.g., via mean-field approximations, without explicit simulation of the solvent molecules.^{105,106} However, the implicit treatment of cosolvent systems, like aqueous urea solutions, is not yet established, and certainly not possible without an in-depth understanding of the properties of these solutions — which is but one goal of this thesis (Chap. 4). Additionally, the ability of current implicit solvent models to describe already the effect of water alone with sufficient accuracy for folding studies is questionable.^{107,108} Therefore, all simulations in this work were performed with explicit solvent.

As for the proteins atoms, initial conditions have to be chosen for the solvent atoms. Velocities are again randomly assigned from a Maxwell-Boltzmann distribution. Since structural and energetic properties of the solvent are usually converged within a few hundred picoseconds, the choice of atomic positions is not as crucial as for the protein. In this work, equilibrated solvent configurations were generated in the simulations described in Chapter 4, which were subsequently used as templates for the simulations with peptides and proteins.

Energy minimization and Equilibration. The atomic positions for protein and solvent atoms obtained from this procedure are not yet suited to be used as starting structure for simulation. First, protein structures resolved by X-ray crystallography and NMR spectroscopy have a limited resolution of typically $\approx 1.0 - 3.0$ Å. Thus, these structures can suffer from van der Waals overlaps or bond-angle deformations. Second, van der Waals overlaps might also be introduced by the successive addition of solvent molecules and ions. The resulting unphysically high forces could destabilize the whole simulation system. To avoid this, all simulation systems in this work were subjected to a short energy minimization prior to the MD simulation using a steepest descent method. The resulting changes in atomic positions were very small (on the order of $10^{-2} - 10^{-1}$ nm), but in some cases decreased the maximum forces by several orders of magnitude, mainly due to the r^{-12} dependence of the Lennard-Jones potential.

After energy minimization, the system configuration can still contain unphysical arrangement of atoms or molecules. For instance, solvent molecules adjacent to charged groups or ions may not be properly oriented, or protein side chains may have non-native orientations resulting from low temperature or other experimental conditions. Therefore, a short

equilibration MD run was performed for each system prior to the actual production run for data collection. The solvent systems in Chapter 4 were equilibrated for 0.5 ns. For the simulations including peptides or proteins, equilibration was performed in two phases: first, a 0.5 ns solvent equilibration with position restrains on the peptide backbone atoms, and second, a 1 ns equilibration without position restrains.

Treatment of non-bonded interactions. The two non-bonded force terms acting between all atom pairs (Coulomb and Lennard-Jones interactions) scale with $\mathcal{O}(N^2)$ and make up the most time-consuming part of the force calculation. Hence, the computational effort can quickly become too large to handle for a large system like a protein in explicit solvent environment. Methods have been developed to reduce the computational effort for these forces. Those that have been applied in the present work will be briefly introduced here.

The Lennard-Jones potential declines with r^{-6} , and can therefore be neglected for sufficiently large r . Typically, only interactions between particles within a “cut-off” distance of ≈ 1 nm are taken into account.¹⁰⁹ This approach has also been followed for the simulations described in this work.

The treatment of the long-range Coulomb-interaction is not as easy. Since the potential declines with only r^{-1} , the use of a cut-off method with a computationally feasible cut-off distance is a too rough approximation that is known to cause severe artifacts in the structure and dynamics of proteins.^{110,111,112}

To avoid such artifacts, *Ewald summation*¹¹³ was used. This method, which is based on calculation of forces on a grid in reciprocal space, was first introduced to calculate long-range interactions of the periodic images in crystals. Therefore, it is particularly suited to describe systems with periodic boundaries (see below). An improved extension of *Ewald summation* is the *particle-mesh Ewald* (PME)^{114,115} method, which scales with $\mathcal{O}(N \log N)$. For the simulations in this work, Lennard-Jones and Coulomb interactions between pairs of atoms within a distance of 1.0 nm were calculated explicitly, and the PME method was used for the electrostatic interactions above 1.0 nm.

System boundaries. In MD simulations of biological systems, the simulation system has to be many orders of magnitude smaller than it is in corresponding experiments or in nature to be computationally manageable. Hence, appropriate boundary conditions have to be applied to minimize artifacts arising from the system boundaries, and to prevent evaporation of the molecules out of the simulation box. A commonly used method for this purpose is the application of periodic boundary conditions. Here, an atom which leaves the system on one side of the simulation volume enters it simultaneously on the opposite side. This way, the simulation system is surrounded by periodic images of itself, and does not have any surface at all. This method is limited to system shapes that can space-fillingly be surrounded by translational copies of itself. In particular, the unit cell can be a cuboid,

a dodecahedron, or a truncated octahedron. While the use of periodic boundaries is an effective approach to eliminate surface artifacts, other artifacts may arise from the artificial periodicity.^{116,117,111} These artifacts are reduced by increasing the box size. In this work, rectangular simulation boxes were used and the size of the box was chosen such that a minimum distance of 1.5 nm was kept between the box boundary and the nearest protein atom.

Temperature and pressure coupling. By integrating the Newtonian equations of motion, energy is conserved and hence, the system is sampled in the microcanonical ensemble. In experiments or *in vivo*, however, proteins are coupled to the large heat reservoir of their solvent environment. Hence, it is more realistic to sample configurations from a canonical ensemble instead, with constant temperature. For this purpose, several methods have been developed to couple the simulation system to a thermostat and mimic the effect of a heat reservoir.^{118,119,120} The coupling to a heat reservoir has an additional advantage: the numerical integration of the equations of motion and the approximated treatment of long-range non-bonded forces introduce random forces by discretization and rounding effects, which heat up the system. With a temperature reservoir, this heating is prevented. In this work, the *Berendsen thermostat*¹¹⁸ was used, which adjusts the simulation temperature T to the reference temperature T_0 of the heat bath by rescaling the atomic velocities according to

$$v' = v \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right)} \quad , \quad (3.6)$$

where Δt is the integration time step and τ_T the coupling constant, which was set to 0.1 ps here.

In addition to the heat bath coupling, biological systems are subjected to a constant pressure of usually 1 atm. To mimic this condition, the simulation system is coupled to a barostat. Here, the *Berendsen barostat*¹¹⁸ was used, which controls the pressure by rescaling atomic positions analogous to the velocity scaling for the temperature control. Pressure coupling was used with a coupling constant of 0.1 ps for all simulations.

Software and Hardware. All simulations were performed using the software package *GROMACS*^{88,89,90} (versions 3.1.4, 3.2.1, and 3.3) on *LINUX* clusters with *Opteron* and *Xeon* processors. The computation time for 1 ns dynamics of a typical system, for example, in the simulations of the CI2 protein presented in Chapter 6, was 4 days on a single CPU. The computation time can be decreased by performing the calculation on several processors in parallel. However, the benefit of parallelization is limited by the necessary data exchange between the CPUs. Dual- and Quadro-CPU architecture was used to achieve optimal scaling with 2 or 4 CPUs, respectively. For some very extensive simulations (e.g., the 500 ps simulations of the CI2 in Chap. 6, or the microsecond-simulations of the Trp-Cage in Chap. 8), fast myrinet network architecture was employed to enable the parallel use of

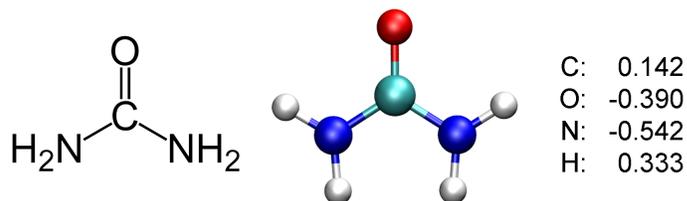


Figure 3.2: Chemical structure of urea, and partial charges of the employed urea force field.

up to 12 CPUs without significant performance loss due to network communication. This way, most of the protein simulations could be performed in about one year.

Force Field. A variety of different force fields have been developed (e.g., CHARMM,¹⁰⁹ GROMOS,¹²¹ AMBER,¹²² OPLS,¹²³ OPLS-AA¹²⁴) for the description of different classes of molecules, like, for instance, proteins, DNA, or carbohydrates. A recent comparison of different force fields is given in Ref. 125. While no force field can simply be ranked to be “the best”, some physical and chemical properties are described better by certain force fields than by others, often depending on the different ways the force fields have been parameterized. Thus, the choice of force field depends on the particular question at hand and property of interest.

In this work, the OPLS-AA force field,¹²⁴ which has been developed especially for proteins, was used for all simulations. It contains explicit description of every atom, in contrast to “united-atom” force fields (e.g. GROMOS, OPLS), which have a combined single-particle representation for small subgroups like CH_2 or CH_3 groups. While united-atom force fields offer the advantage of fewer particles, and thus less computational effort, an all-atom force field was nevertheless used here, since for the investigation of molecular interactions between urea and proteins, atomic detail is regarded to be imperative. The TIP4P¹²⁶ water model and the urea-model of Smith et al.,¹²⁷ which is a flexible version of the original OPLS model by Duffy et al.,²² were employed. The partial charges of the urea model are shown in Fig. 3.2; those of the OPLS-AA force field for all amino acids are shown in the Appendix.

3.3 Relevant Observables — Trajectory Analysis

The trajectory of a simulation specifies the positions and velocities of every atom of the simulation system at every time step. From this enormous set of data, the interesting observables can now be calculated. However, the identification of relevant observables or the choice of suitable reaction coordinates for the description of conformational transitions is not always trivial or unambiguous. In particular, as explained in the previous chapter, folding/unfolding transitions affect multiple different structural aspects of the protein.

Hence, several different observables are used in combination here, to address different structural aspects.

The observables used in this work will be introduced below. First, observables are described which are commonly used for the analysis of protein conformational changes. Subsequently, observables and methods for analysis which were specifically introduced in this work will be described.

Root-mean-square-deviation. The root-mean-square-deviation (rmsd) of a structure X (atom coordinates \vec{x}_i) with respect to a reference structure X^0 (atom coordinates \vec{x}_i^0) quantifies the structural difference between X and X^0 :

$$rmsd = \min_{\{T,R\}} \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \vec{x}_i^0)^2} . \quad (3.7)$$

The minimum is taken for all translations T and rotations R , i.e., structure X is “fitted” to structure X^0 . Over simulation time, the rmsd $X(t)$ of a protein with respect to the starting structure (crystal structure) X^0 indicates conformational changes. Typically, an rmsd of $\approx 2 - 3$ Å is caused by the thermal fluctuations, whereas larger values correspond to conformational changes. For the structural characterization of folding/unfolding transitions, the conformation of the backbone alone — the fold — is of relevance. Hence, the rmsd was calculated in most cases for the backbone- or C_α -atoms only. Further, it can be instructive to calculate the rmsd only for a subset of atoms to locate areas of structural changes.

While the rmsd is an observable which is generally used to describe conformational transitions, the next three observables are more specific to folding/unfolding events, because they directly address the different structural aspects of folding/unfolding transitions.

Solvent accessible surface area. The hydrophobic effect is nowadays regarded to be the main driving force for folding. Since the strength of the hydrophobic effect approximately scales linearly with the exposed hydrophobic surface, this surface is expected to change significantly during folding/unfolding. Hence, the hydrophobic part of the solvent accessible surface area (the “SAS”), is particularly important for the description of (un)folding processes. Usually, not the total SAS but only the solvent accessible *hydrophobic* surface area is used. If not noted otherwise, SAS refers to the hydrophobic part of the surface area in this thesis. For the calculation of the SAS, the double cubic lattice method¹²⁸ was used with a 0.14 nm probe radius.

Secondary structure. As explained in the previous chapter, changes in the secondary structure of a protein are one of the main structural changes during folding/unfolding,

and hence, secondary structure is an important observable here. For the determination of secondary structure, the *GROMACS* tool *do_dssp* was used, which employs the commonly used DSSP classification (see Chap. 2). The *native secondary structure* of a protein can be defined either as that of the crystal structure or as the average secondary structure in a reference simulation of the native state. Here, the latter has the advantage that fluctuations of the native state are captured, which allows a more direct comparison with the unfolding simulations. Therefore, reference simulations of the folded state were performed for each protein to define the native secondary structure.

A suitable reaction coordinate for (un)folding on the level of secondary structure is the *fraction of secondary structure content*, which is defined as the fraction of all residues that are in their native secondary structure configuration. For this, only clearly defined secondary structure elements like helices, β -sheets, and turn-elements were considered, in contrast to, e.g., random-coil or bend elements.

Native contacts. The contacts between proteins residues are, as described in Chapter 2, a means to characterize tertiary structure. Residues were defined to be in contact if the distance between the closest pair of atoms was not larger than 0.4 nm. In analogy to the definition of native secondary structure, contacts can be defined as native if they are present in the crystal structure or if they are present more than 50% of the time in a reference simulation of the native structure. For the same reasons as described above, the latter approach was used.

The fraction of native contacts, often labelled “Q”,¹²⁹ can be used as a reaction coordinate for (un)folding which reflects tertiary structure changes². It is defined as the number of native contacts between residues divided by the number of contacts in the native state. Trivial contacts between residues (n to n and n to $n \pm 1$) were not considered for the calculation of the native contact fraction.

Hydrogen bonds. As implemented in the *GROMACS* tool *g_hbond*,¹³⁰ hydrogen bonds were defined from the geometry between acceptor A, donor D, and the hydrogen H. In particular, a hydrogen bond was regarded to exist between a pair of possible donor atom D and acceptor atom A if the distance between D and A was not larger than 0.35 nm and the angle between D–A and D–H, where H is the hydrogen atom, was not larger than 30°.¹³⁰ Energies of hydrogen-bonds were estimated using the empirical function of Espinosa et al. $E = -2.5 \cdot 10^{-4} \cdot e^{-36 \cdot d}$ kJ/mol,¹³¹ where d denotes the distance between hydrogen and acceptor atom in nm.

The estimated energies are for isolated hydrogen bonds and are certainly not identical with the free energy contribution of these hydrogen bonds for proteins in solution.¹³² In this thesis, however, this formula is only used as a semiquantitative measure of the hydrogen

²Since the formation of secondary structure elements also involves formation of residue contacts, the native contact fraction is not only correlated to the tertiary structure level.

bond strength.^{133,134,135,136,137} In particular, conclusions are only based on the monotonic dependence of hydrogen bond energy on distance (in the considered distance range), and not on accurate numbers, such that this simple treatment should suffice.

The methods described so far are commonly employed for the analysis and characterization of protein dynamics. For the investigations in this work, further analysis methods were introduced to gain additional insight.

Contact coefficient. To investigate the interactions between urea and proteins with molecular detail and in a quantitative manner, the “contact coefficient” C_{UW} is introduced here. It quantifies preferences of the individual amino acids for contact with either urea or water molecules. For an amino acid X, the contact coefficient C_{UW} is defined as:

$$C_{UW} = \frac{N_{X-U}}{N_{X-W}} \cdot \frac{M_W}{M_U} \quad , \quad (3.8)$$

where N_{X-U} and N_{X-W} are the numbers of atomic contacts of amino acid X with urea and water molecules, respectively. Atoms were defined to be in contact if closer than 0.35 nm. To achieve concentration-independence, C_{UW} is normalized using the total numbers of atoms belonging to urea molecules (M_U) or to water molecules (M_W) in the system. Hence, an amino acid with a contact coefficient of $C_{UW} = 1.0$ has no interaction preference for either urea or water. Values above 1.0 indicate preferential interaction with urea, values below 1.0 indicate preferential interaction with water. Since C_{UW} rests on interaction frequencies, it relates to the first peak in the radial distribution function, and, therefore, the short range part of the interaction free energy for the contact formation.

The statistical error of C_{UW} values were determined from the autocorrelation time of the instantaneous contact coefficient. The autocorrelation time was found to be ≈ 100 ps for the analysis on the residue level and ≈ 10 ps for the analysis on the atomic level in Chapter 5. Lifetimes of contacts were calculated for comparison, and showed an exponential distribution with a similar time-constant. For all simulations of which contact coefficients were calculated, the autocorrelation time of 100 ps was used to determine the number of statistically independent frames within the simulation, and thus for the statistical error estimate of the average contact coefficients.

Estimation of free energy profiles from a Markov model. For the simulations with proteins (Chapters 6, 7, and 8), the free energy of the protein was estimated as a function of the solvent accessible surface area (SAS). To this end, the following procedure was used:

1. The SAS was discretized into N equidistant bins $B_{1\dots N}$ between the minimum and the maximum SAS observed in the respective simulations.
2. The Markov state M_i was defined as the ensemble of all protein conformations with an SAS in the respective bin B_i .
3. For every time step, the protein was assigned to a Markov state M_i , according to the SAS of that conformation.
4. The times required for transitions between adjacent Markov states were extracted.
5. From the transition times, the corresponding transition rates were calculated with a maximum-likelihood-approach:
 k_i^{up} denotes the “upward transition rate” from state M_i to state M_{i+1} ,
 k_{i+1}^{down} denotes the “downward transition rate” from state M_{i+1} to state M_i .
6. From the ratio of the upward and downward transition rates between state M_i and state M_{i+1} , the free energy difference ΔG between these states was calculated according to

$$G_{i+1} - G_i = \Delta G = -\log\left(\frac{k_i^{up}}{k_{i+1}^{down}}\right)$$

This approach includes some assumptions and approximations. First, for quantitatively accurate calculations, the states $B_{1\dots N}$ had to be Markovian, which is most likely not the case. In particular, since protein conformations are projected here on a one-dimensional reaction coordinate, the states $B_{1\dots N}$ comprise structurally very heterogeneous conformations, and hence, transition probabilities will not be history-independent. This is particularly true for larger SAS values, which have contributions from structurally very heterogeneous conformations.

It is not expected that this approach reproduces accurate free energy values, but rather qualitative features of the free energy landscape. Thus, it will mostly serve to illustrate and support results and from other analyses, rather than be the basis for independent conclusions.

4

Aqueous urea solutions

4.1 Introduction

It has been suggested that urea-induced changes in the water structure are the driving force for protein denaturation, rather than direct interactions of urea with the protein.^{5,6} This Chapter addresses the relevance of these indirect effects of urea on the protein. For this purpose, the structure and energetics of aqueous urea solutions were investigated for the full experimentally accessible range of temperatures and urea concentrations. The approaches taken will be described in detail after a brief overview of the history and current knowledge of the properties of aqueous urea solutions.

Urea water solutions show a number of remarkable properties. In particular, enhanced solubility of hydrocarbons,¹³⁸ decreased micelle formation¹³⁹ and, most importantly, the ability to denature proteins.¹⁴⁰ In the attempt to explain these properties, mainly two models have been proposed in the 1960s on the basis of thermodynamic arguments and still set the framework for ongoing discussions. The SKSS-model,^{141,142,143} proposed by Schellman, Kresheck, Sheraga and Stokes, attributes the properties of aqueous urea solutions to dimerized or oligomerized urea. In contrast, the FF-model,⁶ suggested by Frank and Franks, focuses on the changes in the water network induced by urea and regards it as “structure-breaker” for water. Since the 1960s a wealth of new information has been provided by calorimetry,^{21,24} circular dichroism spectroscopy,^{16,34} neutron scattering,¹⁴⁴ NMR spectroscopy,¹⁴⁵ fluorescence measurements,¹⁴⁶ time resolved optical Kerr effect,¹⁴⁷ IR- and Raman-spectroscopy,^{148,149,150} and a number of molecular dynamics simulation studies.^{151,27,15,127,152,23,153,25,154,7,155,22,26,156,14,157,158,159,160,161,162,163,164,12,165,166,167,168,169,28} For a more comprehensive bibliographical and historical overview, see Refs. 13 and 170.

Today, it is widely accepted that urea exhibits a certain tendency to self-aggregate in aqueous solution.^{152,157} In many other cases, however, experimental data are still somewhat contradictory. One of the most controversial issues is the effect of urea on water structure. In some works (Refs. 5,171,11), urea is suggested to disrupt the natural water structure and is termed “structure-breaker” or “chaotrope”. Others (Refs. 157,7) find urea to enhance the water structure and coined the terms “structure-maker” or “kosmotrope”. Both of the two views attribute the peculiarities of urea to changes in the water structure and are in line with the FF-model. However, others find no or only negligible changes in the water structure and suggest this finding as evidence for the SKSS-model.^{160,172,148,26,156} A number of recent studies also suggest a combination of direct and indirect effects.^{7,24,14,15}

So far, most studies have focussed on radial distribution functions as a means to analyze both, the influence of urea on water structure and urea self-aggregation. These two effects, apparently, are closely related to the question at hand, but in light of the molecular complexity involved, probably fall short of capturing the whole picture. In particular, it is

argued here that not only the distribution functions, but also the energetics of urea-water interactions needs to be investigated in more detail.

These issues will be addressed along three lines. First, the short range structural properties of urea-water mixtures, including the structural characteristics of urea aggregation, will be investigated by calculating generalized density distribution functions including three translational and three rotational molecular degrees of freedom. Further, the degree of urea aggregation will be quantified. Second, the energetics will be addressed in terms of hydrogen bond interactions. Third, structural perturbations imposed by urea molecules on water structure will also be characterized by density distributions. Going beyond previous studies, the temperature and urea concentration dependence will be covered here for the full experimentally accessible range.

To evaluate the accuracy of the employed urea force field, atom-atom distribution functions, calculated from the simulations, will be compared to neutron scattering data. Additionally, mass densities will be compared to experimental values. In summary, a comprehensive picture of urea water mixtures at the molecular level is obtained, which has implications for putative mechanisms of urea-induced protein denaturation.

4.2 Simulation setup and additional methods

4.2.1 Simulation setup

Urea water solutions were simulated for 11 different concentrations ranging from 0 M to 11.3 M, each at six different temperatures ranging from 280 K to 380 K. To set-up the simulation systems for the various urea concentrations, 0 to 163 (non-overlapping) urea molecules were placed at random positions within the simulation box, which was subsequently filled up with TIP4P water molecules using the genbox routine of the *GROMACS* package. All 66 systems shown in Tab. 4.1 were simulated for 10 ns. The total simulation time, including reference simulations described further below, was about 1 μ s.

4.2.2 Calculation of spatial density distributions

To characterize the short range order of urea and water, six-dimensional (three translational and three rotational degrees of freedom) density distributions $\rho(\Delta x, \Delta y, \Delta z, \alpha_x, \alpha_y, \alpha_z)$ were calculated for water/water, water/urea and urea/urea. To this end, the relative position and orientation of molecules with a center of mass distance smaller than 0.5 nm was computed pairwise. The orientation of one molecule with respect to the other was expressed in axis-angle notation, i.e., a three-dimensional vector $(\alpha_x, \alpha_y, \alpha_z)$ denotes the direction of the rotation-axis, and the length of this vector defines the rotation-angle. For example, $(\pi, 0, 0)$ describes a 180° rotation around the x -axis. The coordinate system was defined as shown in Fig. 4.1.

number of water molecules	number of urea molecules	mole fraction	urea concentration [mol/l]					
			280 K	300 K	320 K	340 K	360 K	380 K
895	0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
844	16	0.02	1.0	1.0	1.0	1.0	0.9	0.9
787	33	0.04	2.1	2.1	2.1	2.0	2.0	1.9
730	49	0.06	3.2	3.2	3.1	3.1	3.0	2.9
690	65	0.09	4.3	4.2	4.2	4.1	4.0	3.9
612	81	0.12	5.6	5.5	5.4	5.3	5.2	5.1
567	98	0.15	6.8	6.7	6.6	6.5	6.4	6.3
518	114	0.18	8.0	7.9	7.8	7.7	7.6	7.4
493	130	0.21	9.0	8.9	8.8	8.6	8.5	8.3
437	146	0.25	10.4	10.2	10.1	9.9	9.8	9.6
401	163	0.29	11.5	11.3	11.2	11.0	10.8	10.7

Table 4.1: Setup of the simulation system. Listed are the number of water and urea molecules in the system, urea mole fractions, and urea concentrations. Due to changes in the box volume, the effective concentrations slightly differ with temperature. The saturation limit at 300 K is at a mole fraction of 0.25.

All position histograms were built using 100 bins in each of the three dimensions and smoothed with a three-dimensional gaussian function of 0.01 nm width,¹⁷³ which was chosen as trade-off between resolution and statistical noise. Orientation histograms for subsets of positions were calculated similarly. All histograms were normalized such that the sum over all bins was 1.

4.2.3 Quantification of urea aggregation

The tendency of urea molecules to self-aggregate was analyzed and quantified in terms of the reduction of total water accessible surface exposed by urea (“interface surface”) with respect to non-aggregated urea. To this aim, a sphere of 0.14 nm radius was used to probe the surface with the Double Cubic Lattice Method.¹²⁸ The interface surface area was used as a measure because its minimization is assumed to be the main driving force for the aggregation. Furthermore, this quantity was found to be more sensitive to the size of the contact area than other measures as Kirkwood-Buff integrals¹⁷⁴ or cluster-analysis (data not shown).

The accurate assessment of aggregation is complicated by the fact that also in the absence of any interaction and, therefore, also aggregation, random contacts between urea molecules would already reduce this solvent exposed surface area. To distinguish this “geometric aggregation” from real aggregation, two types of additional simulations were performed. A first set of simulations was carried out with completely uncharged (i.e., super-hydrophobic) urea molecules to maximize the hydrophobic effect and enforce maximal aggregation. The obtained surface area served to define the 100% aggregation level. To preclude artifacts

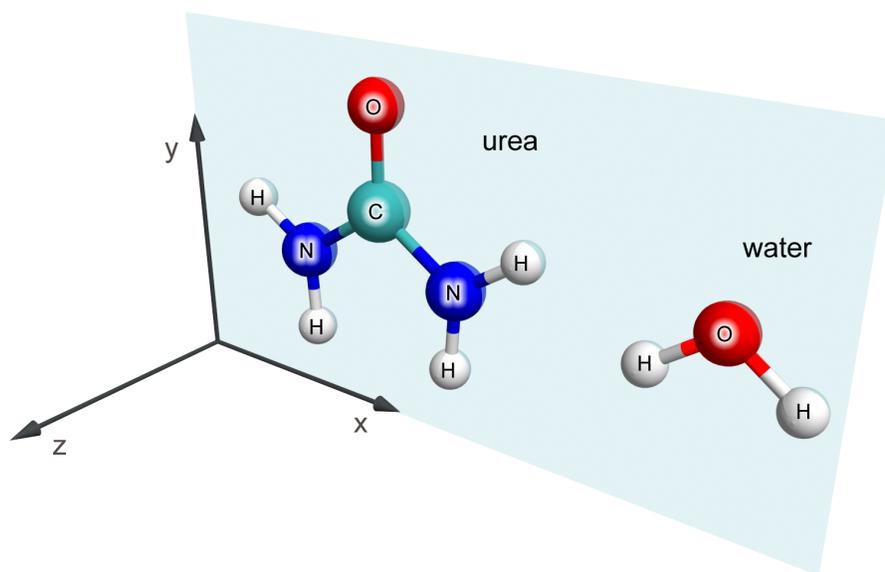


Figure 4.1: Definition of the coordinate system used for calculations of positional and orientational histograms.

caused by periodic boundary conditions (see Appendix 4.5), a boxsize of $(5 \text{ nm})^3$ was used for the higher concentrations.

For a second set of simulations, both, water and urea molecules were completely uncharged to remove any hydrophobic effect from the simulation, such that the opposite extreme of purely stochastic clustering was achieved, which defined the 0% aggregation level. To prevent evaporation of the resulting van der Waals liquid in this case, these simulations were performed under constant volume conditions without pressure coupling.

To assess the statistical accuracy of the estimated interface surface area, its autocorrelation function was calculated, and an autocorrelation time of about 100 ps was obtained. Hence, for each of the 10 ns trajectories, an effective number of 100 independent measurements can be assumed, implying a statistical error of $\frac{1}{\sqrt{N}} = \frac{1}{10}$ of the data obtained from the simulations.

4.2.4 Comparison to experiments, controls

Mass densities were compared to the experimental fitting function presented by Sokolic et al.¹⁵⁴ Radial distribution functions (rdfs) for selected atom pairs were compared to experimentally based data obtained from neutron scattering experiments¹⁴⁴ with empirical potential structure refinement.¹⁷⁵ For this comparison, a simulation was performed with

a mole fraction of 0.2 (500 water molecules, 125 urea molecules) which corresponds to the experimental concentration. The influence of the water model on the coordination geometries was assessed with a pure water simulation at 300 K using the TIP5P model.¹⁷⁶ To detect possible artifacts caused by periodic boundary conditions or finite box size, a larger system of $(6\text{ nm})^3$ boxsize with 508 urea molecules and 5520 water molecules (4.2 M) was simulated at 300 K for 20 ns, and the results obtained from this simulation were compared to those from the $(3\text{ nm})^3$ system of the same concentration.

4.3 Results

4.3.1 Spatial density distributions

To characterize the geometry and short range order of urea-water solutions, spatial density distributions were extracted from 10 ns molecular dynamics simulations. In the following, the results for 300 K and a mole fraction of 0.21 (8.9 M), which is in the common range for denaturation, are presented.

The relative positions of the urea molecules with respect to each other are now discussed. Positions and orientations were determined pairwise, although larger connected networks or clusters of urea were seen, without evidence for dimerization to be preferred over multimerization. Figure 4.2 shows the spatial density distributions within a distance of 0.5 nm from a reference molecule (center of mass). Panels *a* – *c* show the densities in planes through the center of mass of the reference molecule. High densities appear in red, low densities in blue color. The full three-dimensional structure of the preferred coordination positions is shown in panels *d* and *e* via isodensity-surfaces. Panel *f* shows the same data color- and transparency-coded. Five distinct high occupancy regions could be identified. Due to the C_{2v} -symmetry of the urea molecule, only three of these are actually different from each other. According to their positions, these regions were labelled “ σ_x ”, “ σ_y ” and “ σ_z ”, as indicated in Fig. 4.3.

As a next step the orientation relative to the reference urea molecule was examined. The orientational distributions (Fig. 4.3 σ_x , σ_y , σ_z) were relatively localized, implying strong orientational preferences. The highest density corresponds to those orientations that are prevalent at the preferred positions σ_x , σ_y , and σ_z (Fig. 4.3 α , β , γ). The size of the high-density regions can be used to compare the accessible phase space volume and, hence, entropies of the three corresponding conformations.

The orientation histogram of molecules in the σ_x -position (Fig. 4.3 σ_x) exhibits two regions of high density at $(0, 0, \pm\pi)$ as well as two maxima at $(\pm\pi, 0, 0)$. Due to the symmetry of the urea molecule, these four rotations are equivalent and correspond to the same conformation where the urea molecules are rotated by about 180° with respect to the reference molecule around the vector connecting their carbon atoms (z -axis). This conformation (“ α ”) has been termed cyclic urea^{166,160} and is the enthalpically most favorable one, as it is stabilized

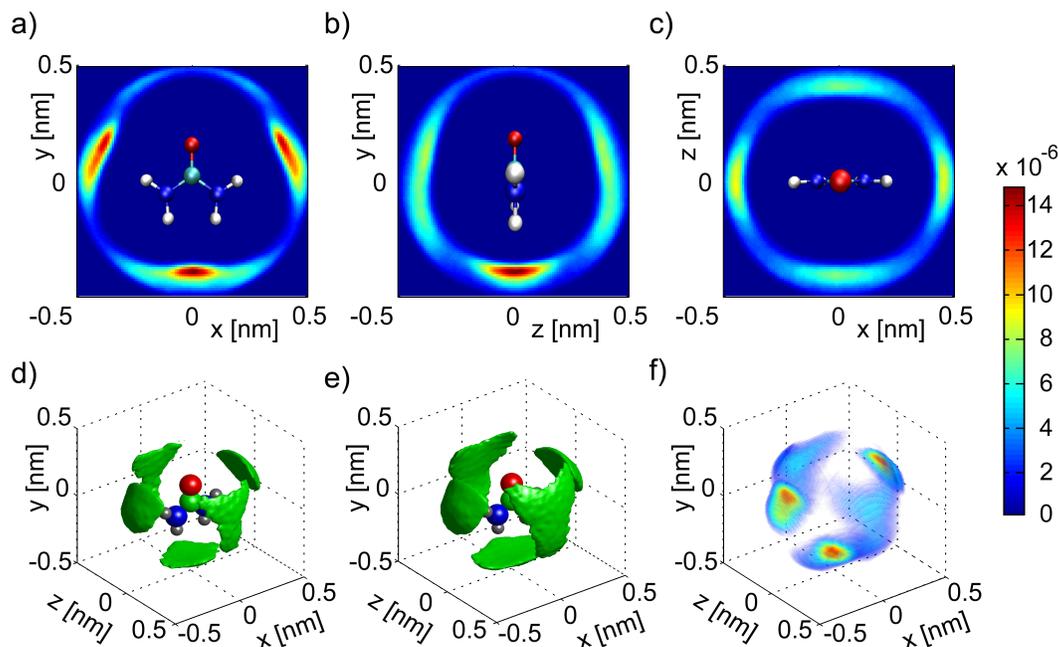


Figure 4.2: Density distributions of urea molecules with respect to a reference urea molecule, a) in the xy -plane, b) in the yz -plane, c) in the xz -plane, d) isosurfaces enclosing 15% of all urea-molecules within a 0.5 nm distance, e) same with 25%, f) color- and transparency-coded density distribution enclosing 50%. Low densities: blue and most transparent, high densities: red and least transparent (see colorbar for a–c,f). Five preferential coordination sites are seen, three of which actually differ from each other.

by two hydrogen bonds (depicted as dotted lines) between O as acceptor and N as donor. Entropically, in contrast, it is the least favorable one of the three conformations, because the two hydrogen bonds restrict the relative mobility of the molecules (as indicated by the short arrows in Fig. 4.3a). Due to the entropic penalty it is expected that this conformation becomes less populated at higher temperature, which was in fact found in the simulations (data not shown).

The two curved volumes in the orientational density distribution of the σ_y -position (Fig. 4.3 σ_y) correspond to a conformation in which the urea molecule is tilted by 90° around the x -axis with respect to the reference molecule and can freely rotate around its z -axis (as indicated by the long arrow in Fig. 4.3 β). This conformation (“ β ”) has weak hydrogen bonds between donor N and acceptors O and N. Due to the free rotability, it is the entropically most favorable one, and its population does not decline at higher temperatures.

Finally, position σ_z (Fig. 4.3 σ_z) shows three curved density clouds for the orientational degrees of freedom. The most populated one also corresponds to conformation β , but here with the other molecule as reference molecule. The other two density clouds represent a third conformation (“ γ ”) in which the urea molecule is rotated by 90° around the y -axis

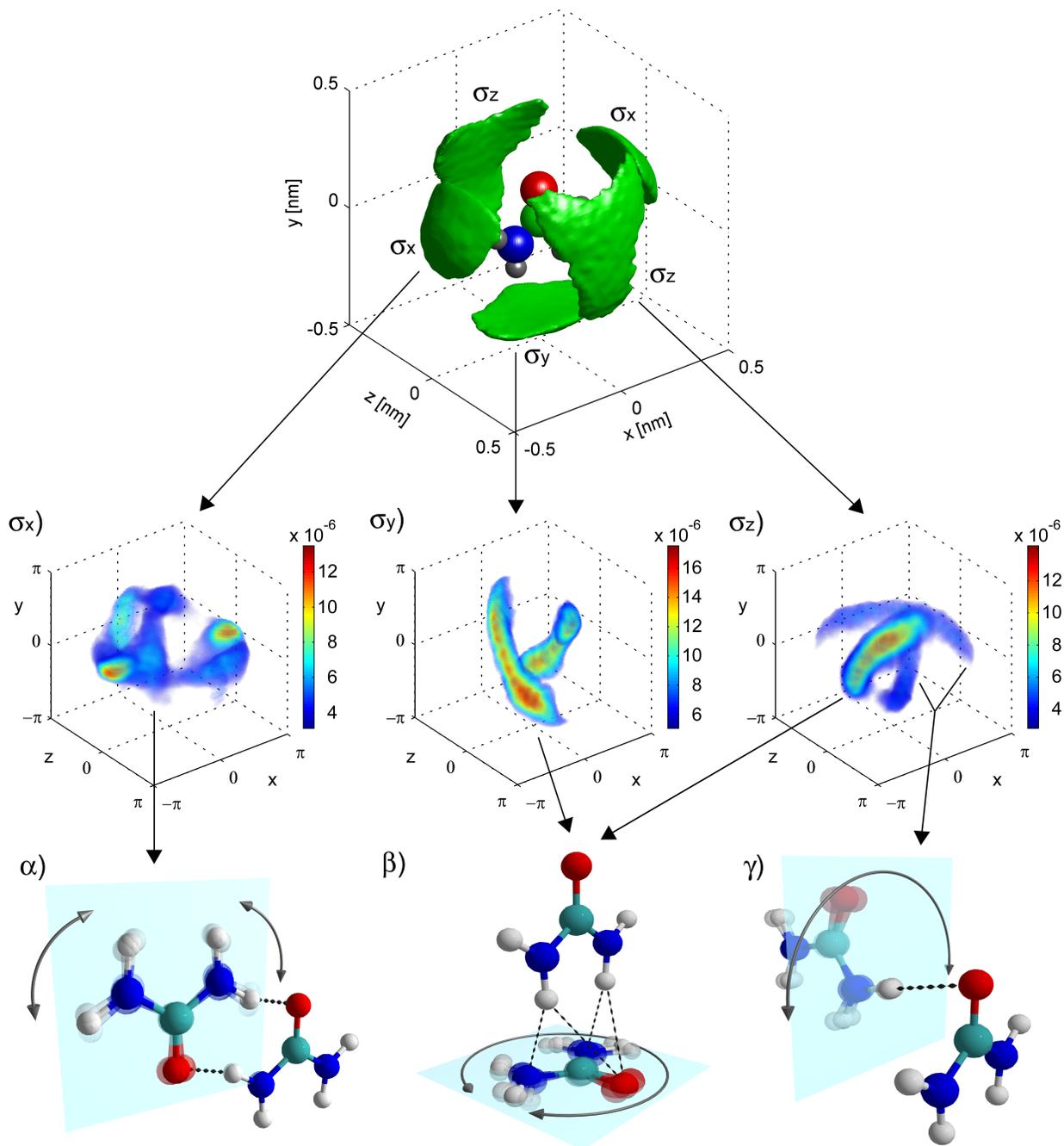


Figure 4.3: Three dominant urea pair conformations are seen in the simulations. Panels x , y , z show the orientational density distribution at the respective positions, x , y , z , in axis-angle representation. Panels α , β , γ illustrate the three different urea pair conformations obtained from the respective orientational density distributions. The length of the arrows in panels α , β , γ depicts the rotability of the molecule in the cyan plane. Hydrogen bonds are marked by black dotted lines. In conformation β , the smaller dots depict weak electrostatic interactions.

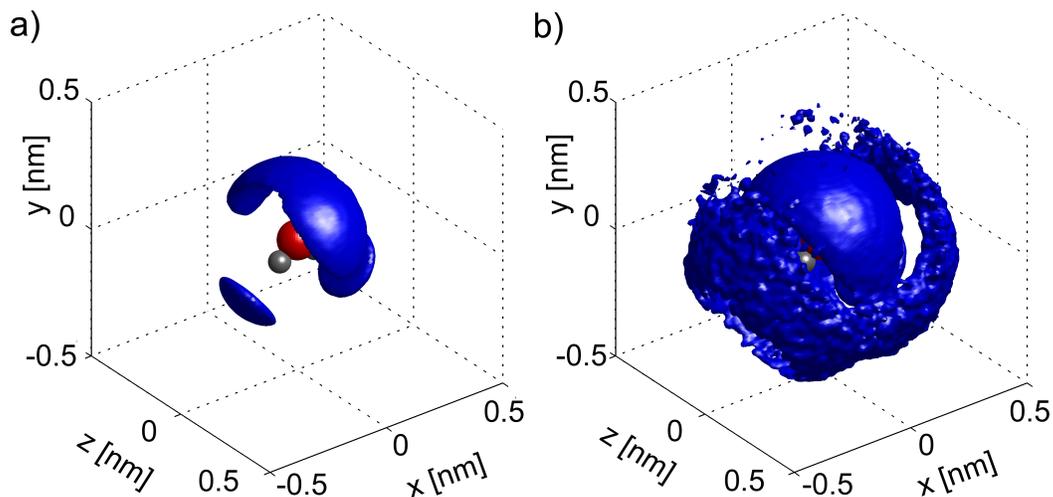


Figure 4.4: Spatial density distribution of water with respect to the reference water molecule, a) first hydration shell (isosurface enclosing 15% of all molecules in a 0.5 nm distance), b) second hydration shell (isosurface enclosing 40% of all molecules in a 0.5 nm distance).

and then inclined by a few degrees around its new z -axis so that a hydrogen bond between N as donor and O as acceptor can form (Fig. 4.3 γ). Urea molecules in this conformation are seen to rotate in a range from -90° to 90° around the hydrogen bond. In terms of enthalpy and entropy it falls between conformations α and β . There is one relatively strong hydrogen bond with O as acceptor, and the rotational entropy is larger than for α but smaller than for β . As expected, the population of this conformation also decreased at higher temperatures, but not as strongly as for conformation α .

Now, the positions of water molecules with respect to other water molecules are discussed. Two hydration shells can be identified in the respective density distribution (Fig. 4.4). The first, inner hydration shell contains the four tetrahedrally arranged sites for hydrogen bonding. As a result of the single point charge representing the two free electron pairs on the oxygen atom in the TIP4P model, the region between the two acceptor positions is also populated. A perfectly tetrahedral density distribution was observed in a test simulation using TIP5P water,¹⁷⁶ where each of the two free electron clouds of the oxygen is represented by a dummy atom (results not shown).

As a measure for short-range order in the water structure, the relative population of the first solvation shell was used. This property was defined as the fraction of water molecules whose center of mass is enclosed in the volume shown in Fig. 4.4, out of all water molecules which are within 0.5 nm center of mass distance. For illustration, this property is expected to be maximal in ice due to the perfectly tetrahedral arrangement of the molecules. This property was found to increase with decreasing temperature as well as increasing urea concentration (Fig. 4.5). In this regard, urea was found to strengthen water structure and make it more ice-like. The hydrogen bond analysis further below will corroborate this

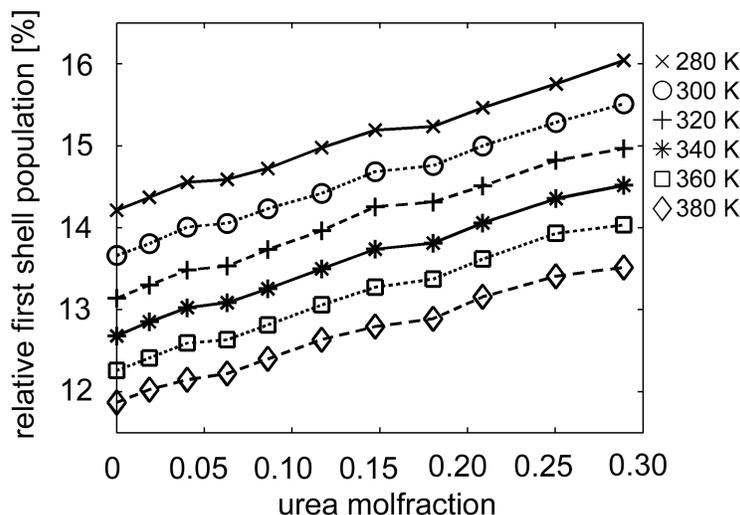


Figure 4.5: Relative population of the first hydration shell around water molecules (normalized to the total number of water molecules within 0.5 nm distance).

finding.

Finally, the density distributions of urea and water with respect to each other are discussed, which are shown in Fig. 4.6a) and b). As can be seen, the preferred coordination positions of urea and water widely overlap. Because of the larger size of the urea molecule, the urea density maxima are located a bit farther away from the reference center.

The large overlap between the coordination geometries of urea and water, together with the aforementioned only minor perturbations in the water structure even at high urea concentrations likely contributes to the high solubility of urea in water.

4.3.2 Hydrogen bonds

Figure 4.7 shows the number of hydrogen bonds per molecule for different urea concentrations. As expected, for all concentrations a decrease of the number of hydrogen bonds with increasing temperature is observed due to the enthalpic nature of the hydrogen bonds.

The number of hydrogen bonds per urea molecule increases with urea concentration (Fig. 4.7a) since urea gains more hydrogen bonds to urea than it loses hydrogen bonds to water. For water, however, the average number of hydrogen bonds per molecule increases only marginally with increasing urea concentration from 3.53 in pure water to 3.56 in 11.3 M urea solution at 300 K (Fig. 4.7b). In this respect, urea substitutes well for water and perturbs the hydrogen bond network only slightly.

Considering not only the numbers, but also the energetics of the hydrogen bonds, a significant effect of urea concentration becomes evident. Fig. 4.8a shows that the water-water

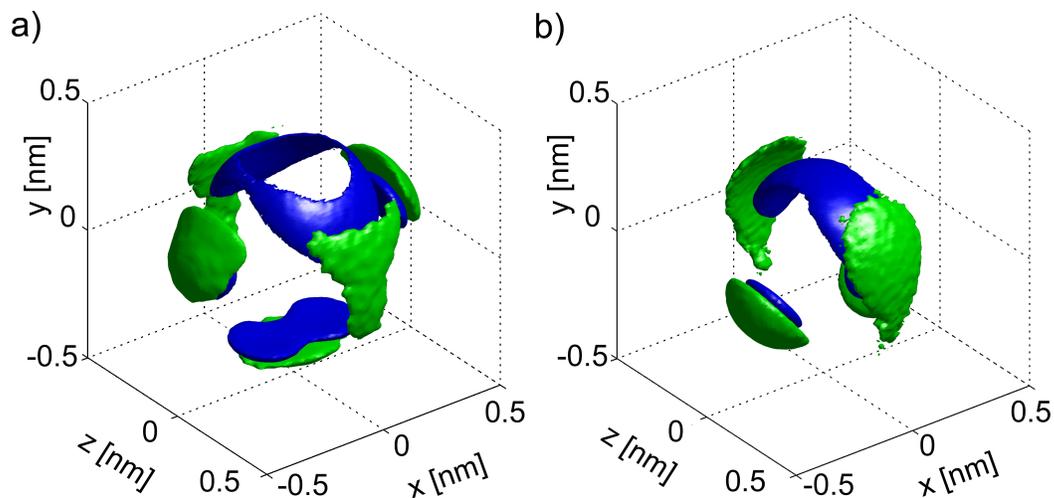


Figure 4.6: Spatial density distributions of urea (green) and water (blue) with respect to a) urea and b) water, displayed as isosurfaces enclosing 15% of all molecules within a 0.5 nm distance.

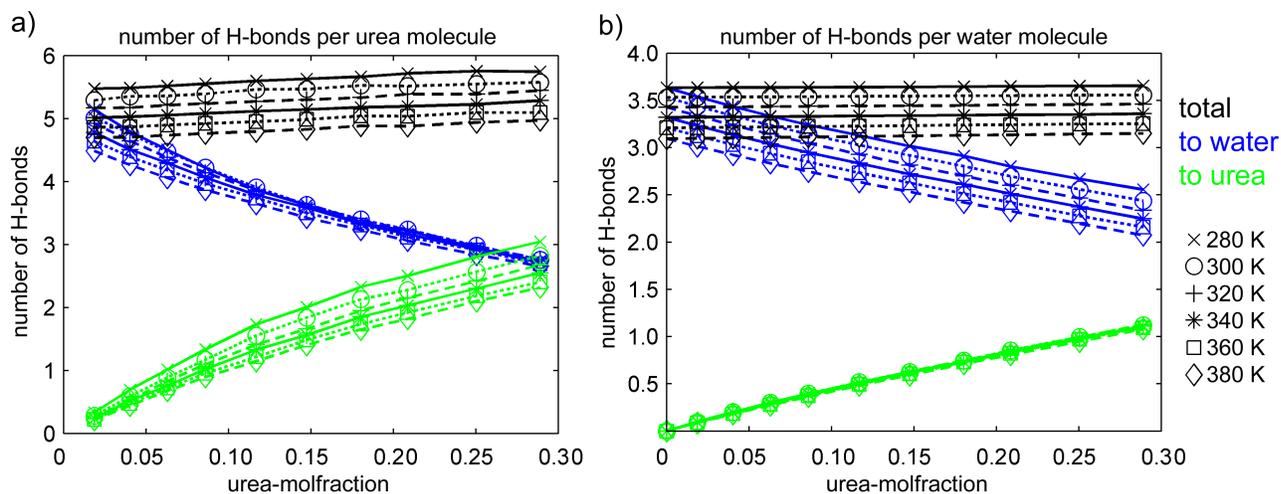


Figure 4.7: Average number of hydrogen bonds, normalized a) per urea molecule, b) per water molecule. Blue: hydrogen bonds to water, green: hydrogen bonds to urea, black: sum.

hydrogen bonds become stronger with increasing urea concentration. This is in line with the aforementioned strengthening of the water structure (see Fig. 4.5) and supports the view of urea being structure-maker rather than structure-breaker for water.

However, despite the strengthening of single water-water hydrogen bonds with increasing urea concentration, the total hydrogen bond energy per volume of the solution decreases significantly (Fig. 4.8b).

An analysis of the individual hydrogen bond energies between all donor-acceptor pairs (Fig. 4.8c) provides further insight and also explains this effect. The strongest hydrogen bonds were formed between water molecules with an average energy of 27.6 kJ/mol (300 K, 8.9 M). Between water and urea, the strongest hydrogen bonds were those formed between the water oxygen atom (O_W) as donor and the urea oxygen atom (O_U) as acceptor with a mean energy of 24.1 kJ/mol. The weakest hydrogen bonds existed between O_W as donor and the urea nitrogen atom (N_U) as acceptor (5.9 kJ/mol), and between N_U as donor and acceptor (6.2 kJ/mol).

This progression water-water, urea-water, urea-urea from strongest to weakest hydrogen bonds has two consequences. First, due to the decreasing number of water-water hydrogen bonds in favor of the less energetic water-urea and urea-urea hydrogen bonds with increasing urea concentration (compare also Fig. 4.7), the total hydrogen bond energy per volume decreases, as shown in Fig. 4.8b. Second, urea self-aggregation due to the hydrophobic effect is expected. The latter will be analyzed in more detail below.

4.3.3 Urea aggregation

Urea self-aggregation was quantified by measuring the interface surface area between water and urea, as described in the Methods section. Figure 4.9a shows the mean interface surface area between urea and water for different hypothetical partial charges.

The lower limit for aggregation (0%) was calculated with both, urea and water completely uncharged and exhibits the largest interface surface area (dotted line in Fig. 4.9a). In the absence of any contacts between urea molecules, the interface surface area is expected to increase linearly, by 1.41 nm² per urea molecule. However, due to the limited volume, random contacts (“stochastic clustering”) between urea molecules occur, and thus the interface surface area increases less than linearly with the concentration. Beyond the mole fraction of 0.2, the interface surface area decreases. This effect becomes obvious by considering that a system with pure urea would not have any interface surface at all. The upper limit for aggregation (100%) was defined using a set of simulations with completely uncharged urea, but water with regular charges (dashed line in Fig. 4.9a).

The two extremes of purely stochastic clustering (dotted line in Fig. 4.9a) and maximal aggregation for uncharged urea (dashed line) are subsequently used to define a scale to quantify the degree of urea aggregation. Maximal aggregation was taken as 100% urea aggregation, and purely stochastic clustering as 0%.

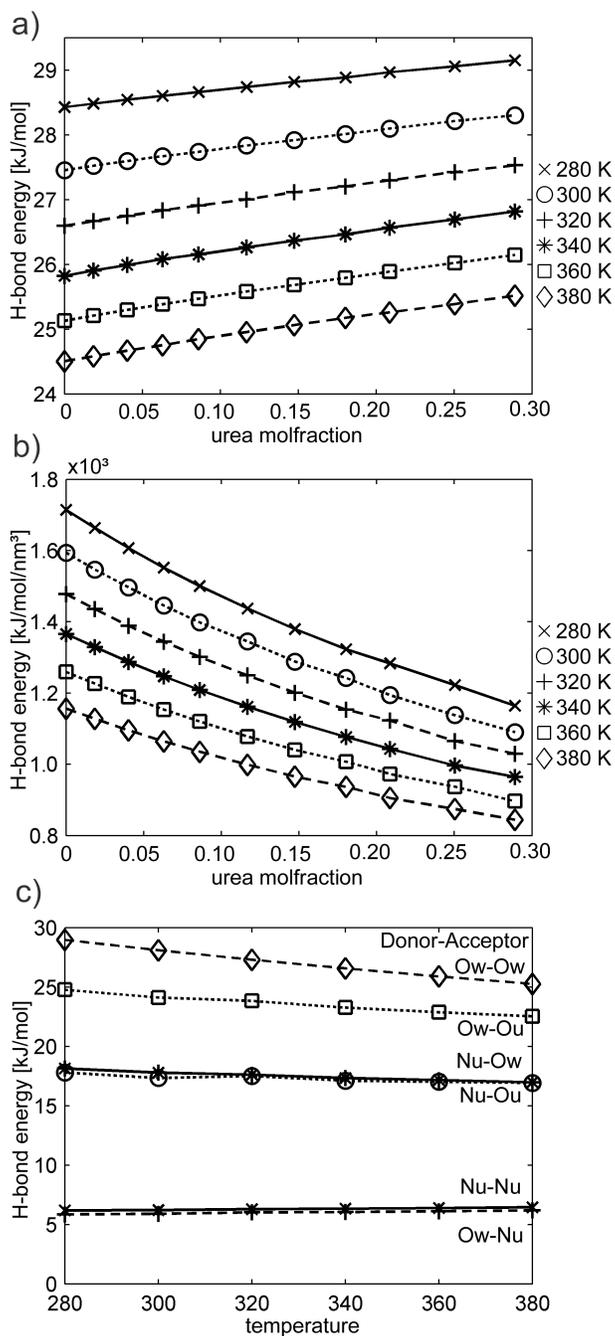


Figure 4.8: a) Water-water hydrogen bond energies. b) Total hydrogen bond energy per volume. c) Mean energy per hydrogen bond for all donor-acceptor combinations (at 0.21 mole fraction).

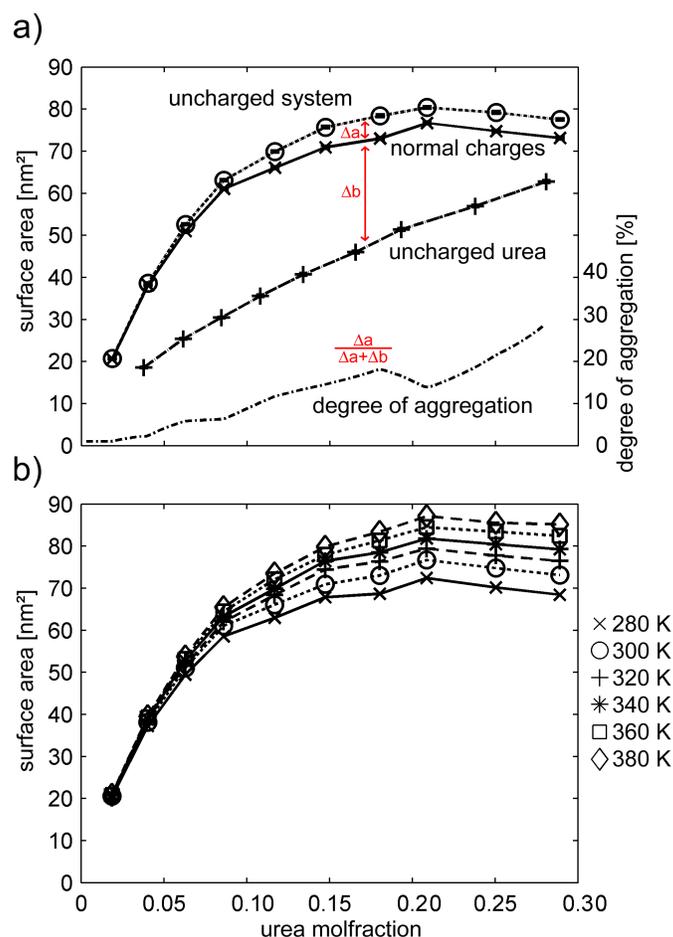


Figure 4.9: Interface surface area a) with regular charges (solid line) within the range of minimal aggregation (uncharged system, dotted line) and maximal aggregation (uncharged urea, dashed line) at 300 K. The lower, dash-dotted line displays the degree of aggregation in percent (right axis). b) for 8.9 M at different temperatures. The statistical error for the surface areas is about 3 nm².

The degree of urea aggregation within this scale is depicted by the lower (dash-dotted) line in Fig. 4.9a. At typical concentrations used for protein denaturation (about 8–9 M, mole fraction of 0.2), an aggregation of ca. 20% is seen. At these concentrations, the difference in surface area was 5.6 nm², corresponding to the surface of 4 urea molecules.

Figure 4.9b shows that urea aggregation decreases with temperature, which corresponds to higher solubility. It is noted that part of the surface area increase also resulted from the increase of volume with temperature, as seen from test simulations with constant volume conditions. This effect, however, was found to be small (data not shown).

In the light of the denaturing effect of urea on proteins, this tendency to self-aggregate might point towards and explain preferential binding to the peptide backbone and less polar parts of the protein which are more exposed in the denatured state.

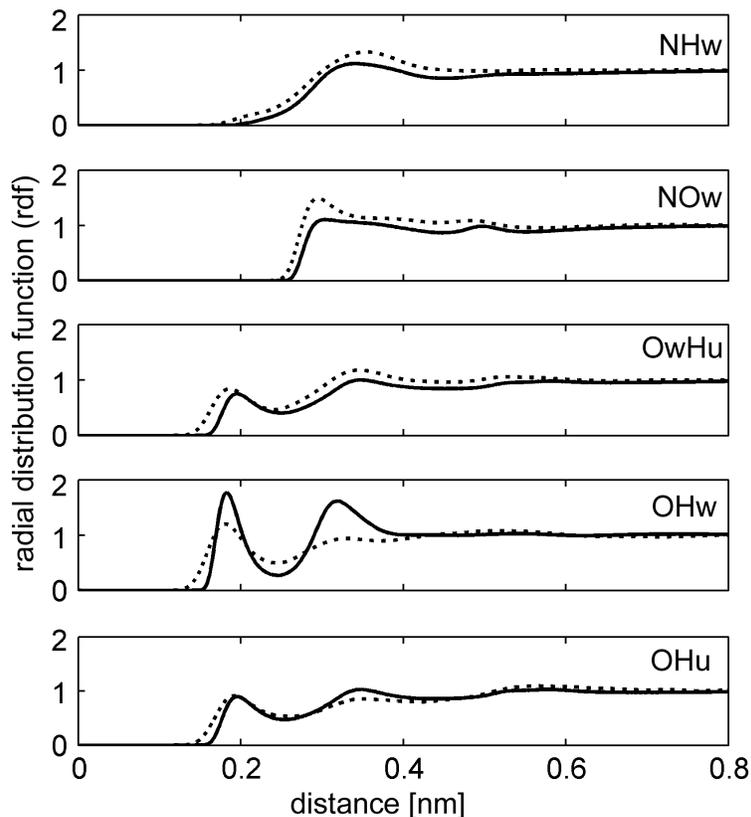


Figure 4.10: Radial distribution functions between atoms involved in hydrogen bonds. Solid lines: simulation data, dotted lines: refined data from neutron scattering experiments (Ref. 144).

4.3.4 Comparisons to experiments, controls

To assess the accuracy of the force field, the simulation data were compared to measured mass densities and radial distribution functions.

The mass densities ρ of the simulations at 300 K deviated by less than 1% from the experimental data using the fit function¹⁵⁴ $\rho = 0.9972 + 0.0158c_{urea} + 0.00009681c_{urea}^2$ for all urea concentrations c_{urea} given in $\frac{mol}{l}$ (data not shown). Radial distribution functions for various atom pairs involved in hydrogen bonding were calculated and compared to neutron scattering data (Fig. 4.10) from Soper and coworkers obtained with empirical potential structure refinement.¹⁴⁴ As can be seen, the positions of the peaks are well reproduced. In some cases, the heights of the peaks deviate. However, the peak heights are of minor relevance for the conclusions here since they are based on the energies of the hydrogen bonds, which correlate with the respective peak positions.

To test if the results are affected by possible periodic boundary or finite size artifacts, a larger system size of $(6 \text{ nm})^3$ with 4.2 M urea at 300 K was simulated and yielded virtually identical structural and energetic properties as compared to the corresponding simulation

with the smaller system size. Therefore, such artifacts can safely be excluded for the present study.

Despite the good agreement with experimental data, it is noted that the choice of force field is critical. In particular, the extent of urea aggregation has been found to be sensitive to force field details.¹⁶⁸ The OPLS urea model in combination with the TIP3P and SPC/E¹⁷⁷ water models has been found to overestimate urea aggregation.¹⁶⁸ In the present study, the combination of TIP4P water model with the refined OPLS urea¹²⁷ was used, which has not been investigated in the study of Weerasinghe et al.¹⁶⁸ However, since the urea model used in the present work is based on the OPLS model, the numbers given here might be somewhat too large.

4.4 Summary and Conclusion

This study of structure and energetics of urea-water systems aimed at a deeper understanding of the special properties of these systems with a special focus on the ability of urea to denature proteins. The aggregation tendency of urea was quantified to ca. 20% on a scale ranging from purely stochastic clustering to full aggregation. Three different pair conformations for urea have been identified. As a result of the respective entropic and enthalpic contributions, the relative population of each conformation depends on temperature to a different extent.

An analysis of the hydrogen bond energies between urea and water revealed that the water-water hydrogen bonds are considerably stronger than those between water and urea, or those between urea and urea. These differences in hydrogen bond energies suggest that urea self-aggregation might be driven mainly by the hydrophobic effect.

Even at high urea concentrations, the mutual density distribution of water molecules was found to be changed surprisingly little. Furthermore, the coordination position distribution between urea surrounding water and water surrounding urea, respectively, was found to be very similar. This finding, together with an almost concentration-independent number of hydrogen bonds per water molecule shows that urea is able to substitute for water in the hydrogen bond network very well in geometric and steric terms.

Taken together, these results indicate that the effect of urea on the water structure is only small. A slightly kosmotropic effect of urea is indicated by stronger water-water hydrogen bonds and a more rigid occupation of the tetrahedral coordination positions with increasing urea concentration. This strengthening of the water structure might facilitate protein denaturation by a mechanism resembling cold denaturation, i.e., the entropic penalty to solvate hydrophobic groups, and thus the hydrophobic effect, is reduced for higher ordered water. However, the structural perturbations were found to be only very small and do not seem sufficient to drive protein denaturation by urea. Instead, since urea is structurally very similar to the peptide backbone, urea self-aggregation might point towards relevant direct interactions of urea with the protein. These will be investigated in the next Chapter.

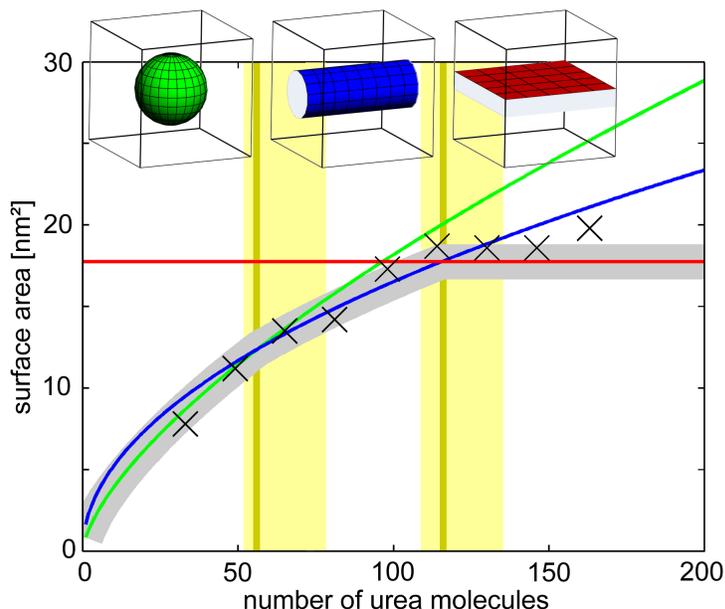


Figure 4.11: Interface area for strong aggregation with periodic boundaries as predicted by theoretical model (green: sphere, blue: cylinder, red: layer, gray: minimum surface, dark yellow: transition concentrations, black crosses: simulation data for 50% regular charges with offset correction). The yellow bars depict the transition concentrations between the shapes observed in the simulation.

4.5 Appendix: Quantification of aggregation under periodic boundary conditions

When studying urea aggregation, the use of a simulation box with periodic boundary conditions, together with a high degree of aggregation led to an artifact which has to be accounted for, and which is described in this Appendix. A sphere has the smallest surface area ($4\pi \cdot r^2$, where r denotes the radius) for a given volume and hence is the preferential shape for a strongly aggregating substance. Above a certain concentration, however, a cylinder spanning the box width (see illustration in Fig. 4.11) exhibits an even smaller surface ($2\pi \cdot r \cdot b$, where r is the radius and b the boxsize, hence the height of the cylinder), since its caps touch each other at the periodic image and thus effectively vanish. Along similar lines, a planar layer offers the smallest surface area ($2 \cdot b^2$, where b is the boxsize) at even higher concentrations.

This behavior was observed in the simulations with reduced urea partial charges, which were performed to quantify the degree of urea self-aggregation. Figure 4.11 shows the calculated surface areas for the three shapes sphere (green line), cylinder (blue line) and layer (red line). The points where the lines intersect mark the transition points from one shape to another (dark yellow lines) and define the minimal surface area (gray line).

A simulation set was performed with only 50% regular urea charges to obtain increased aggregation. These simulations exhibited the described geometric transitions from sphere to cylinder to layer with increasing urea concentration. The observed transition regions are marked as light yellow bars in Figure 4.11 and are in good agreement with the calculated transition points. The interface surface area (black crosses in Fig. 4.11) follows the minimal surface after an offset correction to account for the roughness of the surface, deviations from ideal shapes, partly solvated urea, and box volume fluctuations.

For the situation at hand, this change in “ideal” surface area due to periodic boundary conditions would distort the aggregation scale. Therefore, those urea-clusters that had formed cylinders or layers, which occurred at mole fractions larger than 0.1 in the $(3\text{ nm})^3$ box, were simulated in a larger box of size $(5\text{ nm})^3$ with additional water, such that a spherical shape was retained in the simulation. Solvated urea molecules remote from the sphere were excluded from the calculation of interface surface area, and the concentration was corrected accordingly. This procedure allowed accurate determination of the urea-water surface also above the critical concentration of about 0.1 mole fraction.

5

Direct Interactions of Urea with Amino Acids

5.1 Introduction

The results of the previous Chapter have suggested that indirect interactions of urea with the protein are of only minor relevance for urea-induced denaturation. In this Chapter, therefore, the direct interactions of urea with all 20 natural amino acids are investigated.

While many recent studies support the direct interaction model for urea-induced denaturation,^{17,156,16,27,28,29} it is controversially discussed whether polar or apolar residues or the peptide backbone constitute the main interaction sites for urea. That the peptide backbone is an important interaction site for urea is now widely accepted.^{178,179,180} However, some studies^{21,15,26,16,27,29} stress the importance of urea–protein hydrogen bonds to polar residues. Other studies^{8,30,181,19,20,22,23,24,25,182,28} support the importance of apolar urea–protein contacts weakening the hydrophobic effect. Hence, more detailed insight into the interactions of a denaturant with amino acids is imperative to understand how denaturants work.

This study aims to elucidate and quantify by extended molecular dynamics simulations the interactions of urea with each of the natural 20 amino acids. To this aim, interaction frequencies between urea and the individual amino acids are investigated to decide whether urea interacts preferentially with polar or apolar residues or with the peptide backbone. To quantify residue interaction with urea, a contact coefficient C_{UW} is introduced as a measure for preferential interaction with urea relative to that with water. The C_{UW} analysis will also provide detailed insight into urea–peptide interactions on the atomic level. Additionally, the role of hydrogen bonds between urea or water and the peptide residues is investigated, and hydrogen bond energies are estimated.

To separate sequence dependence and secondary or tertiary structure effects from the immediate interaction between urea and the respective amino acids, all 20 amino acids were investigated by simulations of glycine-capped tripeptides (GXG). The influence of sequence and structure on the immediate or direct interactions of the amino acids are discussed in the conclusions.

5.2 Methods

5.2.1 Simulation setup

Each of the 20 natural amino acids (X) was simulated in a glycine-capped tripeptide (GXG). The glycine-termini with NH_2 and COOH were kept uncharged. For histidine, all three protonation states were considered, resulting in 22 simulations in total, each of 100 ns length. The initial backbone angles of the tripeptides were set to β -sheet conformation

($\phi = -135^\circ$, $\psi = 128^\circ$). Since the autocorrelation time of (ϕ , ψ) was found to be below 1 ns in the simulations, these starting conditions did not impose significant bias. All tripeptides were simulated individually in aqueous urea solution with 1250 water molecules and 250 urea molecules. An appropriate counterion (Na^+ or Cl^-) was added for each charged amino acid. Additionally, a second set of simulations was performed with all electrostatic interactions switched off to estimate steric effects on the calculated contact coefficients, which may arise from different volumes of urea and water molecules.

To set-up the simulation systems, the initial size of the periodic cubic box was set to $(3.9 \text{ nm})^3$ to accommodate 1250 water and 250 urea molecules in addition to the tripeptide. 250 (non-overlapping) urea molecules were placed at random positions into the simulation box containing the tripeptide. Subsequently, the box was filled up with 1250 water molecules. The total simulation time for data collecting was $4.4 \mu\text{s}$.

5.2.2 Force field energies

Force field energies (Lennard-Jones and Coulomb) between amino acid X and urea (E_{X-U}) or water (E_{X-W}) were calculated for all atoms in atomic contact using the same distance cutoff criterium as for the contact coefficients. Energies per atom were then defined as:

$$E_X^{norm} = \frac{E_{X-W}}{N_X} - \frac{E_{X-U}}{N_U} \quad , \quad (5.1)$$

where N_X and N_U denote the number of atomic contacts of residue X with water or urea, respectively.

5.3 Results

5.3.1 Contact coefficients

First, contact coefficients C_{UW} (Fig. 5.1) are discussed. As can be seen, C_{UW} is higher than one, indicating preferential contacts to urea, for all amino acids except the anionic ASP and GLU, which have a significantly lower C_{UW} of about 0.89 each. The cationic ARG and LYS have the second lowest C_{UW} with 1.26 each. The other amino acids exhibit a C_{UW} between 1.44 (THR, SER) and 1.82 (CYS). For each amino acid the C_{UW} of the backbone alone is higher than for the complete amino acid. The C_{UW} of the backbone alone, averaged over all residues, is 1.78 ± 0.18 .

In summary, urea interacts mainly with aromatic and nonpolar residues, as well as with the protein backbone. Polar and especially charged residues interact less frequently with urea, the charged amino acids ASP and GLU show even more interactions with water than with urea. Note that $C_{UW} < 1$ does not necessarily imply a positive free energy of transfer

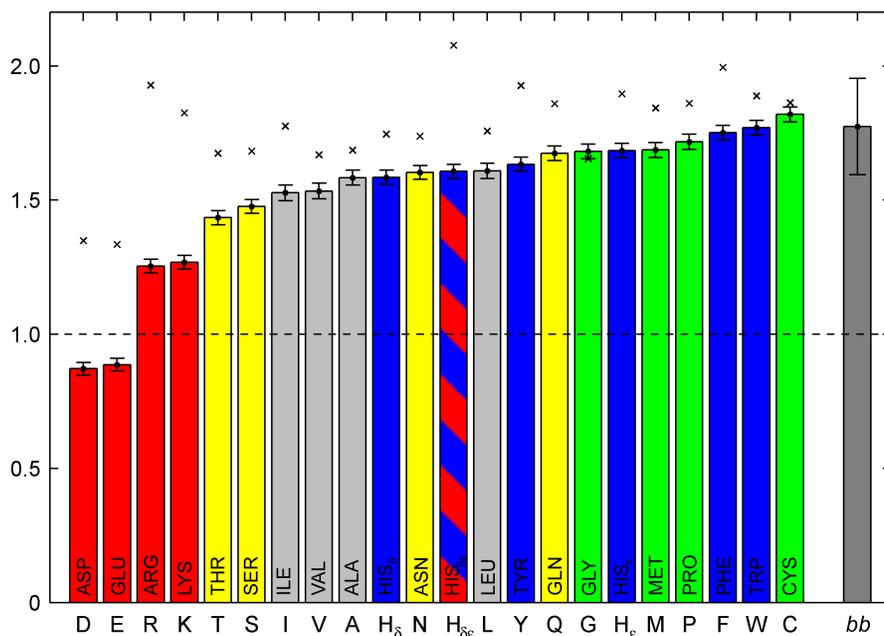


Figure 5.1: Contact coefficient C_{UW} for each amino acid as well as the backbone average (“bb”). Crosses denote the C_{UW} of the backbone alone. High values above one indicate preferential interactions with urea; a value of one corresponds to equal probability to interact with urea or with water. The color characterizes the amino acids; red: charged, yellow: polar, gray: aliphatic, blue: aromatic, green: apolar.

from water to urea solution, which has been found to be negative for hydrophobic as well as for hydrophilic residues.^{183,184,185,10}

To elucidate which parts of the amino acids show contact preferences for either urea or water, C_{UW} was calculated atomwise. This in-depth analysis was further motivated by the difference in average C_{UW} for the backbones and the complete residues. Figure 5.2 shows atomic interaction sites for urea and water for all amino acids. Blue indicates preferential interaction with water (low C_{UW}), green indicates preferential interaction with urea (high C_{UW}), white corresponds to no interaction preference ($C_{UW}=1$).

Again, clear differences in the C_{UW} are seen for the different the amino acids. In particular, the carboxyl groups of ASP and GLU have very low C_{UW} values and represent the main interaction sites for water. For both amino acids, also the sidechain CH_2 groups and even the backbone show reduced interactions with urea due to the charge of the carboxyl group. This effect is slightly more pronounced for ASP than for GLU, since the backbone is separated from the carboxyl group by one CH_2 group less in ASP. ARG and LYS show both, preferential interaction sites for water as well as for urea. The amino groups are the main interaction sites for water, whereas the backbones exhibit high C_{UW} values and are not significantly affected by the charged amino groups due to the long apolar sidechains of both amino acids. Further interaction sites for water are the hydroxyl group of TYR and

the two nitrogen atoms in the HIS $_{\delta\epsilon}$ -ring.

Pronounced interactions with urea are seen for the aromatic rings of PHE, TRP, and TYR as well as for the whole CYS, and the C $_{\beta}$ and C $_{\gamma}$ atoms of ILE, LEU, and MET. These atoms have been assigned a small charge (between -0.06e and -0.12e) in the OPLS-AA force field. The peptide backbone shows preferential interaction with urea for *all* amino acids.

Overall, it is found for the residue as well as for the atomic level that polar parts (with large partial charges) mainly interact with water, while less polar parts (with small partial charges) interact mainly with urea.

5.3.2 Hydrogen bonding

Hydrogen bonds between water, urea, and the amino acids were analyzed, and their strength was estimated via the Espinosa-formula.¹³¹ The average hydrogen bond energies are given in Fig. 5.2 for all sidechain donors and acceptors. As can be seen, all hydrogen bonds to water (blue numbers) are stronger than the corresponding hydrogen bonds to urea (green numbers). Therefore, hydrogen bond sites on the sidechains are preferentially solvated by water rather than by urea molecules. The strongest hydrogen bonds (≈ 46 kJ/mol) are seen between the carboxyl groups of ASP or GLU and water. Since the energies of hydrogen bonds to the backbone do not significantly vary between the amino acids, their average energies is given in Fig. 5.3 and they are not shown individually in Fig. 5.2. Figure 5.3 shows the hydrogen bond energies between protein (backbone), water, and urea. The larger numbers in Fig. 5.3a denote the average hydrogen bond energies for all donor/acceptor combinations. The small number pairs next to the arrows in panel a) refer to hydrogen bonds with donor/acceptor in the direction of the respective arrow. For instance, hydrogen bonds between water and the peptide backbone have an energy of 18.0 kJ/mol with the peptide as donor and water as acceptor, and an energy of 29.8 kJ/mol with water as donor and the peptide backbone as acceptor. On average, intra-water hydrogen bonds are the strongest with an energy of 27.8 kJ/mol, followed by hydrogen bonds between water and the peptide backbone with an energy of 25.8 kJ/mol. For urea, the energy difference between hydrogen bonds to protein or water is very small. Judging from the hydrogen bond energies alone, water seems to be a significantly more favorable hydrogen bond partner than urea for the peptide backbone. That this difference in hydrogen bond energies between protein-urea and protein-water does not lead to preferential solvation of the backbone by water, as one might expect, shows that optimization of peptide-solvent hydrogen bonds is not the determinant for backbone solvation by water or urea in this simple view. This issue will be discussed further below.

Statistical errors are below 0.1 kJ/mol (not to be confused with the accuracy of the empirical formula used to estimate the hydrogen bond energies). Intra-protein hydrogen bond energies were calculated from simulations of the CI2 protein (Chap. 6).

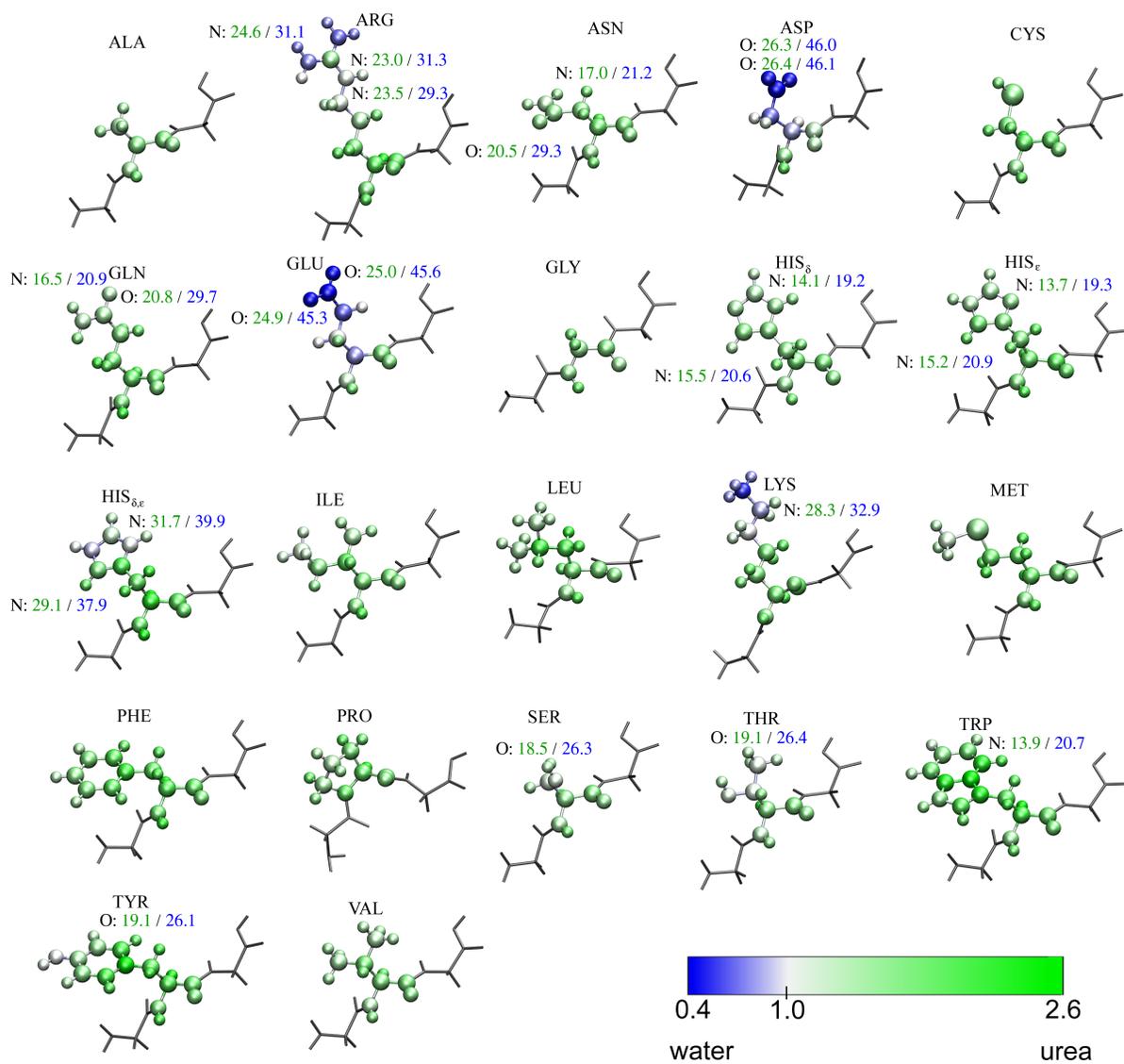


Figure 5.2: Atomic interaction sites for urea and water. Blue indicates low C_{UW} , green high C_{UW} , white corresponds to $C_{UW}=1$. The numbers denote average energies of hydrogen bonds to water (blue) or urea (green) for the respective atoms in kJ/mol.

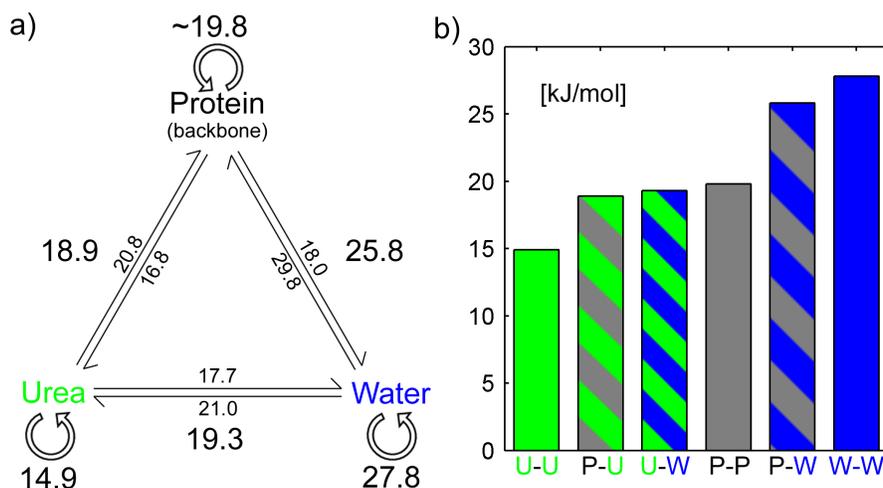


Figure 5.3: Average hydrogen bond energies (in kJ/mol) between the protein backbone, urea, and water. Errors are below 0.1 kJ/mol. The small number pairs next to the arrows in panel a) refer to hydrogen bonds with donor/acceptor in the direction of the respective arrow. The large numbers denote the weighted average. The energy of protein-protein hydrogen bonds was taken from simulations with the CI2 protein (Chap. 6).

5.3.3 Driving forces for preferential interactions

In order to explore the driving forces of the preferential interactions, the force field energies (Coulomb + Lennard-Jones) between the amino acids X and urea/water were calculated (Fig. 5.4a). Charged and polar residues (low C_{UW}) have the largest force field energies with both, water and urea. The Lennard-Jones contributions to the potential energies were all positive and much smaller than the respective Coulomb part.

The enthalpy gain upon substituting urea with water in the solvation shell, measured by the energy difference between X-water and X-urea, is largest for these residues. For residues with a high C_{UW} , the enthalpy gain is much smaller.

Figure 5.4b correlates the potential energy differences E_X^{norm} (Eq. 5.2.2) between X-water and X-urea with the contact coefficients C_{UW} . Indeed, the linear fit indicates a significant correlation (regression coefficient $r^2 = 0.8$). In particular, more negative (i.e. stronger) interaction energies imply low C_{UW} values. This correlation offers the conclusion that the interaction of water molecules with charged and polar residues (low C_{UW}) is dominated by enthalpy contributions, in particular electrostatic ones. Those residues with only a small enthalpy gain are solvated preferentially by urea (high C_{UW}) because displacement of water from less polar protein surface into bulk is enthalpically and entropically favorable. This picture agrees with the results from the hydrogen bond energies discussed above.

Further, a set of 22 simulations was carried out with all electrostatic interactions switched off to estimate the contribution of Coulomb and Lennard-Jones energies to the contact preferences and to extract purely steric contributions. Contact coefficients from these

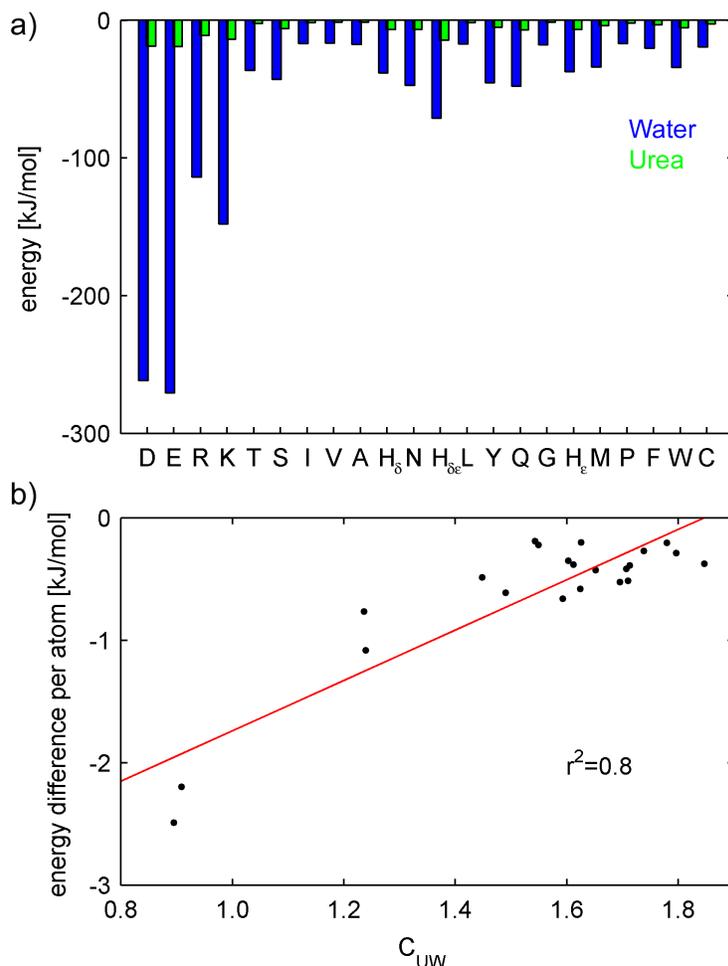


Figure 5.4: Force field energies between urea and water. (a) Total potential energy within cutoff radius (blue: water, green: urea). (b) Energy difference between residue–water and residue–urea per atom versus contact coefficient C_{UW} . The linear fit with a regression coefficient of $r^2 = 0.8$ is drawn in red.

simulations were all quite similar to each other (1.25 ± 0.06). This result further supports the view that electrostatic interactions are — directly via enthalpy or indirectly via the hydrophobic effect — the main determinants of the observed contact preference profile.

5.4 Discussion

These simulations suggest that polarity/apolarity is the main determinant of the interaction preference C_{UW} of amino acids with urea or water. One would therefore expect that C_{UW} correlates to hydrophobicity scales. Indeed, qualitative agreement is found with common scales.^{186,187,188,189,190,191} In particular, amino acids ranked very hydrophobic by

these scales (CYS, PHE, TRP, etc) interact preferentially with urea in the simulations (high C_{UW}), whereas amino acids ranked as very hydrophilic (ASP, GLU, ARG, LYS, etc) interact preferentially with water (low C_{UW}). On the atomic level, good agreement is found with atomic hydrophilicities reported by Kuhn et al.,¹⁹¹ who found oxygen atoms to be most hydrophilic, followed by nitrogen, followed by carbon and sulfur ($O \approx O^- > N^+ > N \gg C \approx S$). As the correlations between these different hydrophobicity scales are in the range of $r^2 = 0.69 - 0.95$ (problems involved in defining and interpreting such scales are well known^{192,193}), one would not expect correlations with C_{UW} higher than this value. The observed correlation coefficient of $r^2 = 0.35 - 0.53$ between C_{UW} and the different hydrophobicity scales therefore suggests that the hydrophobic effect is a key determinant for the contact preferences C_{UW} , but certainly not the only one.

Indeed, comparison with free energies of transfer from water to urea solution ($\Delta G_{W \rightarrow U}$) shows a similar correlation. The C_{UW} values reported here agree with main features from the early study of Nozaki and Tanford,⁸ and from more recent studies by Auton and Bolen.¹⁷⁹ In particular, less polar residues like PHE, TRP, TYR have a large $\Delta G_{W \rightarrow U}$ as well as a high C_{UW} . The correlation of C_{UW} with all $\Delta G_{W \rightarrow U}$ reported in Ref. 8 is $r^2 = 0.32$, and $r^2 = 0.31$ for the data reported in Ref. 179. Hence, transfer free energies $\Delta G_{W \rightarrow U}$ are obviously also related to C_{UW} . Although it can not be excluded that further effects contribute to the observed C_{UW} , these results suggest these two to be the main determinants for the observed contact preferences.

Also investigated was the relation of C_{UW} to m-value contributions, which measure the variation of the free energy of unfolding with denaturant concentration.¹⁹⁴ Indeed, very good agreement is seen with the m-value contributions reported by Auton and Bolen.¹⁷⁹ Their data are particularly suitable for comparison with C_{UW} values, since both are normalized to size or surface area. The regression coefficient with their data (Fig. 2a in Ref. 179) is $r^2 = 0.57$. It is noted, however, that CYS is a unique and puzzling outlier, with completely opposite tendencies, and was excluded from the regression fit. Large m-value contributions are reported for less polar residues like TRP, PHE, and LEU, as well as for the backbone, while the contribution of ASP, GLU, ARG, and LYS is very small. This large correlation shows that the contact coefficient C_{UW} introduced here can obviously be related to the m-value contribution per surface reported in Ref. 179.

By determining contact coefficients from tripeptides (GXG) with neutralized termini, values were provided for individual amino acids. In particular, this approach excludes effects from sequence, secondary or tertiary structure interactions, as well as effects from the N- and C-termini. In a protein, in contrast, the interaction of each amino acid with the surrounding water and urea solvent will additionally depend on local neighboring residues. For example, C_{UW} is expected to be lower than in the GXG tripeptides when X is flanked by a polar residue like GLU, or higher when X is flanked by an apolar residue like TRP. The influence of flanking peptide bonds on hydrophilicity has previously been discussed by Roseman et al.¹⁹⁵

5.5 Conclusions

Aiming at a comprehensive characterization of the interactions of amino acids with aqueous urea solutions, contact preferences for urea with individual amino acids were calculated from molecular dynamics simulations of 22 tripeptides. All amino acids (except ASP and GLU) were found to interact preferentially with urea. A clear spectrum of contact preferences has emerged, ranging from slight preferences for water contacts (ASP and GLU), to high preferences for urea contacts (TRP and CYS). In summary, urea was found to preferentially solvate apolar and aromatic residues as well as the peptide backbone.

These findings suggest a number of important implications for the mechanism of protein denaturation by urea. Under native conditions in water, and mainly due to the hydrophobic effect, apolar and aromatic residues are typically buried within the hydrophobic core of the folded state and not exposed to the solvent. Further, the protein backbone is largely shielded from solvent contact via formation of secondary structure elements. In aqueous urea solution, the contact analysis shows that urea molecules are located at the protein surface, in particular accumulating close to less polar residues and the backbone. The resulting displacement of water molecules from the solvation shell of less polar residues and the backbone into bulk water is favorable both, entropically and enthalpically. The former, because the translational and rotational entropy is increased, as one urea molecule frees about three water molecules; the latter, because these three water molecules can form strong water–water hydrogen bonds.

The resulting overall weakening of the hydrophobic effect renders unfolding of the protein by exposure of the hydrophobic core and dissociation of secondary structure elements energetically favorable. Taken together, these results, as discussed so far, strongly support the view that hydrophobic interactions are the main determinant of urea-induced protein denaturation^{8,24} rather than interactions with polar residues.^{21,15,26,16,27}

This result seems to be at variance with previous findings,^{26,15,9} however, which reported hydrogen bonds to be essential. The results of the hydrogen bond energy analysis now allow to resolve this puzzle by quantifying the different contributions discussed controversially by different authors. In particular, in the simulations at hand, urea was seen to form hydrogen bonds to the peptide backbone which are of similar strength as backbone-backbone hydrogen bonds, but are significantly weaker than backbone-water or water-water hydrogen bonds. Thus, these results confirm the widely held view that hydrogen bonds between urea and the protein backbone are essential,^{9,15,26,16} but do not support the suggestion¹⁵ that a competition of urea with native interactions is the driving force for urea-induced protein denaturation. In particular, it has already been argued,³⁹ that in such competition between native protein-protein hydrogen bonds and protein-solvent hydrogen bonds, protein-urea hydrogen bonds would have to be stronger than protein-water hydrogen bonds to explain denaturation. Indeed, the opposite was found here. Therefore, rather than being the driving force, it is suggested here that urea-protein hydrogen bonds only serve to avoid highly unfavorable unsatisfied hydrogen bond sites of the backbone,^{60,61} while at the same

time shielding it from entropically unfavorable water contact. Note that this finding is not at variance with the result that the weaker urea-protein hydrogen bonds actually tend to replace the stronger water-protein hydrogen bonds, which is suggested here to be due to the entropic effect described above.

The suggested dominance of the entropic effect leads to the speculation that the mere fact that urea molecules are significantly larger than water molecules substantially contributes to its denaturation power.

In summary, this Chapter suggests a synthesis of seemingly opposing viewpoints. Whereas urea-protein hydrogen bonds do not seem to drive the denaturation, they do contribute to the overall energetics. According to the mechanism proposed here, the denaturation power of urea rests on its trade-off between two essential but conflicting features. First, it is apolar enough to solvate apolar groups; second, it is polar enough to form weak hydrogen bonds to the backbone and to incorporate well into the water hydrogen bond network, as found in Chap. 4. Hence, urea can be regarded to denature proteins by interfacing between water and natively buried parts of the protein.

6

**Polar or apolar? — The role of polarity
for the denaturation power of urea**

6.1 Introduction

Based on the results of the previous Chapter, it was suggested that *direct apolar interactions* of urea with less polar residues are the driving force for denaturation, as opposed to polar interactions.

Now, this suggestion will be tested, and the relevance of direct polar and apolar contacts critically probed. For this purpose, a *Gedankenexperiment* is performed, in which urea polarity is varied. If polar contacts like hydrogen bonds between urea and the protein constituted the determinant interaction for denaturation, one would expect urea with increased polarity (hyperpolar urea) to be an even stronger denaturant than real urea. If, in contrast, apolar contacts played the major role for denaturation, one would expect urea with decreased polarity (hypopolar urea) to be the stronger denaturant.

Molecular dynamics simulations offer the unique possibility to perform simulations according to this Gedankenexperiment, and to compare the denaturation strengths of hyperpolar or hypopolar urea to that of “real” urea. By scaling the atomic partial charges of the urea force field, urea polarity is effectively modified, and hyper- or hypo-polar urea obtained. To investigate the effect of urea polarity on its denaturation strength, extensive simulations of the chymotrypsin inhibitor 2 protein¹⁹⁶ were performed in urea with six different partial charge scalings, and the respective effects on the protein stability analyzed. For comparison, additional simulations were performed in water.

6.2 Methods

Chymotrypsin inhibitor 2 (CI2, Fig. 6.1) is one of the best-characterized two-state folders. Its small size of 64 residues, fast ($\approx 3\ \mu\text{s}$) and two-state folding process,^{197,198} content of α - as well as β -structure, and the amount of existing data on it make it a common test system for folding and unfolding studies in experiments^{197,199,200,201,202,203} as well as simulations^{204,205,206,15,207} or theoretical studies.^{208,209}

Of particular interest here is an MD study about urea-induced denaturation of the CI2 protein (Ref. 15), where unfolding was observed within 3 ns — six orders of magnitude faster than in experiments. A secondary objective of the simulations presented here was to test whether the surprisingly fast unfolding reported in Ref. 15 can be reproduced.

6.2.1 Simulation setup

To allow comparison with the simulations reported in Ref. 15, the same CI2 mutant (E33A, E34A) was used and the simulation temperature was set to 333 K, except for one simulation

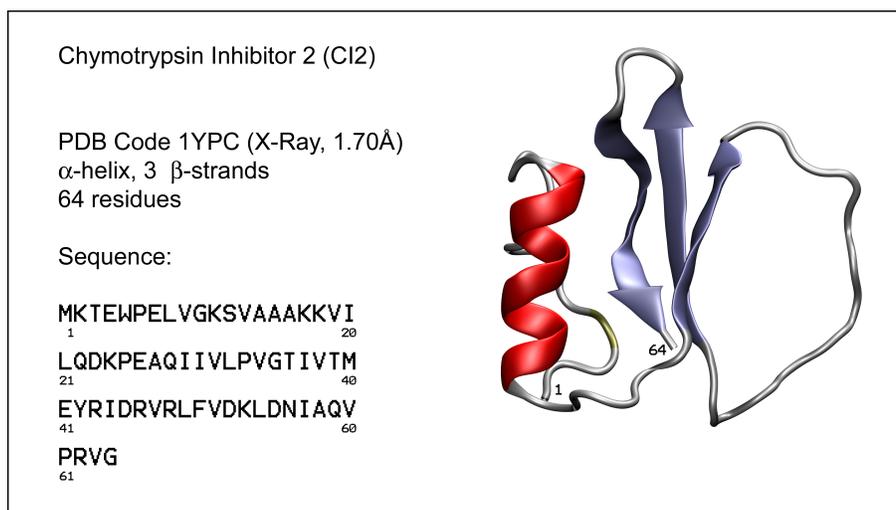


Figure 6.1: The CI2 protein.

at 300 K.

The structure of the CI2 protein was obtained from the Protein Data Bank,⁹⁷ PDB-code 1YPC.⁹⁸ Unresolved side chain atoms for residue MET40 (residue number 59 in the pdb file) were added using the program WHAT IF.²¹⁰ The box-size was chosen such that a minimum distance of 1.5 nm between protein atoms and the box was kept in each direction. For the solvation of the protein, pre-equilibrated structures of water and 8M urea were used (taken from the simulations reported in Chap. 4). Sodium and chloride ions were added to yield a physiological 150 mM ion concentration.

To avoid over-interpretation of possibly anecdotal events, multiple simulation runs were carried out for each parameter set (Tab. 6.1). Two simulations of CI2 in water (physiological solution), three simulations with regular urea charges, two simulations with 25 % urea charges, five simulations with 50 % urea charges, four simulations with 75 % urea charges, two simulations with 150 % urea charges, and two simulations with 200 % urea charges were performed, each at 333 K. In addition, one simulation in water at 300 K was performed to define native contacts and native secondary structure. The total simulation time of all simulations was ca. 7 μ s.

label	solvent	scaling factor for urea partial charges	simulation time [ns]
W ^{300 K}	water (300 K)		100 ns
W ¹	water		285 ns
W ²	water		500 ns
U ¹ _{25%}	8 M Urea	25 %	378 ns
U ² _{25%}	8 M Urea	25 %	300 ns
U ³ _{25%}	8 M Urea	25 %	435 ns
U ¹ _{50%}	8 M Urea	50 %	176 ns
U ² _{50%}	8 M Urea	50 %	357 ns
U ³ _{50%}	8 M Urea	50 %	395 ns
U ⁴ _{50%}	8 M Urea	50 %	296 ns
U ⁵ _{50%}	8 M Urea	50 %	289 ns
U ¹ _{75%}	8 M Urea	75 %	332 ns
U ² _{75%}	8 M Urea	75 %	225 ns
U ³ _{75%}	8 M Urea	75 %	250 ns
U ⁴ _{75%}	8 M Urea	75 %	250 ns
U ¹ _{100%}	8 M Urea	100 %	402 ns
U ² _{100%}	8 M Urea	100 %	285 ns
U ³ _{100%}	8 M Urea	100 %	522 ns
U ¹ _{150%}	8 M Urea	150 %	461 ns
U ² _{150%}	8 M Urea	150 %	500 ns
U ¹ _{200%}	8 M Urea	200 %	277 ns
U ² _{200%}	8 M Urea	200 %	234 ns

Table 6.1: Summary of all 22 simulation runs performed for the CI2 protein.

6.3 Results and Discussion

6.3.1 Dynamics of the native state in water/urea

As two measures for unfolding, Fig. 6.2 a and b show the C_α root-mean-square-deviation (rmsd) and the solvent accessible hydrophobic surface area (SAS), respectively, for the simulations in water (W^{1,2}, blue) and 8M urea solution (U^{1,2}_{100%}, green). As can be seen, the C_α -rmsd fluctuates around an average value of 0.3 nm for both, the simulations in water as well as in aqueous urea. No significant differences between the effects of both solvents on protein structure and dynamics can be seen in the rmsd. Clearly, no unfolding of the protein is observed in the simulations, which has to be expected from the measured millisecond time scale for CI2 denaturation.²⁰¹

Nevertheless, and perhaps unexpectedly, the SAS in aqueous urea is, on average, about 2 nm² larger than in water. As can be seen in Fig. 6.2, this difference is larger than the variance of the SAS for the single simulations. Closer inspection reveals that this

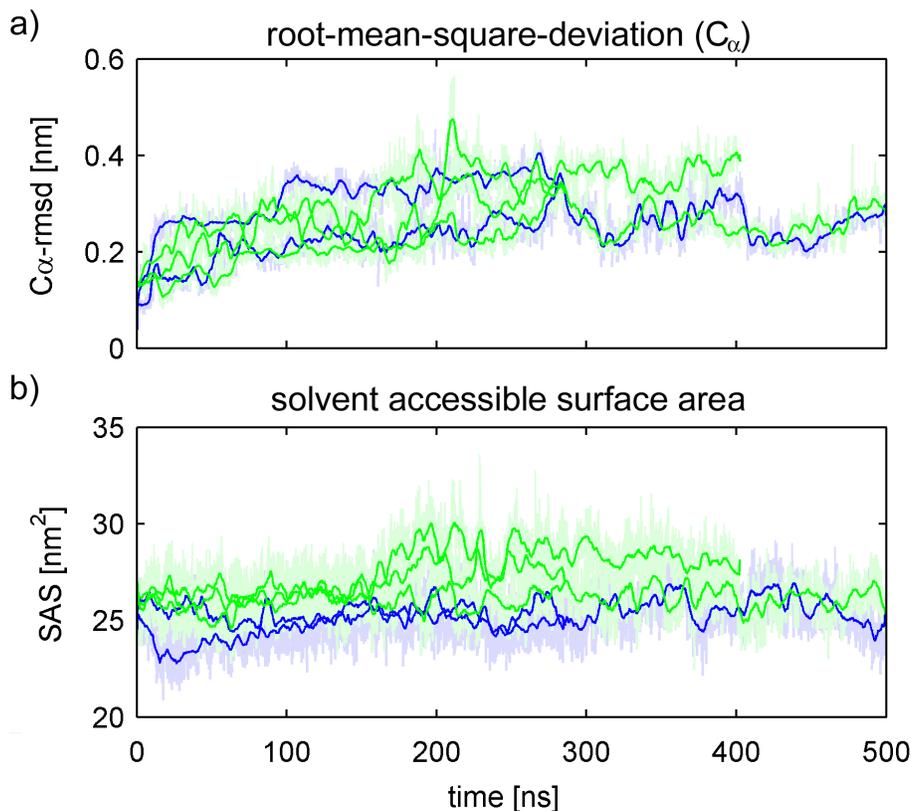


Figure 6.2: CI2 in native conformation; a) C_{α} rmsd, b) SAS for the two simulations in water (blue) and the 3 simulations in aqueous urea with regular charges (green). The solid bold lines show traces smoothed by a running average over 500 ps; dim lines show raw data.

difference results from few specific residues whose side chains are more solvent-exposed in aqueous urea than in water. In particular, MET1, LEU32, ILE44, and PHE50 contribute dominantly to this difference (0.22 nm^2 , 0.17 nm^2 , 0.30 nm^2 , and 0.29 nm^2 , respectively). Since these amino acids are among those which were found to have particularly strong contact preferences for urea (Chap. 5), it is expected that the higher exposure of these side chains is due to favorable interactions with urea molecules.

The contact coefficient C_{UW} was calculated for each amino acid type in the CI2 protein. Figure 6.3 shows the C_{UW} values averaged over time and over the three simulations in aqueous urea solution ($U_{100\%}^{1,2,3}$). Contact coefficients for the CI2 are largely similar to those calculated for tripeptides in Chap. 5. In particular, apolar and aromatic amino acids, as well as the backbone, are again found to have contact preferences for urea, whereas charged amino acids have preferences for water contact. This finding confirms that (a)polarity is clearly a determining factor for the specific interactions of urea with the CI2 protein residues, as expected from the contact preferences of the individual amino acids (Chap. 5). Further, the quantitative differences between both sets of contact coefficients can be used to quantify the influence of sequence and structure on the contact preferences. The correlation

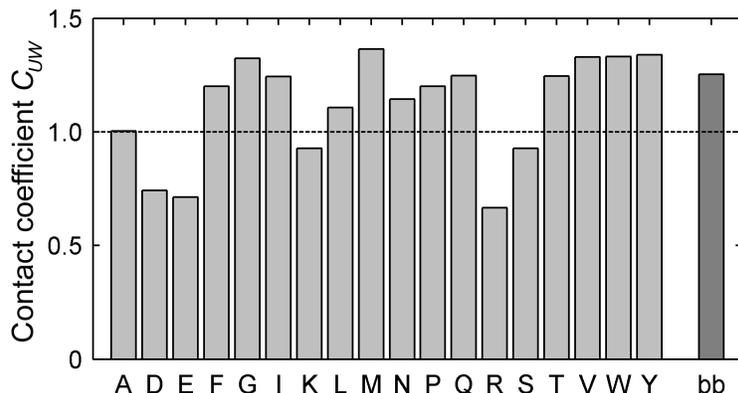


Figure 6.3: Interaction coefficient C_{UW} for each amino acid type in the CI2 and the backbone average (“bb”). Values above 1.0 indicate preferential interaction with urea, values below 1.0 indicate preferential interaction with water.

between both sets is $r^2 = 0.69$, indicating that effects from sequence and structure of the folded CI2 protein make up ca. 30 % of the contact preferences.

6.3.2 Protein stability in hypo- and hyperpolar urea

To investigate the denaturation strengths of hyper- or hypopolar urea, the CI2 protein was simulated in urea with partial charges scaled to values of 25 %, 50 %, 75 %, 150 %, and 200 %. Urea with partial charge scaling of x % will be denoted as “urea $_x$ %”. Fig. 6.4 shows the SAS of the CI2 for these simulations. As can be seen from the SAS, the protein unfolds in all nine simulations with hypopolar urea (urea $_{75}$ % and urea $_{50}$ %, magenta and orange lines, respectively). In contrast, with urea $_{150}$ %, the SAS remains close to the native value and the protein stays stable during all simulations (black lines).

What is the reason for the observed strong denaturation power of hypopolar urea with 75 % and 50 % partial charges? It was expected that decreasing polarity by down-scaling the partial charges should imply stronger interactions with apolar residues of the protein. This is indeed the case, as seen in the interaction coefficient C_{UW} for urea with different partial charge scalings (Fig. 6.5); urea with reduced partial charges shows increased interaction coefficients with certain amino acids, mainly PHE, ILE, LEU, PRO, VAL, TRP, and TYR. Urea $_{100}$ % preferentially interacts with these amino acids, and this preference is enhanced for hypopolar urea. Furthermore, interactions of certain other amino acids, mainly ASP, GLU, LYS, and ARG, are even less favored with hypopolar urea. Overall, lowering urea polarity increases the contrast of the interaction profile with respect to the one of urea $_{100}$ %, and enhances the preferential interactions: preferred interactions become even more frequent, less preferred interactions even less frequent. Obviously, this effect

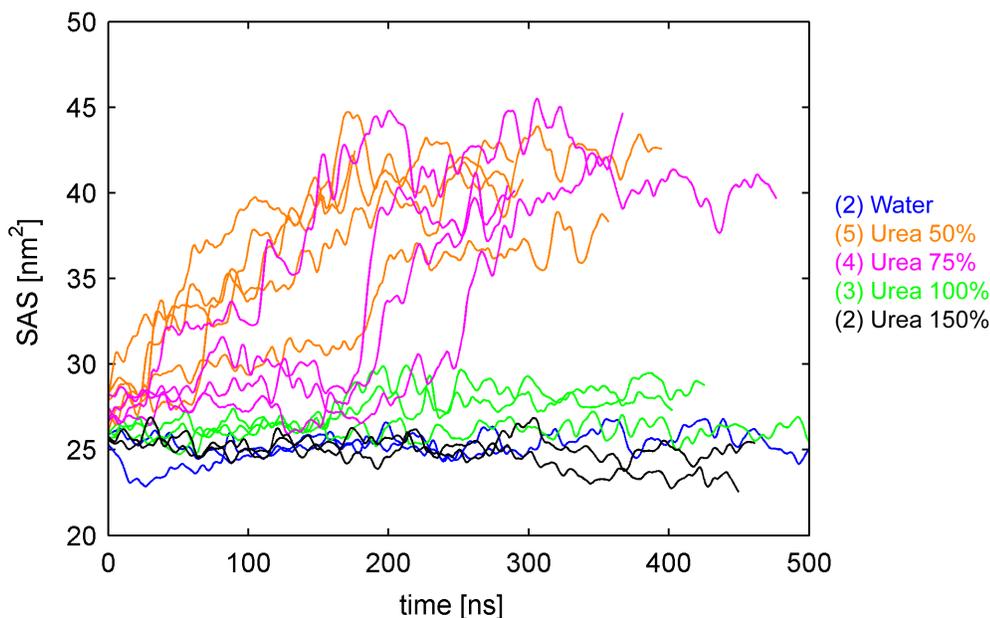


Figure 6.4: Solvent accessible surface area of the protein in all simulations. Blue: water, orange: urea_{50%}, magenta: urea_{75%}, green: urea_{100%}, black: urea_{150%}. The lines show traces smoothed by a running average over 500 ps. The numbers in the legend denote the number of respective simulations.

accelerates and facilitates unfolding of the protein.

Hyperpolar urea_{150%}, in contrast, exhibits fewer interactions with those amino acids preferentially interacting with “regular” urea_{100%}. Interactions with charged residues is even preferred by urea_{150%} over interactions with less polar residues. This inversion of the preferential interaction profile leads to a compactification of the folded state, as indicated by the smaller SAS (Fig. 6.4), rather than to destabilization and unfolding. This results suggests that urea_{150%} would be a weaker denaturant than urea_{100%} in sufficiently long simulations on the millisecond-timescale, or none at all.

The effects of the different solvent environments on protein stability are also reflected in the estimated free energy profiles with the SAS as reaction coordinate (Fig. 6.6). In particular, they show that the minimum of the free energy of the native state in urea_{100%} (green line) is shifted towards higher SAS values with respect to that in water (blue line), and to lower SAS values in urea_{150%} (black line). Since the free energy profiles for the multiple simulations in water ($W^{1,2}$), urea_{100%} ($U_{100\%}^{1,2,3}$), or urea_{150%} ($U_{150\%}^{1,2}$), respectively, were very similar to each other, the free energy profiles were calculated from all simulations together in the respective solvent. For the unfolding simulations in hyperpolar urea, however, the free energy profiles of the individual trajectories were very heterogeneous. Therefore, the averaged free energy profiles of urea_{50%} (orange line) and urea_{75%} (magenta line) provide only limited information, which is indicated by their dashed lines. The free energy profiles of these unfolding simulations will be addressed further below in the discussion of the

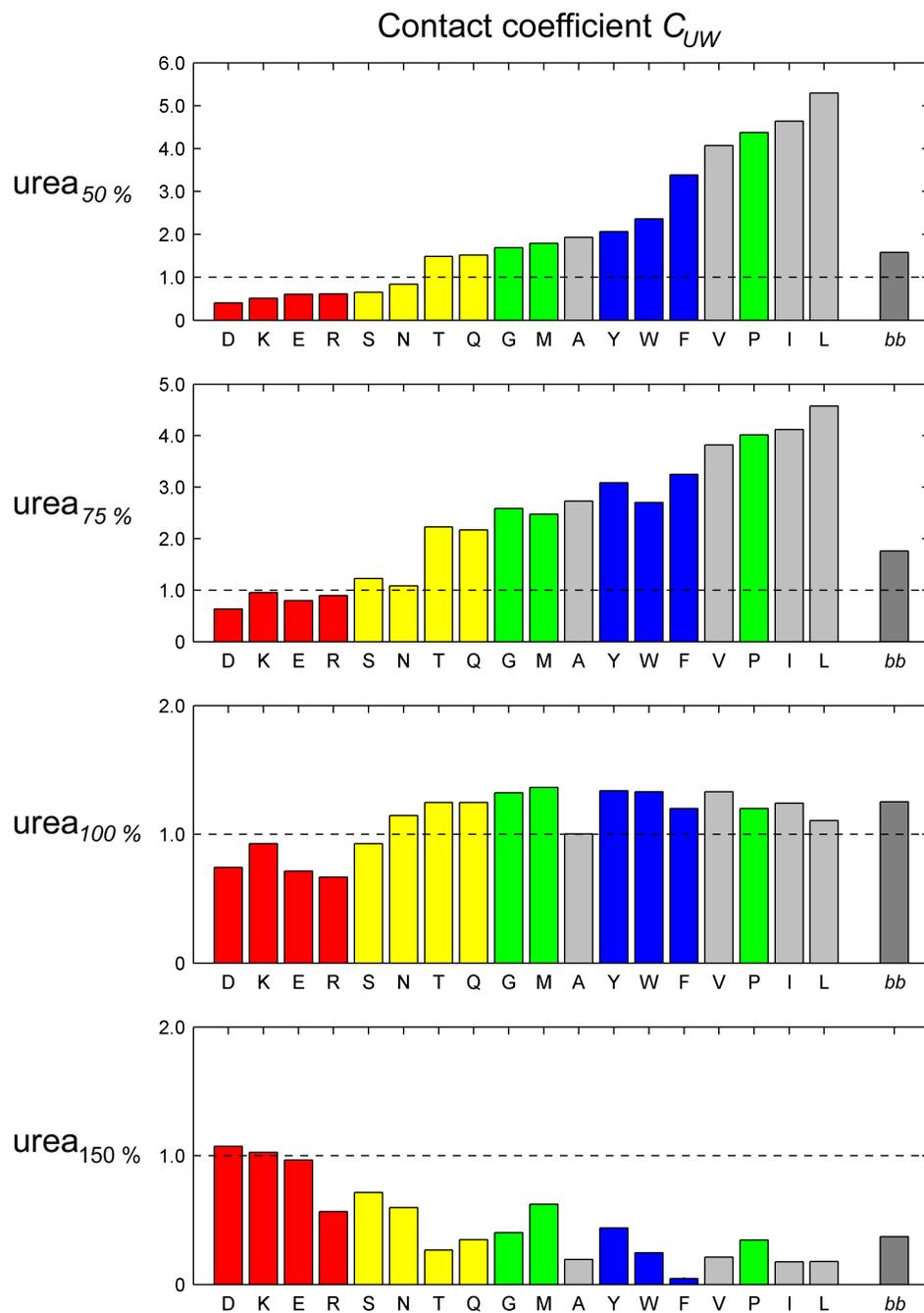


Figure 6.5: Interaction coefficient C_{UW} for all amino acids types in the CI2 protein, as well as the backbone average (“*bb*”). The four panels show C_{UW} for the different urea partial charge scalings (50%, 75%, 100%, 150%). The color characterizes the amino acids; red: charged, yellow: polar, gray: aliphatic, blue: aromatic, green: apolar. For better comparability, all C_{UW} are sorted according to C_{UW} in urea_{50%}.

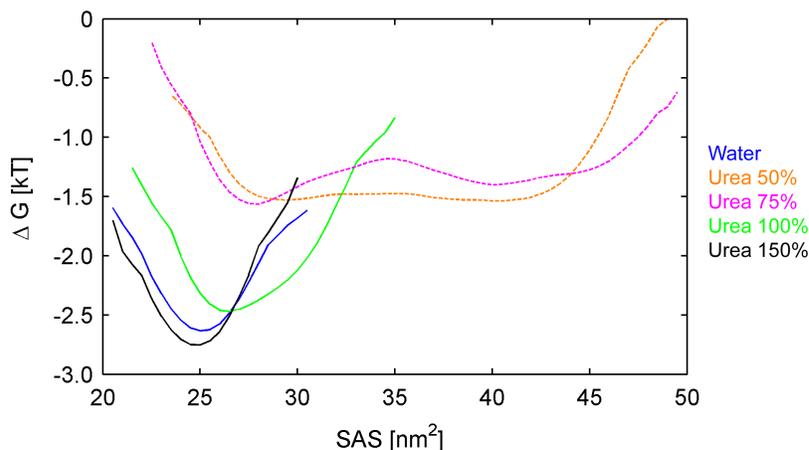


Figure 6.6: Estimated free energy profiles for the CI2 protein in the different solvents, with the SAS as reaction coordinate. Blue: water, orange: urea_{50%}, magenta: urea_{75%}, green: urea_{100%}, black: urea_{150%}. The dashed lines of urea_{50%} and urea_{75%} indicate that these profiles provide only limited information here, since they represent averages calculated from heterogeneous unfolding trajectories.

unfolding trajectories.

Also performed were simulations with “extreme” urea_{25%} and urea_{200%}. However, these simulations exhibit artifacts which render them irrelevant for the present purpose. For partial charges scaled down to 25%, on the one hand, urea shows a strong tendency to self-aggregate to a hydrophobic layer in the periodic simulation box, which does not any more interact with the protein. Urea_{200%}, on the other hand, induces a glass transition in the solvent, with drastically reduced urea diffusion coefficients (from $\approx 2.2 \cdot 10^{-5} \frac{\text{cm}^2}{\text{s}}$ to $< 0.001 \cdot 10^{-5} \frac{\text{cm}^2}{\text{s}}$). As a result of the vanishing mobility, the urea molecules do not interact with the protein either. Underestimation of diffusion coefficients in common force-fields has previously been observed for high ion concentrations.²¹¹

Both effects, urea aggregation and reduced diffusion coefficients, were also observed for the simulations with urea_{50%} and urea_{150%}, respectively, but to significantly lesser extents. Nevertheless, it has to be addressed whether the conclusions drawn here hold despite of these side-effects of urea partial charge scaling.

In particular, one might argue that protein unfolding in urea_{50%} is not a direct consequence of reduced urea polarity, but rather an effect of urea aggregation, for instance due to large differences in local urea concentrations. However, since unfolding is also observed in the simulations with urea_{75%}, where no significant degree of aggregation is seen, all conclusions regarding hypopolar urea — in particular that it is a stronger denaturant than regular urea — can simply be based on the observations with urea_{75%}, and thus hold.

With respect to hyperpolar urea, one might argue that the reduced urea diffusion coefficient for urea_{150%} (from $\approx 2.2 \cdot 10^{-5} \frac{\text{cm}^2}{\text{s}}$ to $\approx 0.1 \cdot 10^{-5} \frac{\text{cm}^2}{\text{s}}$) is the reason that no unfolding occurs

in hyperpolar urea. Two effects of reduced urea mobility have to be considered. First, the reduced mobility of urea molecules might prevent their “correct” placement on the protein surface within the simulation time. Due to the high urea concentration, however, the average diffusion time for a urea molecule to cross the average distance between two urea molecules is only ≈ 0.6 ns for urea_{150%} (for urea_{200%}, it is ≈ 90 ns). Hence, this effect can safely be excluded from consideration. Second, the reduced mobility of urea molecules might slow down conformational changes of the protein due to higher solvent viscosity. This concern is more critical. While the inversion of the interaction profile in conjunction with the compactifying effect of urea_{150%} on the folded protein suggest that hyperpolar urea has no significant destabilizing effect on the protein, these findings do not satisfactorily rule out that the protein might be only seemingly stable in urea_{150%} due to slowed down kinetics. The proof that this is not the case, and that even large conformational changes of a protein can occur in urea_{150%} within these simulation time scales, will follow in Chapter 7.2.

6.3.3 Unfolding pathways in hypopolar urea

In the simulations, the CI2 protein unfolds reproducibly in urea_{75%} (four simulations) and urea_{50%} (five simulations). Figures 6.7 and 6.8 show the unfolding pathways for urea_{50%} and urea_{75%}, respectively, in a representation of fraction of native secondary structure versus fraction of native contacts, as a measure for native tertiary structure. Starting from the folded state (top right), the protein undergoes several conformational changes, eventually leading to denaturation and unfolding in all nine trajectories.

One unfolding pathway (U_{75%}¹) will be described in detail, then common features and differences of the nine unfolding events are discussed. In simulation U_{75%}¹ (Fig. 6.8a), reversible fluctuations of the secondary-structure (β -strand 3, ILE57–ARG62) mark the beginning of unfolding. After 29 ns, a part of the coil region between β -strand 1 and β -strand 2 twists. In particular, THR36, ILE37, and VAL38 flip their sidechain orientation by about 180°. This is immediately followed, at 30 ns, by a flip of the turn region of residues 22–25. This irreversible and fast unfolding step with loss of native contacts is succeeded by a longer phase with reversible fluctuations again: the α -helix (Res. 13–22) loses stability and decays within the next 80 ns. The ALA-rich region (ALA14, ALA15, ALA16) is destroyed last at 110 ns. After this slow phase from 30 ns to 110 ns, the turn (and former α -) region between residues 18–25 flips further away, while the ALA-end of the helix partly rebuilds and decays again several times. At 140 ns, the tertiary structure unfolds further before at 150 ns, finally, the last secondary structure is lost when β -strands 2 (Res. 46–52) and 3 (Res. 56–62) decay.

Now, common features or differences in the individual unfolding trajectories are addressed. Interestingly, a similar stepwise process with alternating phases of loss of secondary and tertiary structure is also observed in most other simulations, as opposed to either a si-

multaneous loss of both structure levels, or, alternatively, complete loss of one structure level ahead of the other. In many cases, fast transitions (roughly 5 ns for a step) due to irreversible loss of native contacts are seen between two regions in the unfolding map that are each sampled for a longer time (20–100 ns and above). These regions represent reversible fluctuations and eventually loss of secondary structure. A step-by-step process for unfolding such as described here is consistent with the nucleation-condensation mechanism of folding for the CI2 protein.^{202,203}

It was investigated whether common transient structures or putative intermediates exist in the unfolding pathways. To this end, for every unfolding trajectory i , the rmsd was calculated with respect to every structure $X_j(t)$ (with a time resolution of $\Delta t = 100$ ps) of each of the other unfolding trajectories j (data not shown). In this analysis, conformations which occur in trajectory i as well as in trajectory j , would be revealed by a minimum in the respective rmsd. Unexpectedly, no pronounced minimum was found, which indicates that no pair of trajectories shares common structures, and that unfolding proceeds structurally different in all nine cases. During unfolding as well as after complete denaturation, the protein explores quite different regions of phase space. This is consistent with the observation that the CI2 protein is a two-state folder without intermediates in the folding process.¹⁹⁷

The differences in the unfolding trajectories are also reflected in the heterogeneity of the respective estimated free energy profiles, as shown in Fig. 6.9 for the SAS as reaction coordinate. In particular, the positions of free energy barriers between folded and unfolded state differ between the different unfolding trajectories. Note that a local minimum in this free energy landscapes does not indicate a metastable (un)folding intermediate, since in this one-dimensional reaction coordinate each SAS value corresponds to a whole ensemble of structurally different conformations. From these free energy profiles, it can also be seen that this method to estimate free energy profiles provides only a rough approximation for the shape of free energy profile, rather than giving accurate numbers. Particularly, one would expect the free energy of the unfolded state to be lower than that of the folded state. For this accuracy, however, longer simulations and more extensive sampling of unfolded conformations would be needed.

Despite different pathways of unfolding, there might be regions of the protein where unfolding is likely to begin. To investigate whether such unfolding “hot-spots” are present in the simulations, the rmsd per residue was calculated for the initial phase of unfolding for each of the simulations $U_{75\%}$ and $U_{50\%}$ (Fig. 6.10). The top row shows the root-mean-square-fluctuations per residue in the native state (simulation W^{300K}), for comparison. As can be seen, unfolding often starts in regions that exhibit large fluctuations already in the native state under physiological conditions. In particular, the C-terminal end of the α -helix (res. Q22) and the adjacent turn-region (res. D23–E26) start to unfold first in some simulations ($U_{75\%}^1, U_{75\%}^2, U_{75\%}^3, U_{50\%}^4$). Also the coil- and turn-regions between β -strands 2 and 3 show marked increase of rmsd in some simulations ($U_{75\%}^1, U_{75\%}^3, U_{50\%}^3, U_{50\%}^4, U_{50\%}^5$). Other unfolding events start in the N-terminal region ($U_{75\%}^3, U_{50\%}^2$) or in the α -helix ($U_{50\%}^3, U_{50\%}^5$).

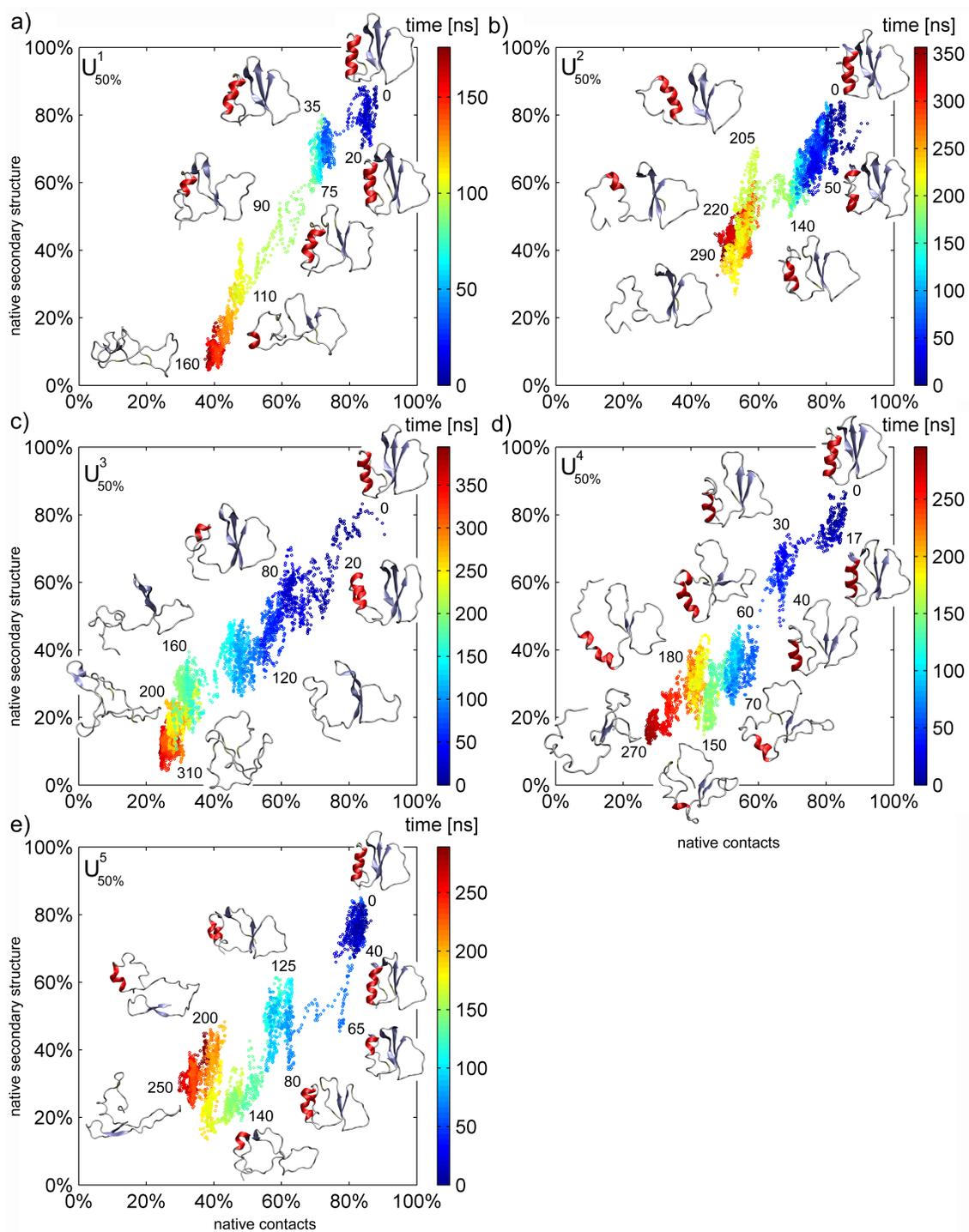


Figure 6.7: Unfolding pathways of the CI2 for the simulations in urea_{50%}, displayed as native secondary structure content versus native contact content. The numbers next to the protein structures denote the respective time of the snapshot in ns.

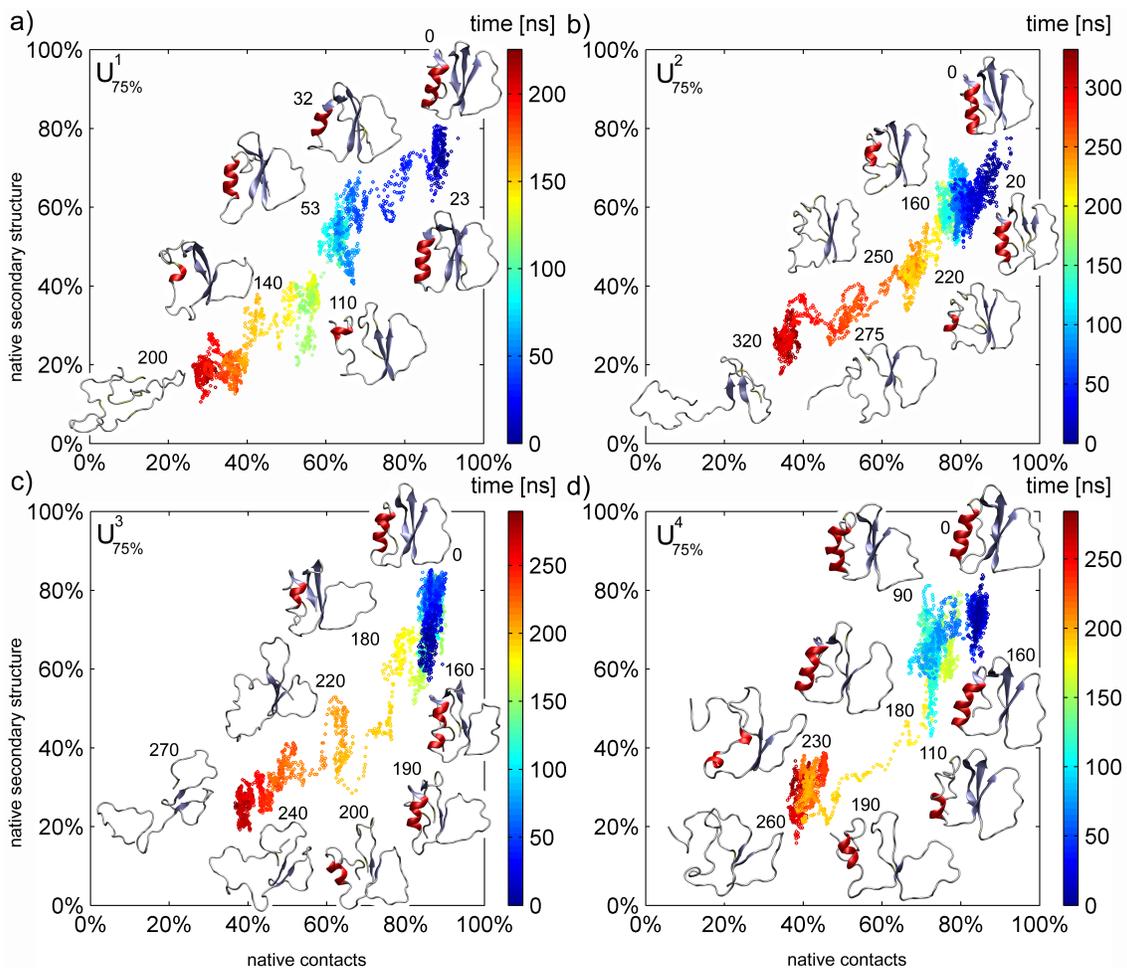


Figure 6.8: Unfolding pathways of the CI2 for the simulations in urea_{75%}, displayed as native secondary structure content versus native contact content. The numbers next to the protein structures denote the respective time of the snapshot in ns.

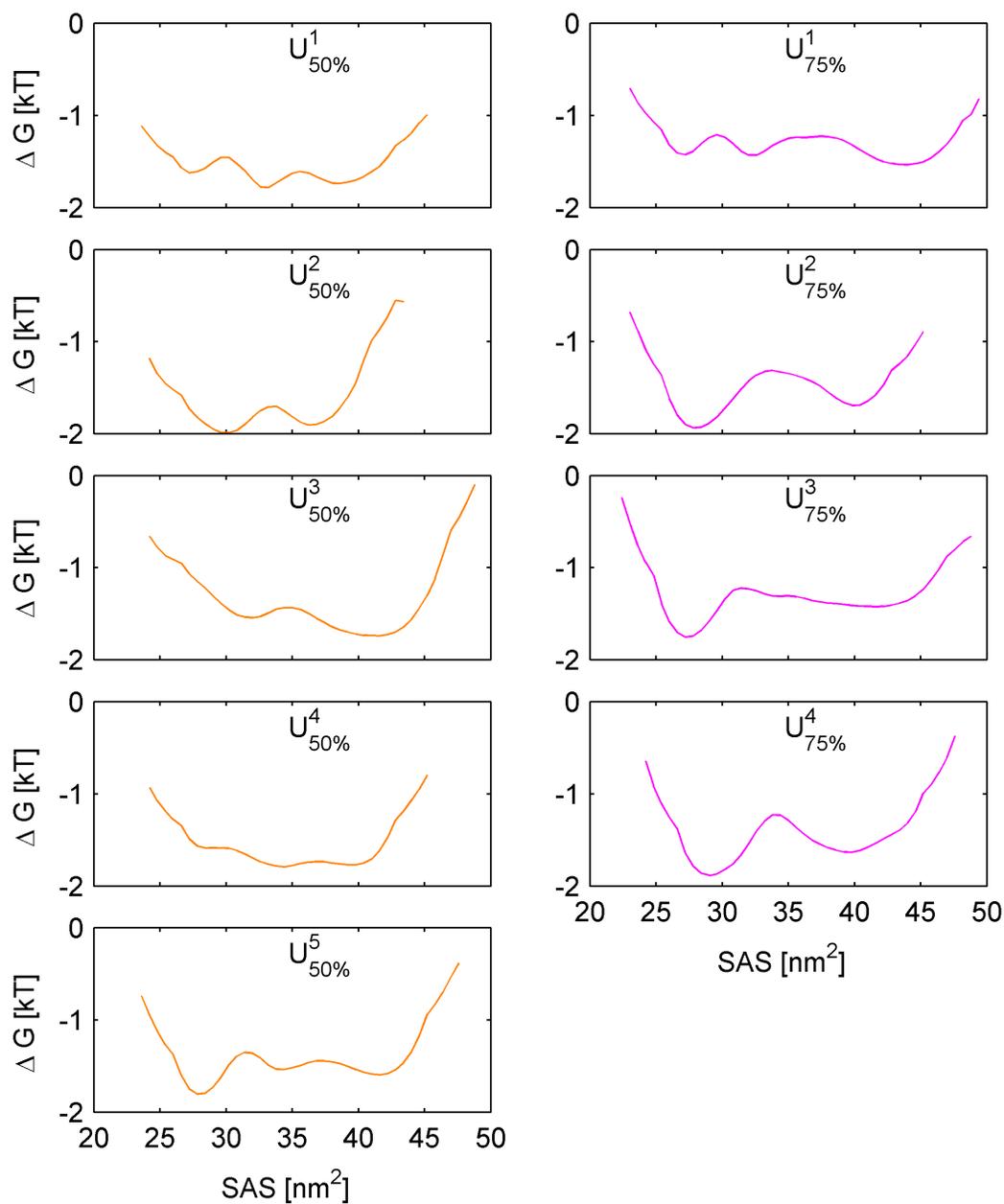


Figure 6.9: Estimated free energy profiles for the unfolding pathways of the CI2 protein in urea_{50%} (orange lines, left) and in urea_{75%} (magenta lines, right), with the SAS as reaction coordinate.

In summary, no unique unfolding “hot-spot” is found but rather several regions where unfolding likely begins.

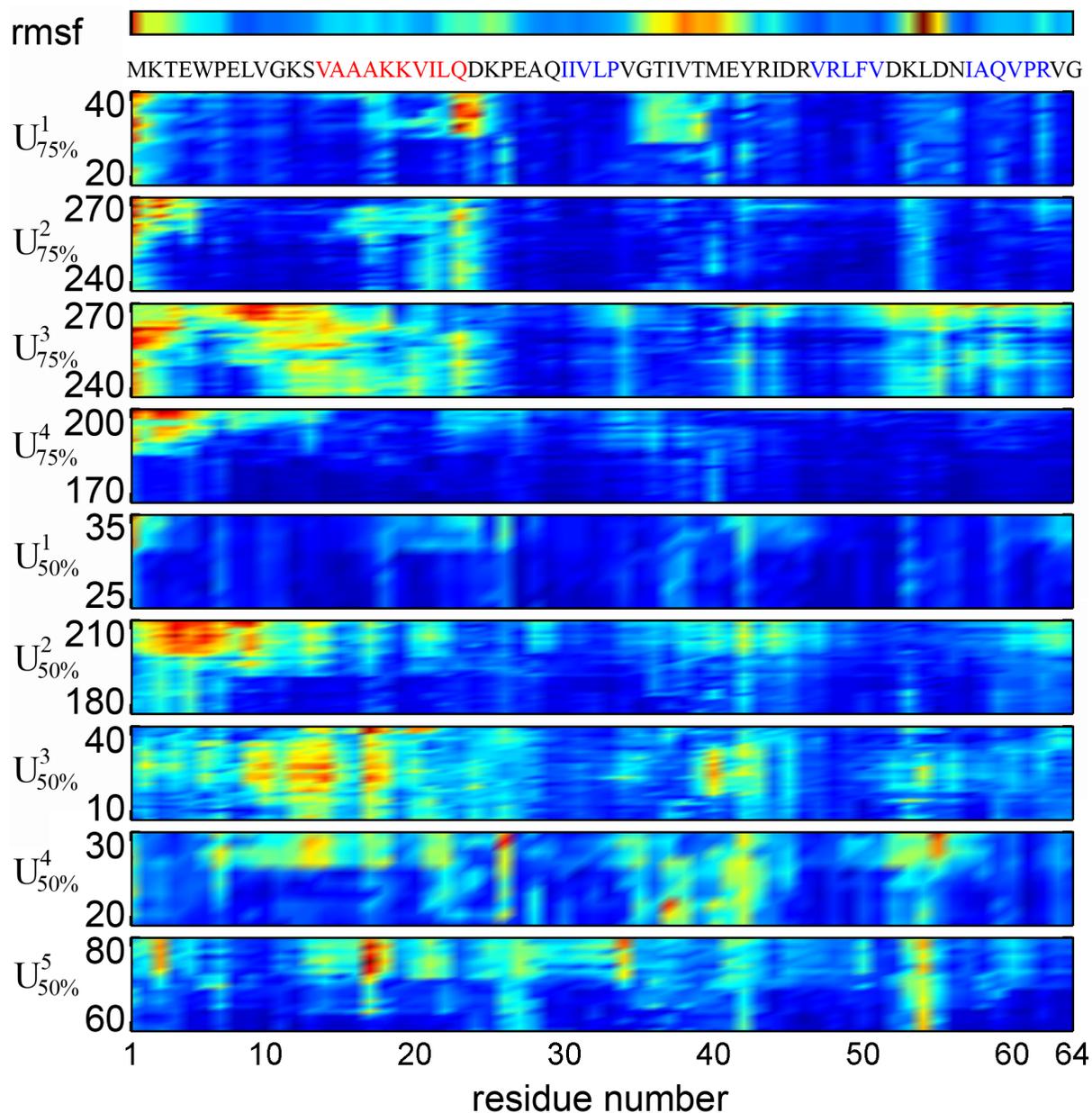


Figure 6.10: Per-residue C_{α} rmsd in the initial unfolding phases. Blue corresponds to low, red to high rmsd. The numbers on the left denote the start and end times of the respective displayed segment in ns. Top row: root-mean-square-fluctuations per residue in the native state. In the one-letter sequence code below, red marks the α -helix and blue the β -strands.

6.3.4 Residual structure in the denatured state

Figure 6.11 shows the frequency of several secondary structure elements as a function of solvent-accessible surface area, which is used here as reaction coordinate. In the simulations in water, SAS-fluctuations in the native state of the protein are, on average, accompanied by conversion of β -structure (blue lines), mostly into random coil (black lines). The simulations in urea_{100%} show a qualitatively similar behavior, with the SAS shifted by a few nm² to larger values, as discussed above. As in the simulations in water, increases in SAS correlate primarily with even more pronounced loss of β -structure and gain of random coil. Also, loss of helix structure is visible for SAS values of above 30 nm². The SAS of the protein in urea_{150%} is lower than in the simulations with urea_{100%}, and comparable to the simulations in water (W). Interestingly, the β -structure content is systematically larger for urea_{150%} than for water. This finding is consistent with the view that hyperpolar urea in fact stabilizes the native state of the protein.

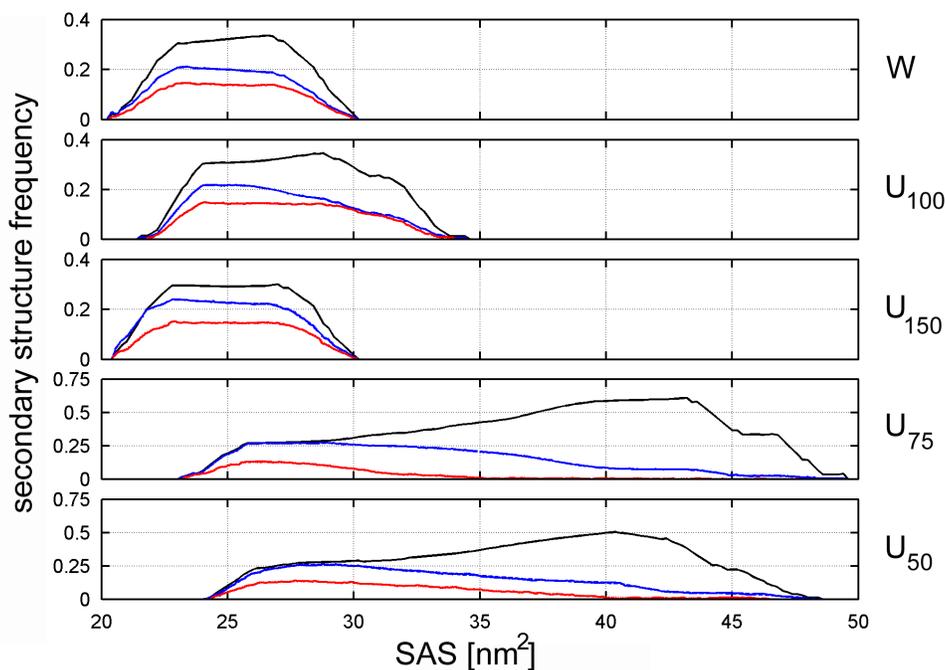


Figure 6.11: Population of secondary structure elements versus solvent-accessible surface area in the different solvents. From top to bottom: water, urea_{100%}, urea_{150%}, urea_{75%}, urea_{50%}. black: coil, blue: β -sheet, red: α -Helix

During unfolding in the simulations with urea_{75%} and urea_{50%}, helical (red lines) and β -structure turn into random coil, and, to some degree, bends. For urea_{75%}, there is less helical and more coil structure than for urea_{50%}. Almost all residual secondary structure in the denatured ensemble is bend according to the DSSP classification scheme.⁶² The presence of only little residual secondary structure in the denatured state of the CI2 protein agrees with experimental results.²¹²

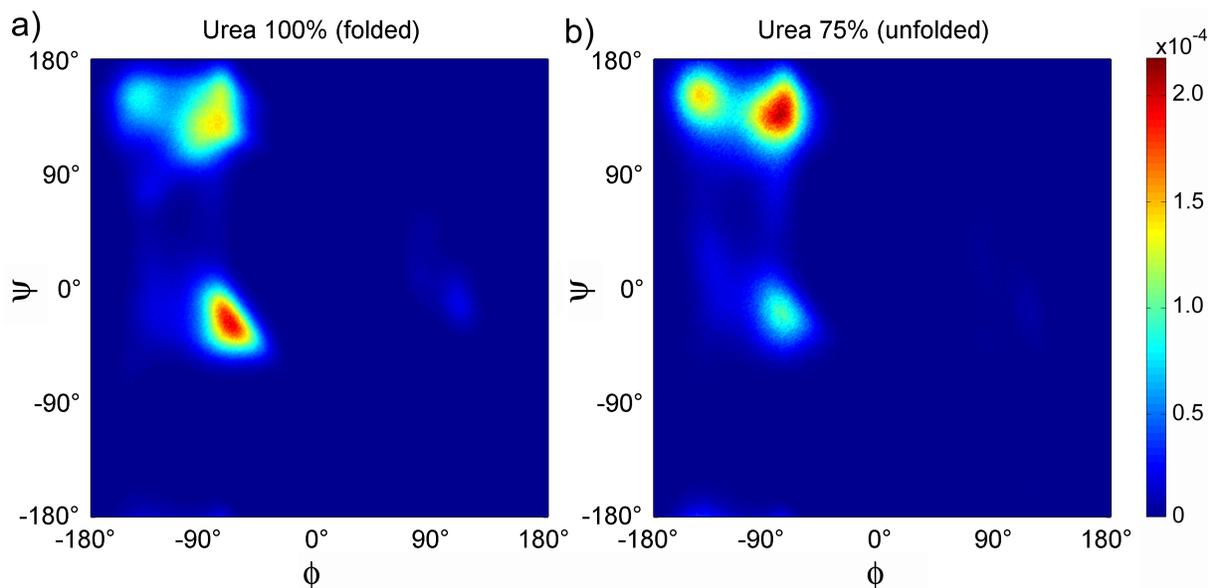


Figure 6.12: Ramachandran plots for a) CI2 in urea_{100%} (folded state), b) CI2 in urea_{75%} (unfolded state).

For the denatured ensemble, polyproline II helix structure (P_{II}) has been suggested, on the basis of CD-spectroscopy,²¹³ as prevalent backbone-configuration ($\phi = -75^\circ, \psi = 145^\circ$). Recently, this suggestion has gained considerable attention due to accumulating evidence for residual structure of denatured proteins.^{32,34,35,36}

To address this issue, the fraction of peptide bond angles in P_{II} configuration, defined by $(\phi, \psi) = (-75 \pm 10^\circ, 145 \pm 10^\circ)$, was calculated. On average, about 6% of the protein was found to be in P_{II} configuration in the simulations in water at 333 K and an almost identical number (7%) for the simulations of the native protein in urea_{100%}. The maximal P_{II} content of the denatured protein in the simulations with urea_{75%} and urea_{50%} was about 13%. In particular, the P_{II} population was large when the SAS of the denatured protein was very high ($>40 \text{ nm}^2$).

Fig. 6.12 shows the Ramachandran-plot for the folded protein in urea_{100%} (panel a) and the denatured protein in urea_{75%} (panel b). These two (ϕ, ψ) distributions are representative for the folded and denatured protein, respectively, and the distributions in the other simulations (native state: water, urea_{150%}; denatured state: urea_{50%}) were mostly similar (data not shown). As expected, the native state predominantly occupies two regions in (ϕ, ψ) space; the α region around $(-70^\circ, -27^\circ)$, and the β -sheet region around $(-83^\circ, 128^\circ, \text{parallel})$ and $(-142^\circ, 149^\circ, \text{antiparallel})$.

For the denatured protein, the P_{II} region was found to be the most populated one, as can be seen in Fig. 6.12b. However, other regions were also significantly populated (mainly helix, $\phi \approx -143^\circ, \psi \approx 152^\circ$ and anti-parallel β -sheet, $\phi \approx -72^\circ, \psi \approx -18^\circ$). This presence but not clear prevalence of the P_{II} configuration corroborates recent results from Scheraga et

al.,³⁶ who argued P_{II} to be one of several possible backbone conformations in the denatured state.

6.4 Summary and conclusion

To elucidate whether polar or, in contrast, apolar interactions of urea with proteins are the key driving force for urea-induced denaturation, *Gedankenexperiment*-simulations were performed, in which the respective denaturation strengths of hyperpolar urea (with strengthened polar interactions) or hypopolar urea (with strengthened apolar interactions) were compared. To this end, the CI2 protein was simulated in water, in regular urea, and in hypo- and hyperpolar urea, which was realized by scaling the partial charges of the urea force field.

In all nine simulations with reduced urea polarity, the protein unfolded. In contrast, it remained stable in the simulations with increased urea polarity, and the folded state was found to be even slightly more compact than in water. These results provide strong evidence that interactions with less polar parts are the main driving force for urea-induced protein denaturation, rather than polar interactions.

For regular urea, the preferences of the individual amino acids for contact with either urea or water were largely similar to those found in Chapter 5 for tripeptides. In particular, less polar residues interacted preferentially with urea, whereas polar and particularly charged residues had larger preferences for interaction with water. As expected, this characteristic interaction profile was amplified for hypopolar urea, and inverted for hyperpolar urea.

The results for urea with regular charges, particularly that the CI2 protein does not unfold within several hundred nanoseconds, agrees with the measured millisecond unfolding time.²⁰¹ Thus, the simulation results from Ref. 15, where complete unfolding was reported to occur within several nanoseconds, could not be reproduced here. Future work might address the discrepancies with systematical variation of the force-field and the treatment of the electrostatics, which are the main differences in the simulation setups.

With respect to the question whether denaturation proceeds via specific pathways or in a structurally rather heterogeneous way, these simulation results suggest the latter. Unfolding of the CI2 was seen to begin in several regions rather than one specific. However, regions with large structural fluctuations under physiological conditions seemed to be particularly likely for the onset of unfolding. Moreover, the nine unfolding pathways in the simulations with urea_{75%} and urea_{50%} turned out to share no common conformations during unfolding, which is in agreement with the fact that CI2 is a two-state-folder without metastable folding intermediates. This heterogeneity of unfolding pathways might give rise to the idea of an “inverse funnel”-like energy landscape for unfolding: the energy-landscape

of the denatured state is relatively flat, and multiple pathways lead from the folded state down to the denatured ensemble.

While no shared conformations were found in the different unfolding pathways, more general common features of the unfolding process were found. In particular, unfolding was observed to proceed with alternating loss of secondary and tertiary structure, often step-wise. This finding is consistent with the coupling between secondary and tertiary structure formation in the nucleation-condensation folding process of the CI2,^{212,202,203} which might suggest that the processes of structure-formation during folding and structure-loss during denaturation can be related to each other.

Finally, regarding residual structure in the denatured state, these simulations confirm that the denatured state of CI2 contains only little residual secondary structure.²¹² Polyproline II helix structure was found here in the denatured CI2, but not to a large extent. This finding supports the recent suggestion that polyproline II is one of several possible backbone conformations in the denatured state.³⁶

7

The effect of urea on partially unfolded proteins

7.1 Regular urea — Hydrophobic collapse lost.

7.1.1 Introduction

The previous Chapters have investigated the molecular driving force of urea-induced protein denaturation, and substantial evidence was found that direct apolar contacts between urea and less polar residues or the peptide backbone are the determinant.

This Chapter now addresses the question how urea affects protein conformations at different states of folding/unfolding. In particular, does urea actively destabilize the folded state, or rather stabilize the denatured state? Calorimetric studies point towards a stabilization of the denatured state,³⁰ but processes are unclear at the molecular level. Per se, MD simulations are perfectly suited to provide insight at the molecular level. However, simulations of the CI2 protein in Chapter 6 have shown that urea has only little effect on the native state even within several hundred nanoseconds, and urea-induced denaturation usually occurs on timescales still orders of magnitude out of reach of state-of-the-art computer simulations at atomic detail. To circumvent this problem, here, a number of partially unfolded structures of the Cold Shock protein Bc-Csp from *bacillus caldolyticus* are generated by a high-temperature unfolding simulation. Subsequently, the dynamics of these partially unfolded structures are investigated in water and in urea at room temperature. This approach allows to study the effect of urea on different states of folding/unfolding with atomic detail within feasible simulation times.

7.1.2 Methods

Simulation setup. The crystal structure of the Cold Shock protein (Fig. 7.1) was obtained from the Protein Data Bank,⁹⁷ PDB-code 1C9O.⁹⁹ The size of the rectangular simulation box was chosen such that a minimum distance of 1.5 nm between protein atoms and the box was kept in each direction. For the solvation of the protein, pre-equilibrated structures of water and 8 M urea were used (Chap. 4). Sodium and chloride ions were added to obtain a physiological ion concentration of 150 mM.

A high-temperature unfolding simulation at 700 K in water was performed to generate a number of partially unfolded conformations. Here, constant volume (NVT ensemble) conditions were applied to prevent solvent evaporation, and the timestep was set to 1 fs to account for the increased velocities of the particles. Figure 7.2 shows the SAS of the pre-unfolding simulation as well as four partially unfolded structures (I–IV) and one completely unfolded structure (D). Together with the native structure (N) and the unfolded structure (D), structures I–IV, were used as starting structures for the actual simulations in water or urea at 300 K, which are listed in Table 7.1. To avoid over-interpretation of possibly anecdotal events, multiple simulation runs were carried out for the starting structures III and IV, on which the main conclusions are based. The total simulation time was $\approx 4.4 \mu\text{s}$.

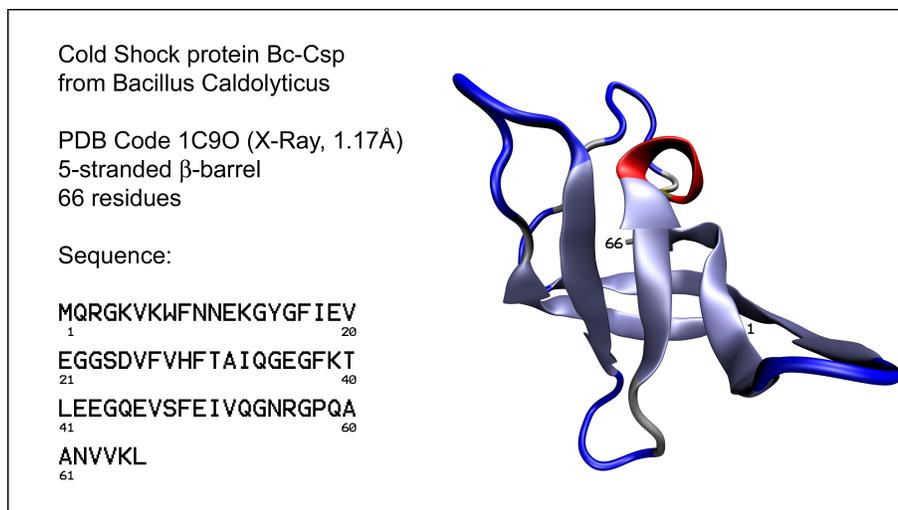


Figure 7.1: The Cold Shock protein Bc-Csp.

starting structure	solvent	number of simulations	simulation time
N (0 ps)	water	1	219 ns
N (0 ps)	8 M urea	1	453 ns
I (750 ps)	water	1	209 ns
I (750 ps)	8 M urea	1	151 ns
II (1000 ps)	water	1	167 ns
II (1000 ps)	8 M urea	1	143 ns
III (1250 ps)	water	3	139 ns / 250 ns / 244 ns
III (1250 ps)	8 M urea	3	119 ns / 254 ns / 255 ns
IV (1500 ps)	water	3	196 ns / 138 ns / 230 ns
IV (1500 ps)	8 M urea	3	142 ns / 274 ns / 276 ns
D (3000 ps)	water	1	308 ns
D (3000 ps)	8 M urea	1	314 ns

Table 7.1: Summary of all 20 simulations performed for the Cold Shock protein, starting from the native structure (N), the partially unfolded structures (I–IV), or the completely unfolded structure (D).

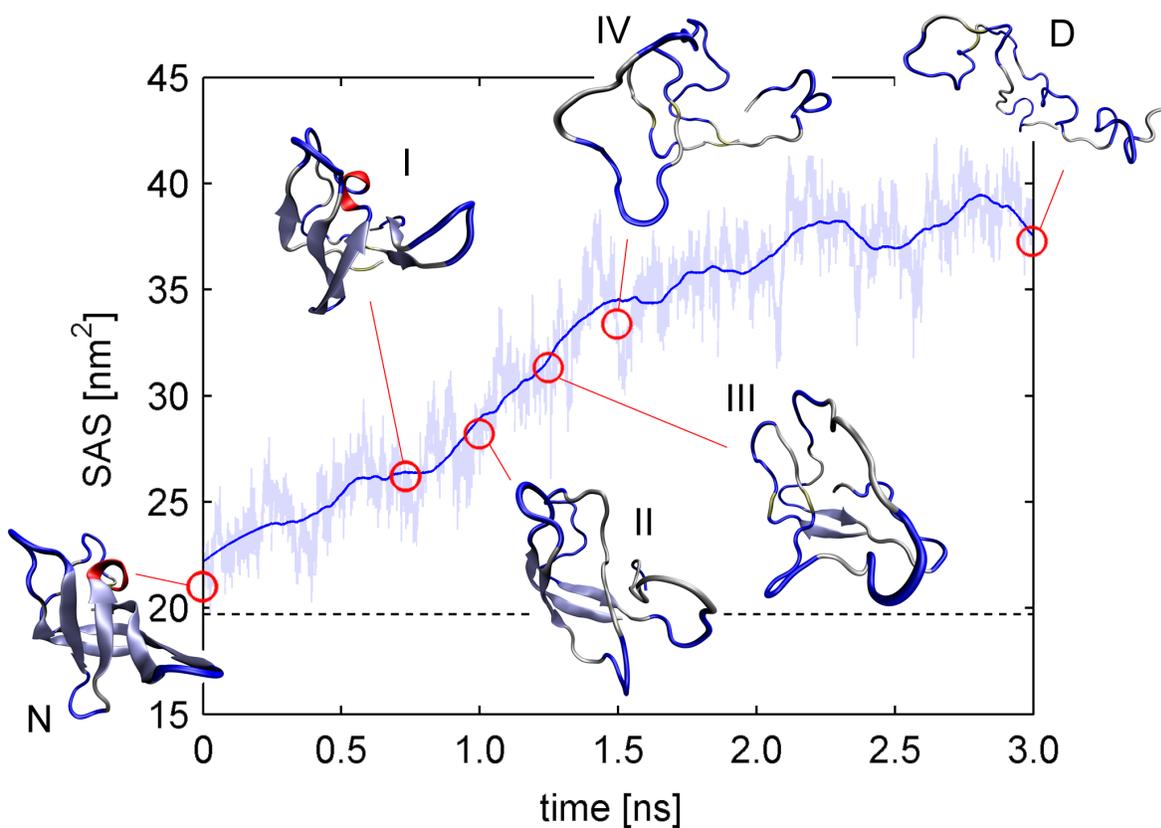


Figure 7.2: SAS of the protein in the high-temperature unfolding simulation. The bold line shows the running average over 250 ps, the dim line shows raw data. Also shown are the snapshots of the four partially unfolded structures (I–IV) selected for subsequent room temperature simulations, together with the native state (N) and the selected completely unfolded structure (D). The dotted line denotes the initial SAS of the folded protein.

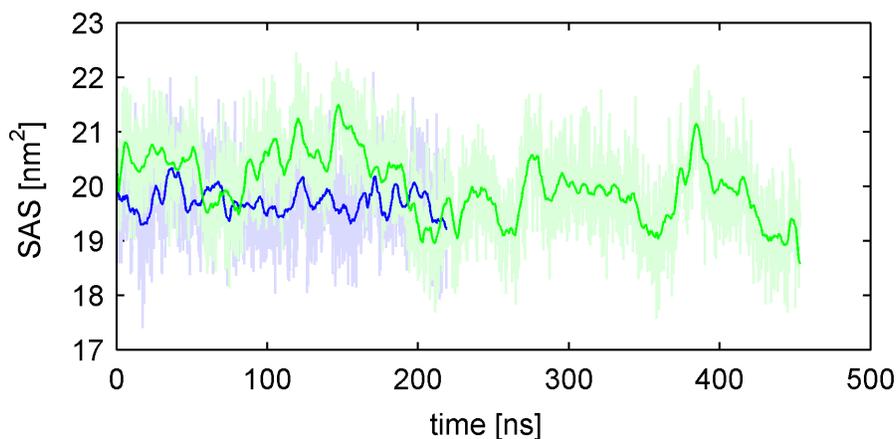


Figure 7.3: Solvent accessible hydrophobic surface area (SAS) of the protein; blue: in water, green: in 8 M urea.

7.1.3 Results and discussion

Dynamics of the native state

Subsequently, the solvent accessible hydrophobic surface area (SAS) of the protein is used as reaction coordinate for the transition from the folded state to the unfolded state. Figure 7.3 shows the SAS for the simulations of the native state, starting from the crystal structure. As can be seen, the SAS of the protein does not increase significantly in the simulation with water (blue lines) or with 8 M urea as solvent (green lines). Thus, the folded state of the protein is stable within the simulation times, as is also confirmed by other measures for conformational changes, such as radius of gyration or root-mean-square-deviation of the backbone atoms (data not shown). As expected from the experimental timescale for denaturation of Cold Shock proteins,^{214,215} no unfolding is observed.

Nevertheless, an effect of urea on the folded protein is seen. As had been observed for the CI2 protein in Chapter 6, the SAS of the protein is higher in urea than in water. Here, for the Cold Shock protein, this difference is on average much smaller ($\approx 0.3 \text{ nm}^2$) than for the CI2 protein ($\approx 2.0 \text{ nm}^2$), and the SAS fluctuations are larger. Further analysis reveals that this SAS difference is mainly caused by transient exposure of PHE27 and ARG56, and the fluctuations in the total SAS correspond to formation and disruption of contact between these residues, which have stacking interaction with each other (Fig. 7.4). Whereas this stacking contact between PHE27 and apolar parts of ARG56 is persistent in the simulation with water (Fig. 7.4a), it is destabilized in urea (Fig. 7.4b), where urea molecules disrupt the stacking interaction, as can be seen from the snapshots in (Fig. 7.4c). Most notably, this stacking contact is also disrupted several times in the simulation in water (note the shorter time scale of the water simulation), but the contact reforms again after a few nanoseconds. This suggests that urea here stabilizes a more open conformation which was adopted due to thermal fluctuations. This destabilization of less polar contacts by urea

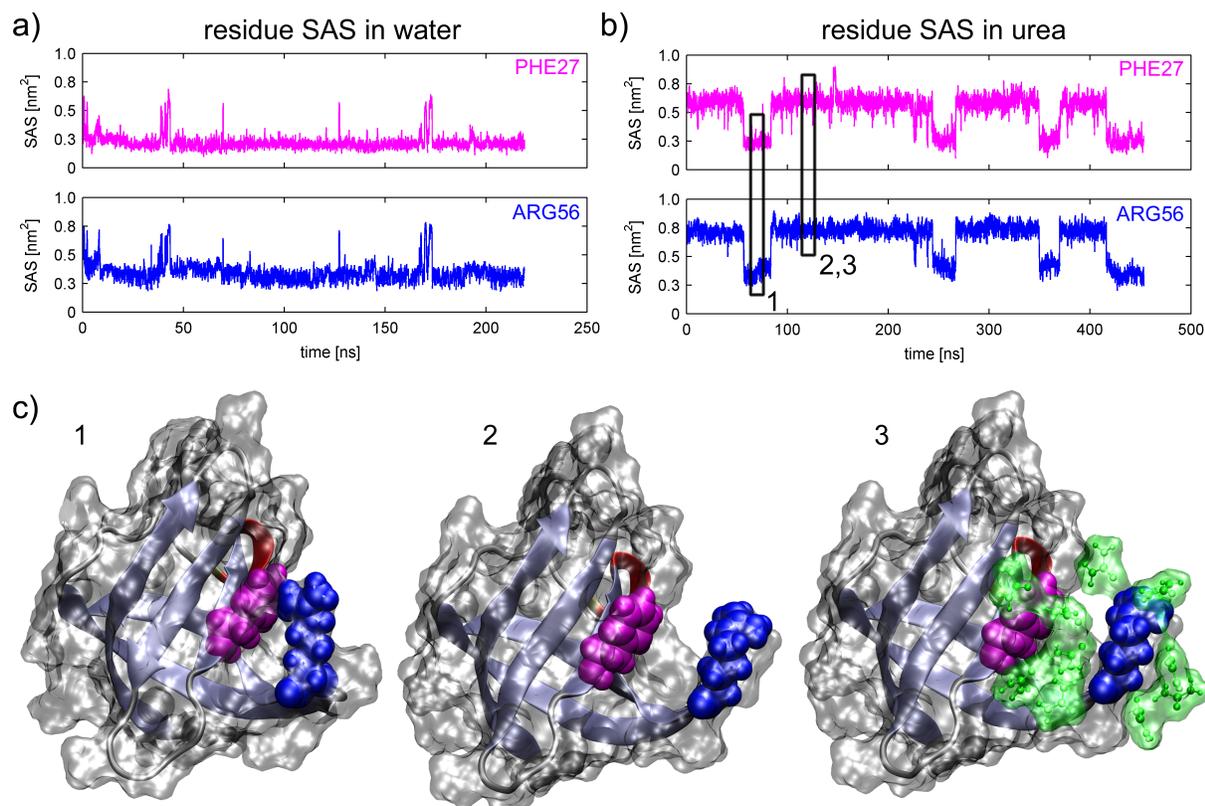


Figure 7.4: The apolar contact between PHE27 and ARG56 is disrupted in urea, which gives rise to the fluctuations in the total SAS. (a) SAS of PHE27 and ARG56 in water, (b) SAS of PHE27 and ARG56 in urea, (c) snapshots of the protein, (1): with PHE27 (magenta) and ARG56 (blue) in contact, (2): with PHE27 and ARG56 not in contact, (3): same as (2) with adjacent urea molecules displayed (green).

might be the first step of unfolding. However, no further destabilizing effects of urea on the protein were observed within the simulation time of ≈ 450 ns.

Dynamics of partially unfolded structures

To investigate the effect of urea at different states of folding/unfolding, partially unfolded structures (I–IV, D) of the Cold Shock protein were generated in a high-temperature unfolding simulation, and subsequently simulated in water and in urea at room temperature.

Solvent accessible surface. The SAS of structures I and II (Fig. 7.5a,b) do not show significant differences between water (blue lines) and aqueous urea solution (green lines). For structure I, the SAS decreases for both solvents during the first 40 ns from the initial value of ≈ 26 nm² (upper dashed line) and then fluctuates around constant averages which

starting structure	$\Delta_{UW}SAS$ [nm ²]
N (0 ps)	0.3 ± 0.1
I (750 ps)	1.7 ± 0.2
II (1000 ps)	1.7 ± 0.2
III (1250 ps)	5.4 ± 0.2
IV (1500 ps)	5.8 ± 0.2
D (3000 ps)	8.8 ± 0.3

Table 7.2: Average SAS differences of the protein in the simulations with urea or with water.

differ from each other by $\Delta_{UW}SAS=1.7\pm 0.2$ nm². Structure II shows a similar behavior of the SAS, except that the initial decrease in SAS is less pronounced for both solvents.

Whereas for structures I and II the $\Delta_{UW}SAS$ is rather small, a pronounced differential effect of the solvent on the SAS is seen for the more unfolded structures III, IV, and D (Fig. 7.5c–e, respectively). Here, the SAS of the protein stays constant or even increases for most of the simulations with urea, while a significant decrease of the SAS occurs is observed for all simulations with water. Correspondingly, as summarized in Tab. 7.2, the average $\Delta_{UW}SAS$ is quite large for the later unfolding stages. In all of these simulations, a hydrophobic collapse is seen for the protein in water, which apparently is prevented by the urea solvent.

Native structure formation. As must be expected, the protein collapse does not lead back to the native state during the few 100 ns simulation time, which is still much shorter than the folding time. However, for certain structures, partial formation of native contacts (left panel in Fig. 7.6) and native secondary structure elements (right panel in Fig. 7.6) is observed. For structures I and II, native contacts form to some extent in water but not in urea (Fig. 7.6a,b), or at least at a slower rate. Hence, the reduction of SAS is caused by partial refolding of the protein, which is more pronounced for conformation I, which is closer to the native conformation, than for conformation II. Interestingly, more native contacts form in the simulations with water than in those with urea. For structures III, IV, and D (Fig. 7.6c–e, respectively), no significant differences between the native contact formation for the two different solvents are seen. Regarding native secondary structure formation, a difference between water and urea is observed for structure III (Fig. 7.6c), and, less pronounced, for structure IV (Fig. 7.6d). In these cases, native secondary structure is partially formed in water but not in urea. For the structures I, II, and D (Fig. 7.6a,b,e), no significant difference between both solvents is seen. However, these results are not as clear or statistically certain as those regarding the hydrophobic collapses. Hence, it is focussed here on the effect of urea on the hydrophobic collapse.

These results are supported by experimental findings. CD-spectroscopic measurements showed that no native state topology is present in collapsed unfolded Csp.²¹⁶ Moreover, a similar collapse to a disordered state as seen here was observed to precede folding of

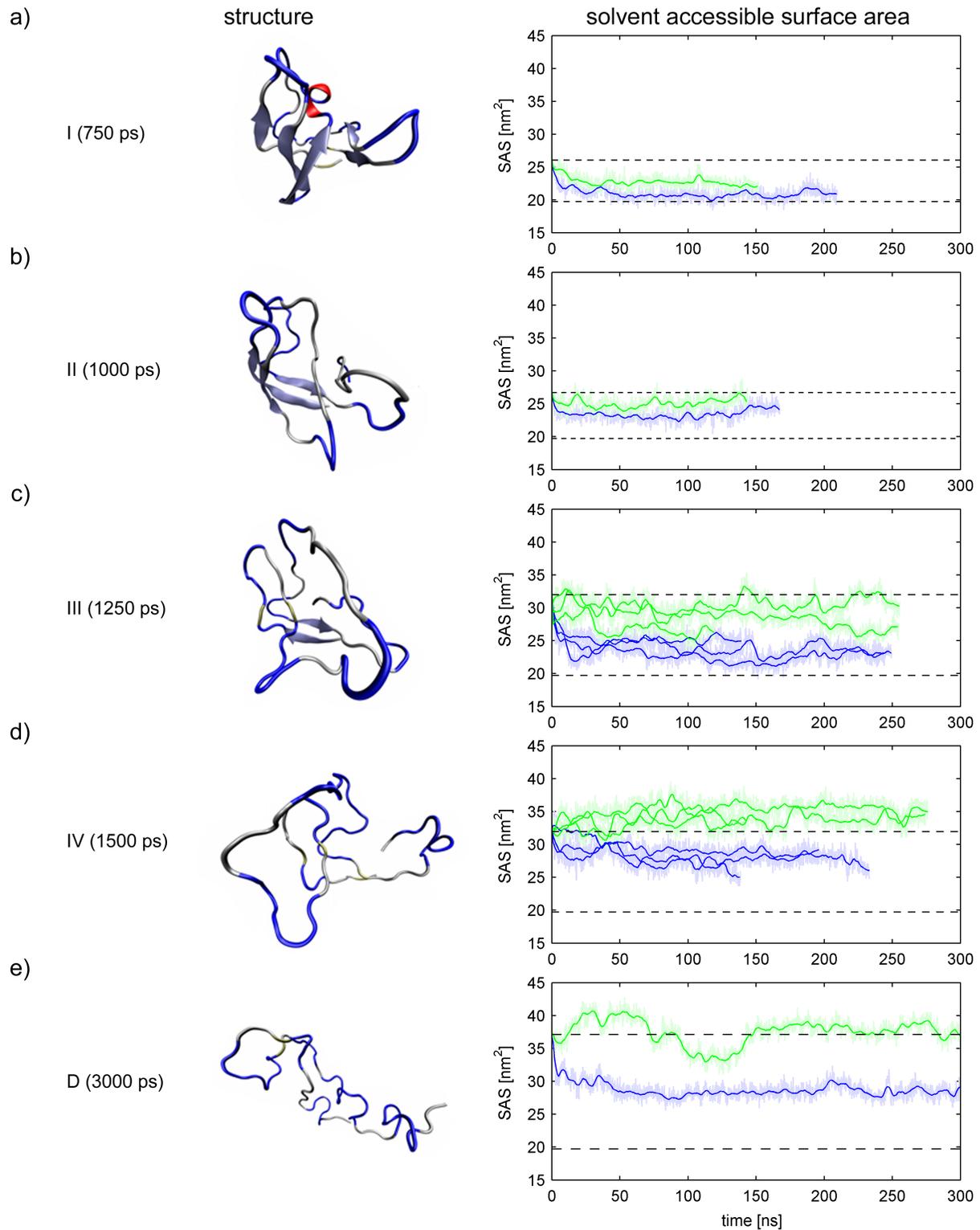


Figure 7.5: SAS for the partially unfolded structures I–IV, and D (a–e, respectively). Left panel: snapshots of the respective structure. Right panel: SAS. The lower dotted lines indicate the SAS of the protein in the crystal structure. The upper dotted lines show the initial SAS of the respective starting structure.

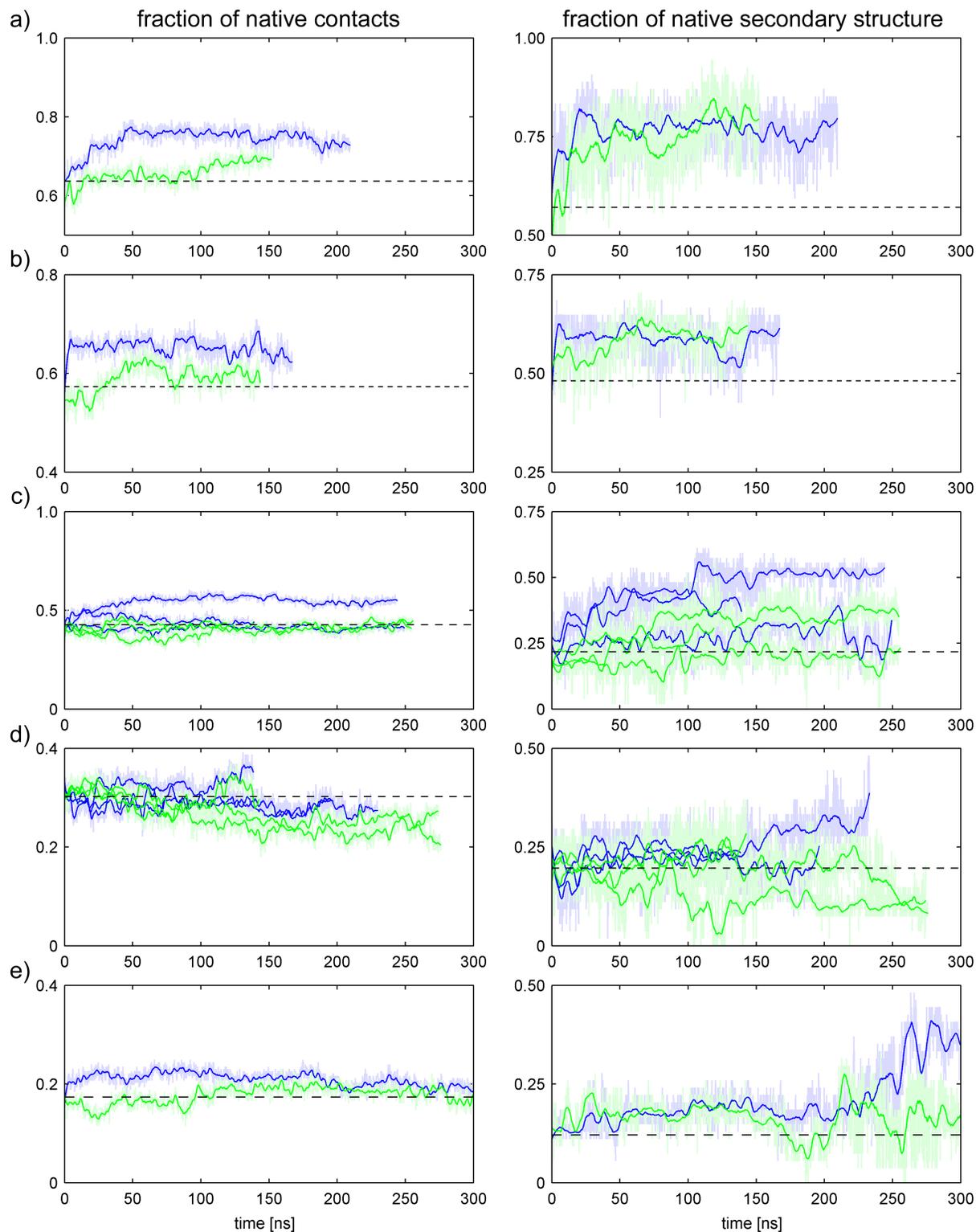


Figure 7.6: Native structure content for the partially unfolded structures I-IV, and D (a-e, respectively). Left panel: fraction of native contacts. Right panel: fraction of native secondary structure. The dotted lines show the initial values of the respective starting structure.

Cold Shock proteins in kinetic experiments.^{217,218} In particular, as in these simulations, the collapse was found to be faster than secondary structure formation.²¹⁹ The timescale of the hydrophobic collapses observed here (20–70 ns) is in excellent agreement with the timescales between 50 and 70 ns found in experiments for the Cold Shock protein²²⁰ or some other proteins of similar size.^{219,221}

Collapses at the residue-level. Further detailed analysis of the hydrophobic collapse events mentioned above revealed that local collapses of parts of the protein are the reason for the decrease in SAS. For illustration, an example is given in Fig. 7.7. Panel a shows the SAS of TRP8 and SER24 of conformation IV from simulations W_1^{1500} and U_1^{1500} . As can be seen, in the simulation with water (blue), the SAS of both residues decreases, whereas it increases or remains constant in the simulation with urea (green). In Fig. 7.7b, which highlights the SAS of both residues in the simulation with water for the time between 40 and 60 ns, a rapid decrease of the SAS to low values is seen between 40 and 55 ns. Fig. 7.7c shows three snapshots of these two residues and the surrounding protein during this fast local hydrophobic collapse. At the start of the collapse, (40 ns, left panel), TRP8 (dark blue) and SER24 (magenta) are separated and largely exposed to the solvent, with the adjacent residues (light gray) being flexible and loosely packed. After 48 ns, this initially open pocket has started closing (middle panel), thereby shielding hydrophobic areas from the solvent, and thus reducing the SAS. At 56 ns, the pocket is nearly closed, and TRP8 and SER24, as well as neighboring residues, are almost completely buried from the solvent. Similar events are seen in all of the simulations in water and represent the main contribution to the observed total SAS decrease. For urea as solvent, no or only rare collapses of this kind are seen.

Interactions between urea and the protein

With the results of the previous chapters in mind, the molecular cause for the impediment of hydrophobic collapses by urea is at hand. Favorable apolar contacts of urea with less polar residues were found to be the driving force of denaturation, and it was suggested that these apolar contacts reduce the hydrophobic effect for less polar residues. Here, the absence of hydrophobic collapses can immediately be explained by a reduced hydrophobic effect. It remains to be checked now, if those preferential contacts between urea and less polar residues are also present in the simulations of the Cold Shock protein presented here. For the folded state, the disruption of an apolar contact between PHE27 and the apolar part of ARG56 was already qualitatively related to the presence of urea molecules around these residues. Now, a quantitative analysis of the interactions between urea and the residues of the Cold Shock protein is performed.

For this purpose, the urea/water contact coefficient C_{UW} was calculated for all residues of the Cold Shock protein. Since the focus here is on the interaction of urea with open, non-collapsed, conformations, contact coefficients were calculated from the trajectory of

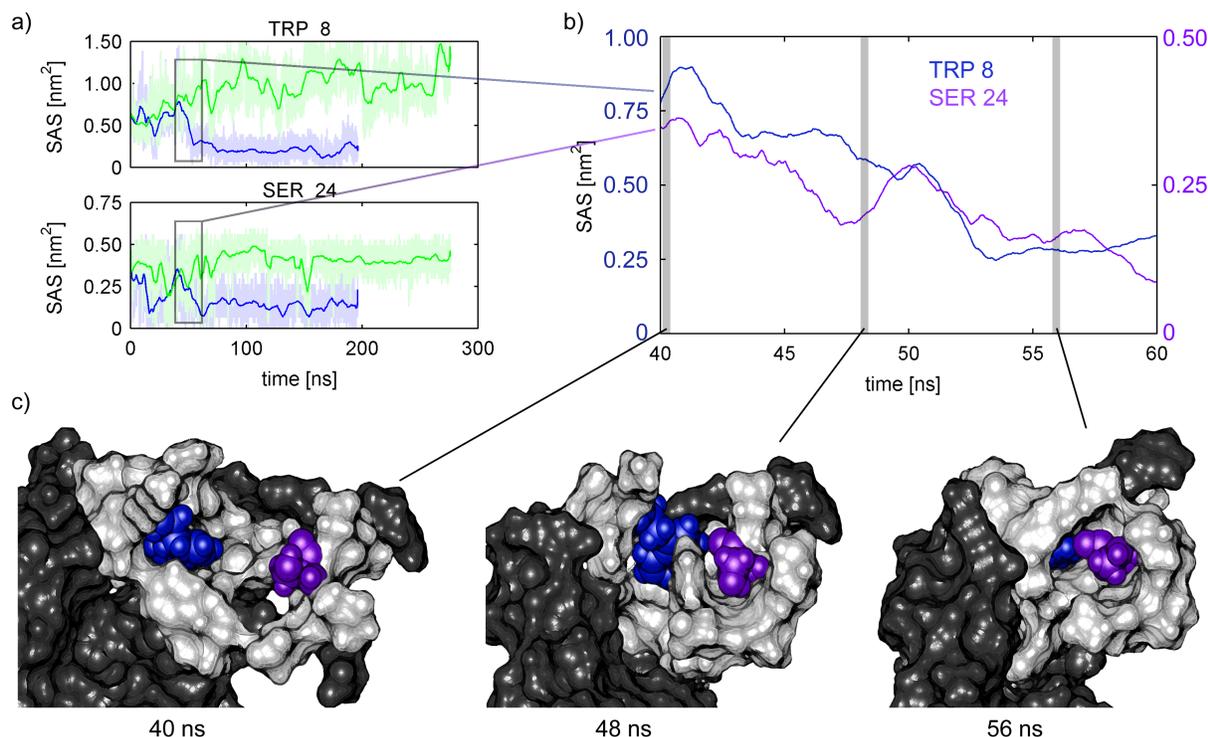


Figure 7.7: a) SAS of TRP8 and SER24 in simulations of structure IV in water (blue lines) and urea (green lines). b) SAS for both residues in water (dark blue: TRP8, magenta: SER24). c) The snapshots show the local hydrophobic collapse event with TRP8 highlighted in dark blue and SER24 highlighted in magenta. Residues in the collapse region are highlighted in light gray.

the unfolded structure (D). Additionally, they were also calculated from the trajectory of the folded structure (N), to assess the influence of structure on the contact coefficients.

The contact coefficients for the unfolded state ($C_{UW}^{(D)}$, Fig. 7.8a) largely agree with those presented in Chapter 5 for glycine-capped tripeptides. In particular, as was expected, less polar residues have strong contact preferences for urea. Also the backbone shows strong preference for urea contact, whereas the preference for urea contacts is lower for charged and also for polar residues. Interestingly, the contact coefficients calculated for the unfolded Cold Shock protein are on average $\approx 16\%$ higher than those calculated for the tripeptides, which is obviously an effect of the sequence. Other sequence effects on the contact coefficients were also observed. For instance, GLY44 has a significantly lower contact coefficient (1.5) than GLY on average (2.1) in the Cold Shock protein, which is an effect of adjacent GLU42, GLU43, and GLU46. A linear regression of the contact coefficients for the unfolded Cold Shock protein with those for the tripeptides shows a correlation coefficient of $r^2 \approx 0.71$. This suggests that effects from sequence and residual structure in the Cold Shock protein make up about 30% of the residue contact preferences in the unfolded Cold Shock protein.

In contrast, the correlation between the contact coefficients calculated for the folded Cold

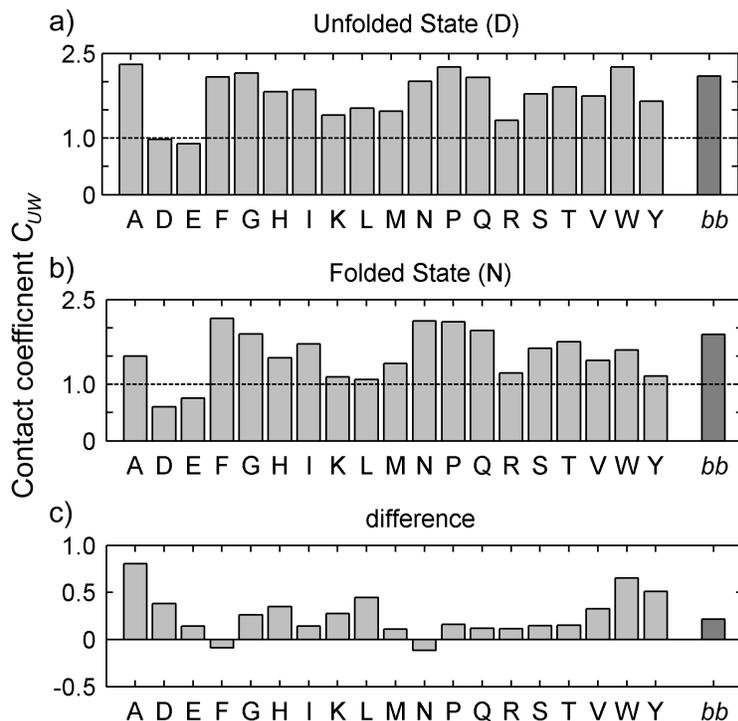


Figure 7.8: Contact coefficient C_{UW} for each residue type and the backbone average (“bb”). (a) for the unfolded state ($C_{UW}^{(D)}$), (b) for the folded state ($C_{UW}^{(N)}$). (c) difference between unfolded state and folded state ($\Delta C_{UW}^{(D)-(N)}$).

Shock protein (Fig. 7.8b) and those of the tripeptides is only $r^2 \approx 0.56$; and further, the correlation between the contact coefficients calculated from the folded or from the unfolded Cold Shock protein is $r^2 \approx 0.74$. These correlations suggest that not only sequence, but also the native structure significantly affects the contact coefficients of the individual residues.

Markov model for urea-induced unfolding

Urea was found to impede hydrophobic collapse events of partially unfolded proteins, in particular of more unfolded structures. To further quantify this effect of urea, a Markov model was derived from all simulations of structures N, I–IV, and D, using the SAS as reaction coordinate. Subsequently, the respective free energy profile was calculated as described in the Methods Chapter. Note, however, that this procedure only gives a rough estimate of the free energy profile, rather than accurate numbers. Figure 7.9 shows the estimated free energy profile of the Cold Shock protein in water (blue line), and in urea (green line). As can be seen, open conformations with a high SAS are energetically unfavorable in water. Hydrophobic collapses reduce the SAS, and, hence, lead to energetically more favorable states. In contrast, no such gradient of the free energy is seen with urea as solvent. Instead, the free energy landscape of the protein in urea is almost flat. Therefore,

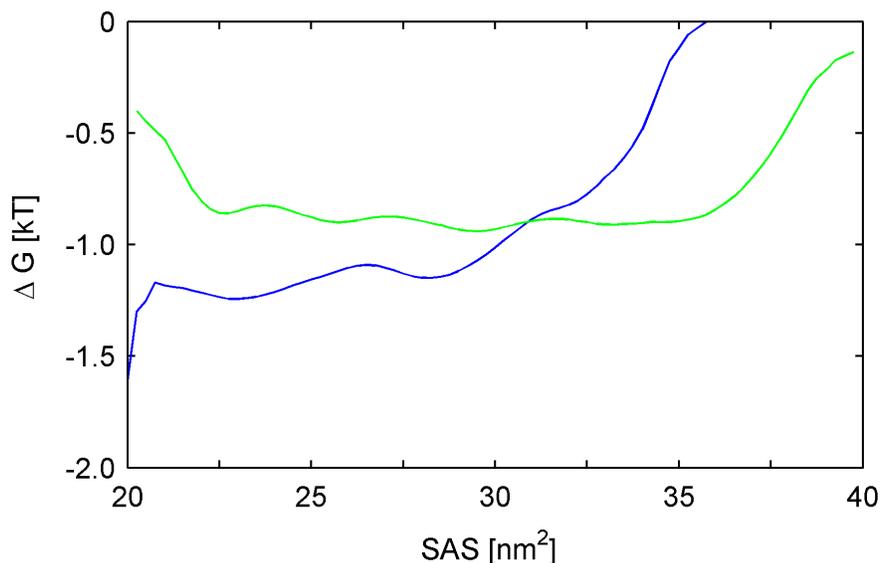


Figure 7.9: The free energy profile of the Cold Shock protein in the two solvents was estimated from a Markov model, with the SAS as reaction coordinate. Shown is the estimated free energy of the Cold Shock protein in water (blue), and in urea (green).

hydrophobic collapses would not lead to a decrease of the free energy, and hence do not or only rarely occur.

7.1.4 Summary and conclusion

To investigate the effect of urea on proteins at different stages in the folding/unfolding process, partially unfolded structures of the Cold Shock protein Bc-Csp were simulated in water and in aqueous urea solution.

Simulations of the native state up to 450 ns showed no unfolding in water or in urea. The main differential effect of the two solvents on the protein was the slightly higher SAS of the protein in urea. On the molecular level, this was caused by a disruption of residue contacts. In total, however, the difference in structure and dynamics of the protein in water or urea were found to be only small.

In the simulations of the partially unfolded conformations, in contrast, significant differential effects of the two solvents on the protein dynamics were observed. While the SAS of the protein decreased clearly in all simulations with water, the changes in SAS were much smaller in aqueous urea. These differences were more pronounced the more unfolded the starting structure was. In the two most unfolded structures investigated, the SAS even increased in urea, indicating further unfolding of the protein. This SAS decrease in water resulted from hydrophobic collapses of the protein, which constitute the first step of

refolding. On the residue level, collapse events correspond to contact formation between residues, which thereby reduce their solvent exposed surface and thus the total SAS of the protein. For urea as solvent, no or only rare collapses of this kind are seen.

Regarding the stage of folding/unfolding where urea acts, these results suggest that urea impedes refolding of partially unfolded proteins rather than that it actively destabilizes the native state. Effectively, however, this shifts the equilibrium towards the denatured state in the same manner as active destabilization of the native state. This interpretation is supported by experimental data for the folding and unfolding kinetics of Cold Shock proteins, which show that the (re)folding rate is strongly dependent on denaturant concentration, whereas the unfolding rate is almost constant.^{222,214,215}

Apart from the impediment of hydrophobic collapses at later stages of unfolding, these simulation results give indications for effects of urea on less unfolded structures. While the impediment of hydrophobic collapses by urea is predominantly observed for more unfolded structures, urea also seemed to impede the re-formation of native secondary structure in medium unfolded structures. Further, urea was found to have the largest influence on the re-formation of native contacts in structures unfolded to a low or medium degree. This sequence of prior hydrophobic collapse, followed by secondary structure formation, and finally the formation of native contacts, is in line with previous suggestions about the sequence of folding steps.²²³

To investigate the molecular cause for the absence of hydrophobic collapses in aqueous urea, interactions between urea molecules and the protein residues were analyzed via calculation of contact coefficients. As expected from the results of the previous chapters, almost all amino acids were found to have contact preferences for urea, and this preference was particularly pronounced for less polar residues as well as for the peptide backbone. This result suggests that favorable direct interactions of urea with less polar residues weaken the hydrophobic effect, and hence the driving force for hydrophobic collapse.

This explanation for the impediment of hydrophobic collapses by urea seems intuitive, and is in line with the view that has emerged from the previous chapters. So far, however, this explanation is only a suggestion. In particular, it has not been shown that the *correlated* observation of preferential apolar contacts and impeded hydrophobic collapses actually implies a *causality* — i.e., that the impediment of hydrophobic collapses *depends* on preferential apolar contacts. This issue will be addressed in the second part of this chapter.

7.2 Hyperpolar urea — Hydrophobic collapse regained

7.2.1 Introduction

In Chapter 5, a broad spectrum of preferences for contact with either urea or water was found for the 20 individual amino acids. Based on this result, it was suggested that it is the apolar contacts between urea and the protein — rather than the polar contacts — which are the key determinant for the denaturation. Indeed, *Gedankenexperiment*-simulations of the CI2 protein in urea with artificially increased or decreased polarity have shown that the latter is an even stronger denaturant than “real” urea, because favorable interactions of urea with apolar residues are even more favorable in hypopolar urea (Chap. 6). Hyperpolar urea, in contrast, did not destabilize the CI2 protein, and the native state was found to be even slightly more compact than in urea with realistic polarity. These results provided strong evidence that apolar contacts are indeed the driving force for denaturation. Further, these results motivated the prediction that hyperpolar urea might in fact be a weaker denaturant than real urea. With the results of Chapter 7.1, particularly the impediment of hydrophobic collapse events in urea, this prediction will now be tested. In particular, it will be checked if hyperpolar urea (150% partial charges) actually is a weaker denaturant than regular urea, in which case hydrophobic collapses should be impeded less in hyperpolar urea than in regular urea as solvent. In addition, this approach also tests whether hydrophobic collapses actually depend on apolar urea-protein contacts, as was suggested in Chapter 7.1.

7.2.2 Simulation setup

Combining the approaches of Chapters 6 and 7.1, partially unfolded structures of the Cold Shock protein were simulated in hyperpolar urea (150% partial charges). To allow direct comparison with the results presented in Chapter 7.1, similar simulation system setups were used. Intermediate structures III and IV from Chap. 7.1 (see Fig. 7.2), for which urea showed the largest effect on the hydrophobic collapse, were both simulated three times, 300 ns each. Urea polarity was increased by scaling the partial charges to 150%, as described in Chap. 6.

7.2.3 Results

Solvent accessible surface area. Figure 7.10a shows the SAS of the protein in the simulations with hyperpolar urea (red lines), in comparison to those with water (blue lines), and with regular urea (green lines). As can be seen, for all simulations starting from structure III (panel a), the SAS decreases significantly within in the first ≈ 25 ns in hyperpolar urea, similarly as it does in water. Further, the average SAS in hyperpolar urea is almost similar to that in water, as summarized in Tab. 7.3. In the simulations of structure IV, the SAS

starting structure	$\Delta_{U_{100}W}SAS$ [nm ²]	$\Delta_{U_{150}W}SAS$ [nm ²]	$\frac{\Delta_{U_{150}W}SAS}{\Delta_{U_{100}W}SAS}$
III (1250 ps)	5.4 ± 0.2	0.2 ± 0.1	4 %
IV (1500 ps)	5.8 ± 0.2	2.1 ± 0.2	36 %

Table 7.3: Average SAS differences between the simulations in regular urea and in water ($\Delta_{U_{100}W}$), as well as in hyperpolar urea and in water ($\Delta_{U_{150}W}$), and the ratio between the differences.

of the protein increases in regular urea, but decreases in hyperpolar urea and in water. However, the decrease of the SAS in hyperpolar urea is rather small, and the average SAS in hyperpolar urea is 2.1 nm² larger than it is in water (36 % of $\Delta_{U_{100}W}$, see Tab. 7.3). These results clearly show that hydrophobic collapses *do* occur in hyperpolar urea. For starting structure III, the collapse is even as complete and as fast as it is in water. For starting structure IV, the collapse in hyperpolar is less complete than in water, but the difference to regular urea is nevertheless significant.

Markov model for the effects of regular and hyperpolar urea. To quantify the differential effect of the three solvents on the hydrophobic collapse, a Markov model was derived with the SAS as reaction coordinate. Fig. 7.11 shows the estimated free energy profile, as calculated from the transition rates between the (pseudo) Markov states. The free energy of the protein in water (blue line) has a minimum at ≈ 23 nm², and states with large SAS are energetically very unfavorable. In comparison, the free energy “well” of the protein in urea (green line) is shifted significantly towards more open conformations (larger SAS), and it is almost flat. The free energy of the protein in hyperpolar urea (red line) resembles that of the protein in water. In particular, non-collapsed states with a large SAS have a very high free energy in both solvents, and are significantly less favorable than in regular urea. Compact states with a small SAS, in contrast, have a low free energy in water and in hyperpolar urea, and are significantly more favorable than in regular urea. In summary, a large agreement of the free energy profile for hyperpolar urea is seen with that for water, but not with that for regular urea. However, differences are also seen between the two free energy profiles. Mainly, the energy barrier between compact states (small SAS) and more exposed states (medium SAS) is much larger in hyperpolar urea than in water.

Taken together, these findings illustrate that the driving force for hydrophobic collapses, which was found to be significantly reduced in urea as compared to water, is found to be strong again in hyperpolar urea. In other words, hyperpolar urea does not, or only insignificantly, reduce the hydrophobic effect for apolar residues.

Native structure. While the main focus of this chapter is the hydrophobic collapse in different solvents, the influence of hyperpolar urea on the re-formation of native structure has also been investigated. Indeed, the fraction of native contacts (Fig. 7.10b) increases in one simulation of structure III simultaneously while the protein collapses (first 40 ns). In

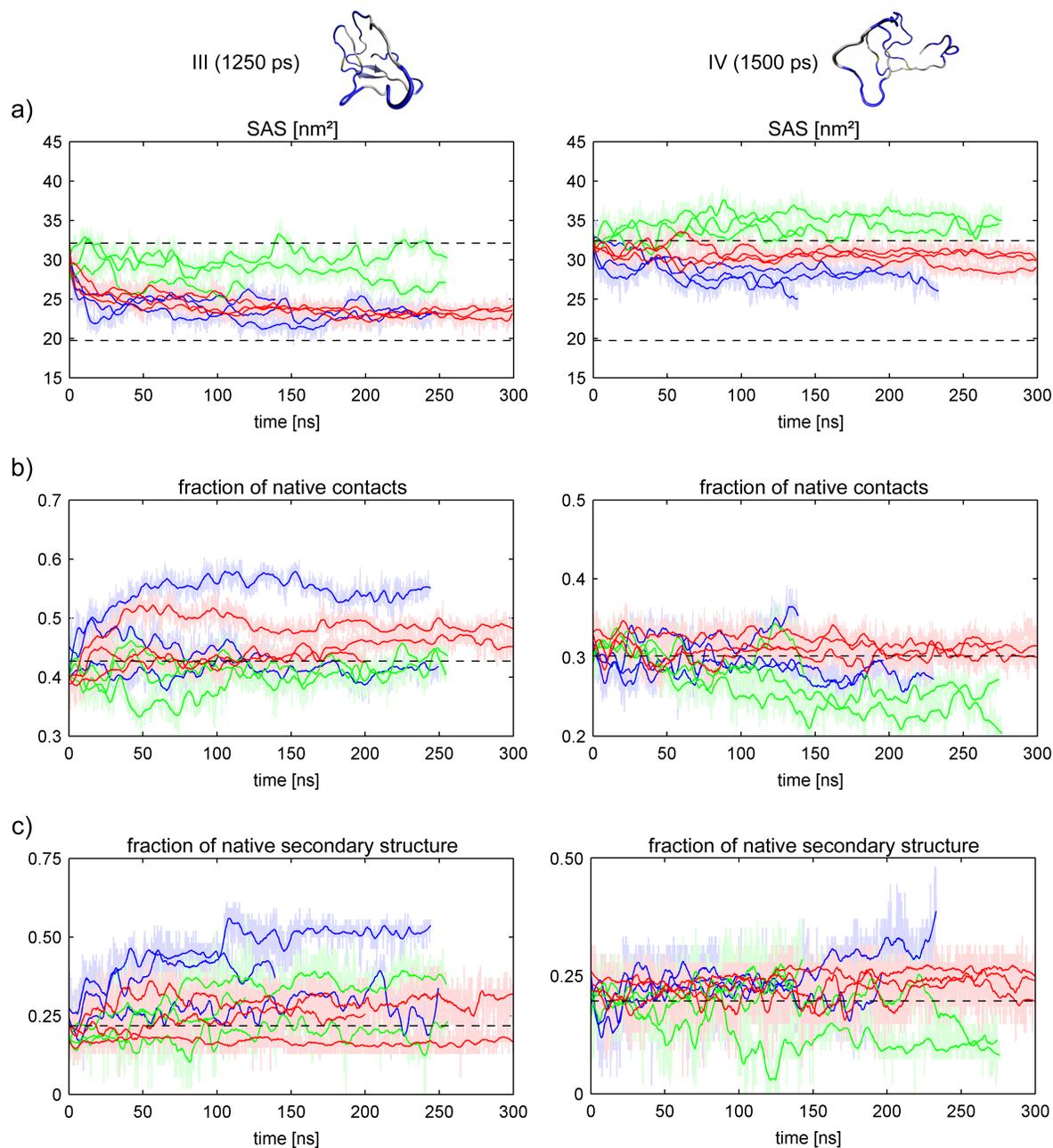


Figure 7.10: Solvent accessible surface area (SAS) and fraction of native structure of the Cold Shock protein in water (blue), regular urea (green), and hyperpolar urea (red) for starting structures III (left) and IV (right). (a) SAS, (b) fraction of native contacts, (c) fraction of native secondary structure. The bold solid lines show traces smoothed with a running average over 500 ps; the dim lines show raw data. In panel a, the upper dashed lines show the initial SAS, and the lower dashed lines show the SAS of the crystal structure, for comparison. The dashed lines in panels b and c show the initial values of the respective starting structures.

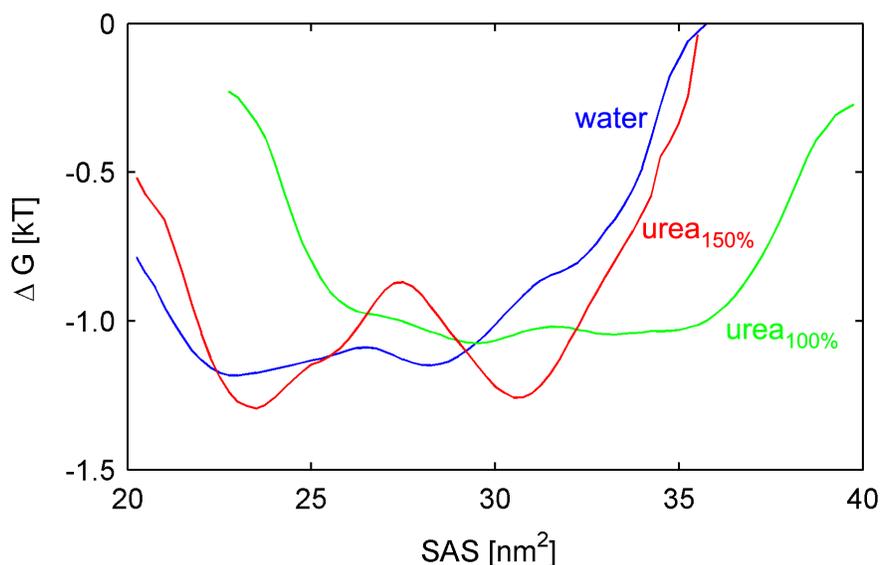


Figure 7.11: The estimated free energy ΔG of the partially unfolded Cold Shock protein, as calculated from a Markov model with the SAS as reaction coordinate. Blue: in water, green: in regular urea, red: in hyperpolar urea.

another simulation of structure III, few native contacts are re-formed at a later stage after the hydrophobic collapse (≈ 170 ns). However, the most significant observation regarding the re-formation of native contacts in hyperpolar is that for both starting structures, the fraction of native contacts is larger in hyperpolar urea than in regular urea. The fraction of native secondary structure (Fig. 7.10c) is, on average, smaller for structure III in hyperpolar urea than in water, although the SAS of structure III is similar for both solvents. Hence, the collapsed structure in hyperpolar urea is equally compact to that in water, but structurally less native. For structure IV, the average fraction of native secondary structure in hyperpolar urea is larger than in regular urea and smaller than in water, but the differences are rather small. In summary, the effect of hyperpolar on the re-formation of native structure is not significant.

Residue interactions with hyperpolar urea. In Chap. 7.1, it was suggested that favorable interactions between urea molecules and apolar residues inhibit hydrophobic collapses of the Cold Shock protein in *regular* urea. In contrast, here, it was found that *hyperpolar* urea does not impede the hydrophobic collapse of the protein. It remains to be checked now if hyperpolar urea indeed has reduced contact preferences to apolar residues, as was found for the CI2 protein (Chap. 6). To this end, the contact coefficient of hyperpolar urea was calculated for the partially unfolded Cold Shock protein.

Figure 7.12 shows the average contact coefficient C_{UW} for each amino acid in the Cold Shock protein, averaged for all six simulations of structures III and IV. The contact coefficients confirms that increasing urea polarity has indeed increased the contact preference

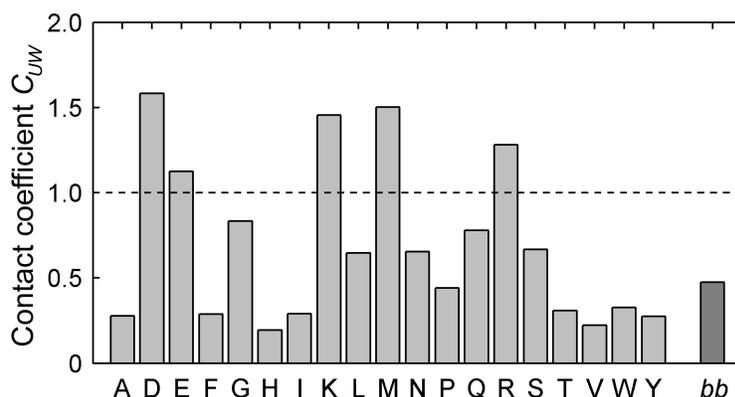


Figure 7.12: Contact coefficients C_{UW} of hyperpolar urea, averaged for each amino acid type in the Cold Shock protein, as well as the backbone (“*bb*”).

for charged and polar amino acids, and decreased the contact preference for apolar amino acids, as was expected.

Further, the average C_{UW} (0.69) is significantly lower than for regular urea in the folded (1.51) or the unfolded state (1.76). Comparison with the C_{UW} values of urea_{150%} for the CI2 protein (Chap. 6) shows large agreement between both data sets, with a regression coefficient of $r^2 = 0.68$. This correlation is significantly larger than that between C_{UW} for the CI2 protein and the Cold Shock protein in regular urea ($r^2 = 0.33$). Hence, the contact preferences of urea_{150%} seem to be less sensitive to structure and sequence than those of regular urea. Nevertheless, effects from sequence are also seen for hyperpolar urea. In particular, the average contact coefficient of MET is surprisingly high here. The reason for this high contact coefficient is that the only MET in the Cold Shock protein (MET1) is adjacent to the charged ARG3, LYS5, and LYS7, which all have high contact coefficients and thus are expected to effectively increase the contact coefficients of adjacent residues.

7.2.4 Conclusion

The main aim of this chapter was to test the hypothesis that increasing the interaction preference of urea for polar residues, and decreasing it for apolar residues, reduces the denaturation power of urea. The partially unfolded structures of the Cold Shock protein were an ideal test system, as urea was found to hinder its refolding by impeding hydrophobic collapses due to favorable interactions with less polar residues (Chap. 7.1). Indeed, these hydrophobic collapses were found to re-occur in hyperpolar urea. Hence, hyperpolar urea actually is a weaker denaturant than urea.

The occurrence of hydrophobic collapses in hyperpolar urea also allows an important conclusion for the simulations of the CI2 protein (Chap. 6). One might argue that the apparent stabilization of the CI2 was just kinetically due to decreased self-diffusion coefficients of

urea_{150%}. However, the hydrophobic collapses observed here for urea_{150%} show that conformational dynamics of the protein in urea_{150%} can occur on timescales an order of magnitude below the simulation times, such that this artifact can be ruled out.

Analysis of contact coefficients confirmed that increasing urea polarity increased the contact preference for polar residues and decreased the contact preference for apolar residues, as expected. Thus, the favorable contacts between urea and apolar residues, which were found to impede hydrophobic collapses and stabilize the denatured state, are eliminated here. Hence, the denatured state is stabilized to significantly lesser extent. Taken together, these results provide independent confirmation that not the polar, but rather the hydrophobic contacts, are the key determinant of the protein denaturation power of urea.

8

Urea-induced denaturation of the Trp-Cage mini-protein on realistic timescales

8.1 Introduction

The previous chapters have suggested a mechanism for urea-induced protein denaturation. While the suggested view has consistently emerged from several different approaches, it is based on indirect approaches rather than direct observation of urea-induced protein denaturation under experimental conditions. In fact, simulation of chemical denaturation has never been achieved without additional destabilization (e.g., by elevated temperature or bias-potentials) on realistic time scales. Therefore, it is not known if current force fields, and also those employed in this thesis, are even able to describe protein denaturation. However, this ability is crucial for the validity of all results and conclusions in this work.

Instead of introducing further aspects to the suggested view, therefore, this final chapter about urea-induced denaturation serves to critically test the ability of the force fields to describe protein denaturation. To this end, microsecond simulations of the Trp-Cage protein¹⁰⁰ in water and in aqueous urea solution were performed.

The Trp-Cage is 20-residue mini protein, which has recently been designed in the search for ultrafast folding proteins.¹⁰⁰ With its small size and a folding time of only $\approx 4 \mu\text{s}$ ²²⁴ under physiological conditions, it is an excellent system to gap the bridge between experiment and simulation.^{70,225} Despite its simple structure, the Trp-Cage possesses a fully built hydrophobic core, consisting of TYR3, TRP6, LEU7, PRO12, PRO17, PRO18, and PRO19. Although no data about the unfolding kinetics of the Trp-Cage in high molar urea concentration have been published yet, its microsecond folding time suggests an unfolding time roughly in the same order of magnitude. Hence, denaturation in high urea concentrations might occur in some of the simulations.

8.2 Simulation setup

The structure of the Trp-Cage protein (Fig. 8.1) was obtained from the PDB databank,⁹⁷ PDB-code 1L2Y.¹⁰⁰ For all simulations, the first of the 38 structures of the NMR ensemble was used. Five simulations were performed in water with a physiological ion concentration, and five simulations were performed in 8M urea solution, all at 300 K. Each simulation had a length of 1 μs .

8.3 Results and Discussion

8.3.1 Denaturation of the Trp-Cage

Figure 8.2 shows the SAS of the protein in the five simulations in water (blue lines) as well as in urea (green lines). For four simulations in urea, a significant increase of the SAS

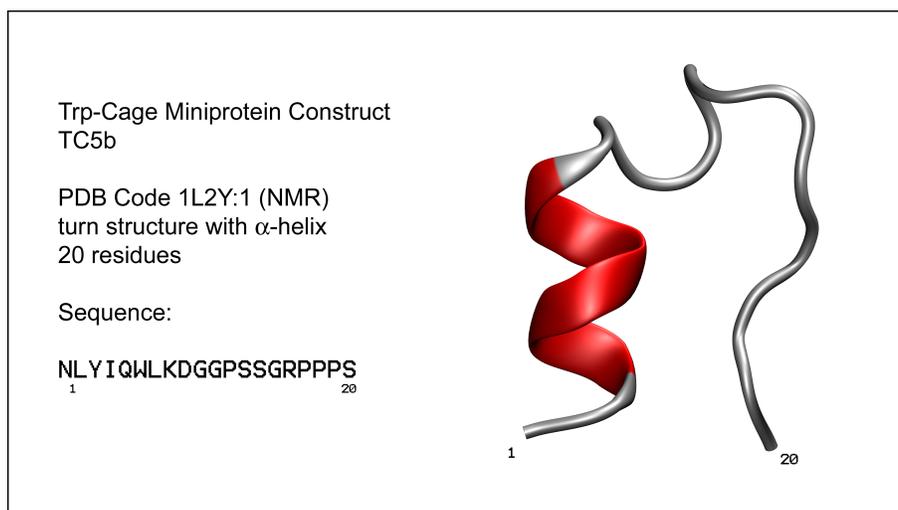


Figure 8.1: The Trp-Cage protein.

is seen from $\approx 10 \text{ nm}^2$ to final values of $13\text{--}15 \text{ nm}^2$. Hence, the Trp-Cage protein indeed unfolds in these simulations. In the fifth simulation in urea, the SAS remains at the native level of $10\text{--}11 \text{ nm}^2$. The reference simulations in water show no unfolding of the Trp-Cage protein, as indicated by the SAS which remains between $10\text{--}11 \text{ nm}^2$. In one simulation, however, the SAS increases temporarily to 12 nm^2 . This observation will be addressed further below.

8.3.2 Characterization of unfolding processes.

Due to the simple structure and small size of the Trp-Cage, only limited insight into unfolding pathways can be obtained here, and the aim of this chapter is to test whether denaturation is seen in the simulations. Therefore, only one of the unfolding trajectories will be discussed briefly. Figure 8.3a shows the SAS (red line) and the $C\alpha$ -rmsd (black line) of the protein for this simulation. The first step of denaturation (at about 300 ns) can be seen in an increase of rmsd rather than SAS (Fig. 8.3b). This rmsd increase results from conformational rearrangement of the turn-region, as displayed in Fig. 8.3c. Next, the α -helical structure is lost and the SAS increases significantly. Then, a transient β -hairpin is formed for $\approx 150 \text{ ns}$ before, finally, all structure is lost and denaturation complete at about 900 ns.

Structurally, the pathways of the other three unfolding events were different. These varieties are also reflected in the estimated free energy profiles. Figure 8.4 shows these profiles, as calculated from a Markov model with the SAS as reaction coordinate, for the unfolding simulation which was described in detail (solid green line), and the other three unfolding simulations in urea (dashed green lines). For comparison, the estimated free energy profile for one simulation in water is also shown (blue line). As has been noted, these profiles are

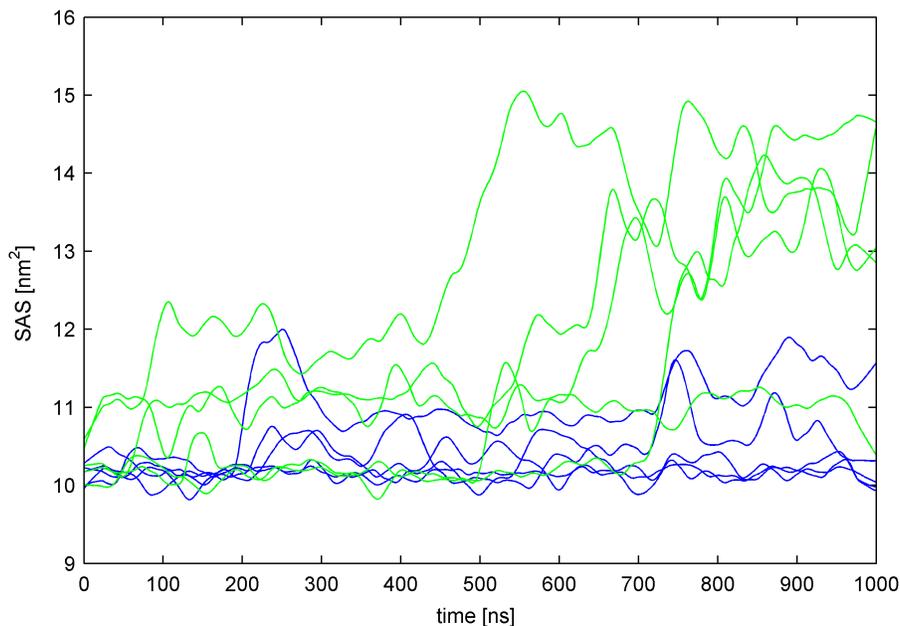


Figure 8.2: SAS of the Trp-Cage protein for simulations in water (blue lines, five simulations) and in 8M urea (green lines, five simulations). The lines show traces smoothed by a running average over 50 ns.

only rough estimates for the free energy landscape, and in particular the free energy of the denatured state is probably inaccurate due to insufficient sampling. Indeed, the free energy of the denatured state in urea is higher than that of the native state, different from what would be expected. Nevertheless, some interesting features can be discerned from this estimate. In particular, a free energy barrier between folded and unfolded state is seen for all unfolding simulations. The heights, as well as the positions of the individual barriers are different from each other, which results from different unfolding pathways in the four simulations. The maximum of each barrier corresponds to the transition state of the respective unfolding reaction. Hence, its location can be used to separate the folded and the unfolded state, and to define unfolding times. The average of the four unfolding times, defined as the time to reach the respective transition state, is $\tau_{unf} = 602 \pm 109$ ns. This average, however, does not take into account that one of the simulations in urea has an unfolding time of $> 1\mu s$, and thus presents a lower limit for the unfolding time.

8.3.3 Interactions of urea with the Trp-Cage.

With this discrimination between folded and unfolded structures, urea contact coefficients were calculated separately for the folded and the unfolded state (Fig. 8.5). Two main features are observed for the contact coefficients of the Trp-Cage. First, the C_{UW} values of the folded and the unfolded state show only small differences to each other. This large similarity indicates that structural effects on the C_{UW} are only minor for the Trp-Cage;

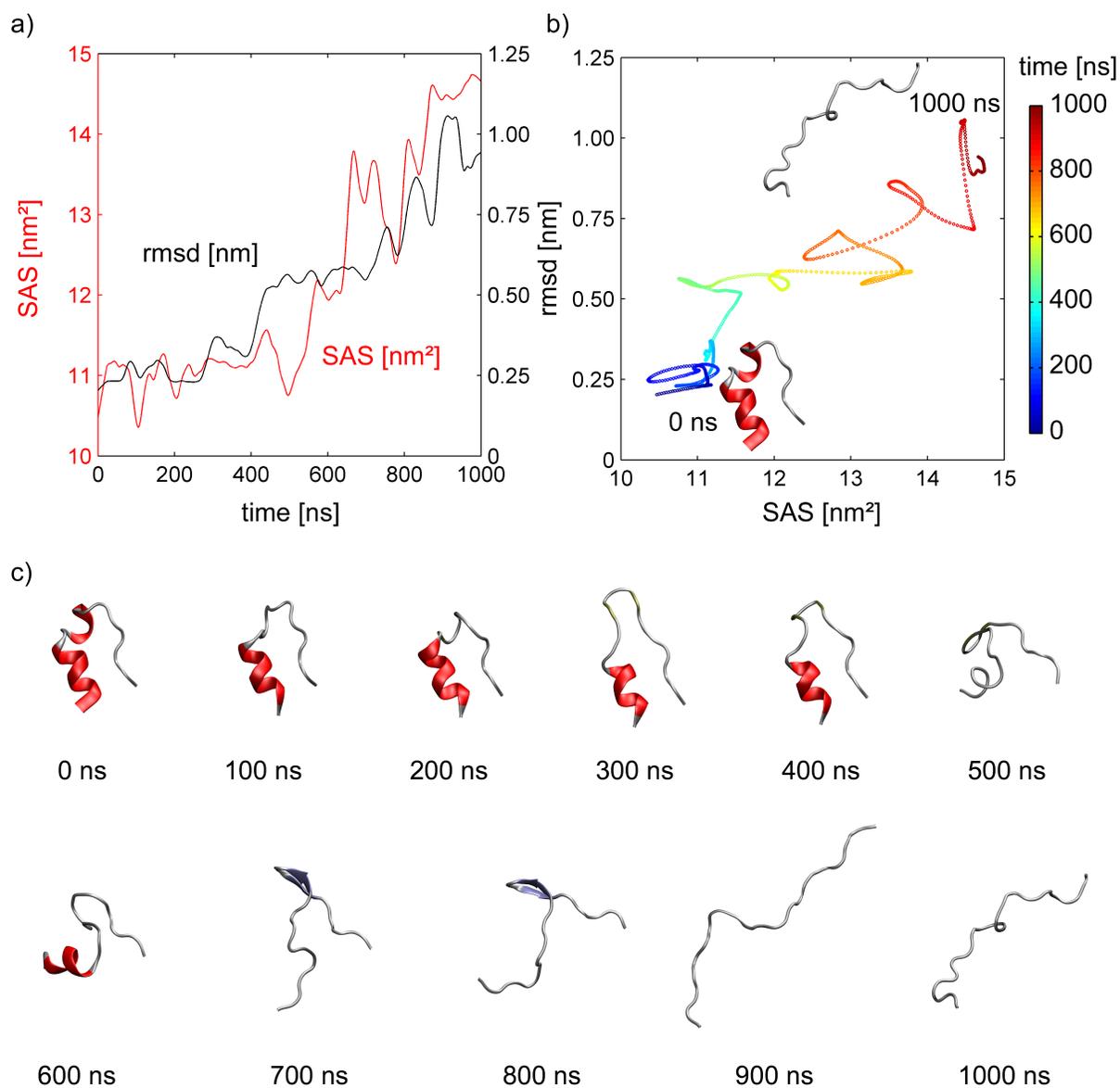


Figure 8.3: The Trp-Cage is denatured by urea in four out of five simulations. (a) SAS (red line) and $\text{C}\alpha$ -rmsd (black line) of the protein for one of these simulations. (b) SAS versus $\text{C}\alpha$ -rmsd, color encodes time. Also shown are the initial folded structure (0 ns) and the final denatured structure (1000 ns). (c) Snapshots of the Trp-Cage during unfolding in this simulation.

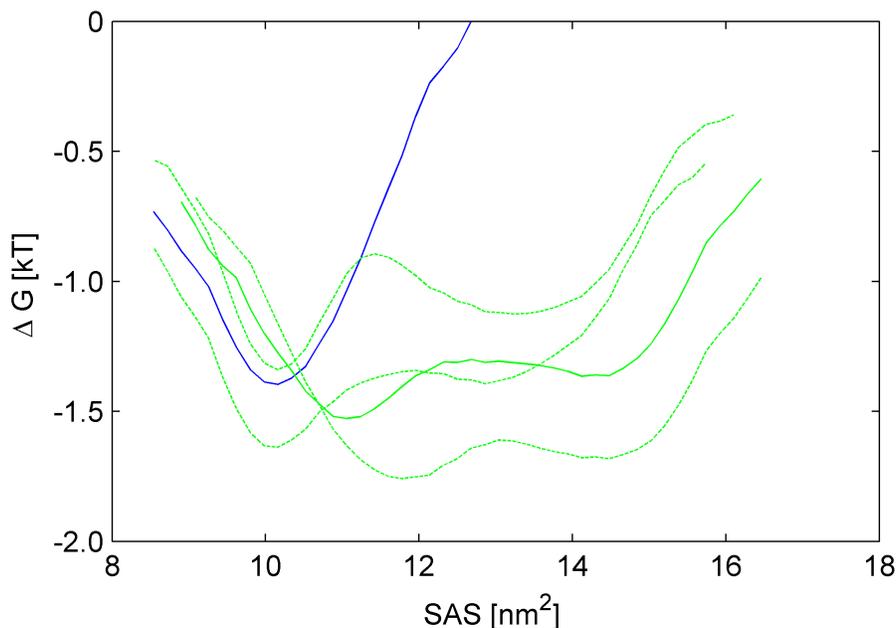


Figure 8.4: Free energy profiles were estimated with the SAS as reaction coordinate. The solid green line shows the free energy profile for the unfolding simulation in urea which has been described, dotted green lines those for the other unfolding simulations in urea. For comparison, the free energy profile of a simulation in water is also shown (blue line).

due to the small size of the protein, most residues are accessible to urea already in the folded state. Second, the C_{UW} values are very high. In particular, the C_{UW} values of the unfolded state are on average 20% higher than the reference C_{UW} values calculated for GXG tripeptides (Chap. 5). This high preference of most residues for interaction with urea probably results from cooperative sequence-effects, since the Trp-Cage contains a relatively large fraction of apolar and aromatic residues. It is expected that both, the accessibility for urea in the folded state, as well as the cooperative sequence-effects, contribute to the sensitivity of the Trp-Cage protein to urea-induced denaturation.

8.3.4 Observation of a folding intermediate.

In one of the simulations in water, an unexpected increase of the SAS was observed. As shown in Fig. 8.6a, the SAS (red line) increases after 200 ns to $\approx 12 \text{ nm}^2$, decreases 50 ns later to $\approx 11 \text{ nm}^2$, and increases again at 750 ns. The rmsd (black line in Fig. 8.6a) provides further insight into these structural changes. It increases simultaneously with the SAS, but instead of decreasing at 250 ns, it remains at a constant level of $\approx 0.7 \text{ nm}$. The rmsd plateau indicates that the protein has adopted a non-native conformation, in which it remains for about 500 ns. This conformation is a metastable folding intermediate which is stabilized by a salt-bridge between ASP9 and ARG16, as has previously been reported²²⁶ from replica-exchange MD simulations.²²⁷ During the conformational transition from the

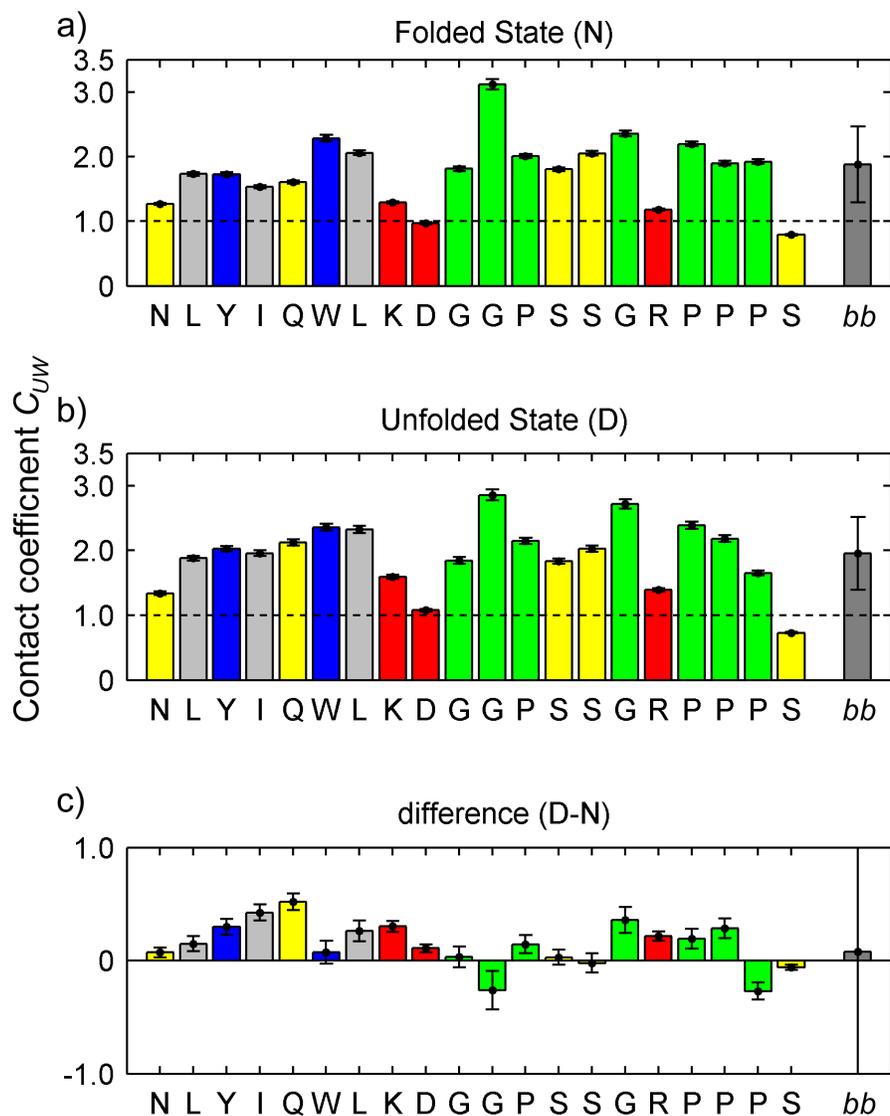


Figure 8.5: Contact coefficients for all residues of the Trp-Cage protein, averaged over all five simulations in urea. (a) for the folded state, (b) for the unfolded state, (c) difference between both. The dashed line at $C_{UW} = 1$ in panels (a) and (b) indicates equal interaction preference for either urea or water. The bars have been color coded according to the classification of the respective amino acid (red: charged, yellow: polar, gray: aliphatic, blue: aromatic, green: apolar). Error bars for the individual amino acids show the statistical error estimate of the average contact coefficient the respective amino acid. The error bars for the backbone show the standard deviation of the average backbone C_{UW} over all residues.

native state to this metastable intermediate, the Trp-Cage transiently increases its SAS before it adopts the compact intermediate state with β -sheet structure (Fig. 8.6d). After 500 ns, refolding towards the native state begins (see Fig. 8.6b). However, the native state is not reached again within the simulation time, and the salt-bridge is still present in the native-like structure, as can be seen from the center-of-mass distance between ASP9 and ARG16 (Fig. 8.6c). This intermediate structure was not populated in the other simulations with water or the simulations with urea.

8.4 Conclusions

The main aim of this chapter was to test if the methods employed in this thesis are able to describe protein denaturation by urea on realistic time scales. For this purpose, microsecond simulations of the ultrafast folding Trp-Cage mini protein were performed in urea, as well as in water as reference. Denaturation was observed in four out of five simulations in urea, whereas the protein did not unfold in water. The estimated average unfolding time of $\approx 0.6 \mu\text{s}$ for the four simulations probably presents a lower limit for the unfolding time, particularly because the unfolding time in the fifth simulation with urea is longer than $1 \mu\text{s}$.

These results are reassuring and provide a validation for the employed methods. In particular, it would have raised serious doubts about the suitability of the methods if the folded state had been found to be instable in water, or if unfolding in urea had been observed several orders of magnitude faster than in experiments — as had previously been reported for simulations of the CI2 protein in Ref. 15.

For the mechanism of urea-induced denaturation, not much further insight has been gained over that of the previous chapters. Rather, various aspects of the suggested mechanism are confirmed from the simulations of the Trp-Cage. In particular, almost all residues were found to have strong preference for contact with urea rather than with water, and especially those residues forming the hydrophobic core were solvated preferentially by urea. This finding accords with the suggested view that favorable direct contacts of urea with less polar residues weaken the hydrophobic effect for natively buried residues, and thus drive unfolding.

Further, the similarity of the contact coefficients for the folded and the unfolded state suggests that structure effects play only a minor role for the contact coefficients of the Trp-Cage, and that most residues are well accessible to urea already in the native state. This large accessibility for urea might contribute to the sensitivity of the Trp-Cage to urea-induced denaturation.

Finally, the unfolding simulations presented here represent the first observation of urea-induced protein denaturation in molecular dynamics simulations on realistic time scales. They demonstrate not only the capability of MD simulations to describe denaturation

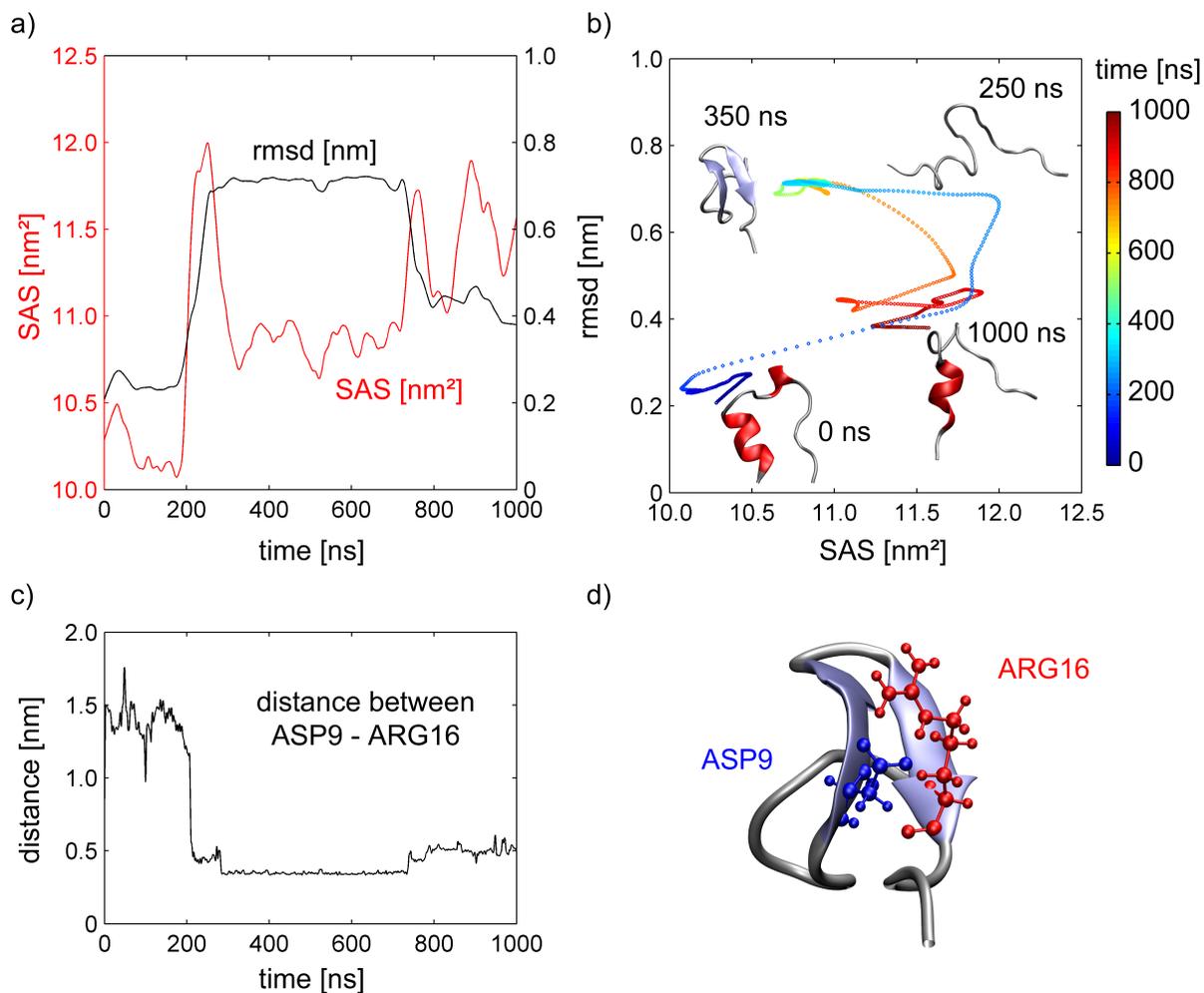


Figure 8.6: A metastable folding intermediate was populated in one simulation with water; this conformation is stabilized by a salt-bridge between ASP9 and ARG16. (a) SAS (red lines) and $C\alpha$ -rmsd of the protein. (b) SAS versus $C\alpha$ -rmsd, color encodes time. Also shown are the native state (0 ns), the transition state to the intermediate state (250 ns), the folding intermediate (350 ns), and a native-like state (900 ns) during refolding to the native state. (c) the center-of-mass distance between the ASP9 and ARG16, which form the salt-bridge of the folding intermediate. (d) representation of the folding intermediate, ASP9 and ARG16 are shown in blue and red, respectively.

processes, but also the general progress with has been achieved in the last few years to successfully bridge the gap between experimental time scales and simulation time scales. For combined studies, ultrafast folding proteins like the Trp-Cage are an ideal test system. Further investigation of the processes described here is planned. In particular, comparison with results from recent NMR experiments²²⁸ on residual structure of the denatured state will give instructive insight on the structural and dynamic properties of the denatured state.

9

Beyond Urea — A comparison of amino acid interaction preferences for Urea, DMSO, Guanidinium, and TMAO

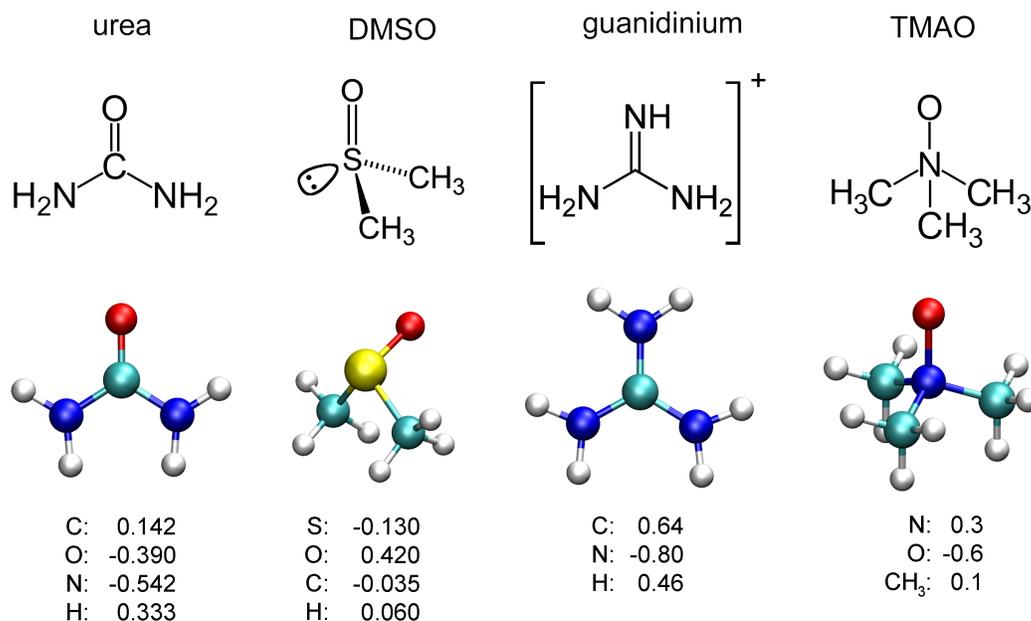


Figure 9.1: Four osmolytes were compared here. The three denaturants urea, DMSO, and the guanidinium ion, as well as the stabilizing osmolyte TMAO. Shown are the chemical structure formulas and the three-dimensional structures of the molecules. The numbers at the bottom denote the assigned partial charges of the respective atoms in the force field.

9.1 Introduction

The denaturation mechanism of urea has been the focus of all previous chapters. While the extensive effort on this subject in the past decades and the ongoing controversial discussions justify the exclusive role urea has been given in this thesis, still, the question arises how other denaturants or even counter-denaturants work. Is the mechanism found for urea transferable to other denaturants? Can common traits of osmolytes with stabilizing or destabilizing effect on proteins be found? To address these questions, and to put the results obtained for urea into a more general frame, this chapter investigates the interactions of the individual amino acids with two other denaturants, DMSO and guanidinium, as well as one counter-denaturant, TMAO.

The chemical structures and three-dimensional representations of the four osmolytes are shown in Fig. 9.1, where also the assigned force field partial charges of the respective osmolyte are listed.

DMSO (dimethylsulfoxide) is commonly used as organic solvent,²²⁹ rather than as protein denaturant. Therefore, the mechanism of DMSO-induced protein denaturation has not received much attention and only little is known. Nevertheless, DMSO is known to denature proteins at high concentrations²³⁰ and it is sometimes employed for this use.^{231,232,233} Here, comparison of DMSO with urea provides insight into common aspects of denaturing

osmolytes.

Guanidinium (Gdm) is, along with urea, the most-frequently used denaturant.^{234,194} Its chemical structure is similar to that of urea, but a third amino group instead of an oxygen atom is bonded to the carbon atom. With the resulting positive charge, guanidinium is a monovalent cation. Hence, it is always accompanied by an anion in an electrically neutral salt. Because ions themselves are known to have a significant effect on protein stability,^{80,235} the ionic character of guanidinium itself as well as the presence of additional anions render the total effect of guanidinium salts on proteins very complex. For instance, at high concentrations GdmCl is an ever stronger denaturant than urea,²³⁶ whereas it has a stabilizing effect on proteins at low concentrations.²³⁷ Further, the denaturation strength of different guanidinium salts differ from each other; for example, GdmSCN is a stronger denaturant than GdmCl, and $(\text{Gdm})_2\text{SO}_4$ has no denaturing effect at all.^{80,81} Due to this complex mixture of single effects, not much is known about the origin of the denaturing effect of guanidinium or even if it is similar to that of urea. While some works suggest that the mechanism of these two denaturants is similar,^{234,238} other findings support the view that the mechanisms differ from each other.^{239,240,241} A promising approach which is supported by several results^{240,242} is to separate the effect of the guanidinium ion into contributions of denaturant-effects and salt-effects. For a detailed understanding of guanidinium-induced denaturation, therefore, a prior understanding of denaturant-effects as well as salt-effects alone is imperative. Nevertheless, the comparison of guanidinium with urea in this work might contribute to the understanding of both denaturants and set the direction for future studies.

TMAO (trimethylamine N-oxide) is, in contrast to the other osmolytes investigated here, not a denaturant but a counter-denaturant. It stabilizes the folded state of proteins against unfolding by denaturants or temperature²⁴³ and even induces folding of some natively unfolded proteins.²⁴⁴ In cells with high stress, TMAO serves as protective osmolyte for proteins.²⁴⁵ The molecular basis for the stabilization of proteins by TMAO is still subject of ongoing research. It has been suggested that the interaction of TMAO with the protein surface, and in particular with the peptide backbone, is unfavorable, and that thus the presence of TMAO forces the protein to reduce its solvent accessible surface by folding. In the past years, much evidence in support of this suggestion has been accumulated.^{246,247,179,248} But also different suggestions exist, for instance, that “an aquated form of this molecule interacts rather strongly” with the protein surface.²⁴⁹ Here, interactions of TMAO with amino acids are investigated at the atomic level. First, this approach addresses the question whether TMAO indeed shows unfavorable interactions with the protein which could explain stabilization of the folded state. Second, the comparison of a counter-denaturant with denaturants will give instructive insight into essential differences between stabilizing and destabilizing osmolytes in general.

9.2 Methods and simulation setup

Since the analysis of contact frequencies between urea and the amino acids turned out to be very instructive, it is employed here as well for the other osmolytes. This analysis has also further advantages: first, the contact coefficients C_{XW} do not only characterize denaturing interactions with the protein, but all kinds of (direct) interactions. Therefore, this analysis is applicable to denaturants as well as counter-denaturants and the results from different osmolytes are directly comparable to each other. Second, the results are independent of sequence and structure, and thus not limited to certain proteins only. Third, high statistical significance is attainable within feasible computation time. However, there are also limitations to this analysis. Mainly, only direct interactions are captured in the analysis. Indirect effects of osmolytes on the protein via alteration of the water structure are outside its scope. For urea, however, indirect effects were found to be of only minor importance (Chap. 4), and the relevance of these indirect effects for osmolyte-protein interactions in general is found to be insignificant by many recent studies.^{24,178,150,148,17,18,156,29} An interesting summary of this emerging view has been made by Chitra and Smith: “While effects on water association are obviously important for determining the properties of cosolvent solutions, using water structure making or breaking arguments to explain the effects of cosolvents on protein stabilization and solubility appears to be too simple and should be avoided in all but a few cases.”¹⁷ Therefore, despite its inability to capture indirect interactions, the contact analysis seems appropriate to investigate osmolyte-protein interactions here.

To allow direct comparison with the results presented in Chap. 5 for urea, the setup for the simulations of guanidinium, DMSO, and TMAO was similar. For DMSO and TMAO, each amino acid was simulated for 100 ns in an aqueous solution of 250 of the respective osmolyte molecules and 1250 water molecules, corresponding to a mole fraction of 0.2 and a concentration of roughly 8 M. For the guanidinium ion, simulations had to be performed at a lower mole fraction, since the high ion concentration at a mole fraction of 0.2 caused an artificial over-stabilization of the systems which led to a glass transition of the solvent. This effect has previously been observed for urea with 200 % partial charge scaling in Chap. 6 and is a known inability of common force fields to correctly describe high ion concentrations.²¹¹ Therefore, the simulations with guanidinium were performed with 125 guanidinium and 125 Cl^- ions, corresponding to a mole fraction of 0.1.

It is not clear that C_{XW} values calculated from simulations at different concentrations are comparable to each other. By definition, the contact coefficient C_{XW} should be independent of the concentration of the osmolyte X. However, the intrinsic contact preferences of the amino acids with either osmolyte X or water might depend on the concentration of X, for instance, because of multi-particle or saturation effects. To check whether C_{XW} really is concentration-independent, an additional simulation of one amino acid (TRP) was performed in an even lower concentration of guanidinium chloride (mole fraction 0.05). The C_{GW} calculated from this simulation was within 5 % of the C_{GW} from the simulation at a mole fraction of 0.1. This agreement suggests that C_{XW} is indeed independent of

concentration and legitimates comparison of C_{XW} obtained at different concentrations.

The four sets of contact coefficients C_{XW} for urea, DMSO, guanidinium, and TMAO are compared to each other, and also to selected common hydrophobicity scales with linear regression fits. In addition to the experimental hydrophobicity scales from the literature, the correlation with the standard deviation of all atomic partial charges (“std(pc)”) is investigated, which can be regarded as an approximation for the hydrophobicity of each amino acid in the force field.

Force field parameters for DMSO and guanidinium were taken from the OPLS-AA force field.¹²⁴ Since TMAO is not parameterized in the OPLS-AA force field, the parameters were taken from Ref. 150, where a united-atom TMAO model, with each of the CH_3 groups represented by one united atom, has been developed for use in conjunction with TIP4P water and the OPLS force field. The assigned partial charges in the force field of the respective osmolyte are listed in Fig. 9.1. The force field parameters of these molecules are not as extensively verified as those used for urea, and hence the results presented here might be less accurate than those presented for urea.

9.3 Results and discussion

Figure 9.2 shows the contact coefficients C_{XW} of the four osmolytes investigated. The interaction profile of urea has already been described in Chap. 5 and will serve as reference for the comparison with DMSO, guanidinium, and TMAO. To facilitate the detection of systematic trends in the C_{XW} profiles, the bars in Fig. 9.2 have been color coded according to the classification of the respective amino acid (red: charged, yellow: polar, gray: aliphatic, blue: aromatic, green: apolar). Error bars for the individual amino acids show the statistical error estimate of the average contact coefficient of the respective amino acid. In contrast, the error bars for the backbone show the standard deviation of the average backbone C_{XW} over all amino acids. Atomic C_{XW} values are shown in Fig. 9.4 (urea), Fig. 9.5 (DMSO), Fig. 9.6 (Gdm), and Fig. 9.7 (TMAO) to illustrate contact preferences at the atomic level. To improve comparability, the same color scale has been used for all of these four Figures.

9.3.1 DMSO

The contact coefficients C_{DW} of all amino acids with DMSO show a very wide spectrum, ranging from ASP and GLU with strong interaction preference for water ($C_{DW} = 0.26$, and $C_{DW} = 0.37$, respectively), to TRP and PHE with strong interaction preference for DMSO ($C_{DW} = 1.89$, and $C_{DW} = 2.10$, respectively). The general trend of the contact preferences is obvious for DMSO; charged and polar residues have strong preferences for solvation by water, whereas aromatic and apolar residues have strong preference for solvation by DMSO.

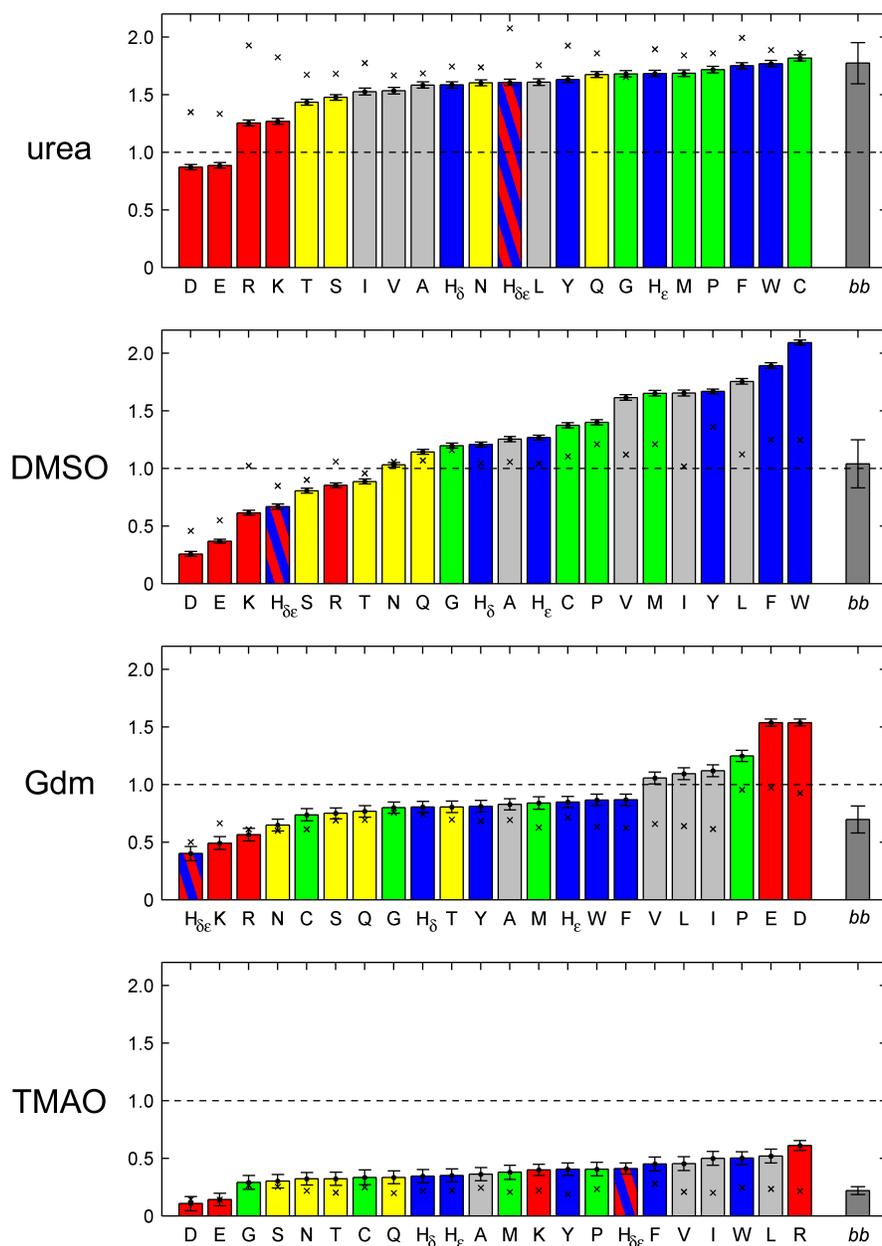


Figure 9.2: Contact coefficients C_{XW} for each amino acid and the backbone average (“ bb ”). High values above one indicate preferential interactions with the respective osmolyte; a value of one corresponds to equal probabilities to interact with osmolyte X or water molecules. The color characterizes the amino acids; red: charged, yellow: polar, gray: aliphatic, blue: aromatic, green: apolar. Crosses denote the C_{XW} of the backbone alone. Error bars for the individual amino acids show the statistical error estimate of the average contact coefficient the respective amino acid. The error bars of the backbone (“ bb ”) shows the standard deviation of the average backbone C_{XW} over all amino acids.

In this respect, the interaction profile of DMSO is similar to that of urea. The correlation coefficient of between C_{UW} and C_{DW} is $r^2 = 0.60$ (see Tab. 9.1). However, two main differences are observable:

First, DMSO has a much more heterogeneous spectrum of contact preferences than urea. In particular, charged and polar amino acids show significant preferences for contact with water rather than DMSO. Accordingly, the average C_{DW} over all amino acids is slightly lower than for urea, and the standard deviation is twice as large (see Tab. 9.2). This clear dependence of C_{DW} on amino acid polarity is also reflected in the correlation with hydrophobicity scales (Tab. 9.1), which is very large (max. $r^2 = 0.66$) for DMSO.

Second, the average C_{DW} of the backbone is significantly lower for DMSO than it is for urea. In fact, the backbone shows no contact preference for either water or DMSO ($C_{DW} = 1.04 \pm 0.21$), whereas in urea a strong preference of the backbone for urea contact over water contact was observed ($C_{UW} = 1.77 \pm 0.18$).

These observations are also evident in the atomic C_{DW} (Fig. 9.5); the charged groups of ARG, ASP, GLU, HIS $_{\delta\epsilon}$, and LYS have contacts almost exclusively with water molecules, whereas the aromatic rings of PHE, TRP, and TYR show very strong preference for DMSO contact. Interestingly, in some amino acids (ARG, LEU, TRP, and TYR, e.g.) large differences in the atomic C_{DW} are seen between neighboring atoms. These local differences suggest that specific interaction sites for DMSO exist in some amino acids.

The main reason for the differences in the interaction preferences of urea and DMSO is, probably, that DMSO has two apolar methyl groups (CH_3) where urea has two amino groups (NH_2) which are able to form hydrogen bonds. In particular, the reduced number of hydrogen sites of DMSO might explain the lower contact preference for polar residues or the peptide backbone, as compared to urea.

9.3.2 Guanidinium

For guanidinium, the spectrum of contact coefficients C_{GW} (Fig. 9.2) shows some qualitative differences to those of urea or DMSO. As expected due to the positive charge of the guanidinium cation, the acidic amino acids ASP and GLU have strong interaction preferences with guanidinium (high C_{GW}), whereas the basic amino acids HIS $_{\delta\epsilon}$, LYS, and ARG have strong interaction preferences for water (low C_{GW}). Apart from these particularly high or low C_{GW} values, only small trends are seen. Most amino acids, and in particular the polar ones, have a small interaction preference for water, whereas a small interaction preference for guanidinium is seen for PRO, ILE, LEU, and VAL. In general, however, the C_{GW} spectrum is rather homogeneous aside from the acidic and basic amino acids. The average C_{GW} over all amino acids is 0.88 ± 0.29 , and the average C_{GW} of the backbone alone is 0.70 ± 0.12 .

At the atomic level, the carboxyl-groups of ASP and GLU can be identified as preferential interaction sites for guanidinium (Fig. 9.6), and an effect on the C_{GW} of neighboring atoms can be seen for the backbones of these residues. While the charged groups of ARG, LYS, and $\text{HIS}_{\delta\epsilon}$ show pronounced contact preference for water, the central carbon-atom (CZ) in the guanidine group of ARG has a significantly larger C_{GW} than its adjacent atoms. This interaction accords with π -stacking interactions which have been discussed for the guanidinium ion and ARG.²⁵⁰ Similar local differences are also found in ILE, LEU, and PRO, which, together with the aromatic rings of TRP and TYR, have atomic sites with contact preference for guanidinium.

Correlation of C_{GW} with C_{UW} , C_{DW} , or some common hydrophobicity scales yields linear regression coefficients of $r^2 \approx 0$. Thus, seemingly no systematic trend in the interaction preferences of amino acids with guanidinium can be found. However, the low correlations are mainly due to the systematic outliers ASP and GLU. As discussed above, the strong preference of these amino acids for interaction with guanidinium is due to the cationic nature of the guanidinium ion. While this salt-effect is significant, it is certainly not the only driving force for interactions between guanidinium and amino acids, and it might obscure other systematics in the correlations. It has been suggested previously to separate the effect of guanidinium into denaturant- and salt-effects.²⁴² Along this line, ASP and GLU have been excluded from the correlation analysis. Indeed, this exclusion reveals systematic trends in the C_{GW} , and all correlation coefficients are increased drastically. The correlation of C_{GW} with C_{DW} , for instance, is increased from $r^2 \approx 0$ to $r^2 = 0.47$, as illustrated in Fig. 9.3. Certainly, the high C_{GW} of ASP and GLU is not the only result of salt-effects of guanidinium, but they obviously contribute dominantly to the correlation deviations. The average correlation of this reduced C_{GW} set with some common hydrophobicity scales is $r^2 = 0.32 \pm 0.14$, and the maximum correlation with one scale is $r^2 = 0.49$ (see Tab. 9.1). While these correlation coefficients are similar to those found for urea, the correlation with the contact coefficients C_{UW} for urea is small ($r^2 = 0.14$) even for the reduced set.

9.3.3 TMAO

The contact coefficients C_{TW} of TMAO, which is a counter-denaturant, are shown in the bottom panel of Fig. 9.2. All amino acids show extraordinarily low contact coefficients with TMAO. The average C_{TW} is 0.37 ± 0.11 , and the backbone average is even lower (0.22 ± 0.03) and outside the standard deviation of the average C_{TW} . Hence, direct interaction of TMAO with all amino acids, and in particular with the peptide backbone, is obviously highly unfavorable.

Deviations in the C_{TW} spectrum are small, but some trends are discernible. The acidic ASP and GLU show the lowest C_{TW} values (0.11 and 0.14, respectively), corresponding to very strong preference for water contacts. As for the other osmolytes, polar amino acids

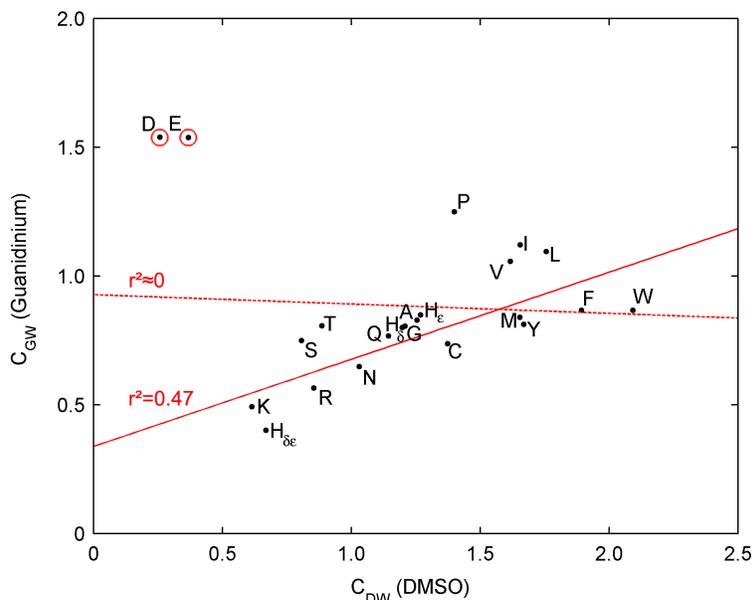


Figure 9.3: Correlation of contact coefficients for guanidinium (C_{GW}) and DMSO (C_{DW}). The dashed line shows a linear fit ($r^2 \approx 0$) of the complete set, the solid line shows the linear fit ($r^2 = 0.47$) with exclusion of ASP and GLU (red circles).

	C_{UW}	C_{DW}	C_{GW}	C_{TW}
C_{UW}		0.60	0.14	0.23
C_{DW}	0.60		0.47	0.39
C_{GW}	0.14	0.47		0.03
C_{TW}	0.23	0.39	0.03	
std(p.c.)	0.35	0.60	0.39	0.39
hydrophobicity (Ref. 186)	0.34	0.39	0.48	0.04
hydrophobicity (Ref. 187)	0.37	0.34	0.16	0.07
hydrophobicity (Ref. 188)	0.33	0.66	0.22	0.19
hydrophobicity (Ref. 189)	0.36	0.66	0.25	0.18
hydrophobicity (Ref. 190)	0.31	0.55	0.32	0.16
hydrophobicity (Ref. 191)	0.53	0.64	0.49	0.14
average	0.36 ± 0.09	0.54 ± 0.14	0.32 ± 0.14	0.13 ± 0.06

Table 9.1: Linear regression coefficients r^2 of correlations of the C_{XW} sets with each other and with selected common hydrophobicity scales. std(pc) denotes the correlation with the standard deviation of the atomic partial charges for each amino acid, which can be regarded as an approximation for the hydrophobicity of the amino acids implemented in the force field. For guanidinium (C_{GW}), ASP and GLU were excluded to separate salt and denaturant effects (see text).

	total average	backbone average
C_{UW}	1.53 ± 0.25	1.77 ± 0.18
C_{DW}	1.21 ± 0.49	1.04 ± 0.21
C_{GW}	0.88 ± 0.29	0.70 ± 0.12
C_{TW}	0.37 ± 0.11	0.22 ± 0.03

Table 9.2: Contact coefficients averaged over all amino acids or over the backbone only.

also show lower contact coefficients than aliphatic, aromatic, or apolar ones. Interestingly, ARG shows the highest C_{TW} of all amino acids. As can be seen from the atomic C_{TW} values (Fig. 9.7), the CZ carbon of the guanidine group in ARG, which also showed high contact coefficients for guanidinium and DMSO, contributes significantly to the high average C_{TW} of ARG.

The correlation of C_{TW} with hydrophobicity scales (average $r^2 = 0.13 \pm 0.06$) is very low, which might be interpreted to suggest that hydrophobicity is not a determinant for the contact coefficients. However, since all amino acids have very low C_{TW} values and show no qualitative differences, trends in the C_{TW} are certainly less relevant than the absolute values. Therefore, the correlation analysis is probably not very instructive for C_{TW} . Notably, for C_{TW} , the correlation with $\text{std}(\text{pc})$ is significantly higher than the correlation with the hydrophobicity scales. This could suggest that the parametrization of the TMAO molecule is inaccurate, but it could also be irrelevant due to the homogeneity of the C_{TW} values.

9.4 Conclusions

Aiming to shed light on the different effects of DMSO, guanidinium, and TMAO on protein stability, contact coefficients were calculated for these osmolytes. The contact coefficients of urea (Chap. 5) were used as a reference to deduce similarities or differences to the denaturation mechanism of urea, which was found to be dominated by favorable direct interactions of urea with the peptide backbone as well as with less polar residues.

DMSO shows contact coefficients relatively similar to those of urea, but contact preferences are more pronounced for DMSO than for urea. In particular, DMSO interacts favorably with hydrophobic amino acids, but charged amino acids prefer solvation by water rather than DMSO. Hence, hydrophobicity seems to be the determinant of the interactions of DMSO with amino acids. Further, the solvation preference of the peptide backbone is the second significant difference between the interactions of the two osmolytes; while urea was found to be a particularly good solvent for the peptide backbone, the backbone shows no

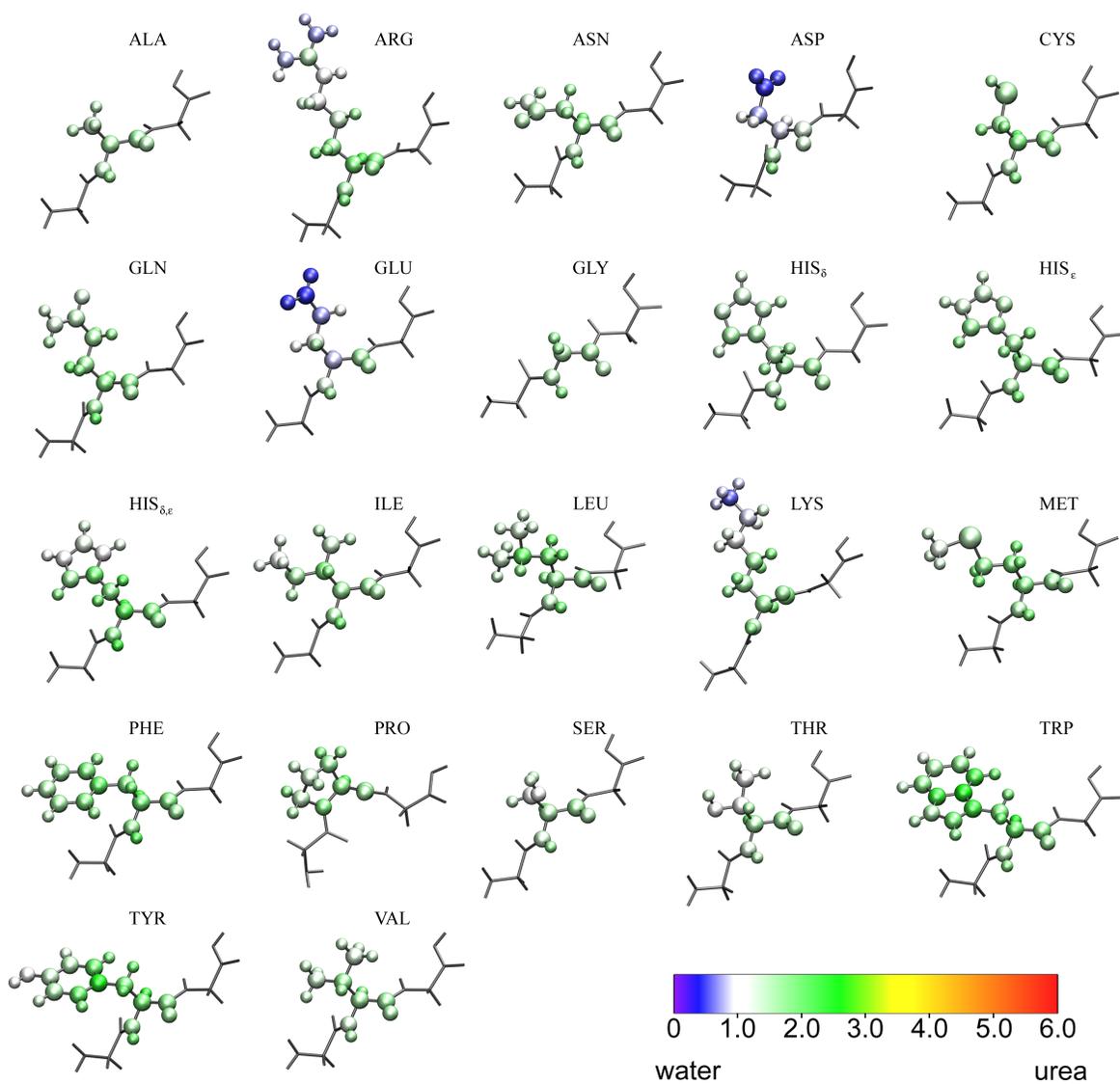


Figure 9.4: Contact coefficients C_{UW} for all amino acids with urea at the atomic level. The color of each atom denotes its C_{UW} value, as shown in the color bar. $C_{UW} = 1$ (white) means no interaction preference for urea or water, $C_{UW} < 1$ and $C_{UW} > 1$ correspond to interaction preferences for water or urea, respectively.

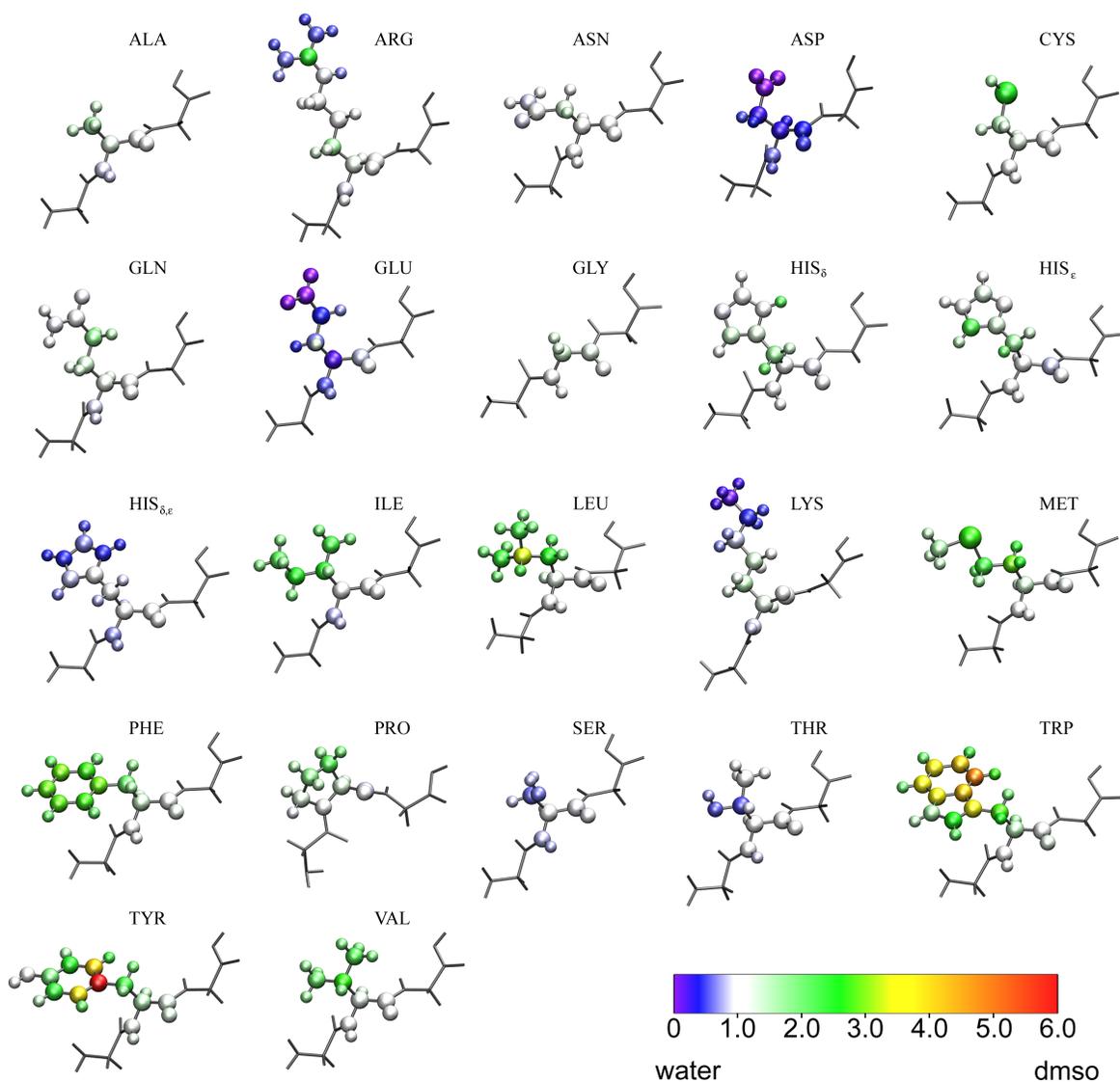


Figure 9.5: Contact coefficients C_{DW} for all amino acids with DMSO at the atomic level. The color of each atom denotes its C_{DW} value, as shown in the color bar. $C_{DW} = 1$ (white) means no interaction preference for DMSO or water, $C_{DW} < 1$ and $C_{DW} > 1$ correspond to interaction preferences for water or DMSO, respectively.

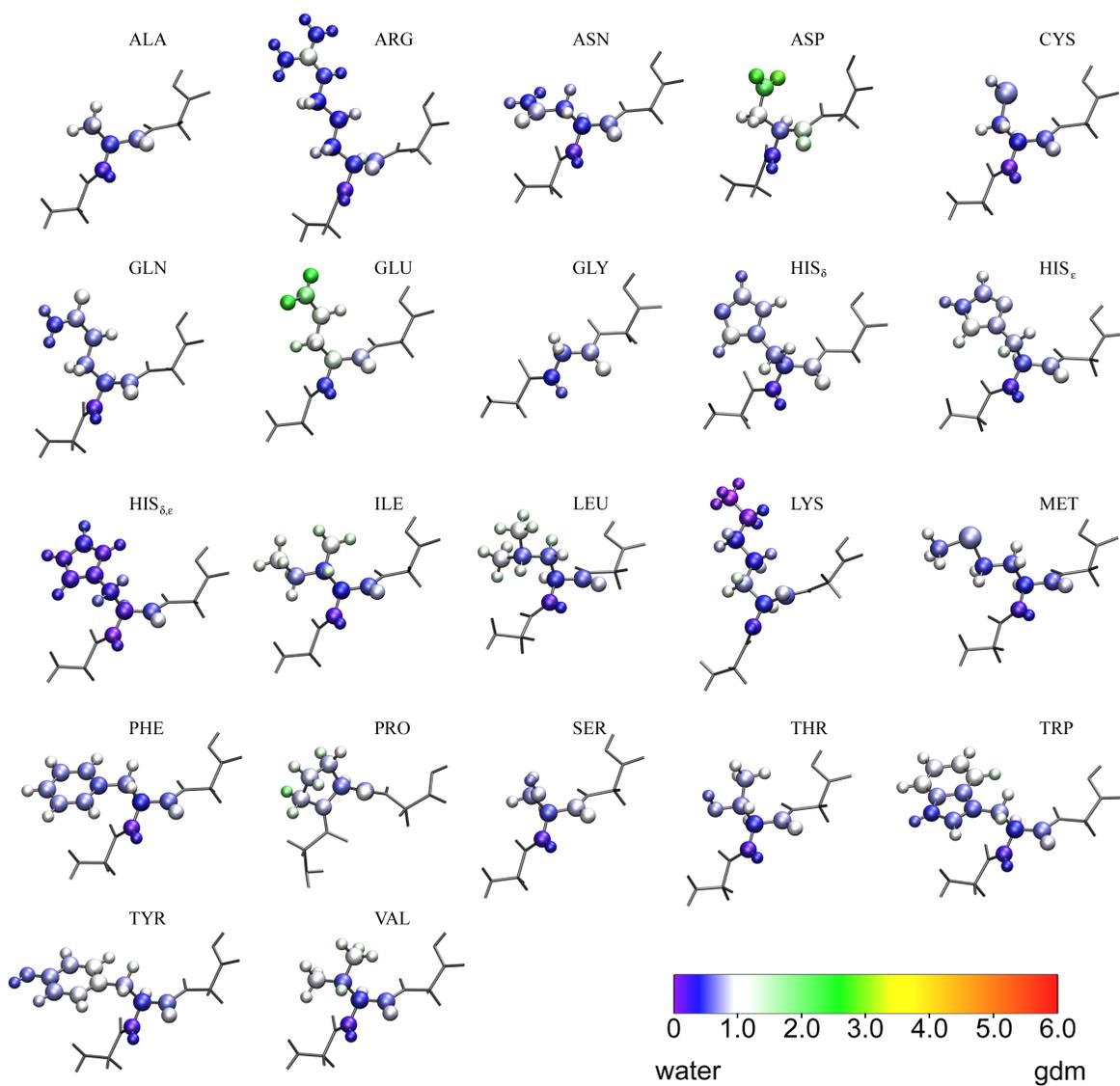


Figure 9.6: Contact coefficients C_{GW} for all amino acids with guanidinium at the atomic level. The color of each atom denotes its C_{GW} value, as shown in the color bar. $C_{GW} = 1$ (white) means no interaction preference for guanidinium or water, $C_{GW} < 1$ and $C_{GW} > 1$ correspond to interaction preferences for water or guanidinium, respectively.

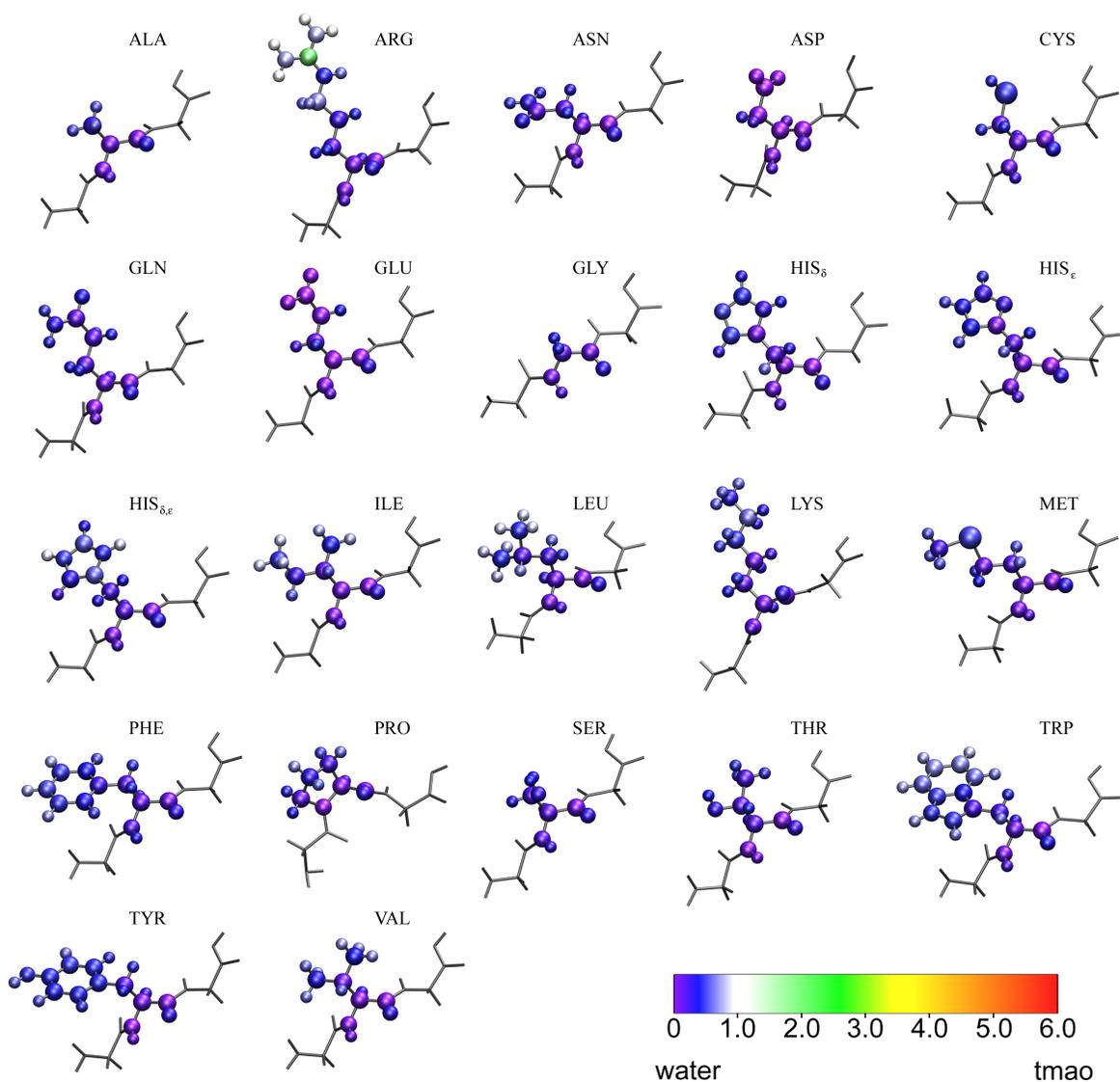


Figure 9.7: Contact coefficients C_{TW} for all amino acids with TMAO at the atomic level. The color of each atom denotes its C_{TW} value, as shown in the color bar. $C_{TW} = 1$ (white) means no interaction preference for TMAO or water, $C_{TW} < 1$ or $C_{TW} > 1$ correspond to interaction preferences for water or TMAO, respectively.

preference for solvation by either water or DMSO. These results suggest that the denaturing effect of DMSO on proteins is based on favorable hydrophobic contacts with less polar residues, rather than preferential solvation of the peptide backbone. In particular, the ability of DMSO to solvate less polar residues reduces the hydrophobic effect and leads to the exposure of residues natively buried in the hydrophobic core. This mechanism would be generally similar to that found for urea, but without favorable backbone interactions. An interesting implication of this difference is that more residual secondary structure elements, which shield the backbone from unfavorable solvent interactions, might exist in DMSO-denatured proteins than in those denatured by urea. Future work could address the verification of this prediction.

Guanidinium, due to its positive charge, shows contact preferences for acidic amino acids, whereas electrostatic repulsion leads to only few contacts with basic amino acids. Aside from these salt-effects, a similar trend in the profile of interaction preferences is found as for DMSO, but it is significantly less pronounced. Further, the backbone shows contact preference for water rather than for guanidinium. It seems that the interaction of guanidinium with amino acids can roughly be decomposed into a contribution from direct electrostatic interaction and a contribution from hydrophobicity. The latter, however, is less pronounced than for urea and DMSO and does not seem sufficient to explain denaturation. Further, the relevance of the strong charge interactions for denaturation is unclear. For proteins which are stabilized by favorable electrostatic interactions between charged or polar residues (salt bridges), guanidinium might interrupt these salt bridges by substituting for the positively charged residue of the salt bridge, and hence destabilize the protein. However, this effect is unlikely to be the basis for guanidinium-induced denaturation, since this mechanism would limit the effectiveness of guanidinium to only few proteins, and other salts (e.g., NaCl) with similar interaction preferences for charged residues do not cause denaturation. Also, indirect effects via alteration of the water structure, which for urea have been found to be not relevant, might play a role for guanidinium due to the polarizing effect of the charge on the water shell. Indirect effects, however, are not captured in the analysis of contact coefficients performed here. Taken together, the results are not conclusive enough to make solid suggestions about the molecular basis for guanidinium-induced denaturation. It seems, however, unlikely that the mechanism is the same as for urea-induced unfolding, as had been suggested.^{234,238} Rather, the mechanism could be a complex combination of different contributions. In the future, more insight will probably be gained when the individual effects of denaturants and salts on proteins are better understood, and when improved force fields are able to accurately describe systems with high ion concentrations.

TMAO, as a counter-denaturant, exhibits an interaction profile which is very conclusive not only for the mechanism of protein-stabilization by TMAO, but also protein-stabilization and -destabilization by osmolytes in general. All amino acids show significant preference for contact with water rather than TMAO. Hence, contact with TMAO is obviously energetically unfavorable for all amino acids, and in particular for the peptide backbone. These results strongly support the view that protein stabilization by TMAO is due to unfavorable

direct interactions of TMAO with the protein, which force it to reduce its solvent accessible surface and thus to fold.

Which common features of denaturants or counter-denaturants can be identified from the comparison of these results? For urea, favorable direct interactions with less polar residues as well as the peptide backbone have been suggested as the driving force for denaturation. Direct interaction of TMAO with amino acids, and in particular with the backbone, was found to be unfavorable. This extreme difference in the interactions of urea and TMAO with amino acids strongly suggests that the effect of an osmolyte on protein stability is dictated by its direct interactions — favorable for a denaturant, unfavorable for a protective osmolyte — with those parts of the protein which are buried in the native state. These parts include less polar residues, which are typically buried in the hydrophobic core of the protein, and the peptide backbone, which is typically protected from solvent contact by formation of secondary structure elements. The determining role of the peptide backbone for the stabilizing or destabilizing effect of an osmolyte is also suggested by other very recent studies.^{251,179,180,248} The results presented here for DMSO, however, suggest that a high preference for hydrophobic contacts without significant contact preference for the backbone can also suffice for a denaturing agent. For urea, the results of the previous chapters suggested that the favorable contacts with less polar residues and the backbone lead to a weakening of the hydrophobic effect for these parts, and it seems plausible that DMSO might have the same effect for hydrophobic residues.

In summary, these results suggest that the ability of cosolvents to enhance solubility for natively buried parts determines their exposure or burial and thus unfolding or folding of the protein. Further, this simple determinant would explain why denaturants and counter-denaturants exert their destabilizing or stabilizing effect on almost all proteins, independent of structure and sequence.

10

Summary, Conclusion, and Outlook

The primary aim of this thesis was to elucidate the molecular basis for urea-induced protein denaturation. In particular, the central question was addressed whether indirect interactions, direct polar interactions, or direct apolar interactions between urea and the protein are the driving force for denaturation. Several different approaches were developed.

The first step aimed to examine the influence of urea on water structure, and thus to assess the relevance of indirect effects of urea on proteins. To this end, structure and energetics of aqueous urea solutions were investigated for a wide range of temperatures and urea concentrations. Urea was found to show a small tendency to self-aggregate, and three distinct urea pair conformations were identified. Analysis of hydrogen bond energies revealed that water-water hydrogen bonds are significantly stronger than those between urea and water or urea and urea. This result suggested the hydrophobic effect as possible cause for urea self-aggregation. The influence of urea on water structure, however, was found to be surprisingly small. In particular, with increasing urea concentration, the strength of water-water hydrogen bonds increased, and the tetrahedral arrangement of water molecules became more rigid — but both effects were only small. Taken together, these results suggest that the effects of urea on water structure are generally small, and thus should not contribute significantly to the denaturation power of urea on proteins.

The next step focussed on the role of direct interactions. To this end, a comprehensive characterization of the direct interactions between urea and all 20 amino acids was undertaken. To characterize the interactions between the solvent and the individual amino acids, and to separate those from effects due to sequence or structure, glycine-capped tripeptides were studied in aqueous urea solution. A contact coefficient was developed and calculated to quantify interaction preferences of the amino acids for either urea or water. All amino acids, except ASP and GLU, were found to interact preferentially with urea. A particularly pronounced preference for urea interaction was observed for aromatic and apolar amino acids as well as for the backbone, and not — as might be expected — for charged or polar amino acids. Further, whereas urea was found to form hydrogen bonds to the backbone, these hydrogen bonds were weaker than those between water and the backbone. This result indicates that the favorable interactions between urea and the backbone are mainly entropically driven. Taken together, these findings suggest that favorable direct contacts between urea and less polar residues as well as the backbone weaken the hydrophobic effect, that stabilizes the folded state, and thus promote protein unfolding.

This suggestion was tested by a *Gedankenexperiment*. If apolar contacts are indeed the driving force for denaturation, then urea with artificially decreased polarity should be an even stronger denaturant than “real” urea. Accordingly, simulations of the CI2 protein in urea with decreased and also with increased polarity were performed to test this hypothesis. Indeed, unfolding of the CI2 was observed for all simulations in hypopolar urea, but not in regular or hyperpolar urea. This finding confirmed that urea drives protein denaturation by favorable hydrophobic contacts. Further, these results motivated the prediction that

hyperpolar urea might actually be a weaker denaturant than regular urea.

In addition to this main issue, these simulations also allowed to address the question whether specific unfolding pathways exist. All nine unfolding events were found to be structurally different from each other and shared no common conformations during unfolding, suggesting a large amount of structural heterogeneity during unfolding. However, on a more general level, the different pathways shared similar features. In particular, unfolding was found to proceed often step-wise with alternating phases of loss of secondary, or, respectively, tertiary structure — as opposed to either a simultaneous loss of both structure levels, or, alternatively, complete loss of one structure level ahead of the other. These results suggest that denaturation of the CI2 protein does not proceed via unique pathways, but in a rather stochastic sequence of events. Further, in the simulations unfolding was observed to start at several different positions of the CI2 protein, with flexible parts being more frequently an unfolding nucleus. With regard to the question of residual structure in denatured proteins, polyproline II conformations were seen, which have recently been suggested as one of several possible backbone conformations of the denatured state.³⁶

Chapter 7 addressed the question of how urea affects proteins at different stages of the folding/unfolding process. To this end, partially unfolded structures of the Cold Shock protein Bc-Csp were simulated in water and in urea. While the effect of urea on less unfolded structures was only small, a significant effect of urea on more unfolded structures was observed. In particular, hydrophobic collapse events of the protein were found to occur in water, but not in urea. At the molecular level, this observation was explained by favorable interactions between urea molecules and less polar residues. Since urea is a better solvent for these residues than water, the hydrophobic effect, and hence the driving force for the hydrophobic collapse, is reduced in urea. These findings support that urea stabilizes denatured and open conformations of proteins, as had been suggested on the basis of calorimetric experiments.³⁰ Moreover, these results provide a molecular picture for the underlying mechanism.

In the second part of Chapter 7, the combination of this approach with that of Chapter 6 offered a further possibility to verify the determinant role of apolar contacts, as opposed to polar contacts — and to finally test the prediction that hyperpolar urea is a weaker denaturant than regular urea. To this end, simulations of partially unfolded structures of the Cold Shock protein were performed with hyperpolar urea. The hydrophobic collapse events which occurred in water, but were impeded in urea, were found to occur again in hyperpolar urea. Hence, the stabilizing effect of urea on denatured conformations is reduced when urea polarity is increased. This result provides further and independent evidence that apolar rather than polar interactions between urea and the protein constitute the key determinant for urea-induced denaturation.

Necessarily, all conclusions described so far rest on the assumption that the molecular dynamics simulations, and, in particular, the employed force fields, describe urea-induced

protein denaturation sufficiently accurately. Therefore, the main aim of Chapter 8 was to test this assumption. For this purpose, several microsecond simulations of the Trp-Cage miniprotein were performed in water as well as in urea. Indeed, urea-induced denaturation was observed in four out of five simulations in urea, whereas the protein remained folded in all simulations with water. Apart from validating the methods applied in this thesis, these results also represent the first simulations of denaturant-induced protein unfolding on experimental time scales.

Finally, Chapter 9 set the results obtained for urea into a wider context, and thus also contributed to an understanding of other denaturants or protective osmolytes. To this end, interactions of all 20 amino acids with DMSO, guanidinium, and TMAO were investigated and compared to those of urea as a reference. For DMSO, interaction preferences were found to be largely similar to those for urea. Interestingly, the differences between the amino acids were even more pronounced than for urea. In particular, apolar and aromatic amino acids showed interaction preference for DMSO, whereas charged and polar amino acids showed preference for interaction with water rather than DMSO. In contrast to urea, however, no preference of the backbone to interact with DMSO was seen.

While these results suggest that the denaturation mechanisms of urea and DMSO are largely similar, the results for guanidinium were less conclusive. Neither the backbone nor less polar amino acids were found to have particular interaction preference for guanidinium. It is speculated here that the main reason for these discrepancies is the fact that guanidinium is positively charged, whereas urea and DMSO are both neutral. Because of its ionic nature, the interactions between guanidinium and all 20 amino acids are likely to be a complex combination of denaturant-effects and salt-effects,²⁴² governed by entropic and electrostatic forces. Therefore, a deeper understanding of ion effects on protein stability, and in particular of the Hofmeister-effect,⁷⁹ has to be obtained before the guanidinium ion can successfully be addressed.

Analysis of the protective osmolyte TMAO, in contrast, has provided very conclusive results. All amino acids showed strong interaction preferences for water over TMAO. This result suggests that TMAO reduces the peptide solubility through unfavorable direct interactions with the amino acids, and in particular with the peptide backbone. Thus, unfolding becomes energetically less favorable and the folded state is stabilized. In general, direct interactions with natively buried parts of the protein seem to determine the positive or negative effect of an osmolyte on protein stability. This simple determinant would explain why denaturants and counter-denaturants exert their destabilizing or stabilizing effect on almost all proteins, independent of structure and sequence.

A comprehensive picture of urea-induced protein denaturation has emerged from the different approaches of this work. Apart from the more specific conclusions described so far, more general conclusions can be drawn. In particular, this work suggests a synthesis of

seemingly opposing viewpoints. Whereas urea-protein hydrogen bonds do not seem to drive the denaturation, they do contribute to the overall energetics. According to the mechanism proposed here, the denaturation power of urea rests on its trade-off between two essential but conflicting features. First, it is apolar enough to solvate apolar groups; second, it is polar enough to form weak hydrogen bonds to the backbone and to incorporate well into the water hydrogen bond network. With these ambivalent characteristics, urea can be regarded to denature proteins by interfacing between water and natively buried parts of the protein, which effectively reduces the hydrophobic effect and promotes unfolding.

With this primary question about the *driving force* for denaturation answered, the *mechanism* can be addressed. Does urea actively destabilize the folded state or rather stabilize the unfolded state? What “picture” should one have in mind of the dynamics of denaturation? Taken together, the results of this work offer a suggestion. Since apolar contacts are the determinant for denaturation, one would expect the effect of urea on the protein to be strongest when many apolar contacts are possible, i.e., when hydrophobic residues are exposed — that is to say, in the unfolded state. Indeed, this was found here for the Cold Shock protein. In particular, hydrophobic collapse events were seen to be impeded in urea, and the differential effects of water and urea on the protein dynamics were more pronounced for more unfolded structures. For the Cold Shock protein, this observation accords with the experimental finding that the unfolding rate is nearly independent of denaturant concentration, whereas the refolding rate decreases exponentially with denaturant concentration. But what about other proteins for which a strong dependence of the unfolding rate on denaturant concentration is observed? Here, the observed effect of urea on the folded state of the CI2 or the Cold Shock protein can contribute to offer a possible explanation. For both proteins, although no unfolding was seen, the SAS of the native state was systematically larger in urea than in water. In particular, it was observed that urea stabilizes a slightly more open conformation of the protein which was adopted due to thermal fluctuations. It is suggested here, that a consecutive stabilization of such subsequent slightly more open conformations might lead to denaturation. This suggestion is supported by the observation that unfolding of the CI2 in hyperpolar urea started preferably in regions which have high fluctuations already under physiological conditions. Taken together, the picture emerges that urea binds to apolar residues when they are transiently exposed due to thermal fluctuations. Because of the reduced hydrophobic effect under direct interaction with urea molecules, the rate of contact re-formation is decreased, and the protein unfolds step-wise.

While this suggested mechanism accords with various individual results of this work, significantly more statistics are required to assess the validity of this suggestion. But as the Irish author George Bernard Shaw said, “Science is always wrong. It never never solves a problem without creating ten more”. And indeed, this is only one of several interesting new questions that arise from the results of this work and pose attractive challenges for future projects.

Probably one of the most striking questions concerns the denaturation mechanism of guanidinium. Can the effect of guanidinium on protein stability be explained by a “simple” combination of denaturant- and salt-effects? And, if this is possible, is the underlying denaturant-effect similar to the one found here for urea? To successfully address the mechanism of guanidinium-induced denaturation, however, more comprehensive insight into the effects of ions on proteins, and in particular the Hofmeister-effect,⁷⁹ is required. This might be achieved with improved force fields with a better ability to correctly describe systems with high ion concentrations.

Further, the comparison of urea with other osmolytes performed here was only one step towards a general understanding of common traits of denaturants or protective osmolytes. In particular, future projects might investigate the reason for the different interaction preferences of DMSO and TMAO. To this end, similarly detailed analyses of the energetics as performed here for urea, might provide further insight.

The residual structure of denatured states is another exciting topic which is currently debated and closely-linked to this work. Here, a combination of simulation approaches and experiments will be most instructive. While molecular dynamics simulations offer the ability to investigate structure and dynamics of the denatured state on an atomistic level, comparison to experimental data from FRET experiments²¹⁷ or NMR-spectroscopy^{252,228} will provide a conclusive picture.

Finally, an extension and refinement of the contact coefficient analysis introduced here seems tempting. In this work, this method has been key to advance the understanding of the interactions between proteins and denaturants or counter-denaturants. However, the effects of sequence and structure on the interaction preferences of the individual amino acids are only poorly understood so far. In particular, could the effect of structure on the contact coefficients, and thus the “accessibility” of residues for urea in the folded state, be related to the sensitivity of the protein to urea-induced denaturation? Might this even allow to predict the sensitivity to urea on the basis of sequence and structure? To investigate these possibilities, far more statistics are required. In particular, the investigation of proteins which are resistant to urea or denaturants in general will give instructive insight into these open questions.

Appendix

Amino acid	One-letter-code	Three-letter-code
Alanine	A	ALA
Arginine	R	ARG
Asparagine	N	ASN
Aspartic acid	D	ASP
Cysteine	C	CYS
Glutamine	Q	GLN
Glutamic acid	E	GLU
Glycine	G	GLY
Histidine	H	HIS
Isoleucine	I	ILE
Leucine	L	LEU
Lysine	K	LYS
Methionine	M	MET
Phenylalanine	F	PHE
Proline	P	PRO
Serine	S	SER
Threonine	T	THR
Tryptophan	W	TRP
Tyrosine	Y	TYR
Valine	V	VAL

Table A.1: The 20 natural amino acids with their common one-letter-code and three-letter-code abbreviations.

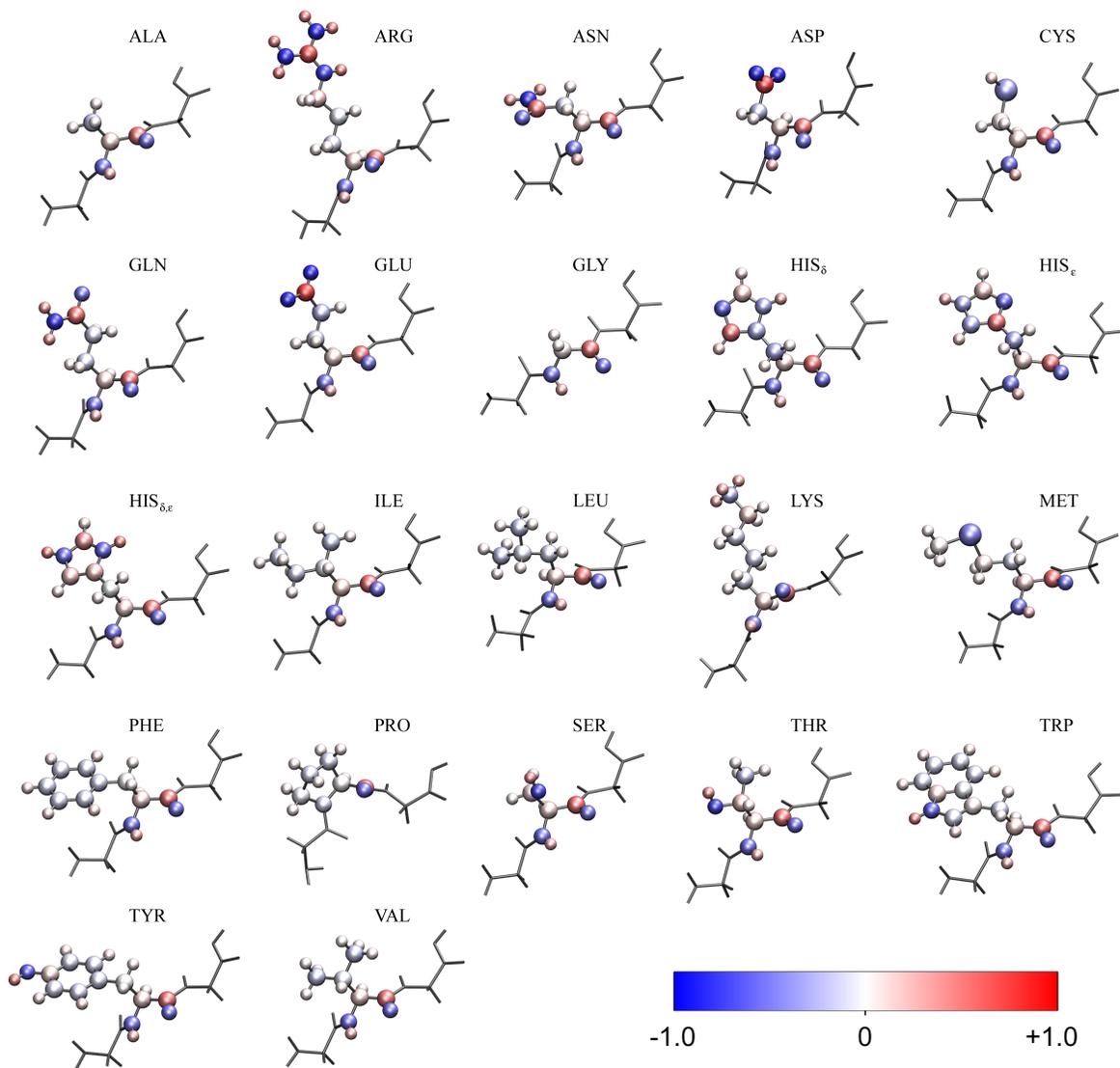


Figure A.1: Atomic partial charges assigned in the OPLS-AA force field to the 20 amino acids (including three protonation states for HIS).

Danksagung

Zu der erfolgreichen Durchführung meiner Doktorarbeit haben viele Menschen beigetragen, von denen ich einigen an dieser Stelle ausdrücklich danken möchte.

Mein besonderer Dank gebührt Prof. Dr. Helmut Grubmüller, unter dessen Leitung die vorliegende Arbeit entstand. Von der Aufnahme in seine Arbeitsgruppe bis hin zur Fertigstellung dieser Arbeit habe ich in ihm einen exzellenten Betreuer gefunden, der mit seinem präzisen Blick auf den Kern der Dinge mir stets Wege zu neuen Betrachtungen aufgewiesen hat aus denen oft neue Erkenntnisse folgten. Mit seinen konstruktiven und kritischen Anregungen sowie innovativen Ideen hat er maßgeblich zum Erfolg dieser Arbeit beigetragen, und gleichzeitig mit seiner lockeren und humorvollen Art für eine wirklich tolle Arbeitsatmosphäre gesorgt. Der große Freiraum meinen eigenen Projektideen nachgehen zu können, aber dabei gleichzeitig bei ihm immer auf ein offenes Ohr, reges Diskussionsinteresse, und hilfreiche Hinweise zu treffen, hat für mich eine ideale und sehr angenehme Arbeitsumgebung geschaffen.

Besonders möchte ich mich auch bei Herrn Prof. Dr. Tim Salditt dafür bedanken, dass er als Doktorvater die Arbeit betreut hat. Darüberhinaus hat er mit wertvollen Anregungen neue Aspekte eingebracht.

Den Mitgliedern der Abteilung für Theoretische und Computergestützte Biophysik am Max-Planck-Institut für Biophysikalische Chemie danke ich für die angenehme und sehr kollegiale Arbeitsatmosphäre und viele fruchtbare Diskussionen. Besonders danke ich an dieser Stelle Bert de Groot, Ulrich Zachariae, Lars Schäfer, und Frauke Gräeter. Für freundschaftlichen Kontakt auch außerhalb der Arbeit danke ich besonders Friedemann Reinhardt, Edgar Luttmann, Maik Götte und Frank Wiederschein. Besonders bedanke ich mich bei Gunnar Schröder sowohl für die kollegiale Freundschaft, als auch für viel Hilfestellung beim anfänglichen Einarbeiten in die Methode der MD Simulationen, und auch für das Lesen von Teilen dieser Arbeit. Ebenfalls bedanke ich mich bei Ira Tremmel, die durch viele Diskussionen und Manuskriptarbeit bei den Veröffentlichungen mir wesentlich geholfen hat. Weiterhin danke ich Eveline Heinemann, Ansgar Esztermann und Martin Fechner für Hilfe bei unzähligen Dingen. Für stimulierende Diskussionen und Tipps danke ich auch Michael Levitt, Vijay Pande, Jane Clark, Ben Schuler, Kim Sharp, Alan Soper, Malcolm Walkinshaw, Matthias Rief, und Hendrik Dietz.

Meinen Eltern danke ich besonders für jegliche Art von Unterstützung, und dass sie meine beruflichen wie auch privaten Interessen, letztere insbesondere in der Musik, stets gefördert haben. Bei Ihnen und auch meinen Geschwistern bedanke ich mich außerdem auch dafür, dass ich mich jederzeit auf sie verlassen kann. Claudia und Hans Seeberg möchte ich für ein wirklich hervorragendes und sehr herzliches Verhältnis danken. Tim Huege und Christoph Schulz danke ich für die langjährige und besondere Freundschaft, und Christoph besonders für die Protogenese.

Ganz besonders danke ich Ann-Kathrin Seeberg — schlichtweg für alles.

Bibliography

- [1] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.
- [2] Cyrus Levinthal. Are there pathways for protein folding? *J. Chim. Phys.*, 65(1):44–45, 1968.
- [3] Ken A. Dill, S. Banu Ozkan, Thomas R. Weikl, John D. Chodera, and Vincent A. Voelz. The protein folding problem: when will it be solved? *Current Opinion in Structural Biology*, 17:342–346, 2007.
- [4] Editorial. So much more to know. *Science*, 309:78–102, 2005.
- [5] J. A. Rupley. The effect of urea and amides upon water structure. *J. Phys. Chem.*, 68(7):2002–2003, 1964.
- [6] Henry S. Frank and Felix Franks. Structural approach to the solvent power of water for hydrocarbons; urea as a structure breaker. *J. Chem. Phys.*, 48(10):4746–4757, 1968.
- [7] Francesco Vanzi, Bhupinder Madan, and Kim Sharp. Effect of the protein denaturants urea and guanidinium on water structure: A structural and thermodynamic study. *J. Am. Chem. Soc.*, 120:10748–10753, 1998.
- [8] Yasuhiko Nozaki and Charles Tanford. Solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.*, 238(12):4074–4081, 1963.
- [9] Dwight R. Robinson and William P. Jencks. The effect of compounds of the urea-guanidinium class on the activity coefficient of acetyltetraglycine ethyl ester and related compound. *J. Am. Chem. Soc.*, 87(11):2462–2470, 1965.
- [10] Mark Roseman and William P. Jencks. Interactions of urea and other polar compounds in water. *J. Am. Chem. Soc.*, 97:631–640, 1975.
- [11] Xavier Hoccart and George Turrell. Raman spectroscopic investigation of the dynamics of ureawater complexes. *J. Chem. Phys.*, 99(11):8498–8503, 1993.

- [12] Abdenacer Idrissi and Franjo Sokolic. A molecular dynamics study of the urea/water mixture. *J. Chem. Phys.*, 112(21):9479–9488, 2000.
- [13] A. Idrissi. Molecular structure and dynamics of liquids: aqueous urea solutions. *Spectrochimica Acta, Part A*, 64(1-2):1–17, 2005.
- [14] Amedeo Caflisch and Martin Karplus. Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure*, 7(5):477–488, 1999.
- [15] Brian J. Bennion and Valerie Daggett. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. USA*, 100(9):5142–5147, 2003.
- [16] Ana Caballero-Herrera, Kerstin Nordstrand, Kurt D. Berndt, and Lennart Nilsson. Effect of urea on peptide conformation in water: Molecular dynamics and experimental characterization. *Biophys. J.*, 89(8):842–857, 2005.
- [17] Rajappa Chitra and Paul E. Smith. Molecular association in solution: A kirkwood-buff analysis of sodium chloride, ammonium sulfate, guanidinium chloride, urea, and 2,2,2-trifluoroethanol in water. *J. Phys. Chem. B*, 106:1491–1500, 2002.
- [18] Joseph D. Batchelor, Alina Olteanu, Ashutosh Tripathy, and Gary J. Pielak. Impact of protein denaturants and stabilizers on water structure. *J. Am. Chem. Soc.*, 106:1958–1961, 2004.
- [19] Norbert Muller. A model for the partial reversal of hydrophobic hydration by addition of a urea-like cosolvent. *J. Phys. Chem.*, 3856-3859(94):3856–3859, 1990.
- [20] Darwin O. V. Alonso and Ken A. Dill. Solvent denaturation and stabilization of globular proteins. *Biochemistry*, 30(24):5974–5985, 1991.
- [21] George I. Makhatadze and Peter L. Privalov. Protein interactions with urea and guanidinium chloride — a calorimetric study. *J. Mol. Biol.*, 226(2):491–505, 1992.
- [22] Erin M. Duffy, Paul J. Kowalczyk, and William L. Jorgensen. Do denaturants interact with aromatic hydrocarbons in water? *J. Am. Chem. Soc.*, 115(20):9271–9275, 1993.
- [23] Jerry Tsai, Mark Gerstein, and Michael Levitt. Keeping the shape but changing the charges: A simulation study of urea and its iso-steric analogs. *J. Chem. Phys.*, 104(23):9417–9430, 1996.
- [24] Qin Zou, Susan M. Habermann-Rottinghaus, and Kenneth P. Murphy. Urea effects on protein stability: Hydrogen bonding and the hydrophobic effect. *Proteins: Structure, Function, and Bioinformatics*, 31(2):107–115, 1998.
- [25] Mitsunori Ikeguchi, Shugo Nakamura, and Kentaro Shimizu. Molecular dynamics study of hydrophobic effects in aqueous urea solutions. *J. Am. Chem. Soc.*, 123(4):677–682, 2001.

- [26] Raymond D. Mountain and D. Thirumalai. Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J. Am. Chem. Soc.*, 125(7):1950–1957, 2003.
- [27] Chris Oostenbrink and Wilfred F. van Gunsteren. Methane clustering in explicit water: effect of urea on hydrophobic interactions. *Phys. Chem. Chem. Phys.*, 7(1):53–58, 2005.
- [28] Maeng-Eun Lee and Nico F.A. van der Vegt. Does urea denature hydrophobic interactions? *J. Am. Chem. Soc.*, 128(15):4948–4949, 2006.
- [29] Edward P. O’Brien, Ruxandra I. Dima, Bernard Brooks, and D. Thirumalai. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions. *J. Am. Chem. Soc.*, 129(23):7346–7353, 2007.
- [30] Charles Tanford. Protein denaturation. c. theoretical models for the mechanism of denaturation. *Adv. Protein Chem.*, 24:1–95, 1970.
- [31] Ken A. Dill and David Shortle. Denatured states of proteins. *Annu. Rev. Biochem.*, 60:795–825, 1991.
- [32] Rohit V. Pappu and George D. Rose. A simple model for polyproline ii structure in unfolded states of alanine-based peptides. *Protein Science*, 11:2437–2455, 2002.
- [33] Nicholas C. Fitzkee, , and George D. Rose. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA*, 101(34):12497–12502, 2004.
- [34] Shelly J. Whittington, Brian W. Chellgren, Veronique M. Hermann, and Trevor P. Creamer. Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry*, 44(16):6269–6275, 2005.
- [35] Bojan Zagrovic, Jan Lipfert, Eric J. Sorin, Ian S. Millett, Wilfred F. van Gunsteren, Sebastian Doniach, and Vijay S. Pande. Unusual compactness of a polyproline type ii structure. *Proc. Natl. Acad. Sci. USA*, 102(33):11698–11703, 2005.
- [36] Joanna Makowska, Sylwia Rodziewicz-Motowido, Katarzyna Bagiska, Jorge A. Vila, Adam Liwo, Lech Chmurzyski, and Harold A. Scheraga. Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl. Acad. Sci. USA*, 103(6):1744–1749, 2006.
- [37] Jeremy M. Berg, John L. Tymoczko, and Lubert Stryer. *Biochemistry, Fifth Edition: International Version*. W. H. Freeman, 2002.
- [38] Johannes Buchner and Thomas Kiefhaber, editors. *Protein Folding Handbook*. Wiley-VCH, 2005.

- [39] Ken A. Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, 1990.
- [40] M. J. Gething and J. Sambrook. Protein folding in the cell. *Nature*, 355:33–45, 1992.
- [41] Joseph D. Bryngelson, Jose Nelson Onuchic, Nicholas D. Socci, and Peter G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Genetics*, 21(3):167–195, 1995.
- [42] Michael Levitt, Merk Gerstein, Enoch Huang, S. Subbiah, and Jerry Tsai. Protein folding: The endgame. *Annu. Rev. Biochem.*, 66:549–579, 1997.
- [43] Ken A. Dill. Polymer principles and protein folding. *Protein Science*, 8(6):1166–1180, 1999.
- [44] Christopher M. Dobson and Martin Karplus. The fundamentals of protein folding: bringing together theory and experiment. *Current Opinion in Structural Biology*, 9(1):92–101, 1999.
- [45] Martin Gruebele. The fast protein folding problem. *Annu. Rev. Phys. Chem.*, 50:485–516, 1999.
- [46] C. J. Tsai, S. Kumar, B. Ma, and R. Nussinov. Folding funnels, binding funnels, and protein function. *Protein Science*, 8(6):1181–1190, 1999.
- [47] Joan-Emma Shea and Charles L Brooks III. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.*, 52:499–535, 2001.
- [48] Jeffrey K. Myers and Terrence G. Oas. Mechanism of fast protein folding. *Ann. Rev. Biochem.*, 71:783–815, 2002.
- [49] Jose Nelson Onuchic and Peter G. Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14:70–75, 2004.
- [50] Robert L. Baldwin. Energetics of protein folding. *J. Mol. Biol.*, 371:283–301, 2007.
- [51] R. Brian Dyer. Ultrafast and downhill protein folding. *Current Opinion in Structural Biology*, 11:38–47, 2007.
- [52] Mallela M.G. Krishna and S. Walter Englander. A unified mechanism for protein folding: Predetermined pathways with optional errors. *Protein Science*, 16(3):449–464, 2007.
- [53] Charles Tanford and Jacqueline A. Reynolds. *Nature's Robots: A History of Proteins*. Oxford University Press, 2003.

- [54] Rosa Uy and Finn Wold. Posttranslational covalent modification of proteins. *Science*, 198(4320):890–896, 1977.
- [55] L. R. Pratt and D. Chandler. Theory of the hydrophobic effect. *J. Chem. Phys.*, 67(8):3683–3704, 1977.
- [56] Lawrence R. Pratt. Molecular theory of hydrophobic effects: "she is too mean to have her name repeated.". *Ann. Rev. Phys. Chem.*, 53:409–436, 2002.
- [57] David Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437:640–647, 2005.
- [58] Hue Sun Chan and Ken A. Dill. Origins of structure in globular proteins. *Proc. Natl. Acad. Sci. USA*, 87:6388–6392, 1990.
- [59] Hue Sun Chan and Ken A. Dill. Polymer principles in protein structure and stability. *Annu. Rev. Biochem.*, 20:447–490, 1991.
- [60] Bryan A. Krantz, Alok K. Srivastava, Sehat Nauli, David Baker, Robert T. Sauer, and Tobin R. Sosnick. Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nature Struct. Biol.*, 9(6):458–463, 2002.
- [61] Patrick J. Fleming and George D. Rose. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Science*, 14:1911–1917, 2005.
- [62] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.
- [63] Cyrus Chothia. One thousand families for the molecular biologist. *Nature*, 357(6379):543–544, 1992.
- [64] D. P. Yee and K. A. Dill. Families and the structural relatedness among globular proteins. *Protein Science*, 2:884–899, 1993.
- [65] Liisa Holm and Chris Sander. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, 233(1):123–138, 1993.
- [66] C. A. Orengo, T. P. Flores, W. R. Taylor, and J. M. Thornton. Identification and classification of protein fold families. *Protein Engineering*, 6(5):485–500, 1993.
- [67] Victor Kunin, Ildefonso Cases, Anton J. Enright, Victor de Lorenzo, and Christos A. Ouzounis. Myriads of protein families, and still counting. *Genome Biology*, 4:401, 2003.
- [68] L. H. Pearl and C. Prodromou. Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.*, 75:271–94, 2006.

- [69] Christopher D. Snow, Houbi Nguyen, Vijay S. Pande, and Martin Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, 420:102–106, 2002.
- [70] Christopher D. Snow, Bojan Zagrovic, and Vijay S. Pande. The trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.*, 124(49):14548–14549, 2002.
- [71] Huan-Xiang Zhou and Feng Dong. Electrostatic contributions to the stability of a thermophilic cold shock protein. *Biophys. J.*, 84:2216–2222, 2003.
- [72] Alexander Schug, Wolfgang Wenzel, and Ulrich H.E. Hansmann. Energy landscape paving simulations of the trp-cage protein. *J. Chem. Phys.*, 122:194711, 2005.
- [73] Christopher D. Snow, Linlin Qiu, Deguo Du, Feng Gai, Stephen J. Hagen, and Vijay S. Pande. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc. Natl. Acad. Sci. USA*, 101(12):4077–4082, 2004.
- [74] Bojan Zagrovic, Christopher D. Snow, Michael R. Shirts, and Vijay S. Pande. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.*, 323:927–937, 2002.
- [75] Guha Jayachandran, V. Vishal, and Vijay S. Pande. Using massively parallel simulation and markovian models to study protein folding: Examining the dynamics of the villin headpiece. *J. Chem. Phys.*, 124:164902(1–12), 2006.
- [76] Jan Kubelka, James Hofrichter, and William A. Eaton. The protein folding speed limit. *Curr. Opin. Struct. Biol.*, 14:76–88, 2004.
- [77] Stephen J Hagen, Linlin Qiu, and Suzette A Pabit. Diffusional limits to the speed of protein folding: fact or friction? *Journal of Physics: Condensed Matter*, 17:S1503–S1514, 2005.
- [78] G. Caldarelli and P. de los Rios. Cold and warm denaturation of proteins. *Journal of Biological Physics*, 27(2-3):229241, 2001.
- [79] F. Hofmeister. Zur Lehre von der Wirkung der Salze. Zweite Mittheilung. *Arch. Exp. Pathol. Ph.*, 24:247–260, 1888.
- [80] Peter H. von Hippel and Kwok-Ying Wong. On the conformational stability of globular proteins. *J. Biol. Chem.*, 240(10):3909–3923, 1965.
- [81] Tsutomu Arakawa and Serge N. Timasheff. Protein stabilization and destabilization by guanidinium salts. *Biochemistry*, 23(25):5924–5929, 1984.
- [82] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

- [83] Wilfred F. van Gunsteren and Herman J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, 29(9):992–1023, 1990.
- [84] Wilfred F. van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P. Geerke, Alice Glaettli, Philippe H. Hünenberger, Mika A. Kastenholtz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F. A. van der Vegt, and Haibo B. Yu. Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed. Engl.*, 45:40644092, 2006.
- [85] Harold A. Scheraga, Mey Khalili, and Adam Liwo. Protein-folding dynamics: Overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.*, 58:57–83, 2007.
- [86] Max Born and Robert Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.*, 389(20):457–484, 1927.
- [87] A. Warshel and M. Levitt. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.
- [88] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.
- [89] Erik Lindahl, Berk Hess, and David van der Spoel. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7:306–317, 2001.
- [90] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. Gromacs: Fast, flexible, and free. *J. Comp. Chem.*, 26(16):1701–1718, 2005.
- [91] E. Hairer, S. P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations I*. Springer-Verlag, Berlin, 1987.
- [92] R. W. Hockney and S. P. Goel. Quiet high-resolution computer models of a plasma. *J. Comp. Phys.*, 14(2):2, 1974.
- [93] E. J. Nyström. Über die numerische Integration von Differentialgleichungen. *Acta Soc. Sci. Fenn.*, 50:1–54, 1925.
- [94] Loup Verlet. Computer “experiments” on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.
- [95] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18(12):1463–1472, 1997.

- [96] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23(3):327–341, 1977.
- [97] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T.N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.
- [98] Y. Harpaz, N. Elmasry, A.R. Fersht, and K. Henrick. Direct observation of better hydration at the n terminus of an α -helix with glycerine rather than alanine as the n-cap residue. *Proc. Natl. Acad. Sci. USA*, 91(1):311–315, 1994.
- [99] Uwe Mueller, Dieter Perl, Franz X. Schmid, and Udo Heinemann. Thermal stability and atomic-resolution crystal structure of the bacillus caldolyticus cold shock protein. *J. Mol. Biol.*, 297:975–988, 2000.
- [100] J.W. Neidigh, R.M. Fesinmeyer, and N.H. Andersen. Designing a 20-residue protein. *Nature Struct. Biol.*, 9:425–430, 2002.
- [101] M. F. Perutz. Electrostatic effects in proteins. *Science*, 201(4362):1187–1191, 1978.
- [102] A. Warshel and S. T. Russel. Calculations of electrostatic interactions in biological systems and in solution. *Q. Rev. Biophys.*, 17:283–422, 1984.
- [103] M. K. Gilson and B. H. Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. *Proteins: Structure, Function, and Genetics*, 4(1):7–18, 1988.
- [104] W. F. van Gunsteren, F. J. Luque, D. Timms, and A. E. Torda. *Molecular Mechanics in Biology: From Structure to Function, Taking Account of Solvation*. Annual Reviews Inc., 1994.
- [105] B. Roux and T. Simonson. Implicit solvent models. *Biophys. Chem.*, 78(1-2):1–20, 1999.
- [106] C. J. Cramer and D. G. Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99:2161–2200, 1999.
- [107] Ruhong Zhou and Bruce J. Berne. Can a continuum solvent model reproduce the free energy landscape of a β -hairpin folding in water? *Proc. Natl. Acad. Sci. USA*, 99(20):12777–12782, 2002.
- [108] Young Min Rhee, Eric J. Sorin, Guha Jayachandran, Erik Lindahl, and Vijay S. Pande. Simulations of the role of water in the protein-folding mechanism. *Proc. Natl. Acad. Sci. USA*, 101(17):64566461, 2004.

- [109] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.
- [110] M. Saito. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. *J. Chem. Phys.*, 101(5):4055–4061, 1994.
- [111] Michael Bergdorf, Christine Peter, and Philippe H. Hünenberger. Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: A continuum electrostatics study. *J. Chem. Phys.*, 119(17):9129–9144, 2003.
- [112] Andrij Baumketner and Joan-Emma Shea. The influence of different treatments of electrostatic interactions on the thermodynamics of folding of peptides. *J. Phys. Chem. B*, 109:21322–21328, 2005.
- [113] P. P. Ewald. Die berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, IV(64):253–287, 1921.
- [114] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $N \cdot \log(N)$ method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.
- [115] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.
- [116] P.H. Hünenberger and J.A. McCammon. Effect of artificial periodicity in simulations of biomolecules under ewald boundary conditions: a continuum electrostatics study. *Biophysical Chemistry*, 78:69–88, 1999.
- [117] Wolfgang Weber, Philippe H. Hünenberger, and J. Andrew McCammon. Molecular dynamics simulations of a polyalanine octapeptide under ewald boundary conditions: Influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B*, 104(15):3668–3675, 2000.
- [118] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with a coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.
- [119] S. Nosé. A molecular-dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.
- [120] W. G. Hoover. Canonical dynamics equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

- [121] Chris Oostenbrink, Alessandra Villa, Alan E. Mark, and Wilfred F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *J. Comp. Chem.*, 25(13):1656–1676, 2004.
- [122] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.*, 7:230–252, 1986.
- [123] William L. Jorgensen and Julian Tirado-Rives. The opls potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.
- [124] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.
- [125] Michael R. Shirts, Jed W. Pitner, William C. Swope, and Vijay S. Pande. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.*, 119(11):5740–5761, 2003.
- [126] William L. Jorgensen, Jayaraman Chandrasekhar, and Jeffrey D. Madura. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.
- [127] Lorna J. Smith, Herman J. C. Berendsen, and Wilfred F. van Gunsteren. Computer simulation of urea-water mixtures: A test of force field parameters for use in biomolecular simulation. *J. Phys. Chem. B*, 108(3):1065–1071, 2004.
- [128] Frank Eisenhaber, Philip Lijnzaad, Patrick Argos, Chris Sander, and Michael Scharf. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comp. Chem.*, 16(3):273–284, 1995.
- [129] Samuel S. Cho, Yaakov Levy, and Peter G. Wolynes. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA*, 103(3):586–591, 2006.
- [130] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, and H.J.C. Berendsen. *Gromacs User Manual version 3.2*. www.gromacs.org, 2004.
- [131] E. Espinosa, E. Molins, and C. Lecomte. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.*, 285(3-4):170–173, 1998.

- [132] Sheh-Yi Sheu, Dah-Yen Yang, H. L. Selzle, and E. W. Schlag. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci. USA*, 100(22):12683-12687, 2003.
- [133] William D. Arnold and Eric Oldfield. The chemical nature of hydrogen bonding in proteins via nmr: ^1J -couplings, chemical shifts, and aim theory. *J. Am. Chem. Soc.*, 122:12835–12841, 2000.
- [134] S. Jenkins and I. Morrison. The chemical character of the intermolecular bonds of seven phases of ice as revealed by ab initio calculation of electron densities. *Chem. Phys. Lett.*, 317:97102, 2000.
- [135] Mark Rozenberg, Aharon Loewenschuss, and Yizhak Marcus. An empirical correlation between stretching vibration redshift and hydrogen bond length. *Phys. Chem. Chem. Phys.*, 2:2699–2702, 2000.
- [136] Ibon Alkorta and Jose Elguero. Theoretical study of strong hydrogen bonds between neutral molecules: The case of amine oxides and phosphine oxides as hydrogen bond acceptors. *J. Phys. Chem. A*, 103:272–279, 1999.
- [137] O. Gálvez, P.C. Gómez, and L.F. Pacios. Variation with the intermolecular distance of properties dependent on the electron density in hydrogen bond dimers. *J. Chem. Phys.*, 115(24):11166–11184, 2001.
- [138] John A. Schellman. A simple model for solvation in mixed solvents: Applications to the stabilization and destabilization of macromolecular structures. *Biophys. Chem.*, 37(1-3):121–140, 1990.
- [139] Walter Bruning and Alfred Holtzer. The effect of urea on hydrophobic bonds: the critical micelle concentration of n-dodecyltrimethylammonium in aqueous solutions of urea. *J. Am. Chem. Soc.*, 83(23):4865–4866, 1961.
- [140] Charles Tanford. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.*, 84(22):4240–4247, 1962.
- [141] J.A. Schellman. The stability of hydrogen-bonded peptide structures in aqueous solution. *Compt. Trav. Lab. Carlsberg ser. Chim.*, 29:230–259, 1955.
- [142] Gordon C. Kresheck and Harold A. Scheraga. The temperature dependence of the enthalpy of formation of the amide hydrogen bond; the urea model. *J. Phys. Chem.*, 69(5):1704–1706, 1965.
- [143] R. H. Stokes. Thermodynamics of aqueous urea solutions. *Aust. J. Chem.*, 20:2087–2100, 1967.
- [144] A.K. Soper, E.W. Castner, and Alenka Luzar. Impact of urea on water structure: a clue to its properties as a denaturant? *Biophys. Chem.*, 105(2-3):649–666, 2003.

- [145] Koji Yoshida, Kazuyasu Ibuki, and Masakatsu Ueno. Pressure and temperature effects on 2H spin-lattice relaxation times and 1H chemical shifts in tert-butyl alcohol- and urea- D_2O solutions. *J. Chem. Phys.*, 108(4):1360–1367, 1998.
- [146] Brian O Nuallain and Stephen G. Mayhew. A comparison of the urea-induced unfolding of apoflavodoxin and flavodoxin from *Desulfovibrio vulgaris*. *Eur. J. Biochem.*, 269:212–223, 2002.
- [147] A. Idrissi, P. Bartolini, M. Ricci, and R. Righini. Time resolved optical kerr effect analysis of urea-water system. *J. Chem. Phys.*, 114(15):6774–6780, 2001.
- [148] Jože Grdadolnik and Yves Maréchal. Urea and urea-water solutions—an infrared study. *J. Mol. Struct.*, 615(1-3):177–189, 2002.
- [149] R. Keuleers, H. O. Desseyn, B. Rousseau, and C. Van Alsenoy. Vibrational analysis of urea. *J. Phys. Chem. A*, 103(24):4621–4630, 1999.
- [150] Kim A. Sharp, Bhupinder Madan, Eric Manas, and Jane M. Vanderkooi. Water structure changes induced by hydrophobic and polar solutes revealed by simulations and infrared spectroscopy. *J. Chem. Phys.*, 114(4):1791–1796, 2001.
- [151] H. Tanaka, Hidekazu Touhara, Koichiro Nakanishi, and Nobuatsu Watanabe. Computer experiment on aqueous solution. IV. molecular dynamics calculation on the hydration of urea in an infinitely dilute aqueous solution with a new urea-water pair potential. *J. Chem. Phys.*, 80(10):5170–5186, 1984.
- [152] A. Idrissi, E. Cinar, S. Longelin, and P. Damay. The effect of temperature on urea-urea interactions in water: a molecular dynamics simulation. *J. Mol. Liq.*, 110(1-3):201–208, 2004.
- [153] D. Tobi, R. Elber, and D. Thirumalai. The dominant interaction between peptide and urea is electrostatic in nature: A molecular dynamics simulation study. *Biopolymers*, 68(3):359–369, 2003.
- [154] Franjo Sokolic, Abdenacer Idrissi, and Aurelien Perera. Concentrated aqueous urea solutions: A molecular dynamics study of different models. *J. Chem. Phys.*, 116(4):1636–1646, 2002.
- [155] Lorna J. Smith, Rachel M. Jones, and Wilfred F. van Gunsteren. Characterization of the denaturation of human α -lactalbumin in urea by molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 58(2):439–449, 2005.
- [156] D.K. Klimov, John E. Straub, and D. Thirumalai. Aqueous urea solution destabilizes $A\beta_{16-22}$ oligomers. *Proc. Natl. Acad. Sci. USA*, 101(41):14760–14765, 2004.
- [157] Rajappa Chitra and Paul E. Smith. Molecular dynamics simulations of the properties of cosolvent solutions. *J. Phys. Chem. B*, 104(24):5854–5864, 2000.

- [158] P.-O. Astrand, A. Wallqvist, and G. Karlstrom. Intermolecular interactions of urea and water. *J. Chim. Phys.*, 88:2457–2464, 1991.
- [159] P.-O. Astrand, A. Wallqvist, and G. Karlstrom. Nonempirical intermolecular potentials for urea-water systems. *J. Chem. Phys.*, 100(2):1262–1273, 1994.
- [160] P.-O. Astrand, A. Wallqvist, and G. Karlstrom. Molecular dynamics simulations of 2 *m* aqueous urea solutions. *J. Phys. Chem.*, 98:8224–8233, 1994.
- [161] Zhiyong Zhang, Yongjin Zhu, and Yunyu Shi. Molecular dynamics simulations of urea and thermal-induced denaturation of s-peptide analogue. *Biophys. Chem.*, 89(2-3):145–162, 2001.
- [162] Qin Zou, Brian J. Bennion, Valerie Daggett, and Kenneth P. Murphy. The molecular mechanism of stabilization of proteins by tmao and its ability to counteract the effects of urea. *J. Am. Chem. Soc.*, 124(7):1192–1202, 2002.
- [163] Daniel Trzesniak, Nico F. A. van der Vegt, and Wilfred F. van Gunsteren. Computer simulation studies on the solvation of aliphatic hydrocarbons in 6.9 M aqueous urea solution. *Phys. Chem. Chem. Phys.*, 6(4):697–702, 2004.
- [164] Franjo Sokolic, Abdenacer Idrissi, and Aurelien Perera. A molecular dynamic study of the structural properties of aqueous urea solutions. *J. Mol. Liq.*, 101(1-3):81–87, 2002.
- [165] E. S. Boek, W. J. Briels, J. van Eerden, and D. Feil. Molecular-dynamics simulations of interfaces between water and crystalline urea. *J. Chem. Phys.*, 96(9):7010–7018, 1992.
- [166] E. S. Boek and W. J. Briels. Molecular dynamics simulations of aqueous urea solutions: Study of dimer stability and solution structure, and calculation of the total nitrogen radial distribution function $g_N(r)$. *J. Chem. Phys.*, 98(2):1422–1427, 1993.
- [167] A. Wallqvist, D. G. Covell, and D. Thirumalai. Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *J. Am. Chem. Soc.*, 120:427–428, 1998.
- [168] Samantha Weerasinghe and Paul E. Smith. A kirkwood-buff derived force field for mixtures of urea and water. *J. Phys. Chem. B*, 107(16):3891–3898, 2003.
- [169] Samantha Weerasinghe and Paul E. Smith. Cavity formation and preferential interactions in urea solutions: Dependence of urea aggregation. *J. Chem. Phys.*, 118(13):5901–5910, 2003.
- [170] John A. Schellman. Fifty years of solvent denaturation. *Biophys. Chem.*, 96:91–101, 2002.

- [171] E.G. Finer, F. Franks, and M.J. Tait. Nuclear magnetic resonance studies of aqueous urea solutions. *J. Am. Chem. Soc.*, 94(13):4424–4429, 1972.
- [172] Julian Tirado-Rives, Modesto Orozco, and William L. Jorgensen. Molecular dynamics simulations of the unfolding of barnase in water and 8 m aqueous urea. *Biochemistry*, 36(24):7313–7329, 1997.
- [173] Wolfgang Härdle and Leopold Simar. *Applied Multivariate Statistical Analysis*. Springer, Berlin, 2003.
- [174] John G. Kirkwood and Frank P. Buff. The statistical mechanical theory of solutions. I. *J. Chem. Phys.*, 19(6):774–777, 1951.
- [175] A. K. Soper. Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water. *Molec. Phys.*, 99(17):1503–1516, 2001.
- [176] Michael W. Mahoney and William L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20):8910–8922, 2000.
- [177] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.
- [178] Elizabeth S. Courtenay, Michael W. Capp, and M. Thomas Record Jr. Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Science*, 10:2485–2497, 2001.
- [179] Matthew Auton and D. Wayne Bolen. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. USA*, 102(42):15065–15068, 2005.
- [180] Andreas Möglich, Florian Krieger, and Thomas Kiefhaber. Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *J. Mol. Biol.*, 345(1):153–162, 2005.
- [181] V. Prakash, C. Loucheux, Stephen Scheuffele, Marina J. Gorbunoff, and Serge N. Timasheff. Interactions of proteins with solvent components in 8 m urea. *Archives of Biochemistry and Biophysics*, 210(2):455–464, 1981.
- [182] Serge N. Timasheff and Guifu Xie. Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys. Chem.*, 105:421–448, 2003.
- [183] Pradip K. Nandi and Dwight R. Robinson. Effects of urea and guanidine hydrochloride on peptide and nonpolar groups. *Biochemistry*, 23:6661–6668, 1984.

- [184] Donald B. Wetlaufer, Sunit K. Malik, Leon Stoller, and Ronald L. Coffin. Nonpolar group participation in the denaturation of proteins by urea and guanidinium salts. model compound studies. *J. Am. Chem. Soc.*, 86:508–514, 1964.
- [185] G. C. Kresheck and L. Benjamin. Calorimetric studies of the hydrophobic nature of several protein constituents and ovalbumin in water and in aqueous urea. *J. Phys. Chem.*, 68:2476–2486, 1964.
- [186] Y. Nozaki and C. Tanford. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.*, 246(7):2211–2217, 1971.
- [187] J. Janin. Surface and inside volumes in globular proteins. *Nature*, 277:491–492, 1979.
- [188] H. Robert Guy. Amino acid side-chain partition energies and distribution of energies in soluble proteins. *Biophys. J.*, 47:61–70, 1985.
- [189] George D. Rose, Ari R. Geselowitz, Glenn J. Lesser, Richard H. Lee, and Micheal H. Zehfus. Hydrophobicity of amino acid residues in globular proteins. *Science*, 229(4716):834–838, 1985.
- [190] Domenico Bordo and Patrick Argos. Suggestions for safe residue substitutions in site-directed mutagenesis. *J. Mol. Biol.*, 217(4):721–729, 1991.
- [191] Leslie A. Kuhn, Craig A. Swanson, Michael E. Pique, John A. Trainer, and Elizabeth D. Getzoff. Atomic and residue hydrophilicity in the context of folded protein structures. *Proteins: Structure, Function, and Genetics*, 23:536–547, 1995.
- [192] Marvin Charton and Barbara I. Charton. The structural dependence of amino acid hydrophobicity parameters. *Journal of Theoretical Biology*, 99(4):629–644, 1982.
- [193] James L. Cornette, Kemp B. Cease, Hanah Margalit, John L. Spouge, Jay A. Berzofsky, and Charles DeLisi. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.*, 195(3):659–685, 1987.
- [194] Jeffrey K. Myers, C. Nick Pace, and J. Martin Scholtz. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Science*, 4:2138–2148, 1995.
- [195] Mark A. Roseman. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.*, 200:513–522, 1985.
- [196] C. A. McPhalen and M. N. G. James. Crystal and molecular structure of the serine proteinase inhibitor ci-2 from barley seeds. *Biochemistry*, 26(1):261 – 269, 1987.
- [197] Sophie E. Jackson and Alan R. Fersht. Folding of chymotrypsin inhibitor 2. 1. evidence for a two-state transition. *Biochemistry*, 30(43):10428–10435, 1991.

- [198] K. L. Maxwell, D. Wildes, A. Zarrine-Afsar, M. A. De Los Rios, A. G. Brown, C. T. Friel, L. Hedberg, J. C. Horng, D. Bona, E. J. Miller, A. Vallée-Bélisle, E. R. Main, F. Bemporad, L. Qiu, K. Teilum, N. D. Vu, A. M. Edwards, I. Ruczinski, F. M. Poulsen, B. B. Kragelund, S. W. Michnick, F. Chiti, Y. Bai, S. J. Hagen, L. Serrano, M. Oliveberg, D. P. Raleigh, P. Wittung-Stafshede, S. E. Radford, S. E. Jackson, T. R. Sosnick, S. Marqusee, A. R. Davidson, and K. W. Plaxco. Protein folding: defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Science*, 14(3):602–616, 2005.
- [199] Sophie E. Jackson and Alan R. Fersht. Folding of chymotrypsin inhibitor 2. 2. influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry*, 30(43):10436–10443, 1991.
- [200] Sophie E. Jackson, Marco Moracci, Nadia elMasry, Christopher M. Johnson, and Alan R. Fersht. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry*, 32(42):11259–11269, 1993.
- [201] Sophie E. Jackson, Nadia elMasry, and Alan R. Fersht. Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. *Biochemistry*, 32(42):11270–11278, 1993.
- [202] Alan R. Fersht. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA*, 92(24):10869–10873, 1995.
- [203] Laura S. Itzhaki, Daniel E. Otzen, and Alan R. Fersht. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.*, 254(2):260–288, 1995.
- [204] L. Reich and T.R. Weikl. Substructural cooperativity and parallel versus sequential events during protein unfolding. *Proteins: Structure, Function, and Bioinformatics*, 63(4):1052–1058, 2006.
- [205] Aujun Li and Valerie Daggett. Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA*, 91:10430–10434, 1994.
- [206] Aijun Li and Valerie Daggett. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.*, 257:412–429, 1996.
- [207] Ryan Day and Valerie Daggett. Sensitivity of the folding/unfolding transition state ensemble of chymotrypsin inhibitor 2 to changes in temperature and solvent. *Protein Science*, 14:1242–1252, 2005.

- [208] C. Merlo, K.A. Dill, and T.R. Weikl. Phi values in protein-folding kinetics have energetic and structural components. *Proc. Natl. Acad. Sci. USA*, 102(29):10171–10175, 2005.
- [209] T.R. Weikl and K.A. Dill. Transition-states in protein folding kinetics: The structural interpretation of phi values. *J. Mol. Biol.*, 365(5):1578–1586, 2006.
- [210] G. Vriend. What if: A molecular modeling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.
- [211] Berk Hess, Christian Holm, and Nico van der Vegt. Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Chem. Phys.*, 124:164509, 2006.
- [212] Bengt Nölting, Ralph Golbik, José L. Neira, Andrés S. Soler-Gonzalez, Gideon Schreiber, and Alan R. Fersht. The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. USA*, 94(3):826–830, 1997.
- [213] M. Lois Tiffany and S. Krimm. New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers*, 6(9):1379–1382, 1968.
- [214] K. L. Reid, H. M. Rodriguez, B. J. Hillier, and L. M. Gregoret. Stability and folding properties of a model β -sheet protein, escherichia coli cspa. *Protein Science*, 7:470–479, 1998.
- [215] Benjamin Schuler, Werner Kremer, Hans Robert Kalbitzer, and Rainer Jaenicke. Role of entropy in protein thermostability: Folding kinetics of a hyperthermophilic cold shock protein at high temperatures using ^{19}F NMR. *Biochemistry*, 41:11670–11680, 2002.
- [216] Armin Hoffmann, Avinash Kane, Daniel Nettels, David E. Hertzog, Peter Baumgärtel, Jan Lengefeld, Gerd Reichardt, David A. Horsley, Robert Seckler, Olgica Bakajin, and Benjamin Schuler. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA*, 104(1):105–110, 2007.
- [217] Everett A. Lipman, Benjamin Schuler, Olgica Bakajin, and William A. Eaton. Single-molecule measurement of protein folding kinetics. *Science*, 301:1233–1235, 2003.
- [218] Christine Magg and Franz X. Schmid. Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.*, 335:1309–1323, 2004.
- [219] Mourad Sadqi, Lisa J. Lapidus, and Victor Muñoz. How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. USA*, 100(21):12117–12122, 2003.
- [220] Daniel Nettels, Irina V. Gopich, Armin Hoffmann, and Benjamin Schuler. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA*, 104(8):2655–2660, 2007.

- [221] Stephen J. Hagen and William A. Eaton. Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.*, 301:1019–1027, 2000.
- [222] Thomas Schindler, Michael Herrler, Mohamed A. Marahiel, and Franz X. Schmid. Extremely rapid protein folding in the absence of intermediates. *Nature Struct. Biol.*, 2(8):663–673, 1995.
- [223] Christopher M. Dobson, Philip A. Evans, and Sheena E. Radford. Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.*, 19(1):31–37, 1994.
- [224] Linlin Qiu, Suzette A. Pabit, Adrian E. Roitberg, and Stephen J. Hagen. Smaller and faster: The 20-residue trp-cage protein folds in 4 μ s. *J. Am. Chem. Soc.*, 124:12952–12953, 2002.
- [225] J. Juraszek and P. G. Bolhuis. Sampling the multiple folding mechanisms of trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA*, 103(43):1585915864, 2006.
- [226] Ruhong Zhou. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA*, 100(23):1328013285, 2003.
- [227] K. Hukushima and K. Nemoto. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.
- [228] K. Hun Mok, Lars T. Kuhn, Martin Goez, Iain J. Day, Jasper C. Lin, Niels H. Andersen, and P.J. Hore. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature*, 447:106–109, 2007.
- [229] Stanford Moore. Amino acid analysis: Aqueous dimethyl sulfoxide as solvent for the ninhydrin reaction. *J. Biol. Chem.*, 243(23):6281–6283, 1968.
- [230] Michael Jackson and Henry H. Mantsch. Beware of proteins in dms. *Biochimica et Biophysica Acta*, 1078(2):231–235, 1991.
- [231] Michael Kotik, Sheena E. Radford, and Christopher M. Dobson. Comparison of the refolding hen lysozyme from dimethyl sulfoxide and guanidinium chloride. *Biochemistry*, 34(5):1714–1724, 1995.
- [232] H. Iwase, M. Hirai, S. Arai, S. Mitsuya, S. Shimizu, T. Otomo, and M. Furusaka. Comparison of dms-induced denaturation of hen egg-white lysozyme and bovine α -lactalbumin. *Journal of Physics and Chemistry of Solids*, 60:1379–1381, 1999.
- [233] Shekhar C. Mande and M. Elizabeth Sobhia. Structural characterization of protein-denaturant interactions: crystal structures of hen egg-white lysozyme in complex with dms and guanidinium chloride. *Protein Engineering*, 13(2):133–141, 2000.

- [234] Jr. Raymond F. Greene and C. Nick Pace. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α -chymotrypsin, and β -lactoglobulin. *J. Biol. Chem.*, 249(17):5388–5393, 1974.
- [235] W. Kunz, , P. Lo Nostro, and B. W. Ninham. The present state of affairs with Hofmeister effects. *Current Opinion in Colloid & Interface Science*, 9(1-2):1–18, 2004.
- [236] Christopher E. Dempsey, Thomas J. Piggot, and Philip E. Mason. Dissecting contributions to the denaturant sensitivities of proteins. *Biochemistry*, 44(2):775–781, 2005.
- [237] Lorenz M. Mayr and Franz X. Schmid. Stabilization of a protein by guanidinium chloride. *Biochemistry*, 32(31):7994–7998, 1993.
- [238] Raymond D. Mountain and D. Thirumalai. Alteration in water structure induced by guanidinium and sodium ions. *J. Phys. Chem. B*, 108(51):19711–19716, 2004.
- [239] Nihmat A. Morjana, Barry J. McKeone, and Hiram F. Gilbert. Guanine hydrochloride stabilization of a partially unfolded intermediate during the reversible denaturation of protein disulfide isomerase. *Proc. Natl. Acad. Sci. USA*, 90:2107–2111, 1993.
- [240] O. D. Monera, C. M. Kay, and R. S. Hodges. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Science*, 3:1984–1991, 1994.
- [241] Md. Sohail Akhtar, Atta Ahmad, and Vinod Bhakuni. Guanidinium chloride- and urea-induced unfolding of the dimeric enzyme glucose oxidase. *Biochemistry*, 41:3819–3827, 2002.
- [242] Jeni S. Smith and J. Martin Scholtz. Guanidine hydrochloride unfolding of peptide helices: Separation of denaturant and salt effects. *Biochemistry*, 35(22):7292–7297, 1996.
- [243] T. Arakawa and S.N. Timasheff. The stabilization of proteins by osmolytes. *Biophys. J.*, 47:411–414, 1985.
- [244] Ilia Baskakov and D.W. Bolen. Forcing thermodynamically unfolded proteins to fold. *J. Biol. Chem.*, 273(9):4831–4834, 1998.
- [245] Paul H. Yancey, Mary E. Clark, Steven C. Hand, R. David Bowlus, and George N. Somero. Living with water stress: Evolution of osmolyte systems. *Science*, 217(4566):1214–1222, 1982.

- [246] Aijun Wang and D. W. Bolen. A naturally occurring protective system in urea-rich cells: Mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry*, 36(30):9101–9108, 1997.
- [247] D. W. Bolen and Ilia V. Baskakov. The osmophobic effect: Natural selection of a thermodynamic force in protein folding. *J. Mol. Biol.*, 310:955–963, 2001.
- [248] Timothy O. Street, D. Wayne Bolen, and George D. Rose. A molecular mechanism for osmolyte-induced protein stability. *Proc. Natl. Acad. Sci. USA*, 103(38):13997–14002, 2006.
- [249] Karen G. Hovagimyan and John T. Gerig. Interactions of trimethylamine n-oxide and water with *cyclo*-alanylglycine. *J. Phys. Chem. B*, 109(50):24142–24151, 2005.
- [250] Philip E. Mason, John W. Brady, George W. Neilson, and Christopher E. Dempsey. The interaction of guanidinium with a model peptide. *Biophys. J.*, 93(1):L04–L06, 2007.
- [251] Yufeng Liu and D. W. Bolen. The peptide bond plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry*, 34:12884–12891, 1995.
- [252] David Shortle and Michael S. Ackerman. Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, 293(5529):487–489, 2001.

Lebenslauf

Martin Stumpe

Name:	Martin Christian Stumpe
Geburtsdatum:	24.9.1976
Geburtsort:	Bad Pyrmont
Staatsangehörigkeit:	deutsch
Aktuelle Anschrift:	Südring 6, 37079 Göttingen
Familienstand:	ledig
Eltern:	Dieter Stumpe Irmgard Stumpe, geb. Matthias
1983 – 1987	St. Marien Grundschule Lügde
1987 – 1996	Städtisches Gymnasium Blomberg Abitur im Juni 1996
08/1996 – 09/1997	Zivildienst im Bathildis-Krankenhaus Bad Pyrmont
10/1997 – 05/2003	Studium der Physik an der Westfälischen Wilhelms-Universität Münster Diplom im Mai 2003, Titel der Diplomarbeit: "Inhaltsadressiertes Auslesen in einem volumenholographischen Speichersystem mit Phasencodierung"
10/2003 – 10/2007	Promotion am Max-Planck-Institut für biophysikalische Chemie, Göttingen, über das Thema "On the molecular basis of urea-induced protein unfolding", Bewertung mit <i>summa cum laude</i>
11/2007 –	PostDoc am Max-Planck-Institut für biophysikalische Chemie, Göttingen