Force field molecular dynamics at high hydrostatic pressures: Water, osmolytes, and peptides

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften (Dr. rer. nat.)

der Fakultät für Chemie und Pharmazie

der Universität Regensburg

vorgelegt von

Christoph Hölzl

aus Ingolstadt

im Jahr 2018

Promotionsgesuch eingereicht am: 13.12.2018

Die Arbeit wurde angeleitet von: Prof. Dr. Dominik Horinek

Prüfungsausschuss:	Vorsitzender:	Prof. Dr. Joachim Wegener
	1. Gutachter:	Prof. Dr. Dominik Horinek
	2. Gutachter:	Prof. Dr. Stefan M. Kast
	weiterer Prüfer:	Prof. Dr. Richard Buchner

Abstract

Understanding water and aqueous solutions of biomolecules at high hydrostatic pressure is important not only for the research of deep sea life, but also for gaining insight into properties relevant to ambient conditions. Some small organic molecules, called osmolytes, can stabilize or destabilize the native conformation of proteins in cells against external stresses like temperature and pressure. We use force field molecular dynamics simulations to study the influence of selected osmolytes on water and proteins solutions with respect to their response to high hydrostatic pressure. This is achieved using different approaches, which are aimed to lay the foundations for understanding and being able to simulate osmolytes at high pressures and then applying these fundamental results to more realistic systems:

1. We present a detailed analysis of the structure of pure water close to a hydrophobic alkane monolayer, which serves as a model system for studying hydrophobic solvation, up to 10 kilobars of pressure. The structural properties of interest include the frequency and geometry of rings in the hydrogen bond network, from which we deduce that the instantaneous structure of water at this hydrophobic interface increases its likeness to hexagonal ice with increasing pressure.

2. We show that it is necessary to modify the force field parameters of the protecting osmolyte trimethylamine-N-oxide (TMAO) in order to correctly reproduce its solvation structure at high pressures, and we present a systematic procedure for scaling the partial charges. For the most relevant protein denaturant, urea, we develop a new force field for aqueous solutions which significantly improves upon existing literature models at normal pressure and in the high pressure range.

3. We apply the transfer model to periodic glycine and alanine homopeptides in different secondary structures to show that helical conformations of these peptides are stabilized by TMAO even at 5 kilobars relative to an extended structure.

4. We simulate folding of the Trp-cage miniprotein in water and in TMAO solution at normal and high pressures. We find some evidence that the protecting effect of TMAO on the tertiary structure of Trp-cage is weaker but still present at 10 kbar.

Glossary

Acronyms

aiMD	*ab initio* molecular dynamics
CD	circular dichroism
EC-RISM	embedded cluster reference interaction site model
ff	force field
H-bond	hydrogen bond
KBI	Kirkwood-Buff integral (G_{ij})
LJ	Lennard-Jones
MD	molecular dynamics
OTS	octadecyltrichlorosilane
PDB	Protein Data Bank
PMF	potential of mean force
QC	quantum chemistry
RDF	radial distribution function g_{ij}
RMSD	root mean squared deviation
SAM	self-assembled monolayer
SASA	solvent-accessible surface area
TFE	transfer free energy
TMAO	trimethylamine-N-oxide
vdW	van der Waals

Symbols

ϵ	dielectric constant
ϵ_i	parameter in the LJ potential of atom type i (unit: kJ/mol)
G_{ij}	Kirkwood-Buff integral (KBI)

g_{ij}	radial distribution function (RDF)
κ_{T}	isothermal compressibility
k	Boltzmann constant
μ_i	chemical potential
$\mathbf{p}$	momentum vector
P	probability or probability density
p	system pressure
$q_{3/4}$	tetrahedral/trigonal-pyramidal order parameter
q_i	point charge
ρ	mass density
$\mathbf{r}$	position vector
R_g	radius of gyration
ρ_i	number density of component i
σ_i	parameter in the LJ potential of atom type i (unit: nm)
T	Temperature
U	potential energy
V	volume of the simulation box or subbox
V_i	partial molar volume of component i
χ_{R}	ring twist
y_i	activity coefficient in the molarity scale

Contents

Chapter 1

Introduction

At the time when life first developed on earth at least 3.5 billion years ago, the conditions on the planet were extreme. It is a distinct possibility that life originated in the deep oceans, e.g. around hydrothermal vents,[1] where the conditions were still harsh, but more consistent. This means that early life forms must have had mechanisms to protect proteins against these external stresses, since their ability to function is very sensitive to their environment. Many factors are involved in driving the folding and preventing the denaturation of proteins. Water as the omnipresent solvent plays a major role by driving the hydrophobic collapse of proteins.[2] The solvent mixture in the cytoplasm contains a multitude of components that influence the stability of proteins.[3] Therefore, it is imperative that we have an understanding of how water and those cosolvents interact with proteins under high hydrostatic pressure. Using pressure as the main thermodynamic variable is not only useful for studying the biophysics of deep sea organisms, but has also become an invaluable tool for determining properties of biomolecules under ambient conditions, including the equilibrium and dynamics of protein conformations that are relevant to folding and unfolding.[4–6] In this thesis, we approach this topic from different angles using force field molecular dynamics simulations. What follows are brief introductions to the individual chapters.

Oil and water are not miscible and form separate phases instead. This phenomenon is widely known as the hydrophobic effect, but despite making continuous progress,[7] the molecular interactions which give rise to the macroscopic effects are still not fully understood. Besides protein folding, hydrophobic interactions with water are important in many other areas, including the self-assembly of micelles[8] and other colloidal properties.[7] There are multiple systems one has to consider when trying to explain the hydrophobic effect, since it manifests differently depending on the size of the hydrophobic particle or phase.[9] For small hydrophobic molecules, the decades-old theory that water forms an ice-like cage around the solute[10] was later refuted, but a moderate slowdown of the water rotational dynamics was observed in IR spectroscopy experiments[11] and simulations.[12] Although it is now commonly accepted that the solvation shell of hydrophobic solutes is not crystalline ice, there is evidence that water has certain ice-like properties at some interfaces. Vibrational sum-frequency spectroscopy (VSFG) measurements revealed distinct signals for the stretch vibration of the water O-H bond around antifreeze proteins[13] and at extended hydrophobic interfaces.[14] Water orientations extracted from simulations also support the view of dangling hydrogen bonds (H-bonds) at simple hydrophobic interfaces consistent with an ice-like layer.[15,16] Since the structure of water is very complex, a wide selection of parameters has been developed to characterize different aspects of water. The simplest

order parameters measure the local tetrahedral order around each water molecule,[17,18] but they are unable to determine whether a water molecule is part of an ice crystal. A combination of three parameters has been used to distinguish between liquid water, hexagonal ice and clathrates.[19] A new algorithm has recently been developed which can detect hexagonal and cubic ice as well as the most common clathrate types using the correlation of the local order around each water molecule.[20] However, since these parameters assume that each water molecule has approximately four neighbors, they will most likely fail when applied to interfaces, and we require different methods to identify the structural properties of interfacial water. Water at hydrophobic interfaces has a layer of reduced density, commonly called the hydrophobic depletion layer or 'gap', which has a size on the order of Angströms for alkanes[21] and perfluorinated alkanes.[22] In previous work we showed that the hydrophobic gap at an octadecyltrichlorosilane (OTS) self-assembled monolayer (SAM) is significantly compressible.[23] In chapter 3, we investigate the structure of water at this interface on different length scales and quantify its similarity to that of hexagonal ice at different pressures.

The term osmolyte refers to the subset of cosolute molecules that exhibit osmotic activity. It includes many small organic molecules that regulate the osmotic pressure in the cells of all living organisms against external stresses. The group of organic osmolytes includes methylamines, amino acids, sugars, polyols, and urea.[3] Through evolution, all organisms use a combination of osmolytes that best protects their cells against stresses such as high salinity, temperature, and pressure, while at the same time assuring that these osmolytes are compatible with the cell interior and do not impede the function of proteins in the cytoplasm.[24] Individually, osmolytes can be separated into two categories: There are osmolytes that protect proteins against denaturation by external stresses, which are called protecting osmolytes or osmoprotectants.[25] The second category are osmolytes with a destabilizing effect on the native folded state of proteins, commonly called denaturing osmolytes. The most commonly studied protecting osmolyte is trimethylamine-N-oxide (TMAO), because it is one of the most effective protecting agents against protein denaturation by temperature, pressure, and denaturing osmolytes.[25–27] Currently, there is no full consensus about the molecular mechanism by which TMAO stabilizes proteins. Zou et al. suggest an indirect mechanism in which TMAO alters the water structure around proteins by strengthening water-water H-bonds. Unfolding a protein involves breaking intermolecular H-bonds and forming more protein-water H-bonds, which would require breaking the stronger water-water H-bonds in the presence of TMAO.[28] However, there is also evidence for a direct mechanism in which the exclusion of TMAO from the protein surface leads to a stabilization of the native protein conformation.[29] For a detailed discussion of osmolyte protection mechanisms we refer to a review by Canchi and Garcia.[30] One extraordinary effect of TMAO is its protection of proteins against pressure denaturation. It was found that in salt water fish (specifically teleosts), there is a linear relation between the content of TMAO in the muscle tissue and the fishes' native depth.[31] Out of all denaturing osmolytes, urea is by far the most relevant in biology, and its denaturation mechanism is now commonly accepted to be a direct favorable interaction with the protein.[32–34] Therefore, TMAO and urea were chosen as the objects of study and the topic of chapter 4. In nature, these two osmolytes often occur together in a urea to TMAO ratio of 2:1. The destabilizing and stabilizing effects of both components are additive and approximately cancel out, which means this mixture can be used in high stress environments in adequately high concentrations without interfering with protein function.[3,35,36] In this thesis, the main focus lies on using pressure as a variable to better understand the effect of osmolytes on proteins and water. While there are experimental studies on osmolyte effects on the pres-

sure denaturation of proteins and peptides,[37–41] simulations are significantly hindered by the fact that force field approaches are dependent on sets of parameters that have not been validated outside of ambient conditions, whereas *ab initio* methods are too expensive even for small peptide systems. Chapter 4 exclusively deals with binary osmolyte/water mixtures of TMAO and urea. We introduce the necessary modifications to an existing TMAO model for simulations at high pressure, and we fully develop a new force field for urea which performs well under ambient and high pressure conditions.

Our main interest in osmolytes is the study of their molecular-scale protein denaturation and stabilization mechanisms. Recently, a simulation study of the folding/unfolding equilibrium of a hydrophobic model polymer has shown that the stabilizing effect of TMAO may not monotonically increase with TMAO concentration, but that TMAO could even facilitate the unfolding of hydrophobic polymers at very high concentrations.[42] In chapter 5, we apply the tools developed in the previous chapter to study the effect of TMAO concentration and pressure on the stability of different helical structures of model homopeptides of glycine and alanine.

One of the main long-term goals of molecular dynamics simulations is gaining the ability to study the folding process of proteins. This is not only relevant to understanding their biophysical and biochemical functions in nature, but also to gain insight into the origin and ramifications of folding defects, which are believed to cause diseases such as cystic fibrosis[43] and Alzheimer's.[44] It was found that protecting osmolytes such as TMAO can prevent folding defects,[45,46] which raises the question whether this effect differs from the molecular mechanism by which these osmolytes stabilize against temperature and pressure denaturation. The main challenge for simulations are the timescales on which most proteins fold, which is usually on the order of milliseconds. Immense effort has been put into designing the fastest-folding protein, which resulted in the 20-residue Trp-cage miniprotein TC5b[47] with a folding time on the order of 4 μs.[48] It has already been intensely studied as a prime model system for globular proteins using experiments and simulations in water as well as osmolyte solutions.[32,48–55] The interplay of urea and pressure on Trp-cage stability has been investigated using replica-exchange MD (REMD) simulations[32] and, more recently, a combination of high-pressure NMR spectroscopy and simulations.[56] However, the studies combining TMAO and Trp-cage have so far been limited to normal pressure conditions.[54,55] In chapter 6, we present results of folding simulations in TMAO solutions at ambient and high pressures.

Chapter 2

Methods and fundamentals

2.1 Simulation details

All simulations in this work were performed using the following methods, unless they are explicitly mentioned in the methods section of a subsequent chapter. For the molecular dynamics (MD) simulations the `GROMACS` program package with versions ranging from 4.6.5 to 2016.4 (all compiled in single precision) was used. In molecular dynamics, Newton's equations of motion of a many-body system are numerically integrated using a finite difference method, in this case the leap-frog algorithm. In our nonpolarizable force field MD simulations, all interactions are described by analytical two-body potentials. The Lennard-Jones (LJ) potential models the van der Waals interactions (eq. 2.1):

$$U_{ij}^{\mathrm{LJ}}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.1}$$

Here i and j are the two particles, r is the variable distance between the two particles, ϵ is the potential minimum depth, and σ is the distance where the potential vanishes. We use the Lorentz-Berthelot combination rules to calculate σ and ϵ between different atom types:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{2.2}$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \cdot \epsilon_j} \tag{2.3}$$

For electrostatic interactions we use the Coulomb potential. The LJ interactions and the real-space Coulomb interactions were cut off at 1 nm, and the long-range electrostatics was computed using smooth particle-mesh Ewald summation[57] with a lattice spacing of 0.12 nm. In isotropic systems, the effect of a finite LJ cutoff range was accounted for using a long range dispersion correction for the energy and virial as implemented in `GROMACS`.

In order to simulate thermodynamic ensembles other than the microcanonical (NVE) ensemble, thermostat and barostat algorithms were applied. For temperature control, the stochastic velocity rescaling algorithm (sometimes called 'canonical sampling through velocity rescaling' or CSVR)[58] was used with a time constant of 1 ps to fix the temperature to 300 K. For pressure control, two barostat methods were used: The Berendsen barostat,[59] which does not reproduce the correct pressure fluctuations of the NPT ensemble, was applied in equilibration simulations with a time constant of 1 ps, since it relaxes the system volume to its equilibrium value very quickly. The Parrinello-Rahman barostat[60] does con-

tain the correct pressure fluctuations, and was used in subsequent production simulations. Before equilibration, the potential energy of the starting geometry was minimized using a simple steepest descent algorithm.

With one exception, we chose the TIP4P/2005 water model for all simulations in this work. This model has been tested and characterized in a large range of conditions ranging up to extreme temperatures and pressures, and perform very well regarding most measures.[61,62]

2.2 Hydrogen bonds

The way we determine hydrogen bonding in our simulations is based on a purely geometric criterion. A variety of such H-bond definitions have been used in the literature, which will not be discussed here.[63] We use a relatively new criterion introduced in ref. 64. This criterion has been proven to properly identify H-bonds in water at high pressures up to 10 kbar. An H-bond is defined to exist when

$$r_{\mathrm{H\cdots A}} < -0.171\,\mathrm{nm} \cdot \cos\left(\theta_{\mathrm{D-H\cdots A}}\right) + 0.137\,\mathrm{nm} \tag{2.4}$$

where D and A are the donor and acceptor atoms respectively, $r_{\mathrm{H\cdots A}}$ is the distance between the donated hydrogen and the acceptor, and $\theta_{\mathrm{D-H\cdots A}}$ is the angle between the donor-hydrogen bond and the hydrogen-acceptor vector.

2.3 Kirkwood-Buff theory

In their theory of mixtures,[65] Kirkwood and Buff derived relationships between the radial distribution functions (RDFs) and the particle number fluctuations for solutions in the grand canonical (μVT) ensemble. The fluctuations were then related to macroscopic thermodynamic properties such as the derivative of the chemical potential, partial molar volume, and compressibility, thus establishing a direct connection of the integrals over the RDFs, hereafter called Kirkwood-Buff integrals (KBIs), and the thermodynamics of the system. In this work we apply Kirkwood-Buff theory only to binary mixtures, and we will use the notation and relations found in ref. 66. The KBI G_{ij} is defined in eq. 2.5:

$$G_{ij} = \int_0^\infty 4\pi r^2 \left[g_{ij}^{\mu VT}(r) - 1\right] \mathrm{d}r \tag{2.5}$$

Here, r is the distance between the centers of mass of molecules of type i and j, and $g_{ij}^{\mu VT}$ is the RDF of the two components in an open system with constant volume and temperature. In a binary mixture of molecules A and B, the thermodynamic properties can be expressed using the auxiliary quantities η and ξ:

$$\eta = \rho_A + \rho_B + \rho_A\rho_B(G_{AA} + G_{BB} - 2G_{AB}) \tag{2.6}$$

$$\xi = 1 + \rho_A G_{AA} + \rho_B G_{BB} + \rho_A\rho_B(G_{AA}G_{BB} - G_{AB}^2) \tag{2.7}$$

Here ρ are the number densities of the components. The partial molar volumes are then

$$\bar{V}_A = \frac{1 + \rho_B(G_{BB} - G_{AB})}{\eta} \tag{2.8}$$

$$\bar{V}_B = \frac{1 + \rho_A(G_{AA} - G_{AB})}{\eta} \tag{2.9}$$

and the isothermal compressibility can be calculated as

$$\kappa_{\mathrm{T}} = \frac{\xi}{\mathrm{k}T\eta} \tag{2.10}$$

We also obtain the density derivative of the chemical potential of the components

$$\left(\frac{\partial \mu_A}{\partial \rho_A}\right)_{p,T} = \frac{\mathrm{k}T}{\rho_A(1 + \rho_A G_{AA} - \rho_A G_{AB})} \tag{2.11}$$

where p and T are pressure and temperature, and k is the Boltzmann constant. Since we will specifically be dealing with activity coefficients in the molarity scale y_A, we use the following expression for the logarithmic derivative of the activity coefficient of component A with respect to the number density (or molar concentration) of component A:

$$y_{AA} = \left(\frac{\partial \ln y_A}{\partial \ln \rho_A}\right)_{p,T} = \frac{\rho_A(G_{AB} - G_{AA})}{1 + \rho_A(G_{AA} - G_{AB})} \tag{2.12}$$

While this theory is incredibly useful, its practical application in simulations suffers from one major problem: It is only exact for the RDFs in the grand canonical ensemble. Although it is possible to simulate this ensemble directly using e.g. grand canonical Monte Carlo, this method is severely limited in its application to low densities since the molecule insertion moves during the simulation need to have a reasonably high probability.[67] Instead, it is common practice to simulate the system of interest in the NVT or NpT ensembles. The obtained RDFs are then empirically corrected to account for the differences between the ensembles outside of the thermodynamic limit. The simplest correction is to integrate eq. 2.5 up to a finite distance, and assume that $g_{ij} = 1$ for larger distances. This way the divergent integral is assigned a finite value. Other ad-hoc corrections are the shift or division of the whole RDF by the difference of the long-range limit of g_{ij} from unity. A more sophisticated method uses the argument that, in a closed system, a local excess of particles causes a depletion elsewhere.[68] Recent progress has been made by combining this correction with a different one, which resulted in the best convergence of KBIs obtained from RDFs so far.[69]

In this work we use a different approach developed by Cortes-Huerto et al.,[70] which forgoes the use of RDFs and uses the relationship between the KBIs and the particle number fluctuations:

$$G_{ij} = V\left(\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right) \tag{2.13}$$

V is the volume of the system, $\langle N\rangle$ are the average particle numbers, and δ_{ij} is the Kronecker delta. This relation is also only valid for an open system. In practice, we calculate the fluctuations in a subvolume V inside the simulation box with fixed N and total volume V_0. The KBIs of the subsystem are then a function of both V and V_0:

$$G_{ij}(V, V_0) = V\left(\frac{\langle N_i N_j\rangle_{V,V_0} - \langle N_i\rangle_{V,V_0}\langle N_j\rangle_{V,V_0}}{\langle N_i\rangle_{V,V_0}\langle N_j\rangle_{V,V_0}} - \frac{\delta_{ij}}{\langle N_i\rangle_{V,V_0}}\right) \tag{2.14}$$

All the particle number averages now depend on both the volume of the system and that of the selected subbox. The authors then derive a relation between the KBIs in the subbox and in the thermodynamic limit G_{ij}^{∞}:

$$G_{ij}(V, V_0) = G_{ij}^{\infty}\left(1 - \frac{V}{V_0}\right) - \frac{V\delta_{ij}}{V_0\rho_i} + \frac{\alpha_{ij}}{V^{\frac{1}{3}}} \tag{2.15}$$

There are two kinds of effects that cause the differences between eq. 2.13 and eq. 2.14: Ensemble effects due to the fixed number of particles in the closed system, and effects resulting from the finite size of the subvolume V. The corrections for ensemble effects in eq. 2.15 scale with the ratio $\frac{V}{V_0}$, and the finite size correction is proportional to the unknown parameter α. By renaming

$$\lambda = \left(\frac{V}{V_0}\right)^{\frac{1}{3}} \tag{2.16}$$

and multiplying eq. 2.15 with λ, we obtain

$$\lambda G_{ij}(\lambda) = \lambda G_{ij}^{\infty}\left(1-\lambda^3\right) - \lambda^4 \frac{\delta_{ij}}{\rho_i} + \frac{\alpha_{ij}}{V_0^{\frac{1}{3}}} \tag{2.17}$$

In the region of $\lambda^3 \approx 0$ this reduces to

$$\lambda G_{ij}(\lambda) = \lambda G_{ij}^{\infty} + \frac{\alpha_{ij}}{V_0^{\frac{1}{3}}} \tag{2.18}$$

which is linear in λ. We can now calculate $G_{ij}(\lambda)$ from one simulation with volume V_0 by analyzing subboxes with different volumes. From the intercept of the linear fit of $\lambda G_{ij}(\lambda)$ we obtain α, and the slope gives us the KBI in the thermodynamic limit G_{ij}^{∞}. As an example, fig. 2.1 shows $\lambda G_{ij}(\lambda)$ directly calculated from eq. 2.14, and the function in eq. 2.17 using the parameters obtained from the linear fit.

In our implementation of this method, the KBIs of multiple subvolumes are calculated and then averaged for each λ value. The placement of the subboxes is either random or on a regular grid, where the subboxes can overlap and wrap around the periodic boundaries of the system. For all results in this work, we use the grid method with 512 subboxes per λ value, although the subbox placement does not influence the KBIs within the uncertainties as long as the system is sufficiently sampled. The λ range for the linear fit needs to be adjusted depending on the total system size. For a box with edge lengths of 4 nm we use $0.1 \leq \lambda \leq 0.2$ for aqueous urea solutions and $0.1 \leq \lambda \leq 0.3$ for TMAO solutions. Adjustments may be necessary depending on the solute and box size.

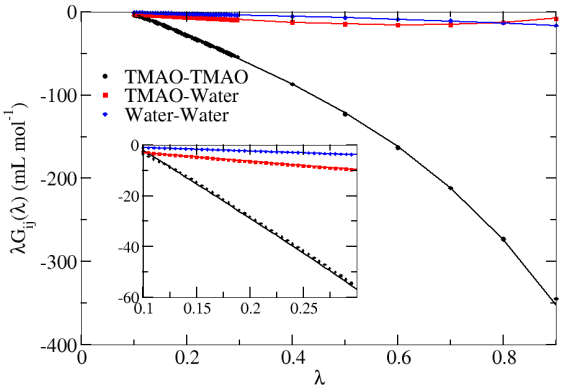


Figure 2.1.: Example for the KBI calculation method in ref. 70: $\lambda G_{ij}(\lambda)$ of the three possible combinations of species in a 2 mol/L aqueous TMAO solution (results taken from section 4.1). The symbols are the values calculated from the fluctuations (eq. 2.14), and the lines are eq. 2.17. G_{ij}^{∞} and α_{ij} are determined from a linear fit in the range $0.1 \leq \lambda \leq 0.3$ (the range shown in the inset). For each λ, eq. 2.14 was evaluated for 512 subboxes placed on a grid, and then the average over the subboxes was taken.

2.4 Free energy sampling methods

2.4.1 Umbrella sampling

We are often interested in the free energy F of a system. It is defined in statistical thermodynamics as

$$F = -\beta^{-1} \ln Q(N, V, T) \tag{2.19}$$

where $\beta = (\mathrm{k}T)^{-1}$, k is the Boltzmann constant, T the temperature, Q the canonical partition function, and N the number of particles in the system. Unfortunately, Q is an integral over the $6N$-dimensional phase space, which is impossible to solve even for moderately large N:

$$Q = \frac{1}{N!\, h^{3N}} \int \mathrm{d}^{3N}\mathbf{r}\, \mathrm{d}^{3N}\mathbf{p}\, \exp\left(-\beta H(\mathbf{r}, \mathbf{p})\right) \tag{2.20}$$

with the Planck constant h, the Hamiltonian H, and the $3N$-dimensional position and momentum vectors $\mathbf{r}$ and $\mathbf{p}$. Here we have shown the partition function for a system of indistinguishable particles, which is the origin of the prefactor $N!^{-1}$ (which would be different for other system compositions).

Although the free energy is not directly accessible, it is possible to calculate differences in free energy along a chosen reaction coordinate, called the potential of mean force (PMF). Ref. 71 gives an introduction to free energy methods in general, and umbrella sampling in particular, which we will summarize in the following section. The partition function can be written as a function of any reaction coordinate ξ:

$$Q(\xi) = \frac{\int \mathrm{d}^{3N}\mathbf{r}\, \delta(\xi'(\mathbf{r}) - \xi)\, \exp(-\beta U(\mathbf{r}))}{\int \mathrm{d}^{3N}\mathbf{r}\, \exp(-\beta U(\mathbf{r}))} \tag{2.21}$$

δ is the Dirac delta distribution. Here the momenta $\mathbf{p}$ were integrated out, and all prefactors were discarded, since they are an additive constant in the free energy and we only care about differences along ξ. U is the potential energy, which is assumed to only depend on the positions. In principle, one could run a regular MD simulation to sample the whole range of ξ and calculate the free energy differences from the probability distribution of ξ. In practice, this approach would require extremely long simulations to ensure sufficient sampling of free energy barriers along the reaction coordinate. In umbrella sampling, which was developed by Torrie and Valleau,[72] an artificial bias potential $w(\xi)$ is added to the unbiased potential energy $U^{\mathrm{u}}(\mathbf{r})$ of the system:

$$U^{\mathrm{b}}(\mathbf{r}) = U^{\mathrm{u}}(\mathbf{r}) + w_i(\xi) \tag{2.22}$$

The index i refers to one specific bias potential, which samples one region of ξ space, usually called a 'window'. The most common bias potential is harmonic, which we use throughout this work:

$$w_i(\xi) = \frac{\mathrm{K}}{2}(\xi_i - \xi)^2 \tag{2.23}$$

ξ_i is the equilibrium position for the reaction coordinate in window i, and K is the force constant.
The probability distribution of ξ in a window is a biased distribution $P_i^{\mathrm{b}}(\xi)$. The unbiased free energy of this window can be calculated except for the free energy term F_i, which is independent of ξ (see ref. 71 for details):

$$F_i(\xi) = -\beta^{-1} \ln P_i^{\mathrm{b}}(\xi) - w_i(\xi) + F_i \tag{2.24}$$

If the range of ξ is large enough to require simulations of multiple windows, the free energy profiles cannot easily be connected, since poor sampling at the borders of the windows leads to large fluctuations of $F_i(\xi)$ far away from the equilibrium position ξ_i. One solution to this problem is the weighted histogram analysis (WHAM) method,[73] which introduces weights $p_i(\xi)$ by which the individual unbiased $P_i^{\mathrm{u}}(\xi)$ are scaled to minimize the total variance of the global unbiased probability $P^{\mathrm{u}}(\xi)$.

$$P^{\mathrm{u}}(\xi) = \sum_{i=1}^{N_{\mathrm{windows}}} p_i(\xi)\, P_i^{\mathrm{u}}(\xi) \tag{2.25}$$

From the weights $p_i(\xi)$ the constants F_i can be obtained. But since $P_i^{\mathrm{u}}(\xi)$ depends on F_i, the $p_i(\xi)$ have to be optimized iteratively. Finally, the PMF is calculated:

$$F(\xi) = -\beta^{-1} \ln P^{\mathrm{u}}(\xi) \tag{2.26}$$

We exclusively use the WHAM implementation in `GROMACS`, which also includes error estimates for the PMF.[74]

2.4.2 Temperature replica-exchange MD

When we are interested in sampling high-energy regions of phase space, but there is no simple reaction coordinate, other methods have to be used. One solution is the temperature replica-exchange scheme, which was initially adapted for the use in molecular dynamics simulations by Sugita and Okamoto.[75] Multiple copies of the same system, the 'replicas', are simulated in parallel at different temperatures in the NVT or NPT ensemble. For a certain number of simulation steps, the equations of motion of all replicas are propagated independently from each other. Then, at fixed intervals, each replica can exchange its temperature with one replica with the next lower or higher temperature. By running this simulation scheme for a sufficiently long time, a replica stuck in a local free energy minimum will be exchanged to higher and higher temperatures until the free energy barriers can be crossed with significant probability. In this way the averages and energy landscapes at a certain temperature can be calculated even if the trajectory at one temperature is not continuous and made of short segments of simulations of different replicas.

Next, we will briefly review the details of this method as laid out in ref. 75. A state X of a number of M replicas is defined as the sets of all positions and momenta for each replica, called x_m^i, where the superscript i is the index of the replica, and m is the index of the current temperature of replica i, and there exists a (bijective) map between the two ($m = m(i)$ and $i = i(m)$).

$$x_m^i = (\mathbf{r}^i, \mathbf{p}^i)_m \tag{2.27}$$

Here $\mathbf{r}$ and $\mathbf{p}$ are the $3N$-dimensional position and momentum vectors. A state X is then defined by the M points in phase space x_m^i, each of which is assigned to a distinct temperature:

$$X = (\dots, x_m^i, \dots) \tag{2.28}$$

As long as the replicas do not interact with each other, the probability P of finding the

system in state X is just the product of the canonical phase space densities of all replicas:

$$P(X) = \exp\left\{-\sum_{i=1}^{M} \beta_m H(\mathbf{r}^i, \mathbf{p}^i)\right\} \tag{2.29}$$

At an exchange step, two replicas can exchange temperatures. The state X becomes X':

$$X = (\dots, x_m^i, x_n^j, \dots) \tag{2.30}$$

$$X' = (\dots, x_m^{j'}, x_n^{i'}, \dots) \tag{2.31}$$

Since the exchanged replicas now have different temperatures, the instantaneous momenta at the time of the exchange are scaled to be consistent with the kinetic energy at the new temperature:

$$\mathbf{p}^{i'} = \sqrt{\frac{T_n}{T_m}}\mathbf{p}^i \tag{2.32}$$

$$\mathbf{p}^{j'} = \sqrt{\frac{T_m}{T_n}}\mathbf{p}^j \tag{2.33}$$

Next it will be described how the probability for an exchange step is determined. The detailed balance condition is assumed, i.e. the exchange probabilities w are that of a system in equilibrium:

$$P(X)\, w(X \to X') = P(X')\, w(X' \to X) \tag{2.34}$$

From eq. 2.29, the ratio of the exchange probability to its reverse process can be calculated:

$$\frac{w(X \to X')}{w(X' \to X)} = \frac{P(X')}{P(X)}$$

$$= \exp\{-\Delta\} \tag{2.35}$$

$$\Delta = (\beta_n - \beta_m)\left(U(\mathbf{r}^i) - U(\mathbf{r}^j)\right) \tag{2.36}$$

Here we used the facts that all contributions from replicas not involved in the exchange cancel out, and that the kinetic energy contributions of the exchanging replicas cancel due to the momentum scaling in eqs. 2.32 and 2.33. There are multiple possible choices for $w(X \to X')$ that satisfy this condition. The exchange probability used here is the well-known Metropolis criterion from the Monte Carlo simulation scheme with the same name:[76]

$$w(X \to X') = \begin{cases} 1 & \text{if } \Delta \leq 0 \\ \exp\{-\Delta\} & \text{if } \Delta > 0 \end{cases} \tag{2.37}$$

This means that if the replica with a lower temperature has a higher potential energy U, the exchange step is always accepted, and there is a nonzero probability for the exchange otherwise. In this work, we use the `GROMACS` implementation of the replica-exchange method, where a replica can only be involved in one exchange attempt at a time. This means that in alternating exchange steps, a replica is either the low or high temperature in an exchange attempt. For sufficiently high exchange probabilities, the potential energy distributions of neighboring temperatures need to have a reasonably large overlap. We use the method

from ref. 77 to estimate the temperatures that lead to exchange probabilities in the range $0.1 - 0.3$.

2.5 *Ab initio* molecular dynamics

Ideally, we want our force field molecular dynamics simulations to reproduce results obtained from quantum chemistry. We will therefore use *ab initio* molecular dynamics (aiMD) simulations performed by Dr. Sho Imoto and Jan Noetzel (refs. 78 and 79) as reference data. Whenever we refer to aiMD in this work, we mean Born-Oppenheimer dynamics simulations performed with the `QUICKSTEP` module[80] of the `CP2K` program package.[81] The level of theory is density functional theory (DFT) in the generalized gradient approximation (GGA) using the revised Perdew-Burke-Ernzerhof (RPBE) functional[82] in combination with the D3 dispersion correction.[83] A detailed write-up of the technical details used in the simulations of a single TMAO in 107 water can be found in ref. 78. The unpublished data for one urea in 110 water[79] follows the same procedure.

2.6 EC-RISM

In this work, we use results obtained from embedded cluster reference interaction site model (EC-RISM) calculations, which were performed by Patrick Kibies (see refs. 78 and 84). This method combines the reference interaction site model in three-dimensional space (3D-RISM) with a quantum-chemical treatment of a solute molecule to calculate equilibrium properties, e.g. solvent distributions, solvation free energies, and pKa values with chemical accuracy.[85] One parameter in EC-RISM integral equation theory is the solvent susceptibility, which can in theory be determined directly from molecular dynamics simulations of the pure solvent. However, due to technical reasons, this is currently not feasible in 3D-RISM, and the susceptibilities were instead determined from 1D-RISM using the radial distribution functions from MD simulations. Specifically, the results from MD simulations are used in the bridge function term, which is necessary for the closure of the integral equation. The EC-RISM calculation itself is a self-consistent quantum chemistry calculation with the solvent being represented as charges on a grid. The initial charges from 3D-RISM polarize the solute, which in turn changes the optimal solvent charges. This cycle is repeated until self-consistency of the solute and solvent geometries.

All EC-RISM results in this work use solvent susceptibilities extracted from simulations of SPC/E water[86] at 298.15 K at various pressures. Currently, the use of four-point water models (e.g. TIP4P/2005) is not possible for technical reasons. The level of theory was B3LYP/6-311+G(d,p)/EC-RISM. Dipole moments and partial charges were calculated by fitting the partial charges of the solute to the electrostatic potential of the optimized geometry using the ChelpG method.[87] Further technical details of the EC-RISM calculations for TMAO can be found in ref. 78. The unpublished results for urea[84] were calculated analogously.

Chapter 3

Water at hydrophobic interfaces

In this chapter, we perform a detailed molecular-level analysis of the water structure at an alkane self-assembled monolayer (SAM) with pressure as the variable. We present previously published parameters for measuring the local and intermediate-range order of water at interfaces.[88] For a general introduction to this topic we refer to chapter 1.

3.1 Methods

3.1.1 Simulation Details

The simulations of the SAM/water interface are identical to previous unpublished[89] and published[23] work. We modeled the SAM after a SAM of octadecyltrichlorosilane (OTS) on silica, where the silica was omitted since its influence on the investigated properties is negligible. Instead, the first carbon atoms of 36 octadecane chains were fixed via a freeze group on a hexagonal grid with a lattice constant of 5.06 Å, which is the distance between the oxygens of the interfacial silanol-OH groups on silica.[90] The AMBER03[91] LJ-parameters were used in conjunction with the OPLS-AA[92] charges, which introduced the weak electrostatic interactions not present in the AMBER03 force field. However, we later confirmed that the small partial charges do not measurably influence the properties studied in this work. The SAM was solvated in 2096 TIP4P/2005 water molecules. The rectangular box had a fixed xy-base with measurements of $3.036 \cdot 2.629\,\mathrm{nm}^2$, and the box length in z direction varied between 10.1 nm at 1 bar (snapshot in fig. 3.1) and 8.3 nm at 10 kbar. Semiisotropic Berendsen pressure coupling[59] with a time constant of 1 ps was used to scale the z component of the box, mainly because using the Parrinello-Rahman barostat[60] caused unexpected behavior in combination with freeze groups in `GROMACS` 4.6.5. The lengths of all bonds involving hydrogen atoms were constrained, which enabled the use of a 1.5 ps time step for integrating the equations of motion. This system was simulated for 210 ns at 300 K and the seven pressures 1 bar, 1 kbar, 2 kbar, 3 kbar, 4 kbar, 5 kbar, and 10 kbar. The SAM/water interface was also simulated for 15 ns at 250 K, 350 K, and 400 K at 1 bar. The trajectory of a simulation of the water/vacuum interface from previously unpublished work[89] is also analyzed in this work. The system consists of 1073 water molecules in a $3.036 \cdot 2.629 \cdot 4\,\mathrm{nm}^3$ slab, which was equilibrated at 1 bar and 300 K. The box was then extended by adding a 6 nm vacuum phase in the z direction. This system was simulated in the NVT ensemble for 210 ns. For the contact angle calculation we also simulated the SAM in vacuum for 150 ns. Additionally, two bulk ice phases were simulated using anisotropic Parrinello-Rahman pressure coupling with a 3 ps time constant. An ice Ih crystal made of 432 water molecules was simulated for 5 ns at 200 K. Using the `genice`

program,[93] an ice VI crystal with 640 water molecules was generated and subsequently simulated for 5 ns at 25 kbar.

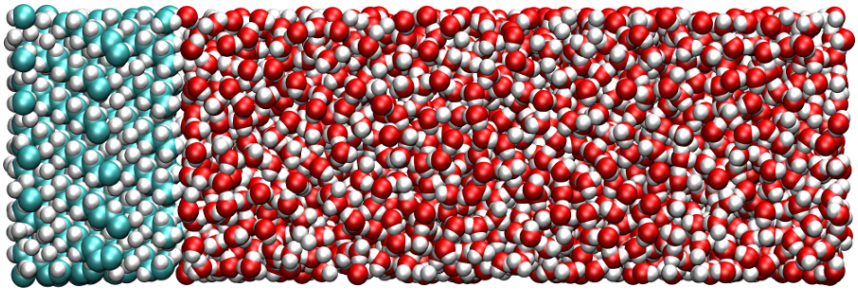


Figure 3.1.: Simulation snapshots of the SAM/water interface at 1 bar. The snapshot was visualized using VMD.[94]

3.1.2 Contact angle

Here, we briefly show that our model SAM reproduces the experimental contact angle of water at an OTS SAM on silica. We use Young's equation (eq. 3.1) to calculate the contact angle θ from the surface tensions γ of the SAM-vacuum, SAM-water, and water-vacuum systems:

$$\cos\theta = \frac{\gamma_{\text{SAM-vacuum}} - \gamma_{\text{SAM-water}}}{\gamma_{\text{water-vacuum}}} \tag{3.1}$$

We chose to calculate the surface tension from the difference of the lateral and normal pressures (calculated from the virial) in the system integrated over the box length L_z:[67]

$$\gamma = \frac{1}{2}\int_0^{L_z} \mathrm{d}z \left(p_{zz}(z) - \frac{p_{xx}(z) + p_{yy}(z)}{2} \right) \tag{3.2}$$

To avoid the calculation of the profile of local virial tensors, we make the assumption that the difference between the lateral and normal pressures not only vanishes in the liquid and gas bulk phases, but also in the anisotropic SAM bulk phase. Additionally, we make the approximation that this difference is constant at the interfaces, which is reasonable if the local pressure difference has a shape similar to that of the LJ-fluid in ref. 95. Then, the surface tension only depends on ensemble averages of the diagonal elements of the pressure:

$$\gamma = \frac{L_z}{2}\left(\langle p_{zz} \rangle - \frac{\langle p_{xx} \rangle + \langle p_{yy} \rangle}{2} \right) \tag{3.3}$$

Our results for the SAM-vacuum system is $-192.06\,\text{bar}\cdot\text{nm}$, the value for the water-vacuum surface is $616.15\,\text{bar}\cdot\text{nm}$, and the SAM-water interface has a tension of $30.00\,\text{bar}\cdot\text{nm}$. The resulting contact angle of 111° is in excellent agreement with the experimental value of 112°.[96] We therefore assume that our model is a good approximation of the OTS-SAM on silica even without explicitly modeling the substrate.

3.1.3 Order parameters

In this work, we use parameters to quantify the local order around water molecules. The tetrahedral order parameter q_4 (eq. 3.4) was introduced by Errington and Debenedetti,[18] which is a modification of the initial order parameter S_g by Chau and Hardwick.[17]

$$q_4(k) = 1 - \frac{3}{8} \sum_{i=1}^{3} \sum_{j=i+1}^{4} \left(\cos(\psi_{ikj}) + \frac{1}{3} \right)^2 \tag{3.4}$$

q_4 quantifies the angular tetrahedral order of the four nearest water molecules around a water molecule k by summing over the squared deviations of the angles ψ between two of the four closest water with the central water k. The double sum includes all possible pair combinations out of the four closest neighbors. For simplification purposes, the positions of the oxygen atoms are used as the positions of the water molecules. q_4 is normalized such that it is zero for a random distribution of positions of the nearest neighbors and one for a perfectly tetrahedral coordination (all $\cos\psi = -\frac{1}{3}$). q_3 (eq. 3.5) is defined almost identically to q_4. The main difference is that the summation runs only over the three closest neighbors. We changed the normalization factor such that $q_3 = 0$ and $q_3 = 1$ refer to random and tetrahedral angle distributions. With q_3 it is possible to investigate the tetrahedrality of the water network when neighbors are missing, i.e. at an interface.

$$q_3(k) = 1 - \frac{3}{4} \sum_{i=1}^{3} \sum_{j=i+1}^{4} \left(\cos(\psi_{ikj}) + \frac{1}{3} \right)^2 \tag{3.5}$$

3.1.4 Rings in the H-bond network

In this work we apply graph theory to study the hydrogen bond network in water. This approach has already been known since the early days of water simulations.[97] In recent years, Matsumoto and coworkers used this method to analyze rings in the hydrogen bond network of supercooled water.[98] We use their method and criterion for detecting rings in the graph.[99] A detailed explanation of the ring detection algorithm can be found in appendix A.

After all rings are detected, we go back from the abstract graph representation of the hydrogen bond network to the spatial coordinate representation to analyze the geometry of the rings. We use the following definition from ref. 98: A ring has a twist χ_R (eq. 3.6),

$$\chi_R = \frac{1}{N} \left| \sum_{i=0}^{N-1} \sin(3\,\theta_{ijkl}) \right| \tag{3.6}$$

where N is the size of the ring, and $ijkl$ are the internal water indices of this ring with $j = (i+1) \bmod N$, $k = (i+2) \bmod N$, and $l = (i+3) \bmod N$. θ_{ijkl} is the torsional angle around the H-bond between the water molecules j and k. Figure 3.2 shows the cases where a single term in eq. 3.6 is zero. The sum over all terms vanishes when all water molecules in the ring have a perfectly tetrahedral H-bond geometry, e.g. in six-membered rings in chair- or boat conformation. Alternatively, the nonzero twists can cancel out even in asymmetric ring geometries.

Figure 3.2.: View projected along a H-bond for orientations where the torsion $\sin(3\,\theta_{ijkl})$ is zero. The lines are H-bonds assuming that each water molecule has four H-bonds. The torsion is zero for both staggered (a) and eclipsed (b) conformations of the H-bonds and takes nonzero values between -1 and 1 otherwise.

3.2 Results

3.2.1 First solvation shell properties

First we will analyze the orientations of water molecules as a function of the distance from the alkane SAM/water interface. We define the angle ξ as the angle between the water dipole moment and the surface normal vector which points towards the water phase. τ is the angle between the water O-H bonds and the surface normal. Figure 3.3 shows the 2D distributions of ξ and τ normalized to the bulk distribution for pressures from 1 bar to 5 kbar. The distance from the interface increases from the top to bottom rows. Throughout all distributions, the most probable orientation is at $(\xi, \tau) = (0, 0)$, which means that the water molecules are oriented in parallel to the interface. The peaks at $(-1.0, -0.5)$ and $(0.5, -0.5)$ closest to the interface mean that one O-H bond points directly towards the interface. Further away from the interface, no preferred orientation besides parallel orientations at $(0, 0)$ are found. In the next layer, two maxima at $(-0.5, 0.5)$ and $(1.0, 0.5)$ emerge. Here, the orientations from the closest layer are inverted such that one O-H bond points away from the interface. Qualitatively, the distributions at all pressures from 1 bar to 5 kbar agree with each other. Only the probabilities of the preferred orientations increase significantly with pressure, which signifies that the water molecules are more strongly ordered at high pressures. The orientations we found are not only consistent with existing literature at normal pressure,[14,15] but also with the molecular orientations in a crystal of ice Ih assuming the surface normal is the c axis of the ice crystal. This fact hints at ice-like ordering of water at hydrophobic interfaces which is amplified in a high pressure environment.

In order to further characterize the properties of water at the SAM/water interface, the properties of the first solvation shell of water are analyzed. Figure 3.4 contains laterally averaged profiles of multiple first-shell properties at the interface. Panel (a) are the mass densities of the SAM and water components at pressures up to 10 kbar. From the profiles and the bulk densities in the inset it is clear that the SAM is compressed very little when increasing the pressure from 1 bar to 10 kbar. However, the density of water at the first maximum at the interface increases significantly with this pressure change. Also, the distance between SAM and the water maximum decreases, which is a result already found in previous work,[23] where it was shown that the hydrophobic depletion layer at this

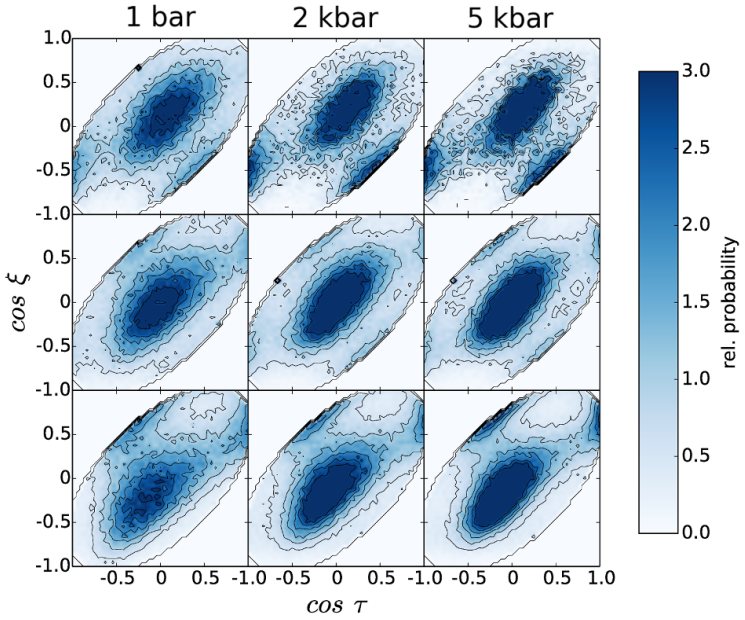


Figure 3.3.: 2D angle orientations at the SAM/water interface. ξ is the angle between the water dipole moment and the surface normal vector (pointing towards the water phase). τ is the angle between the water O-H bonds and the surface normal. The distributions shown here are normalized to the bulk angle distributions. The distributions are shown for pressures increasing from 1 bar (left) to 5 kbar (right). The distance from the interface increases from the top to the bottom panels. At 1 bar, the panels are slices of 0.5 Å. The slice thickness was scaled with the compression of the box in z direction, leading to 0.465 Å slices at 2 kbar and 0.436 Å at 10 kbar. The slices in the middle and bottom rows have the coordinate $z = 0$ in fig. 3.4 (dashed line) in common.
Reproduced from ref. 88 with permission from the PCCP Owner Societies.

interface is significantly compressible. In panel (b) we show that the number of water molecules inside the first hydration shell decreases below four in similar fashion for all pressures, even though the bulk values differ significantly between 5.3 at 1 bar and 7.0 at 10 kbar. In contrast, the number of hydrogen bonds per water falls off at the interface and is almost identical in bulk and at the interface over the whole pressure range. Only a minor increase at 10 kbar can be observed compared to normal pressure. In conjunction with the number of water molecules in the solvation shell, this means that the majority of the water that is added upon compression is not part of the H-bond network but mostly fills the tetrahedral voids, in agreement with literature.[64] Directly at the interface, the number of accepted H-bonds is greater than the number of donated H-bonds. This means there are dangling H-bonds at the interface, which agrees with the angle distributions in fig. 3.3. The profile of the tetrahedral order parameter q_4 (eq. 3.4) is shown in panel (c). When the pressure is increased, the bulk value of q_4 decreases. This can be explained by the fact that the water molecules fill the tetrahedral voids in the H-bond network as shown in panel (b). Therefore, the closest four neighbors are less likely to be tetrahedrally coordinated at higher pressures. When approaching the interface, q_4 decreases monotonically at all pressures, since there is no fourth neighbor available at the interface, thereby making a tetrahedral coordination impossible. In order to be able to quantify the tetrahedrality of the remaining three closest neighbors we define the trigonal-pyramidal order parameter q_3 (eq. 3.5). The qualitative behavior of the bulk values of q_3 in panel (d) is identical to q_4.

For normal pressure, there is no noteworthy difference between the two order parameters at the interface either. However, at pressures above 2 kbar, q_3 exhibits a local maximum at $z = -0.1$ which increases with pressure up to 5 kbar and then decreases when going up to 10 kbar. At this position, q_3 of all elevated pressures is higher than the value at 1 bar. This means that in this layer close to the interface, which has low water density (see a), the remaining water molecules form a layer that has a tetrahedral order more resistant to compression than the bulk phase. The interpretation of this result is difficult without more detailed information about the water structure in this system, which will be analyzed in the remainder of this chapter.

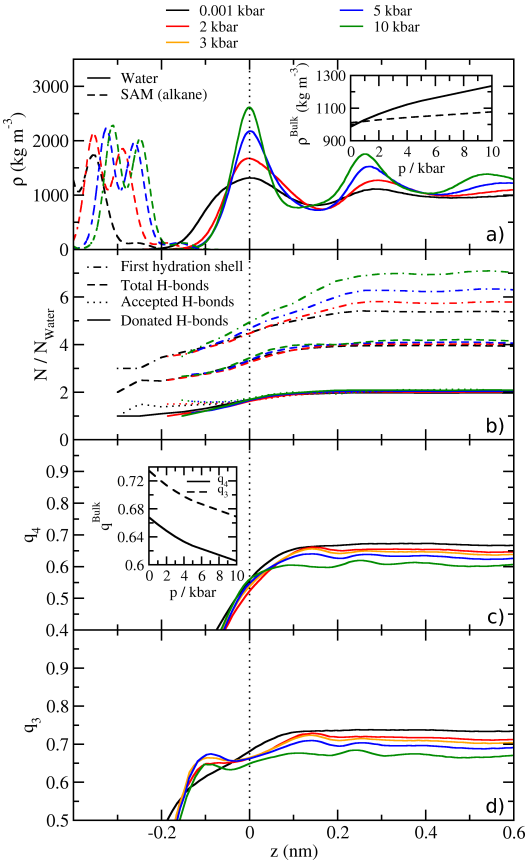


Figure 3.4.: Laterally averaged properties of water at the SAM/water interface. (a) The mass densities of the SAM (dashed lines) and water (full lines). The dotted line is a guide to the eye at the position of the first water maximum. The inset displays the bulk densities of both components as a function of pressure. (b) The number of water molecules in the first hydration shell around water (dotted and dashed lines), defined as the number of water molecules closer than the minimum of the bulk water-water RDF at 1 bar (0.322 nm). Other data sets are the number of accepted (dotted), donated (dashed), and total (full lines) number of hydrogen bonds per water molecule. (c) The tetrahedral order parameter[18] q_4 (eq. 3.4). (d) The trigonal-pyramidal order parameter q_3 as defined in eq. 3.5. The inset in (c) shows bulk values for q_4 and q_3.

3.2.2 Hydrogen bond rings

After investigating the first-shell solvation properties at the SAM/water interface, the structure of water at an intermediate range will be analyzed. In this section, we describe

the topology of the H-bond network, specifically rings formed by H-bonded water molecules as defined in sec. 3.1.4. First the results for four- to eight-membered H-bond rings at the vacuum/water interface are presented, which serve as a reference for the SAM/water results. Figure 3.5 shows that the density profiles of the rings of all sizes assume a maximum in the vicinity of the interface. These maxima are not correlated to the mass density of the water phase. For four- to seven-membered rings, an increase of the absolute ring densities is observed over the whole coordinate range. The eight-membered ring density profile takes values between six- and seven-rings in bulk and decreases to the lowest density out of all observed ring types close to the interface. In bulk, the densities range from 4.6 four-membered rings per nm^3 and 27.4 nm^3 seven-rings. Although significantly larger ring sizes can also be observed, their data contained no meaningful information and is not shown. Next, the geometry of the rings is analyzed. As a measure we use the twist of a H-bond ring χ_R as defined in eq. 3.6, which determines how distorted the ring geometry is. Therefore, the twist is a measure of the 'tension' of the H-bonds in the ring. In fig. 3.6, the distribution of χ_R is shown for multiple layers of water close to the interface. The region we investigated is much closer to the vacuum phase than the ring density maximum and has a mass density significantly lower than the bulk. The twist distributions of four-membered rings are very narrow and centered around $\chi_R = 0$. Only a very minor decrease of the twist is observed when going from the interfacial region to the bulk phase. This is due to the limited flexibility of the H-bonds before a four-membered ring breaks. We observe significantly broader distributions for five-rings with a greater effect of the interface on the twist. This means that five-rings have a greater flexibility and thus a higher tendency to adjust their geometry at the interface towards tetrahedrally coordinated H-bonds. With an increase in ring size to six, the twist distribution is again broadened, as is the decrease of the twist of interfacial six-rings. The four layers span a region of 2 Å, but over this range no change in χ_R can be determined within the uncertainty. However, the bulk distribution is distinctly shifted to larger twists. When going to even larger ring sizes, there is no measurable correlation between position and ring twist for seven- and eight-membered rings. Only further broadening of the distributions takes place, with the eight-rings having a slightly narrower twist distribution than the seven-rings. This analysis is mostly inconclusive because the instantaneous vacuum/water interface can be very different from the time-averaged interface due to capillary waves.[100]

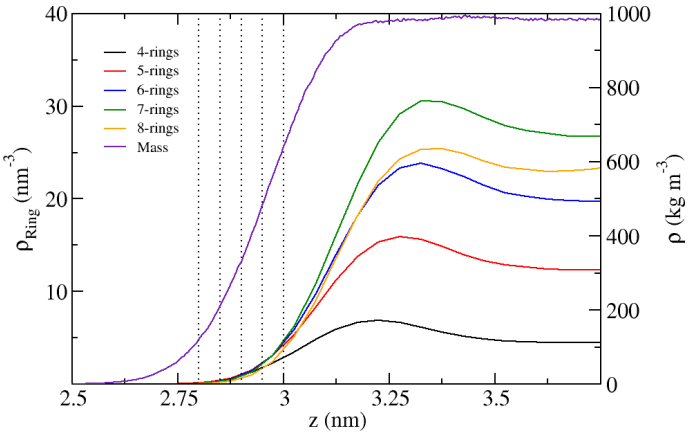


Figure 3.5.: Profiles of H-bond ring densities for ring sizes ranging from 4 to 8 at the vacuum/water interface with the mass density profile for reference. The position of a ring is determined as the center of mass of all water molecules in the ring. The dotted lines mark the 0.5 Å slices for further analysis (see below).

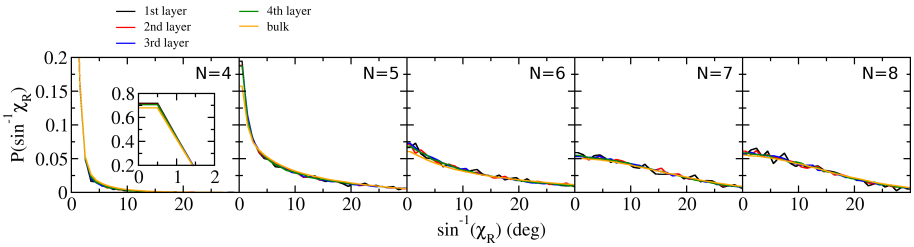


Figure 3.6.: Distributions of the ring twist χ_R of four- to eight-membered rings ($N = 4 - 8$) in the water hydrogen bond network at the vacuum/water interface. The layers (see fig. 3.5) are increasing in distance from the interface in steps of 0.5 Å. The inset shows the maximum of the four-membered ring twist. Only the region from $0 - 30°$ is shown, and the region up to 90° was omitted.

Now we will apply the same methods to the SAM/water interface while also introducing pressure as an additional variable. All ring density profiles from four- to eight-membered rings for pressures between 1 bar and 10 kbar are displayed in fig. 3.7 along with the mass densities. We observe a much more distinct layer structure in all density profiles compared to the vacuum/water interface. The water density oscillates with decreasing amplitude for increasing distance from the interface. The ring densities show the same structure, except the maxima are shifted such that they lie in between the mass density maxima. This tells us that there is a trivial component to the ring densities caused by an increase in overall water density. If the local water density is higher, the total number of H-bonds increases even if the number of H-bonds per water stays constant (as shown previously in fig. 3.4). What we observe here are rings in the H-bond network forming randomly with a higher probability simply due to the high concentration of H-bonds in the density maxima. Therefore, the probability of finding a ring with its center of mass between the two water density maxima is increased. From this we can infer that there is a preference of rings to align perpendicularly to the interface, because this is the direction in which the water density is structured. This finding is qualitatively true for all ring sizes that we analyzed. Additionally, the ring density increases continuously with pressure at the interface as well as in bulk (see insets in fig. 3.7), which is also an expected result considering the previous argument. It is difficult to quantify how the probability of forming rings increases with density, and neither the relative nor the absolute ring densities change monotonously with the ring size. Most of the rings in the high density regions are still at least $1 - 2$ Å away from the interface. Next, we focus the discussion on the ring twist χ_R at the low-density interfacial region. Figure 3.8 contains the twist distributions of four 0.5 Å layers starting at approximately $z = -0.1$ nm, and the bulk phase for varying pressures. The narrow distributions of the four-ring twists are almost identical to the twists at the vacuum/water interface and barely change with the distance from the interface. For five- and six-rings, the twists are significantly lower at the interface than in the bulk. The same effect is still present but very minor for seven- and eight-rings. By comparing the distributions over the pressure range up to 10 kbar, we find that only five- and six-rings show a significant pressure dependence. The twist of five-rings decreases from 1 bar to 5 kbar, but increases again at 10 kbar. Only the six-membered ring twist decreases steadily with pressure. We conclude that pressure appears to increase the preference for low-twist conformations of six-rings at the SAM/water interface. The fact that this effect is only observed for six-rings leads us to compare these rings to those found in ice Ih, which is exclusively made of six-rings. The goal of the subsequent analysis is to compare the rings at the interface in liquid water and the rings in a single layer of hexagonal ice.

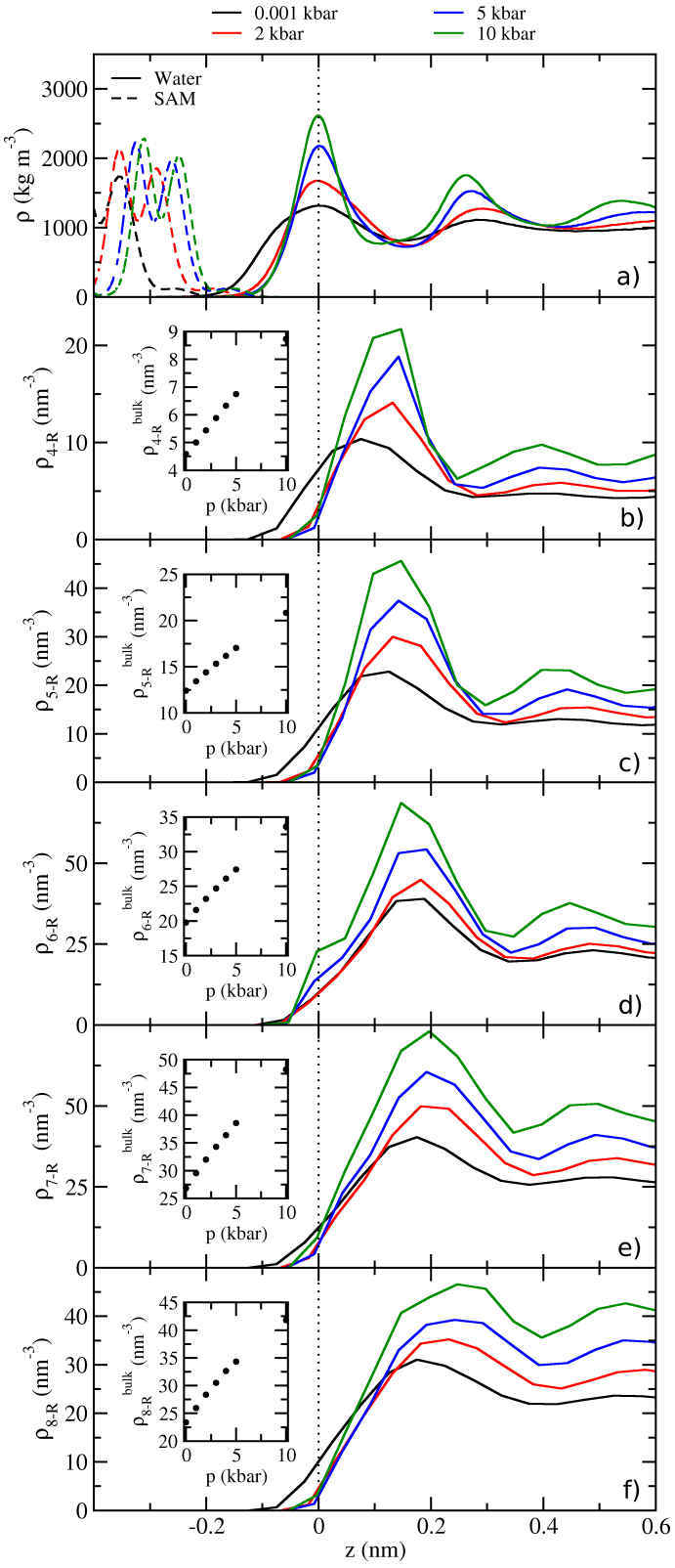
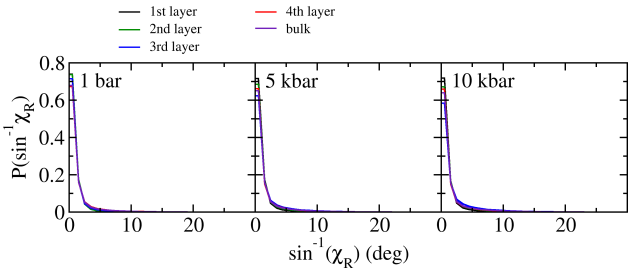
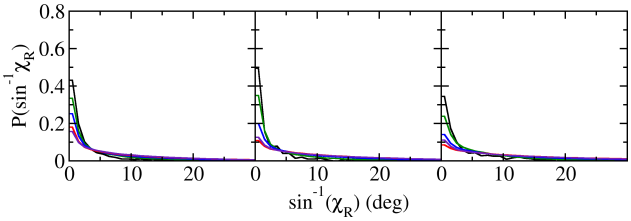


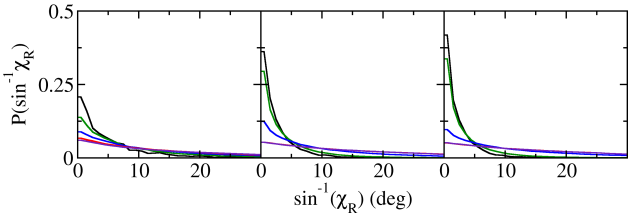
Figure 3.7.: Profiles of H-bond ring densities for ring sizes from 4 (b) to 8 (f) at the SAM/water interface at pressures between 1 bar and 10 kbar with the bulk ring densities in the inset of each panel. (a) is the mass density profile for reference.



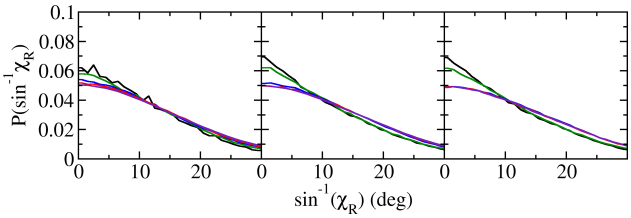
(a) Four-membered rings



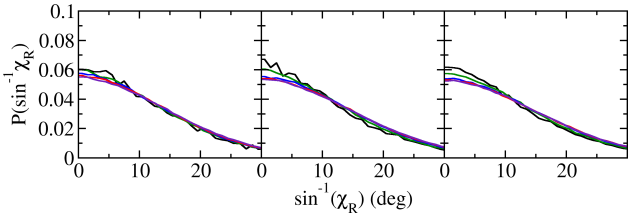
(b) Five-membered rings



(c) Six-membered rings



(d) Seven-membered rings



(e) Eight-membered rings

Figure 3.8.: Distributions of the ring twist χ_R of rings in the water hydrogen bond network. The distance of the layers from the interface increases in steps of 0.5 Å from approximately $z = -0.1$ nm to $z = 0.1$ nm. From left to right, the pressures are 1 bar, 5 kbar, and 10 kbar.

As a first step, we determine the density of six-rings from a simulation of an ice Ih crystal along the c-axis. This axis is chosen because the chair-like six-rings which form layers in ice (fig. 3.9 a) are oriented perpendicularly to the c-axis, and because the rings at the SAM/water interface have the same orientation with respect to the interface, which enables a direct comparison of the rings in the two systems. Ice Ih contains a second type of ring,

which is in a boat-conformation (fig. 3.9 b) and which connects the layers of chair-like rings . Since the two types of rings are spatially separated along the c-axis, we are able to analyze the properties of a single layer of chair-like rings (fig. 3.10). The twist distribution inside this layer of rings is very narrow with 99.9 % of all rings having a twist below 6.5°. From now on we define all rings with a twist below this threshold to be 'ice-like'. Now we can identify these rings with a locally tetrahedral geometry in liquid water. Fig. 3.11 shows a snapshot of a single ice-like ring at the SAM/water interface. Next, the amount of ice-like rings at the interface can be quantified. Instead of determining the absolute density of ice-like rings, we care about the excess density $\rho_{\mathrm{R}}^{\mathrm{ex}}$ of ice-like rings at the interface with respect to the bulk phase:

$$\rho_{\mathrm{R}}^{\mathrm{ex}}(z) = \rho_{\mathrm{R}}(z) \int_0^{\chi_{\mathrm{c}}} [P(\chi_{\mathrm{R}}, z) - P_{\mathrm{bulk}}(\chi_{\mathrm{R}})]\, \mathrm{d}\chi_{\mathrm{R}} \tag{3.7}$$

Here ρ_{R} are the total ring densities, P are the ring twist probability densities, and $\chi_{\mathrm{c}} = 6.5°$ is the upper cutoff for the twist of ice-like rings.

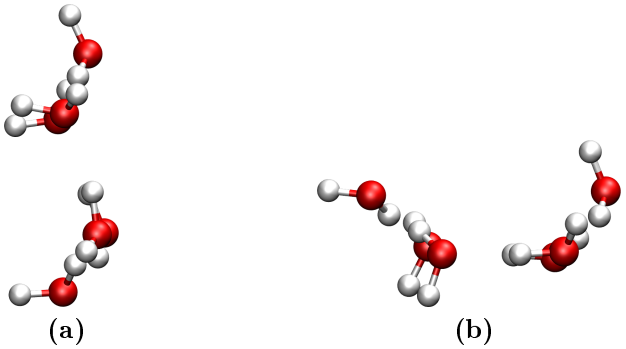


Figure 3.9.: Snapshots of six-membered rings in bulk ice Ih at 200 K and 1 bar. a) A ring in chair conformation perpendicular to the c-axis. b) A ring in boat conformation. These rings connect the layers of chair-like rings in ice Ih.

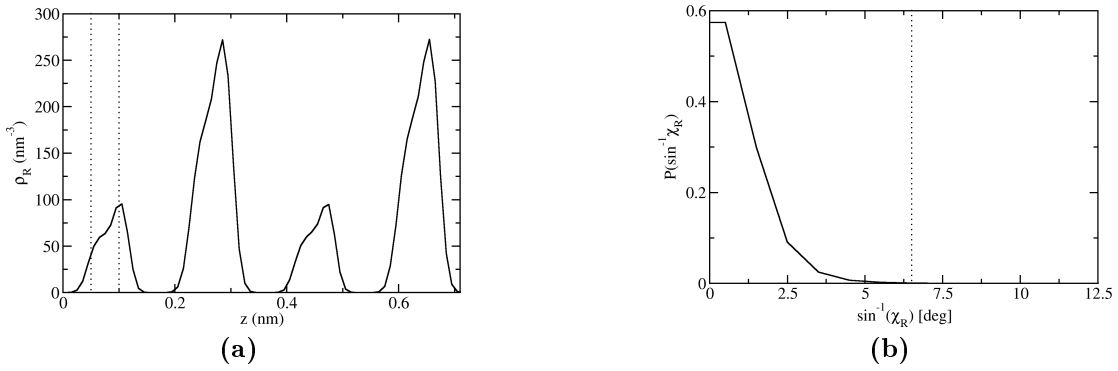


Figure 3.10.: a) Six-membered ring density profile along the c-axis in bulk hexagonal ice Ih at 200 K and 1 bar. The dotted lines mark a layer of rings in chair conformation b) Ring twist distribution of the layer of chair-like rings in a).

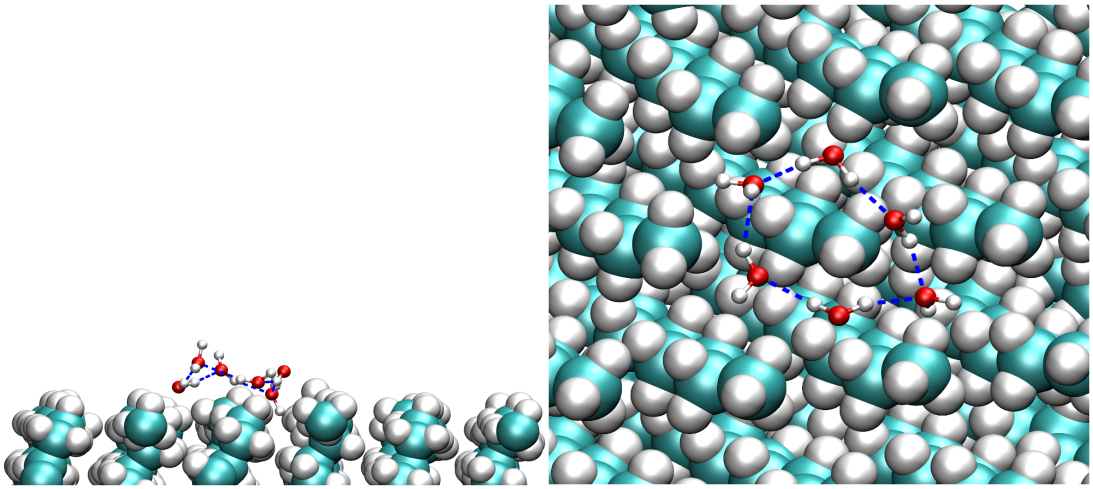


Figure 3.11.: Snapshots of an ice-like (low twist) six-membered ring close to the SAM at 300 K and 1 bar. Blue dashed lines symbolize hydrogen bonds. Reproduced from the supporting information of ref. 88 with permission from the PCCP Owner Societies.

The profiles of $\rho_{\mathrm{R}}^{\mathrm{ex}}$ (fig. 3.12) show that the maximum of the excess ring density is not related to the maximum of the total ring density, but is exactly at the water density peak. At 1 bar, $\rho_{\mathrm{R}}^{\mathrm{ex}}$ is close to zero with a maximum of $0.2\,\mathrm{nm}^{-3}$. However, the ring excess is very sensitive to pressure and increases to $12.2\,\mathrm{nm}^{-3}$ at 10 kbar, which is a significant fraction of the total ring density at that position (45 %). This means that pressure increases the ice-likeness of the six-membered rings at this hydrophobic/water interface. To put the value of $12.2\,\mathrm{nm}^{-3}$ in perspective, the total ring density maximum in a layer of ice Ih is about $95\,\mathrm{nm}^{-3}$. By this measure, the time-averaged water structure becomes significantly ice-like at 10 kbar. It is also interesting to see if there are similarities between the influence of pressure and temperature on these water properties. It becomes apparent (fig. 3.14) that the relation of the properties with pressure and temperature are inverted: At 250 K, water at the SAM/water interface has a higher density, more H-bonds, higher order parameters q_4 and q_3, more six-membered rings, and also a higher ring excess density. However, the temperature effect is relatively weak even at already supercooled temperatures. Before concluding this chapter, it is important to make sure that the pressure-induced structural changes we observe are not due to a phase transition of the water phase. In reality, the phase transition of water to ice VI occurs at 300 K and 10 kbar. In simulations of the TIP4P/2005 water model, this phase transition also exists but is shifted to 25 kbar.[62] However, since ice VI has a tetragonal crystal structure, a perfect crystal contains no six-membered rings. Even when small defects are present, the amount of six-membered rings is insignificant (see fig. 3.13).

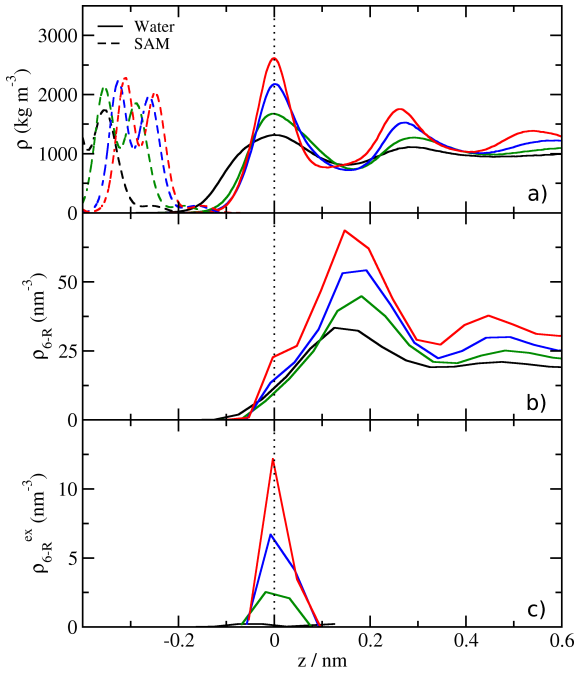


Figure 3.12.: The mass (a) and 6-ring densities (b) are shown again for reference. (c) Excess density of six-membered rings with respect to the bulk phase as defined in eq. 3.7.

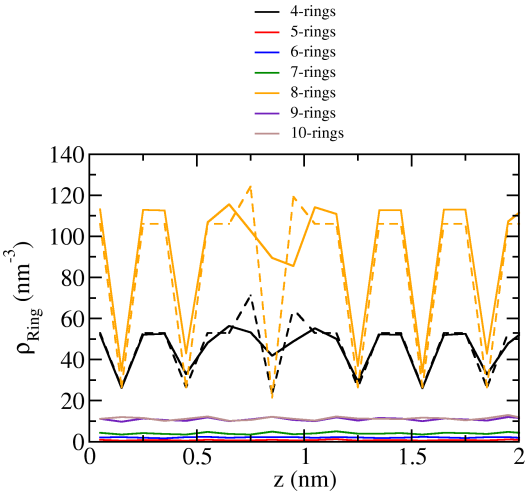


Figure 3.13.: Ring densities in tetragonal ice VI for four- to ten-membered rings. The dashed lines are the profiles of the starting geometry generated with `genice`.[93] The full lines are the profiles of a simulation at 300 K and 25 kbar.

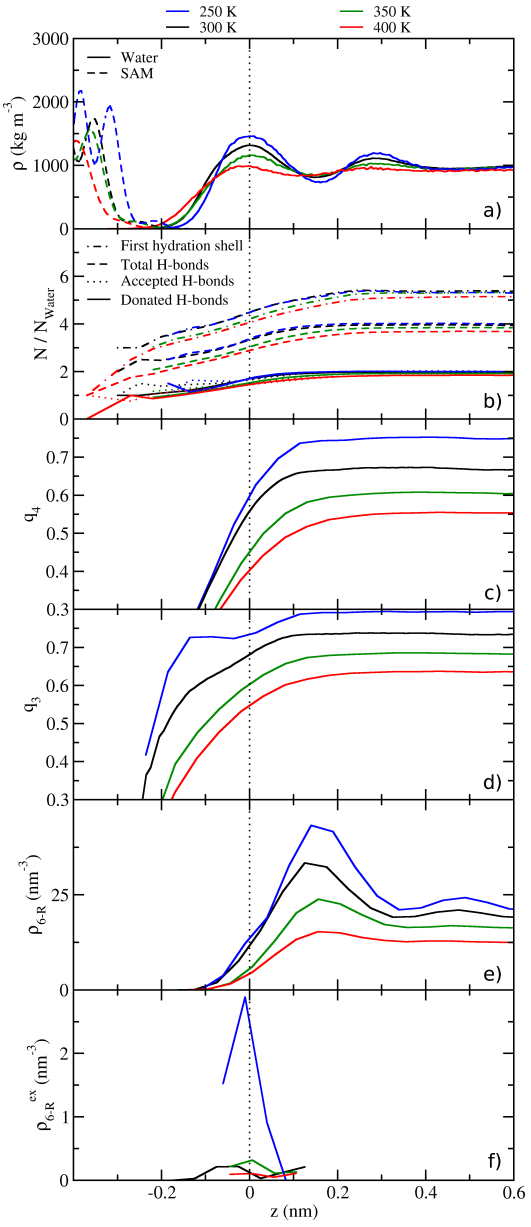


Figure 3.14.: Temperature-dependence of the first-shell water properties (b), tetrahedral (c) and trigonal-pyramidal (d) order parameters, six-membered ring densities (e), and ring excess (f).

3.3 Conclusions and outlook

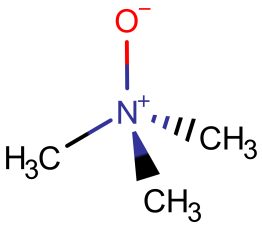
In this chapter we performed a detailed analysis of the influence of pressure on multiple properties of water at a hydrophobic interface, in this case a model of a self-assembled alkane monolayer. We found that the orientations of individual water molecules at the interface are consistent with the orientations in a single ice layer, and that these orientations have a higher relative probability at high pressures. The newly introduced trigonal-pyramidal order parameter q_3 exhibits unexpected behavior, in that it hints towards a layer of water molecules very close to the interface that have a local tetrahedral order which is resistant to pressurization compared to the bulk phase. In order to investigate the structure of water on an intermediate scale, we applied basic graph theory to detect rings in the H-bond network. Not only did we quantitatively analyze these ring structures in water, we also determined the distributions of the ring geometries and found that rings at the interface are more similar to those in hexagonal ice. Most importantly, the ice-likeness of the rings at the interface is proportional to the pressure in the system. We have to note that we only made statements about the ice-likeness of the time-averaged properties in water. At no point do we claim to find actual ice at room temperature.

First and foremost, future work should include an analysis of the water and/or H-bond ring dynamics as a function of distance from the interface. This approach could lead to a better understanding of the counterintuitive finding that some water properties become more similar to hexagonal ice with pressure, even though bulk ice Ih melts under these conditions. Another path towards further understanding of the hydrophobic effect under pressure would be the application of the same methodology to small spherical hydrophobic particles. One promising direction of research would be the characterization of the formation and stability of clathrates or their precursor structures under pressure, especially since the effect of pressure on methane clathrates is highly relevant to environmental science.[101]

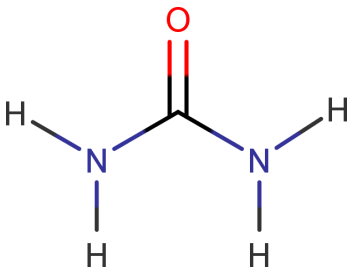
Chapter 4

Aqueous osmolyte solutions from ambient to high pressures

In this chapter, we focus on the development and characterization of force field parameters for the simulation of binary osmolyte/water mixtures at variable pressure. An introduction to osmolytes is given in chapter 1. For trimethylamine-N-oxide (TMAO), we have previously developed a force field for use at ambient conditions.[78,89] Here, we present a method to adapt existing force field parameters for application in high pressure situations based on electronic structure calculations. In the case of urea, we demonstrate that the urea force fields commonly used in the literature are inadequate for the description of many solvation properties of aqueous urea at 1 bar. Therefore, we present a new parameterization of urea, which surprisingly requires no adaptation for the application to the high pressure case.



(a) TMAO



(b) Urea

Figure 4.1.: Structures of trimethlyamine-N-oxide (TMAO, a) and urea (b).

4.1 Trimethylamine-N-oxide (TMAO)

4.1.1 Force field overview

First we will present a brief and incomplete overview of the history of TMAO force fields. The first TMAO model to be used in MD simulations of aqueous solutions was developed in 1995 by Noto et al. It is based on a fully rigid geometry obtained from Hartree-Fock (HF) calculations in vacuum.[102] A couple of years later, the first flexible model was developed by Kast et al.[103] In their approach, the partial charges and torsional potentials were calculated directly from higher-level vacuum QC (MP2) calculations. The remaining nonbonded parameters were obtained from quantum chemistry and crystal structure data of TMAO

hydrates. Finally, the values for the bond length and angle equilibrium values and force constants were obtained from the optimized QC vacuum geometry and resulting vibrational frequencies. This force field was the standard for many years and is still in use today. However, it does not adequately reproduce TMAO-Protein interactions in water. This led to a modification of the Kast force field based on experimental osmotic coefficient data of TMAO/water mixtures. Canchi et al. scaled the charges of the Kast model to achieve the best agreement with experimental data.[54] At the same time, multiple groups worked on improvements on the Kast model in parallel. Shea et al. developed fully atomistic and coarse-grained models optimized for the use in conjunction with the OPLS-AA force field.[92] Separate versions were parameterized for three different water models. In a different approach, Schneck et al. optimized the Kast parameters to simultaneously reproduce the activity coefficients in aqueous solution and the transfer free energies of a model polyglycine to a TMAO solution.[104] The variable parameters were the N-O dipole and the σ of the methyl carbon atoms. Based on this work, we further refined the Schneck model under the constraint that the density of aqueous solutions is also accurately reproduced.[89] Our force field was later published with additional work, which will be presented in this chapter.[78] Therein, a procedure to systematically adapt the force field for the use at elevated pressures was developed. One recent force field by Usui et al. contains dummy charges to represent the lone pairs of oxygen, which leads to more realistic hydrogen bond orientations and dynamics.[105]

4.1.2 Methods

For the recalculation of the KBIs of aqueous TMAO solutions at 1 bar, the trajectories from ref. 89 were reused. The simulation parameters differ in two minor points from the default parameters in this work: A time step of 1.5 ps was used, and only the C-H bond lengths in TMAO were constrained using LINCS. We initially assumed that the N-O bond should be kept flexible, but since performing this work we have found no measurable influence of the N-O bond constraints on our results. Therefore, full bond length constraints are used in later chapters. The original system compositions of the boxes with edge lengths of approximately 4 nm are listed in table 4.1. Each system was simulated for 150 ns. In order to best compare our force field simulations with *ab initio* MD simulations by Imoto[64,78] (see sec. 2.5), we used a box that was exactly twice the length of the aiMD box.

c_{TMAO} (mol/L)	N_{TMAO}	N_{Water}
0.5	20	2109
0.8	30	2050
1.0	40	2000
1.3	50	1945
1.6	60	1893
1.8	70	1836
2.1	80	1782
2.4	90	1725
2.7	100	1678
2.9	110	1627
3.2	120	1572
4.1	150	1411

Table 4.1.: System compositions of TMAO/water mixtures from ref. 89, which were later published in ref. 78.

We can not simulate the same system as in the aiMD simulations, since the LJ interactions would have to be cut off at much shorter lengths due to the periodic boundary conditions. Instead, eight TMAO molecules were placed on a grid, where each solute is in the center of its subcell (see fig. 4.2). The positions of the nitrogen atoms were restrained using a harmonic potential with a force constant of $10^5\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. With this approach, the 8 TMAO were simulated in 856 water in the NVT ensemble for two box sizes. A box length of 2.98764 nm corresponds to the experimental density of 0.5 mol/L TMAO at 1 bar. The system with a size of 2.78446 nm is simulated at the density at 10 kbar, which was extrapolated from measurements up to 700 bar performed by the group of Prof. Winter (see ref. 78). The density extrapolation technique is described in appendix C.

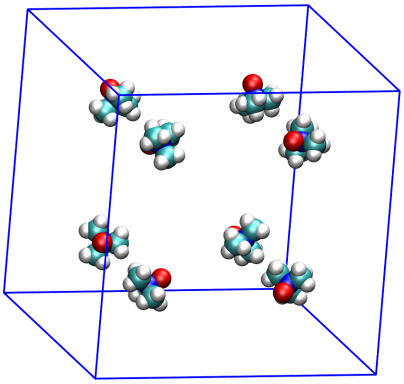


Figure 4.2.: Snapshot of 8 TMAO molecules on a grid. The water molecules are not shown. The box size is twice the size of the aiMD simulation at 1 bar.

Umbrella sampling simulations (see sec. 2.4.1) of two TMAO molecules in 1073 water were performed at 1 bar and 10 kbar. The reaction coordinate is the distance between the centers of mass of the two TMAO. 51 (49) umbrella windows spanning the range $0.24 - 1.24\,(0.24 - 1.20)\,\mathrm{nm}$ were simulated at 1 bar (10 kbar) for 50 ns each with one additional window at the first minimum of the PMF, where the force constant of the harmonic potential was set to $10^3\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. The distance of the two molecules was written every 10 steps. The time step is back to our default setting of 2 ps, which is enabled by applying constraints over all bonds in the system.

4.1.3 Properties at normal pressure

Our TMAO force field, which has been developed in previous work,[89] will be characterized in more detail in the following section. We will only show the properties of our TMAO force field that have not been calculated in the initial work (see ref. 78 for additional properties). First we characterize the TMAO-TMAO interactions in concentrated solutions by recalculating the activity coefficient derivatives y_{TT} of ref. 78 from KBIs using a new, more rigorous method (see sec. 2.3 or ref. 70). The results are shown in fig. 4.3.

$$y_{\mathrm{TT}} = \left(\frac{\partial \ln y_{\mathrm{T}}}{\partial \ln c_{\mathrm{T}}}\right)_{p,T} \tag{4.1}$$

Here the subscript T refers to TMAO, except for the italicized T, which refers to constant temperature. y_{T} is the activity coefficient of TMAO, and c_{T} is the molar concentration. The concentration derivative of the activity coefficient of TMAO increases monotonously with concentration over the whole range. Although y_{TT} of the force field is too low compared to experimental data, the agreement is still very good in the concentration range

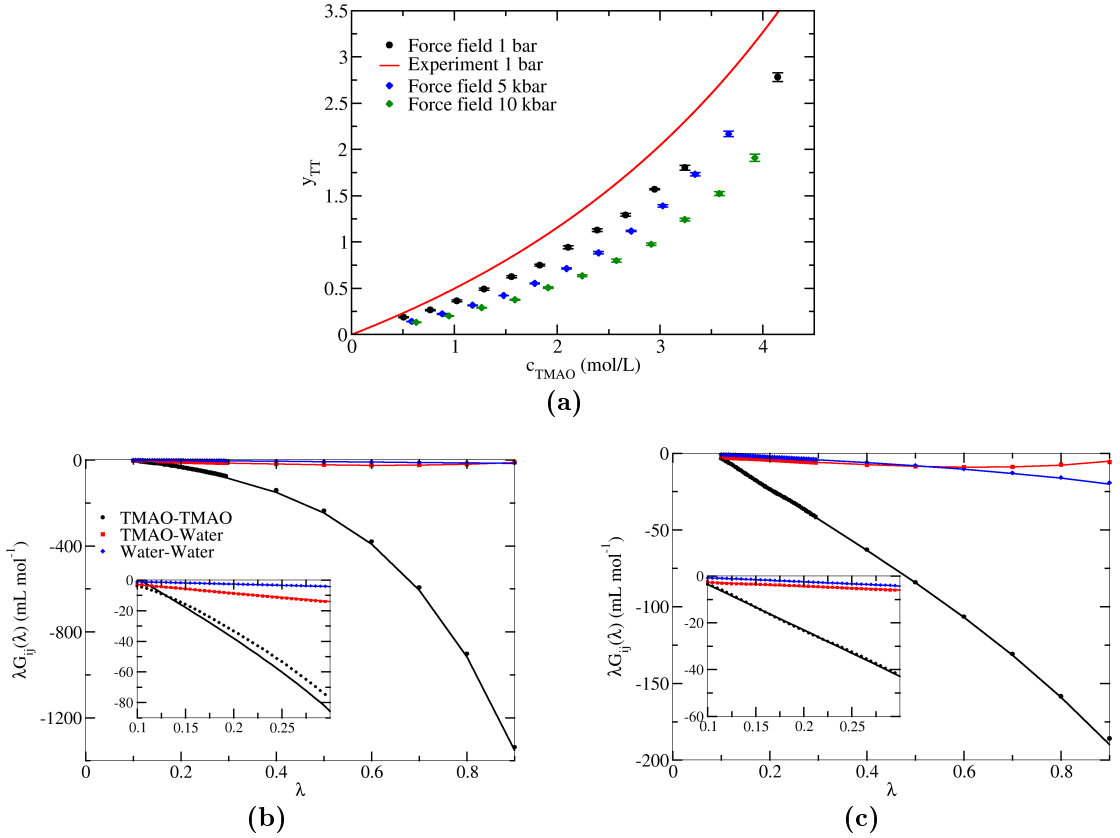


Figure 4.3.: (a) The logarithmic concentration derivative of the TMAO activity coefficient as a function of the TMAO concentration. Experimental values were converted from vapor pressure osmometry data in ref. 106. (b-c) The KBIs of subvolumes (see sec. 2.3) for 0.5 (b) and 4.1 (c) mol/L at 1 bar.

that is usually relevant. We also note that the agreement with experiments was slightly better when using a different correction for the KBIs,[78] which means that the choice of the treatment of finite size and ensemble effects in Kirkwood-Buff theory can be important. The results of the high pressure force fields (see sec. 4.1.4) are shown since they will be used in later chapters.

Next, we move on from solute-solute interactions to the structure of water around TMAO. The simulation was set up in a way where we can compare our force field to the *ab initio* simulations of one TMAO in water. Figure 4.4 contains the RDFs of oxygen and carbon in TMAO to hydrogen and oxygen in water for the effectively infinitely dilute case due to restraints on the TMAO positions. A comparison between the two methods yields excellent agreement for the water distribution around the oxygen and carbon atoms of TMAO, except for very short and very long distances. The RDFs in aiMD do not approach the expected asymptotic limit of 1 due to the finite size effect of the very small system. At short distances, the first peaks of the water density around TMAO-oxygen of the force field are narrower and slightly shifted to larger distances compared to aiMD. This is an expected shortcoming of simple point charge models used in force field MD, which can not reproduce the close-range electrostatics of electronic structure methods. However, the first-shell coordination numbers of water are already practically identical between the two methods, which means that the force field is suited for describing the average solvation shell properties beyond very close contacts.

Lastly, we mention the molecular dipoles of TMAO in the dilute case, which are 6.43 D

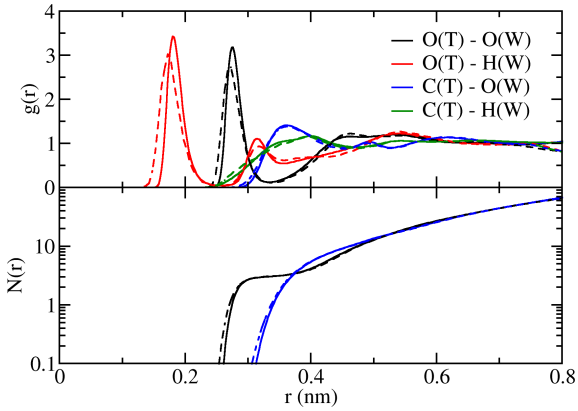


Figure 4.4.: Atomic RDFs $g(r)$ between TMAO (T) and water (W) and coordination numbers $N(r)$ of water oxygens around TMAO at approximately infinite dilution at 1 bar and 300 K. Full lines are the results of force field MD, and dashed lines are *ab initio* MD data from the group of Prof. Marx (ref. 78).

for ffMD, and 8.55 ± 0.07 D in aiMD. This discrepancy is explained by the fact that in the nonpolarizable force field, electrostatic screening effects have to be included in the static partial charges. Leontyev and Stuchebrukhov provided a quantitative treatment based on the 'molecular dynamics in electronic continuum' (MDEC) concept.[107] When the solvent in MD simulations is assumed to be a homogeneous dielectric medium, the effective (screened) dipole moment can be calculated if the high-frequency limit, i.e. the electronic component, of the dielectric constant of the solvent is known:

$$\mu = \mu_{\text{eff}} \cdot \sqrt{\epsilon_{\text{el}}} \tag{4.2}$$

For water at ambient conditions, the experimental value is $\epsilon_{\text{el}} = 1.78$.[107] When we use eq. 4.2 and assume that the dipole moment of the force field is the effective dipole μ_{eff}, we get $\mu = 8.58$ D, which is in excellent agreement with the *ab initio* results.

There exists a plethora of additional force field properties that could be calculated. For a detailed study which compares the properties of the TMAO/water system for many combinations of force fields (including this one), we refer to work by Markthaler et al.[108]

4.1.4 High-pressure adaptation

It has been shown that, at normal pressure, our TMAO force field performs very well regarding many metrics, from the solvation structure to macroscopic thermodynamic properties like the density and activity coefficients. One main goal of this work is to apply the force field at pressures of several kilobars. First we compare NVT simulations of the force field with aiMD results at the density extrapolated from experimental data (using the method in sec. C). Figure 4.5 contains the probabilities of finding a certain number of water-TMAO H-bonds per TMAO molecule for the system where close TMAO-TMAO contacts are prevented. The histograms at 1 bar are in excellent agreement within the estimated error of aiMD. However, at the increased density equivalent to a pressure of 10 kbar, the same force field forms too few H-bonds on average. Imoto et al. have previously shown that the dipole moment of bulk water does not change much between 1 bar and 10 kbar.[64] In contrast, the dipole moment of TMAO in aiMD simulations increases by a considerable amount, while the molecular geometry is practically unchanged.[78] Therefore, we need to make modifications to the force field parameters. Since the TIP4P/2005 wa-

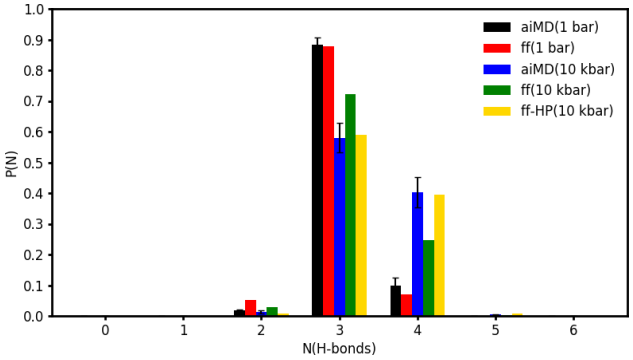


Figure 4.5.: Probability distributions for the number of water-TMAO hydrogen bonds per TMAO molecule at effectively infinite dilution at 1 bar and 10 kbar. For 10 kbar, the force field optimized for 1 bar as well as the charge-scaled ff-HP version (eq. 4.3) are shown.

ter model has been proven to reproduce the properties of water and ice phases over large regions of the phase diagram, we decided to leave it unchanged and focus our attention on the solute parameters. Ideally, we want to systematically modify the TMAO model without a full reparameterization. To keep the changes to the force field as simple as possible, we decided to only modify the partial charges to account for the polarization at elevated pressure. Our approach is to scale the partial charges in a way that reproduces the relative molecular dipole moment change between 1 bar and the target pressure. Instead of *ab initio* MD, which is computationally extremely expensive, we obtained the molecular dipole moment of solvated TMAO as a function of pressure from EC-RISM calculations, a different electronic structure based method (see sec. 2.6). This allows a quick and efficient rescaling of the force field charges via the following relation:

$$q_i^{\text{ff-HP}}(p) = q_i^{\text{ff}} + \alpha q_{\text{O}}^{\text{ff}} \frac{\Delta q_i^{\text{EC-RISM}}(p)}{\Delta q_{\text{O}}^{\text{EC-RISM}}(p)} \tag{4.3}$$

Here ff-HP refers to the high-pressure modification of the force field, i is the atom in TMAO, and Δq are the charge differences between 1 bar and the pressure p. We arbitrarily chose the oxygen atom as a reference for the scaling procedure. α is an optimization parameter that should be chosen such that the new set of charges reproduces the relative change in the TMAO dipole moment from EC-RISM calculations (see fig. 4.6). The scaling approach in eq. 4.3 ensures that the ratios of the charge increments are the same as in EC-RISM, while at the same time conserving the total charge. The small nonzero total charge resulting from rounding was compensated by changing the oxygen charge. With this methodology, our high-pressure force field (ff-HP) correctly reproduces the number of H-bonds of the aiMD simulations at 10 kbar (fig. 4.5). The RDFs and coordination numbers of the ff-HP at 10 kbar (fig. 4.7) are still in good agreement overall, even though the first minimum of the O(T)-O(W) RDF and the second minimum of the O(T)-H(W) RDF are too low for the force field, which manifests in a small deviation of the coordination numbers.

In addition to aiMD, we will use experimental data to validate the force field at high pressures. In fig. 4.8 we compare the densities of simulations of pure TIP4P/2005 water and 0.5 mol/L TMAO solution with experimental data. The experimentally accessible pressure range was 1 bar to 700 bar, and we extrapolated the densities up to 10 kbar (see sec. C). Both the unmodified force field and the ff-HP version accurately reproduce the experimental densities. At 10 kbar, the absolute and relative deviations of simulation results from experiments are $4.5\,\mathrm{g\,L^{-1}}$ (0.36 %) for the normal pressure ff and $5.2\,\mathrm{g\,L^{-1}}$ (0.42 %)

for ff-HP. A major contribution to this overestimation of the density is the TIP4P/2005 water model, which accounts for $2.9\,\mathrm{g\,L^{-1}}$ (0.23 %). This means that we can accurately reproduce the partial molar volume of TMAO at high pressures.

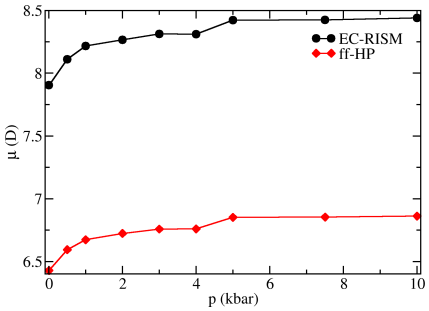


Figure 4.6.: The molecular dipole moment of TMAO as a function of pressure calculated using the EC-RISM method (black) and the force field dipoles obtained from the charge-scaling method using eq. 4.3 (red). The lines are a guide to the eye.

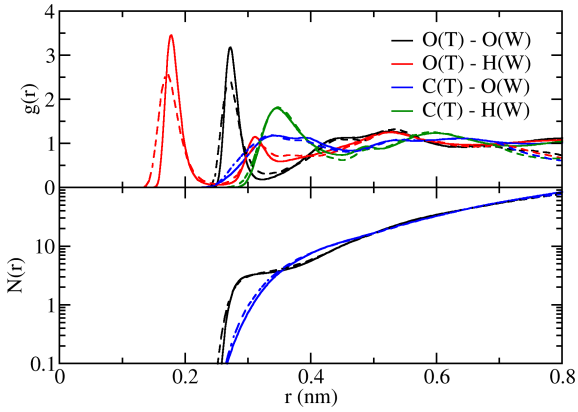


Figure 4.7.: TMAO-water RDFs and water coordination numbers for aiMD (dashed lines) and the charge-scaled ff-HP (eq. 4.3) at the 10 kbar density. Close TMAO-TMAO contacts were prevented in ffMD via position restraints.

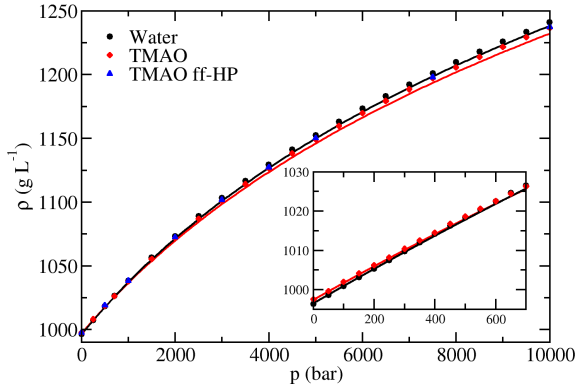


Figure 4.8.: The densities of water (black) and 0.5 mol/L TMAO solution (blue, concentration refers to 1 bar) as a function of pressure. The symbols are force field NpT simulations, and lines are experimental results by the group of Prof. Winter up to 700 bar (see the inset). For higher pressures, the lines represent the extrapolated densities using the method described in appendix C.

By all investigated metrics, the TMAO force field and its high pressure modifications perform well with respect to quantum chemistry and experiments. All force field parameters with the scaled charges for the high pressure modifications at multiple pressures can be found in appendix B.2.

4.1.5 TMAO-TMAO potential of mean force

In this brief section, we present the potential of mean force (PMF) between two TMAO molecules as a function of their distance. The PMFs of the force field at 1 bar and the ff-HP variant at 10 kbar (fig. 4.9) were calculated from umbrella sampling simulations using the weighted histogram analysis method (WHAM, see sec. 2.4.1). The entropic contribution to the free energy, which depends on the TMAO-TMAO distance r, is $-\mathrm{k}T \ln 4\pi r^2$ (which can be simplified to $-2\mathrm{k}T \ln r$ by discarding the constant terms) since the probability of finding two molecules at distance r scales with the surface area of a sphere. This term was subtracted from the PMF to obtain a constant value for large distances. The first local

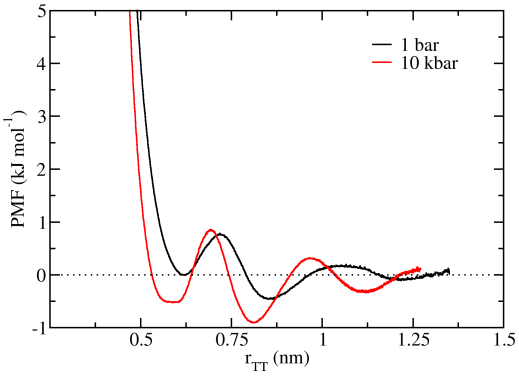


Figure 4.9.: Potential of mean force calculated from umbrella sampling as a function of TMAO-TMAO center of mass distance. Black: 1 bar force field. Red: ff-HP(10 kbar) at 10 kbar.

minimum of the PMF at 1 bar at a TMAO-TMAO separation of 0.61 nm is very close to a free energy of zero. Only around the second minimum at 0.85 nm do interactions become energetically favorable. The ff-HP at 10 kbar shows a very similar behavior, although the minima are shifted to shorter distances (0.58 and 0.81 nm) due to the compression of the system. Interestingly, the first minimum at 10 kbar has an unusually flat, basin-like structure. The pressure increase also leads to lower free energies at the minima. Due to the long-range oscillations of the PMF not decaying to zero at the length limitations of the simulation box, the quantitative pressure effect carries some uncertainty, but it is approximately 0.5 kJ/mol at the minima. The fact that the free energy at the first minimum is negative at 10 kbar does not mean that the TMAO molecules attract each other at high pressures. First we see that the minimum is still at a distance larger than 0.5 nm, which means that the TMAO are still separated by a layer of water. Furthermore, the activity coefficient derivative y_{TT} (figure 4.3) is smaller at 10 kbar than at 1 bar but remains positive over the whole concentration range.

The fact that close TMAO-TMAO interactions are energetically unfavorable has been studied by Schneck et al. at normal pressure using the predecessor of our force field.[104] Their PMF of two TMAO molecules is basically identical to the profile in fig. 4.9. The authors coined the term 'dipolar/hydrophobic frustration' for the main reason of TMAO repulsion. At close distances, the optimal orientation of the molecules regarding electrostatics would be the alignment of the N-O dipoles. In this case, the interaction between

the polar group of one molecule and the methyl group of the other is repulsive. There simply is no way to orient the TMAO molecules for both contributions to be favorable at short distance, which leads to the separation of TMAO in water. Schneck et al. argue that this mechanism is also the reason for the exclusion of TMAO from the peptide backbone, which is suspected to be the explanation for the protein stabilization by TMAO.

4.2 Urea

4.2.1 Force field overview

The first well-known urea model for the simulation of aqueous solutions was a rigid model developed by Kuharski and Rossky in 1984.[109] Almost 10 years later, the flexible OPLS force field by Duffy et al. was optimized based on quantum chemistry calculations of different urea and water geometries.[110] A couple of years later, Tsai et al. introduced two different force fields.[111] One of them was parameterized for simulations with proteins, the other was derived from vacuum QC calculations. There also exists a parameter set based on the force fields included in the AMBER program package.[112] In 2003, Weerasinghe and Smith published a force field that was optimized to reproduce the Kirkwood-Buff integrals in aqueous solutions with the SPC/E water model.[113] This Kirkwood-Buff force field (KBFF) has become the most widely used model by far and still is the standard for simulations of aqueous urea. In the following section we will demonstrate the need for a new force field beyond those in the existing literature by comparing the water structure around several urea models with *ab initio* simulations. Our optimization procedure based on a differential evolution algorithm (DEA) using experimental and aiMD data is presented, followed by a discussion of structural and thermodynamic properties at normal and high pressures.

4.2.2 Simulation methods

The simulations for the comparison with aiMD results (1 urea in 110 water) were set up in the same fashion as the TMAO simulations (sec. 4.1.2) by doubling the box length in all three dimensions and fixing the positions of the urea carbons on a grid at the center of their subcells. The box size for the NVT simulation with 8 urea and 880, which is equivalent to 0.5 mol/L, was 2.99968 nm. It was determined from density measurements at 1 bar by the group of Prof. Sadowski.[114] The box size for the newly developed force field at the 10 kbar density was calculated as 2.79138 nm from a fit of the density vs pressure (eq. C.1) of the same force field. We chose to use the simulations for the calculation of the density because the experimental data, which was available up to 500 bar, could not reliably be extrapolated up to 10 kbar. The trajectory lengths were 90 ns for the AMBER, OPLS, and KBFF urea models at 1 bar and 45 ns for the new force field at all pressures, which turned out to be sufficient sampling for the H-bond statistics and SDFs. For OPLS, the σ LJ parameters between different atom types were calculated by taking the geometric instead of the arithmetic mean: $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$

During the force field optimization, the following three systems were simulated in each iteration:

1. The system with 8 urea in 880 water at 300 K at the 1 bar density for 2 ns.

2. 77 urea in 1941 water in a box with a length of about 4 nm (urea concentration approximately 2 mol/L) at 1 bar (NpT). After an equilibration period of 40 ps the system was simulated for 1 ns.

3. 308 urea in 1362 in a 4 nm box (urea concentration approximately 8 mol/L) with identical setup to the 2 mol/L system.

For the optimized force field, 150 ns long NpT simulations of the urea/water system at 1 bar and 298.15 K were performed for 32 concentrations with the compositions listed in table 4.2. The simulations of the densities of 0.5 and 1 mol/L systems at varying pressures were 10 ns long. Potentials of mean force of two urea molecules in 1073 water as a function of intermolecular distance were calculated via umbrella sampling. At 1 bar, 46 windows with potential minima at distances between 0.26 and 1.24 nm were simulated for 50 ns each. For the 10 kbar PMF, 39 windows between 0.26 and 1.02 nm were simulated.

c_{Urea} (mol/L)	N_{Urea}	N_{Water}
1.0	39	2037
1.3	49	2012
1.5	59	1987
1.8	69	1962
2.0	77	1941
2.3	87	1916
2.5	97	1891
2.8	107	1866
3.0	116	1846
3.3	126	1821
3.5	136	1796
3.8	146	1771
4.0	154	1750
4.3	164	1725
4.5	174	1700
4.8	184	1675
5.0	193	1653
5.3	203	1628
5.5	213	1603
5.8	223	1578
6.0	231	1557
6.3	241	1532
6.5	251	1507
6.8	261	1482
7.0	270	1458
7.3	280	1433
7.5	290	1408
7.8	300	1383
8.0	308	1362
8.3	318	1337
8.5	328	1312
8.8	338	1287

Table 4.2.: System compositions of urea/water mixtures for NpT simulations at 1 bar.

4.2.3 Solvation structure of existing force fields

Before attempting simulations at high pressure, we need to ensure that the force field already performs well at ambient conditions. Here we present the solvation properties of some of the previously mentioned models and compare them to reference data, i.e. *ab initio* MD results by Imoto et al. at the experimental density.[79] We chose the AMBER,[112] OPLS,[110] and KBFF[113] force fields and used them in combination with the TIP4P/2005 water model. Since they were optimized with different water parameters, we checked the influence of water by calculating all properties of the KBFF for the SPC/E water model[86] as well.

As a first criterion, we compare the spatial distribution functions (SDFs) of water oxygen and hydrogen atoms of the force fields. Specifically, figure 4.10 shows the isosurfaces which separate regions with twice the bulk water density. The electronic structure theory based simulations have a ring of high water density around the oxygen, which signifies the hydrogen bonding structure. At the hydrogens, there are lobes of high density for the donated H-bonds. Small areas with high water density parallel to the molecular urea plane point towards the nitrogens being weak H-bond acceptors. The AMBER force field reproduces this solvation structure almost exactly, with only minor differences in the ring shape of the isosurface around the oxygen. In the case of OPLS, the water density around the hydrogens is comparable to aiMD, but is far too low around the oxygen. The regions with high oxygen density do not extend across the molecular plane. Interestingly, the most popular force field, KBFF, is not able to reproduce the solvation structure with either the TIP4P/2005 or the SPC/E water model. No significant solvent effect is observed. We see from the high hydrogen density centered around the oxygen that the directionality of the hydrogen bonds is completely different from the QC results. The water oxygen isosurface does not form a ring structure, and has a region with low density orthogonal to the molecular plane.

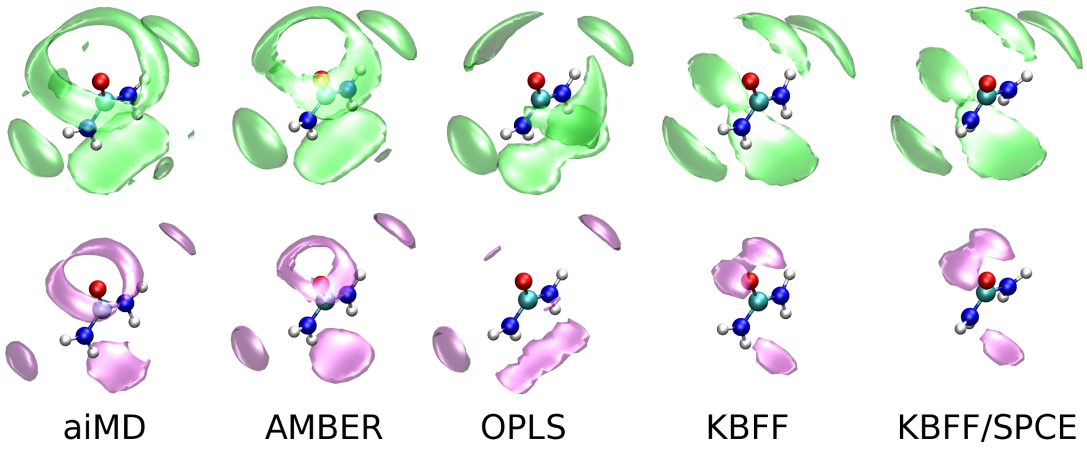


Figure 4.10.: Spatial distribution functions of water oxygen (green) and hydrogen (purple) around urea for different force fields and the *ab initio* simulation as a reference. The SDFs were first symmetrized w. r. t. the vertical and horizontal symmetry plane and then smoothed with a Gaussian with $\sigma = 2\,\text{Å}$. The chosen isovalue is twice the bulk density. SDFs were generated with TRAVIS.[115]

Next, we move on to a quantitative analysis of the solvation properties by calculating the probability distributions of the numbers of water/urea H-bonds per functional group, which are shown in figure 4.12. We find that all force fields accept too few H-bonds at the oxygen, with KBFF in TIP4P/2005 having the fewest, and OPLS being the closest to aiMD. The

AMBER ff accepts approximately the right amount of H-bonds at the nitrogens, whereas OPLS forms too many H-bonds, and KBFF shows almost no tendency for bonding at all. The histograms for the hydrogens were separated into cis- (pointing towards oxygen) and trans hydrogens. In aiMD, there are slightly fewer H-bonds at the cis-hydrogen. The fact that one- and twofold H-bonded trans-hydrogens are more likely than for the cis-hydrogens can be explained by the spatial distributions. Since the high water density region around trans-hydrogens is one continuous lobe in the SDFs, and the trans-hydrogens bonds are oriented in parallel, one water can H-bond to both donors at the same time. Therefore, trans H-bonds are correlated and have a higher probability per donor site. The AMBER and OPLS models are the best performing force fields regarding the donor H-bonds, whereas KBFF again forms too few H-bonds. When we combine the spatial and quantitative results, the AMBER ff is the clear favorite for correctly reproducing urea/water interactions in the dilute system. Before coming to a conclusion about the best currently available force field we will check one important property of highly concentrated urea solutions. In figure 4.11 we show the logarithmic derivative of the urea activity coefficient y_{UU} at approximately 8 mol/L urea. The only force field that is close to the experimental data is KBFF in SPC/E water. This is to be expected, since it was optimized at these conditions. We observe a significant dependence of y_{UU} on the water model for KBFF when the TIP4P/2005 force field is used instead. The AMBER force field does not even reproduce the sign of y_{UU}, thereby making it unsuitable for simulations of concentrated urea solutions. We also observe slow urea-urea aggregation, which is a known problem of this force field[116] as well as other models mentioned previously.[117]

In conclusion, we find that there is no single force field which compares favorably to *ab initio* and experimental data over the full concentration range we are interested in.

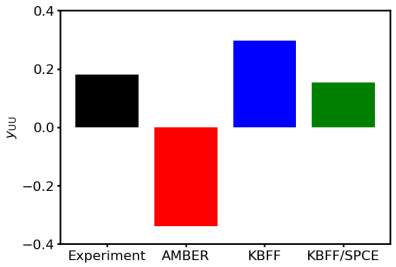


Figure 4.11.: The derivative of the activity coefficient of urea $y_{\mathrm{UU}} = \left(\frac{\partial \ln y_{\mathrm{U}}}{\partial \ln c_{\mathrm{U}}}\right)_{p,T}$ at 8 mol/L for multiple force fields and experimental data by Scatchard et al.[118]

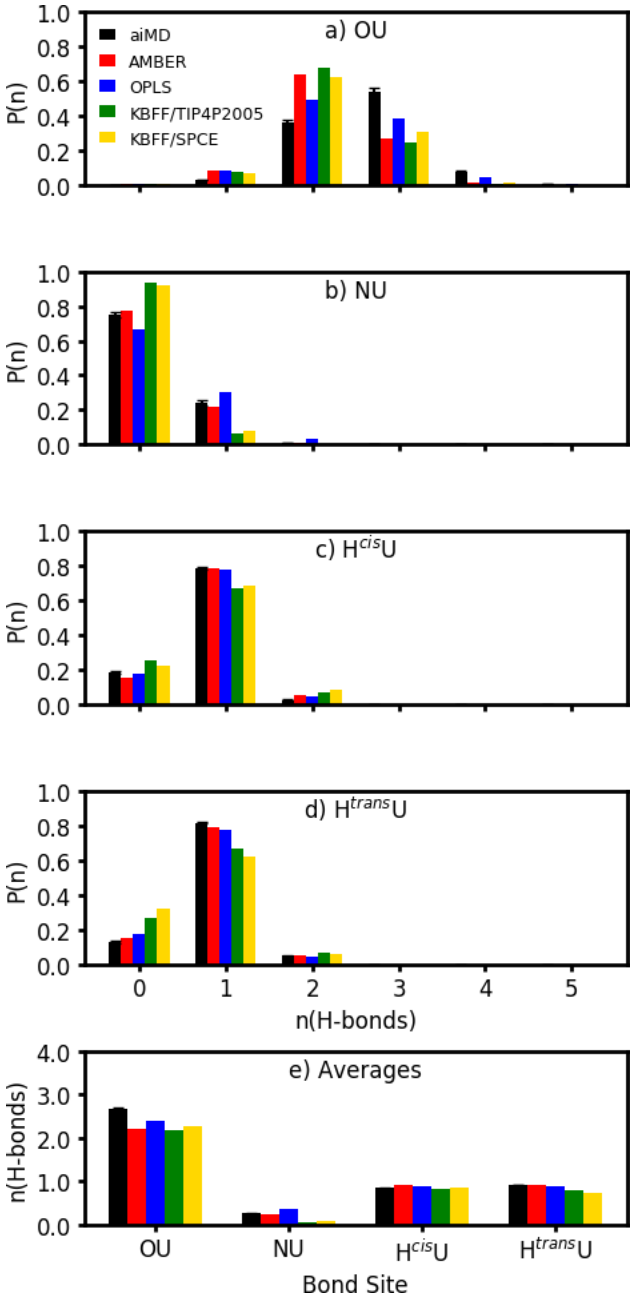


Figure 4.12.: Probabilities $P(n)$ of finding n urea/water H-bonds at specific donor and acceptor sites of urea. Probabilities are given per site. The simulations were performed at 300 K at the constant density that corresponds to 1 bar.

4.2.4 Force field optimization procedure

We have demonstrated that the existing urea force fields used throughout the literature do not reproduce the properties of urea at 1 bar. In the following section, we will elaborate on our approach for the development of a new force field that significantly improves upon currently available ones. Instead of starting from scratch, we use a hybrid of the AMBER and KBFF force fields as a starting point. For the bonded interactions we used the bond lengths and dihedral potential of AMBER and the angle potentials of KBFF. In order to keep the space of optimization variables as small as possible, we made some assumptions and educated guesses concerning the nonbonded interactions: We assumed that the ratios of the σ LJ parameters of AMBER already have reasonable values, so we only scaled all σ by the same factor. In the interest of further reducing the degrees of freedom, we kept the ϵ values constant during the optimization, but this required a good choice of starting values. When comparing the ϵ of AMBER and KBFF, we found that the values for the carbon atom differ by only 0.06 kJ/mol, but the oxygen and nitrogen values deviate by 0.32 and 0.21 kJ/mol respectively. As a very simple ansatz, we took the arithmetic mean of the two force fields for ϵ_{N} and ϵ_{O} under the assumption that this choice prevents urea-urea aggregation in concentrated solutions. For electrostatics, we only assumed that the partial charges are symmetric with respect to the horizontal and vertical symmetry planes of the molecule. The cis- and trans- hydrogens are not chemically equivalent and were chosen to be separate degrees of freedom in the optimization. This means that in total the parameter space is five-dimensional: The single scaling factor s for all σ, and the charges $q_{\mathrm{O}}, q_{\mathrm{N}}, q_{\mathrm{H^{cis}}}$, and $q_{\mathrm{H^{trans}}}$. The remaining carbon atom charge is fixed due to the charge neutrality condition.

Our goal is to find force field parameters that reproduce both structural properties of quantum chemical simulations and experimental thermodynamic properties. In our approach, we want to minimize the cost function f, which sums over the relative deviations of the force field results from aiMD/experiments scaled with weighting factors C_{ij} (eq. 4.4):

$$\begin{aligned} f(q_{\mathrm{O}}, q_{\mathrm{N}}, q_{\mathrm{H^{cis}}}, q_{\mathrm{H^{trans}}}, s_\sigma) = & \sum_{i=\mathrm{O,N,C,H^{cis},H^{trans}}} C_{1i} \cdot \left| \frac{N^{\mathrm{ff}}(i) - N^{\mathrm{aiMD}}(i)}{N^{\mathrm{aiMD}}(i)} \right| + \\ & \sum_{i=\mathrm{O,NH^{cis}_{don},NH^{trans}_{don}}} C_{2i} \cdot \left| \frac{n^{\mathrm{ff}}_{\mathrm{HB}}(i) - n^{\mathrm{aiMD}}_{\mathrm{HB}}(i)}{n^{\mathrm{aiMD}}_{\mathrm{HB}}(i)} \right| + \\ & \sum_{i=1}^{2} C_{3i} \cdot \left| \frac{\rho^{\mathrm{ff}}(i) - \rho^{\mathrm{exp}}(i)}{\rho^{\mathrm{exp}}(i)} \right| \end{aligned} \tag{4.4}$$

Here N are the first-shell coordination numbers

$$N(i) = \int_0^{r^{\mathrm{aiMD}}_{\mathrm{min}}(i)} 4\pi r^2 g_{i-\mathrm{O(W)}}(r)\mathrm{d}r \tag{4.5}$$

integrated up to the first minimum of the RDF in aiMD, and n_{HB} are the average numbers of hydrogen bonds per donor/acceptor site. The experimental densities ρ^{exp} were obtained from a fit of experimental data by Gucker et al.[119] The weights C_{ij} were chosen as

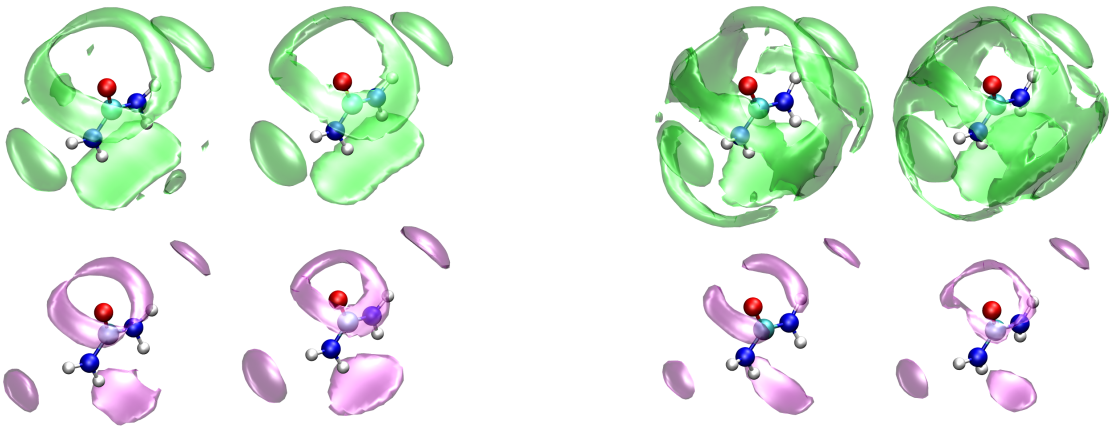
$$C = \begin{bmatrix} 1 & 1 & 1 & 0.5 & 0.5 \\ 1 & 0.5 & 0.5 & & \\ 25 & 25 & & & \end{bmatrix} \tag{4.6}$$

where the column index for atom types is determined from their order in the summation in eq. 4.4. For each set of variables, three simulations are performed (see sec. 4.2.2 for details) and analyzed, from which the quality is calculated as the value of the cost function f. The minimization of f in a five-dimensional space is relatively difficult since the evaluation of f is very time intensive, and analytic gradients are not available. We decided to use the differential evolution algorithm (DEA) introduced by Storn and Price,[120] which performs well in these situations. An explanation of this algorithm and the parameter choices can be found in section E.

4.2.5 Properties of the new force field

4.2.5.1 Structural properties

First we show the performance of the newly optimized force field in comparison to *ab initio* MD at effectively infinite dilution. The spatial distribution functions of water oxygen and hydrogen around urea are shown in fig. 4.13. At 1 bar, the solvation structure of the new force field is a clear improvement over the literature force fields (fig. 4.10), especially around the hydrogens. It still suffers from the same problem of a relatively symmetric isosurface with respect to the rotation around the C-O bond, whereas the aiMD simulations lead to a lower water density perpendicular to the molecular plane. This directionality of the density distribution is most likely only reproduced in wave function based methods and can not be obtained from simple point charge models. When we apply the force field to the extreme pressure of 10 kbar, the ring-like isosurfaces around the oxygen are similar in both methods. An additional lobe between the two hydrogen regions of each NH_2



(a) 1 bar **(b)** 10 kbar

Figure 4.13.: Spatial distributions of water oxygen (green) and hydrogen (purple) around urea for aiMD and the new force field at the 1 bar (a) and 10 kbar (b) densities. The isovalue is twice the bulk density.

group is present compared to 1 bar, which is connected in a cage-like structure in the force field, but not in aiMD. The water hydrogen density around oxygen is not connected in a ring for aiMD, but forms a very narrow isosurface for the force field. Considering that we use a force field with point charges, which should not be able to reproduce the directionality of electrostatics of quantum-chemical methods, the overall agreement of the solvation structures is quite good.

When looking at the radially averaged water oxygen densities (fig. 4.14), we see almost perfect agreement between force field and aiMD at 1 bar, with the exception of the first peak of the O-O RDF. The aiMD peak is broader and lower than for the force field, which is the same observation we previously made in the case of TMAO (fig. 4.4). At 10 kbar, the RDFs are in excellent agreement except for the oxygen atom, where the force field RDF exhibits a small local minimum that does not exist in aiMD. In addition to analyzing the water density around urea, we are also interested in the orientation of water molecules, which manifests in the number of hydrogen bonds (fig. 4.15). For now, only the first four data sets in the figure are relevant (aiMD and ff at 1 bar and 10 kbar). The high pressure variant μ_{bond}-HP-ff is explained in section 4.2.6 and appendix D. At both pressures, the force field has a lower probability than aiMD for finding two H-bonds accepted at oxygen, and a higher likelihood for three H-bonds. The probabilities for four H-bonds are underestimated at 1 bar, and the values at 10 kbar are within the uncertainty of the aiMD simulation. The average numbers are in good agreement at 1 bar and slightly too high at 10 kbar. At both pressures, the average number of H-bonds accepted at nitrogen is in good agreement, but shows a small deviation towards fewer H-bonds. At the cis-hydrogens, the probability of finding a single H-bond is shifted to two at both pressures, but the averages are still very accurate. Around the trans-hydrogens, the probability for zero and two H-bonds is shifted to one H-bond at 1 bar, which balances out to the correct averages. In the case of increased pressure, the probability of zero and one H-bond is lower, which is compensated by the two H-bond probability and therefore gives the correct average. Overall, we found some discrepancies between the H-bond probability distributions between aiMD and force field, but our simulations reproduce the average numbers per functional group very accurately. This concludes the discussion of water/urea interactions at infinite dilution.

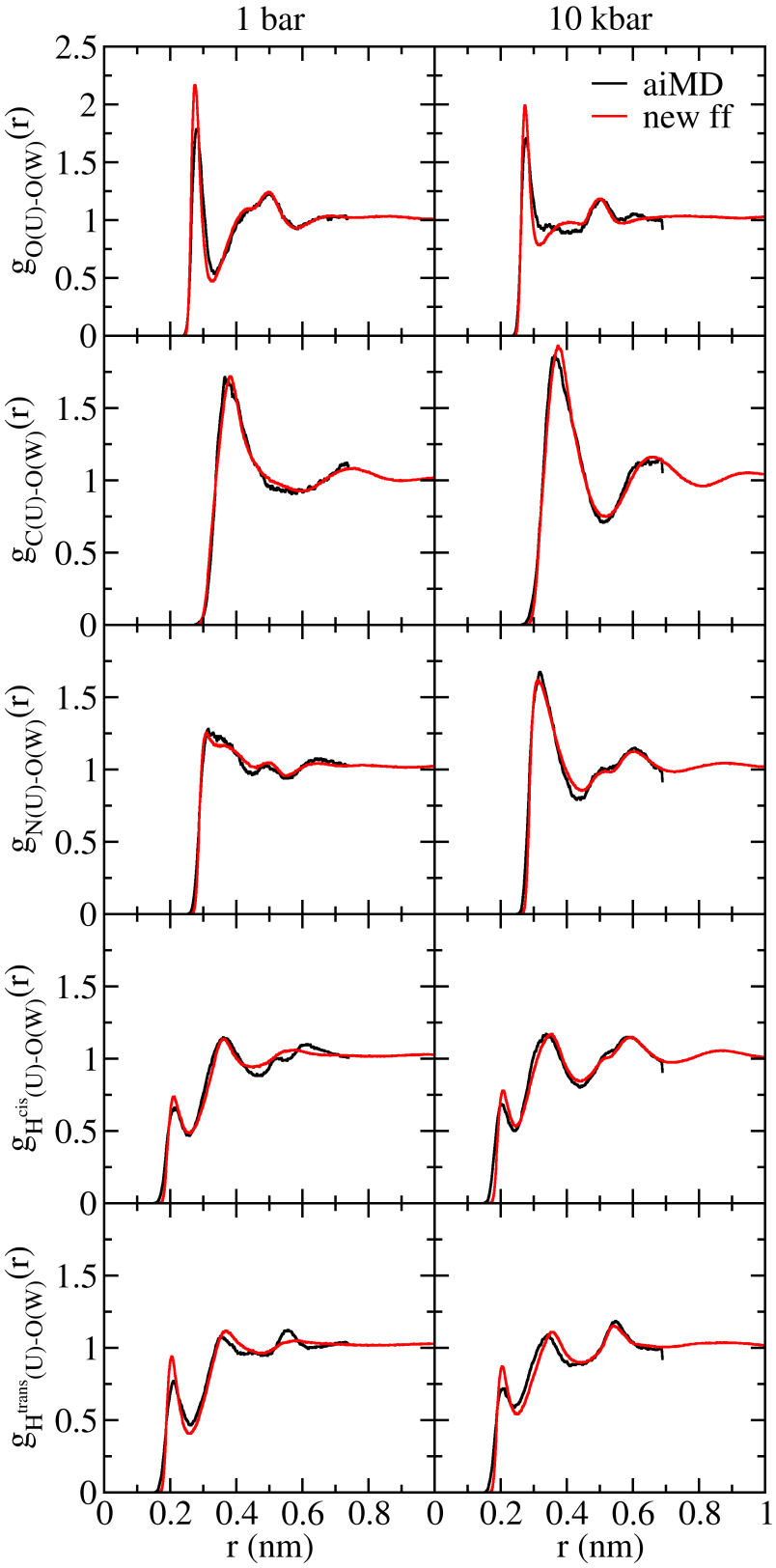


Figure 4.14.: Atomic radial distribution functions between urea and water for aiMD and the new force field at the 1 bar (left) and 10 kbar (right) densities.

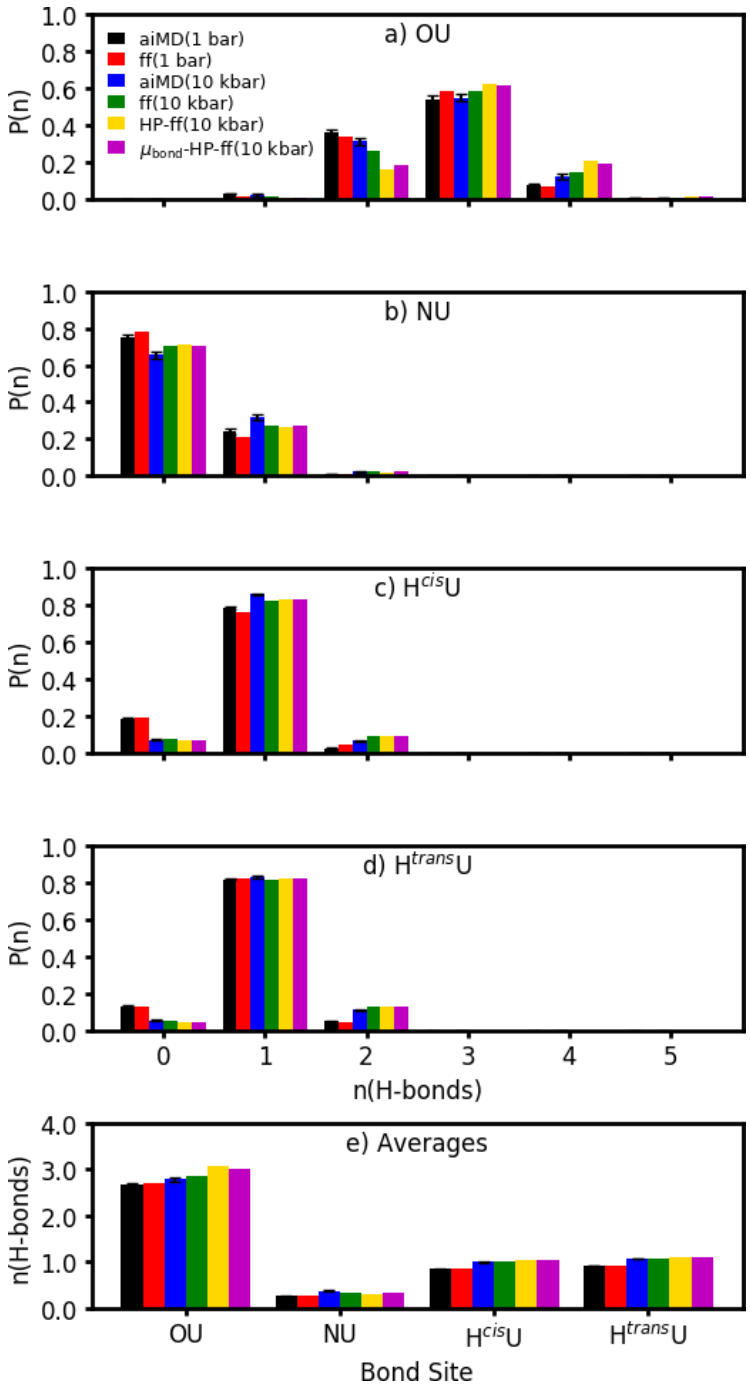


Figure 4.15.: Probabilities $P(n)$ of finding n urea/water H-bonds at specific donor and acceptor sites of urea. Probabilities are given per site. HP-ff is the modification with scaled charges to reproduce the molecular dipole moment increments (see eq. 4.3). In the μ_{bond}-HP-ff variant, the charges were scaled to reproduce the changes of the individual bond dipoles (see section 4.2.6 and appendix D).

4.2.5.2 Thermodynamic properties

Panel a) of figure 4.17 shows that the force field reproduces the experiments by Scatchard et al.[118] very accurately up to 8 mol/L, where the force field overestimates the density by only 1.4 g/mol (0.1 %). From the densities, the apparent molar volumes ${}^{\phi}V_{\mathrm{U}}$ (eq. 4.7) were calculated:

$$ {}^{\phi}V_{\mathrm{U}} = \frac{\rho_{\mathrm{W}} - \rho}{c_{\mathrm{U}} \cdot \rho_{\mathrm{W}}} + \frac{M_{\mathrm{U}}}{\rho_{\mathrm{W}}} \tag{4.7} $$

ρ_{W} is the bulk water density, and M_{U} is the molar mass of urea. As is to be expected from the excellent densities, we find good agreement for ${}^{\phi}V_{\mathrm{U}}$ (b), which increases by only 2 % over the whole concentration range. In panel c, the concentration derivative of the activity coefficient of urea was calculated from the KBIs (eq. 2.12). This quantity is very close to zero even at high concentrations, which shows the almost ideal solution behavior of urea in water. Experiments and simulations are in precise agreement up to moderate concentrations, and even at 9 mol/L the activity coefficients y_{U} (d) differ by at most 0.09. The ability to reproduce activity data over a large concentration range is an important result, because we show that by solely optimizing the solvation structure and density, we obtain a force field that is able to predict the thermodynamics with good accuracy. However, the dielectric constant (e), calculated from the dipole fluctuations of the simulation box,[121] does not match experiments at all. It is a known flaw of the TIP4P/2005 water model that its bulk dielectric constant is only 58.[61] However, the urea force field, in similar fashion to the KBFF model,[113] also fails to reproduce the increase of ϵ with urea concentration and stays almost constant. The translational diffusion coefficient of urea (f) was calculated from the time derivative of the root mean squared displacement:

$$ D_{\mathrm{U}} = \frac{1}{6} \frac{\mathrm{d} < \mathbf{r}_{\mathrm{U}}^2 >}{\mathrm{d}t} \tag{4.8} $$

D_{U} decreases with increasing urea content in experiment[122] and simulation. Over the range with available experimental data, the force field underestimates the diffusion coefficient, which deviates by $6{\cdot}10^3\,\mathrm{cm^2 s^{-1}}$ (5 %) at 1 mol/L. The difference increases to $19{\cdot}10^3\,\mathrm{cm^2 s^{-1}}$ (17 %) at 4 mol/L urea. Whereas the deviations of the force field are acceptable for low concentrations, we have to note that the translational dynamics of our model is too slow in concentrated solutions. In panel (g) we show the adiabatic compressibility of the solutions. First the isothermal compressibility was calculated via finite differences of the densities at 1, 100, and 200 bar:

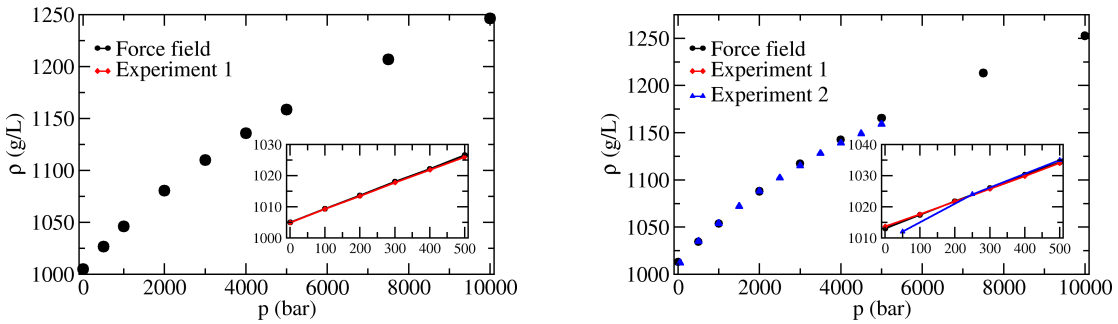
$$ \kappa_{\mathrm{T}} = -\rho \left(\frac{\partial \frac{1}{\rho}}{\partial p} \right)_T \tag{4.9} $$

Then, the κ_T were converted to the adiabatic values by scaling them with the ratio of the heat capacities at constant volume and pressure c_V and c_p (see appendix F for details):

$$ \kappa_S = \kappa_T \frac{c_V}{c_p} \tag{4.10} $$

We find that κ_S is slightly underestimated for our model, but that it generally agrees very well with experiments[123] over the whole concentration range. To conclude the discussion of ambient pressure properties, during our search for a good starting point for high pressure simulations, we also found a force field that is as good or better than the KBFF model[113] at describing activities and most other properties under normal pressure.

Next, we compare the solution densities of the force field to different measurement methods for the case of low to moderate urea concentrations (fig. 4.16). At 0.5 mol/L urea (a), experimental data was only available up to 500 bar (experiment 1). At 500 bar, the simulation overestimated the experimental value by only 0.7 g/L (0.07 %), showing exact agreement within the accuracy of the experiment. Simulated densities up to 10 kbar are shown for completeness, since they were used to compute the system size of the NVT simulations at the 10 kbar density (see above). At 1 mol/L (b), experiment 1 and our force field produce densities that are within 0.7 g/L of each other. The force field again weakly overestimates the density. In the pressure range up to 5 kbar, additional experimental data (experiment 2) is available for 1 mol/L solutions. This method results in a density at 50 bar that is approximately 2 g/L below the other methods. A major contribution to this difference may be the accuracy of 1 g/L of this method. At 500 bar, the difference to the simulations reduces to below 1 g/L, and stays that way up to 2 kbar. For larger pressures, the simulations deviate towards larger densities with a difference of over 6 g/L (0.55 %) at 5 kbar. Overall, the new force field is able to predict the densities up to very high pressures with good accuracy.



(a) 0.5 mol/L at 300 K **(b)** 1 mol/L at 298.15 K

Figure 4.16.: Density of aqueous urea solutions as a function of pressure for 0.5 (a) and 1 mol/L urea (b). The simulations use the new force field that was optimized for 1 bar (see appendix B.3). Experiment 1 is a series of vibrating tube densitometry measurements performed by Michael Knierbein as work for his PhD thesis (see ref. 114). Experiment 2 are X-ray reflectivity experiments, in which density changes relative to the reference in experiment 1 are measured, performed by the group of Dr. Nase (ref. 124). The concentrations are the given for the systems at 1 bar. Insets show the region from 1 to 500 bar.

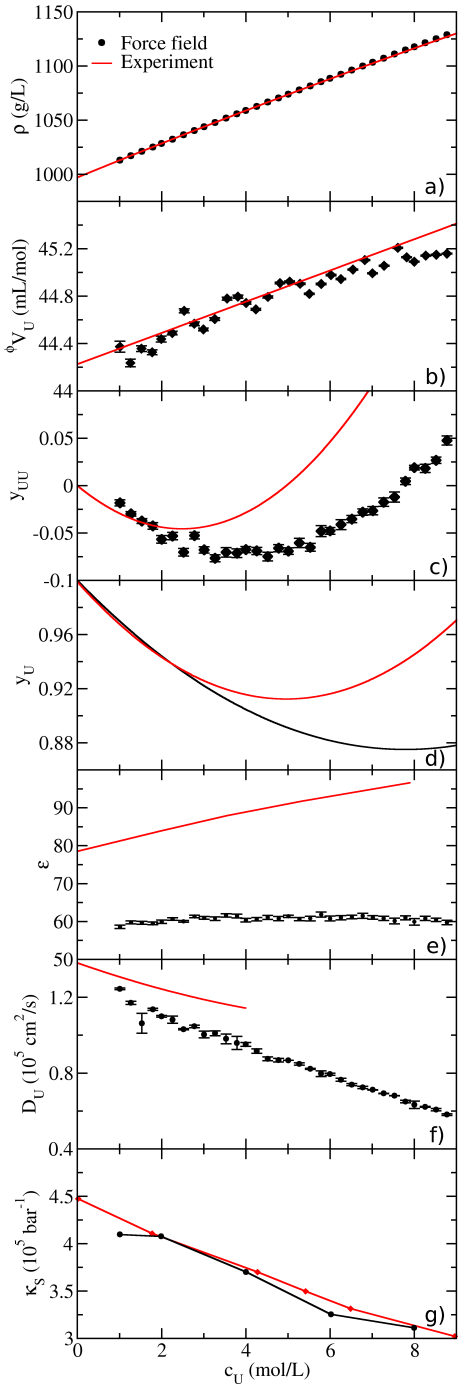


Figure 4.17.: Properties of aqueous urea solutions as a function of the urea concentration c_U at 298.15 K and 1 bar. a) Mass densities ρ. b) Apparent molar volumes $^{\phi}V_U$. c) Logarithmic concentration derivative of the activity coefficient y_{UU} (eq. 2.12). d) Activity coefficient of urea. The simulation data is the analytically integrated quadratic fit of y_{UU}. e) Dielectric constant ϵ. f) Diffusion coefficient of urea D_U. g) Adiabatic (isentropic) compressibility κ_S. The simulations use the new force field that was optimized for 1 bar (see appendix B.3).
Experimental data from ref. 119 was used in panels (a-b). For (c-d), the data was obtained from ref. 118. Experimental dielectric constant data in (e) from ref. 125. Diffusion constant data (f) was taken from ref 122. Experimental compressibilities (g) are from ref. 123.

4.2.6 High pressure force field adaptation

We have found previously that simulating aqueous TMAO at high pressure, while using the same force field as for ambient conditions, will not result in the correct solvation structure, i.e. the number of hydrogen bonds. This was corrected by scaling the charges of the force field to reproduce changes in the molecular dipole moment from quantum chemistry calculations (sec. 4.1.4 or ref. 78). We applied the same scaling method to our urea force field and show the H-bond distributions at 10 kbar in figure 4.15 under the name HP-ff. The HP-ff overestimates the number of H-bonds at oxygen and hydrogen atoms, and underestimates the H-bonds accepted by the nitrogen atoms. It performs worse at reproducing the H-bond numbers of aiMD than the unmodified force field at all sites. One possible explanation for this is that, whereas TMAO only has one major contribution to the dipole moment in the N-O bond, the case of the urea molecule is too complex for only considering the total dipole moment. Instead, we propose a different charge scaling method that reproduces the relative changes of the dipole moments of each individual chemical bond. The general methodology for calculating the charges from EC-RISM results can be found in the appendix (sec. D). The H-bond distributions of this force field modification are shown in fig. 4.15 under the name μ_{bond}-HP-ff. However, we find that this method is only a minor improvement over the HP-ff scaling method. We see that the best way to reproduce the H-bonds from quantum chemical simulations is still to use the force field that was optimized for 1 bar, which works extremely well even at 10 kbar. We therefore decided to use the unmodified urea force field at all pressures throughout this work.

The methods of scaling the charges for adapting force fields for high pressures unexpectedly performed worse than the original force field for urea. For TMAO, the EC-RISM dipole moment increases from 7.90 D at 1 bar to 8.44 D at 10 kbar (6.8 %), whereas it changes from 6.64 D to 7.06 D (6.3 %) in urea. The relative changes are very similar, and even the absolute changes differ by only 0.1 D. At this time, we have no consistent way to tell whether charge scaling is necessary or not. For now, we have to decide on a case-by-case basis.

4.2.7 Urea-urea PMFs

Now that we established which force field to use at high pressures, we are interested in the influence of pressure on the potentials of mean force between two urea molecules in dilute solutions. Figure 4.18 contains the PMFs of our urea force field as a function of the urea-urea distance. The first minimum is at 0.44 nm with a depth of 2.3 kJ/mol at 1 bar. With a pressure increase to 10 kbar, the position shifts to 0.42 nm, and the depth stays constant within the accuracy. The free energy profile only changes noticeably at larger distances, when the urea molecules are separated by at least one water solvation shell. At the first maximum beyond direct urea contacts, the pressure change causes the free energy barrier to shift from 0.62 to 0.59 nm, and the peak height to increase from 0.48 to 1.14 kJ/mol. This trend continues for the decaying oscillations of the PMF for larger distances. It is possible to calculate the RDF from the PMF as follows:

$$g(r) = \exp\left(-\frac{\mathrm{PMF}(r)}{\mathrm{k}T}\right) \tag{4.11}$$

In figure 4.19, we compare the RDFs at infinite dilution with those in concentrated systems. At 1 bar, we find that the first peak at 0.44 nm decreases with intensity from 2.5 to 2.14 with the addition of 8 mol/L urea. The first minimum is lowered slightly from 0.82 to 0.73, and a very minor but measurable positional shift from 0.615 nm to 0.609 nm can be

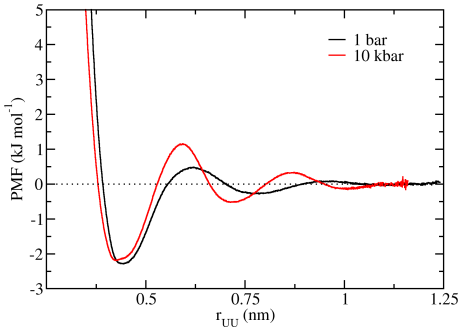


Figure 4.18.: Potential of mean force calculated from umbrella sampling as a function of urea-urea center of mass distance at 300 K. Black: 1 bar. Red: 10 kbar. The new force field (see appendix B.3) was used at both pressures.

observed. The positions of the second maxima shift from 0.77 nm to 0.82 nm and, with the exception of the infinitely dilute system, the maxima increase in intensity with increasing urea concentration. This deviation of the dilute case from the trend is simply explained by the uncertainty of the long-range values for the PMF from umbrella sampling. Although the qualitative behavior of aqueous urea is the same at 10 kbar, the compression causes a notable difference: The first maximum at 0.42 nm, which decreases from 2.4 to 2.2 in intensity, is asymmetric and we assume that it contains two distinct contributions. The first minimum at 0.59 nm is lower than at ambient pressure with values in the range from 0.63 to 0.57. An in urea concentration also shifts the second maximum from 0.71 nm to 0.73 nm, and the intensity decreases from 1.23 to 1.19.

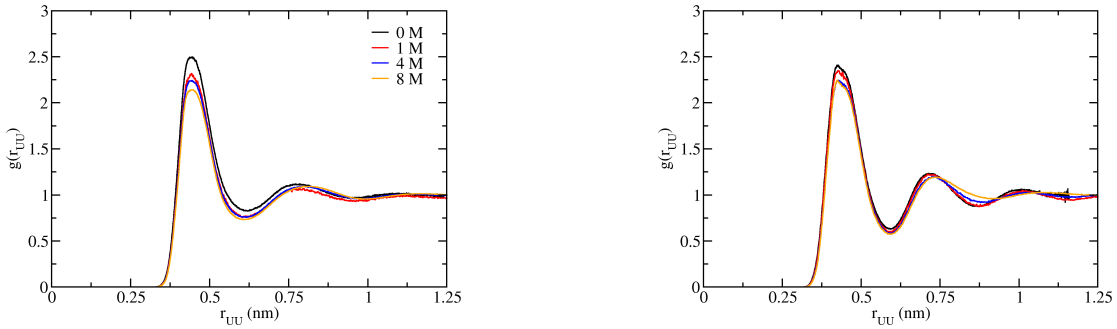


Figure 4.19.: Center of mass urea-urea radial distributions at 1 bar (a) and 10 kbar (b). The RDFs at infinite dilution were calculated from the PMFs using eq. 4.11. The concentrations are the values of the 1 bar systems. The new force field (see appendix B.3) was used at both pressures.

The result that urea molecules are slightly excluded from each other at higher concentrations, which has also been found for other urea models,[117] is not to be confused with an increase in unfavorable urea-urea interactions. At least at 1 bar (fig. 4.17 d), the activity coefficient of urea is slightly negative and mostly decreases with concentration.

In addition to the PMFs, we can extract the relative orientations of the urea molecules to each other from the umbrella sampling simulations. 2D distributions of the angles ϕ and ψ are shown in figures 4.20 (a) for 1 bar and (b) for 10 kbar at different urea-urea separations. Here, ϕ is the angle between the C-O vector of the first urea and the vector connecting the centers of mass of the urea molecules, and ψ refers to the angle

between the two C-O vectors. We numbered four characteristic orientations in (a) and show representative snapshots in subfigure (c). Obviously, the spatial orientation is not fully defined by only two angles. However, with the distance as the third variable we can reliably assign structures to the peaks in the histograms. First we note that the 2D distributions are very similar at both pressures. At short distances, we observe two main orientations with close van der Waals contacts. The most likely orientation is the antiparallel stacking of the ureas (structure 1) due to the favorable electrostatics, which is not present in parallel stacking (structure 2). Around the position of the PMF minimum, the orientations are shifted towards in-plane aligned C-O vectors, but the molecular planes are not necessarily parallel. In the more favorable orientation (structure 4), H-bonds are formed between oxygen and one cis-hydrogen of each urea. The electrostatically less optimal orientation (structure 3) has two trans-hydrogens sharing the oxygen of the second urea as an H-bond acceptor. The high probability region of this structure can be identified in the panels for shorter urea distances, where it is shifted to lower $\cos\psi$ values. Here, there exists no single characteristic orientation, but one example is shown as structure 5. Finally, when going to distances above 5 Å, no strong orientational preference is discernible. This orientational analysis gives us a possible explanation for the urea-urea exclusion at higher concentrations. It may be that the dimer structures 3 and 4 (figure 4.20 c) are less favorable at high concentrations simply because the urea density is too high for solvent-separated dimers to form. For a definitive answer of this question a quantitative study of urea orientations in concentrated solutions will be necessary.

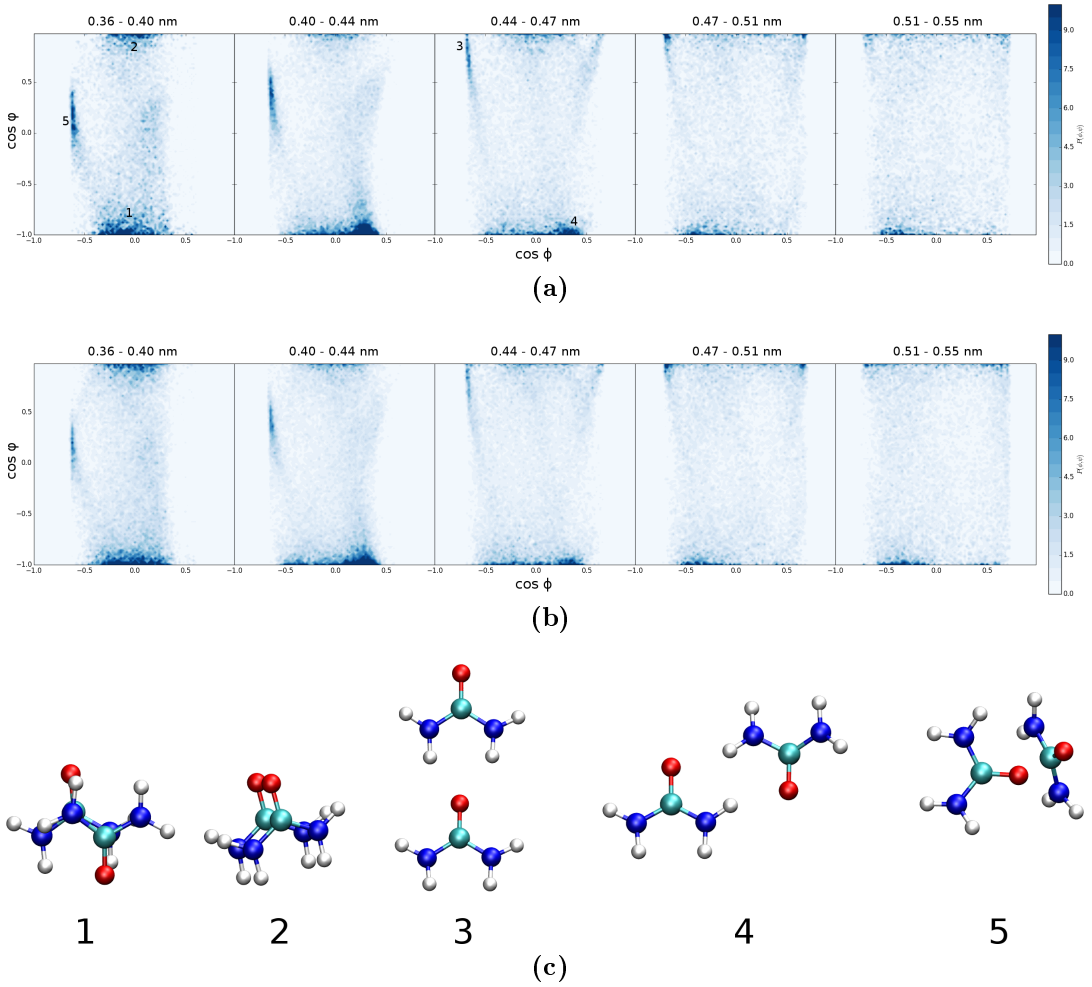


Figure 4.20.: Probability distributions of the angles ϕ and ψ between two urea molecules at 1 bar (a) and 10 kbar (b). ϕ is the angle between the C-O vector of one urea and the vector between the centers of mass of both urea molecules, and ψ is the angle between the two C-O vectors. The urea-urea distance increases for the panels from left to right. The numbers in the distributions of (a) correspond to four characteristic orientations. Snapshots of these orientations are shown in (c). The new force field (see appendix B.3) was used at both pressures.

4.3 Conclusions and outlook

During the course of this chapter, we characterized binary osmolyte/water mixtures of one protecting osmolyte, TMAO, and one denaturing osmolyte, urea, at different pressures. For the case of TMAO, we introduced a scaling procedure for the force field charges to reproduce the correct electrostatics of TMAO at high pressures as calculated from quantum chemical methods. For this method, we assumed that the initial force field is reasonably well optimized with respect to ambient conditions, and apply only small corrections to the point charges, where the magnitude of the adjustment was obtained from EC-RISM calculations. The end result was a set of charges for each pressure, which included the increasing polarization of TMAO with pressure found in *ab initio* molecular dynamics simulations.

Out approach for urea ended up being very different. First we found the need for the development of a new urea force field for ambient conditions. The force field parameters were optimized using a evolutionary algorithm to best reproduce experimental densities as well as solvation structure properties of aiMD simulations. After finding good agreement of many thermodynamic properties between the new model and experiments at 1 bar, it turned out that the force field also works very well at increased pressures without any pressure-dependent adjustments. In fact, we found no charge scaling method that performed better than the original force field. It is unclear why charge scaling works for TMAO but not for urea, especially since both the absolute molecular dipole moments as well as the relative change with pressure in the QC methods are very similar.

For future work, the most important question that needs to be answered is why the scaling of force field charges is necessary to account for the polarization of TMAO, but not for urea. It has to be investigated whether there exists an observable that tells us *a priori* if the solute force field requires modifications. Since the structural properties of the widely used KBFF urea model are not in line with aiMD reference data, conclusions drawn from simulations using this force field should be reinvestigated and validated with the model that was developed in this chapter. This includes studies of the molecular protein denaturation mechanism of urea, e.g. the balance of van der Waals and electrostatic interaction energies with proteins, which could be significantly different with our model. Finally, our models are now ready for application in biologically relevant systems, i.e. aqueous mixtures of TMAO and urea in bulk as well as their use as cosolvents of biomolecules.

Chapter 5

Periodic homopeptides under pressure and in osmolyte solutions

One main goal of this thesis is the investigation of osmolyte effects on proteins at different pressures (see chapter 1 for an introduction). In this chapter, the change in unfolding free energies of model peptides with TMAO concentration are the topic of interest. Horinek and Netz were the first to simulate peptides that extend across the periodic boundary of the simulation box.[116] This approach allows an analysis of fixed secondary structures while avoiding errors introduced by end groups. We use homopeptides of glycine and alanine to determine the effect of TMAO on the unfolding free energies of α- and 3_{10}-helices at normal pressure and at 5 kbar. Additionally, we use a recent implementation of Tanford's transfer model[34,126] to see whether the free energy of transfer (TFE) of a peptide to TMAO solutions is proportional to its solvent-accessible surface area (SASA).

5.1 The transfer model

The transfer model, which was introduced by Tanford in 1964,[126] is a method to systematically calculate the influence of urea on the unfolding free energy of proteins. The unfolding process in water and in an osmolyte solution can be connected via a theoretical thermodynamic cycle that contains the transfer processes of both the folded and unfolded protein from water to the osmolyte mixture (fig. 5.1). For this cycle to be exact, we need

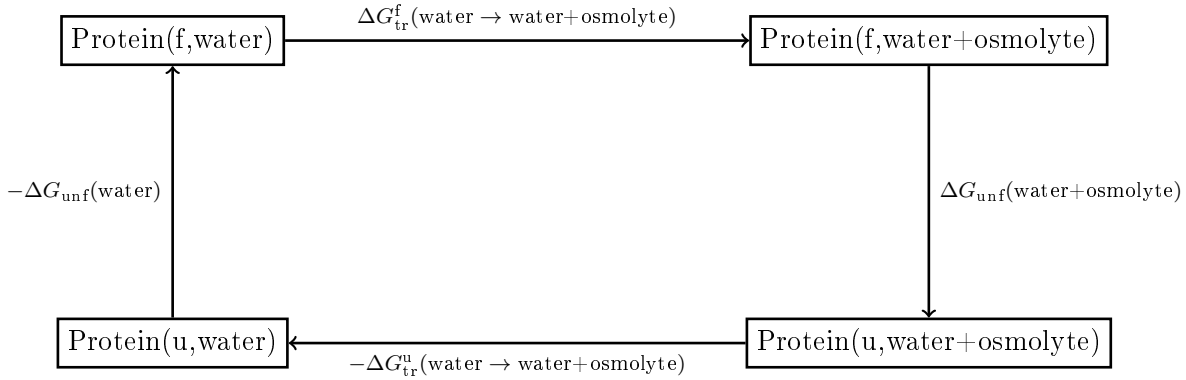


Figure 5.1.: The thermodynamic cycle used to express a difference in unfolding free energies of a protein as a difference of transfer free energies.

to assume that there exists only one unfolded state, and that it is identical in both solvent systems. Furthermore, it is assumed that the protein exists only in two discrete states, unfolded (u) and folded (f). If these assumptions hold, the sum of free energy changes of the cycle is zero, and the difference in unfolding free energies is equal to the difference of the free energies of the transfer of the unfolded and folded protein from water to the osmolyte solution:

$$\Delta\Delta G_{\mathrm{unf}} \equiv \Delta G_{\mathrm{unf}}(\text{water+osmolyte}) - \Delta G_{\mathrm{unf}}(\text{water}) = \Delta G^{\mathrm{u}}_{\mathrm{tr}} - \Delta G^{\mathrm{f}}_{\mathrm{tr}} \tag{5.1}$$

Now the problem is reformulated to a calculation of the transfer free energies (TFEs) of two different protein conformations to osmolyte solutions. In the first general implementation of the transfer model, Auton and Bolen[127,128] calculated TFEs of proteins by adding the TFEs of individual amino acids. The contributions of the protein backbone and side chains, which are again assumed to be additive, are scaled with the area that is in contact with the solvent in the protein conformation of interest, since only the amino acids that interact with the solvent can contribute to the TFE. However, this version of the transfer model is at odds with many studies that show that the backbone and side chains add contributions that favor the unfolded state at higher urea concentrations. Moeser and Horinek improved upon the initial transfer model by introducing the universal-backbone transfer model,[34] which will be applied in this work. The contribution of a single amino acid *aa* to the TFE is calculated by scaling the known values of the TFE per solvent-accessible surface area (SASA) with the SASA of the side chain (*sc*) and the backbone (*bb*) of the amino acid in the protein conformation:

$$\mathrm{TFE} = \sum_{aa} \mathrm{TFE}_{aa} \tag{5.2}$$

$$\mathrm{TFE}_{aa} = \mathrm{SASA}^{sc}_{aa} \cdot \left(\frac{\mathrm{TFE}}{\mathrm{SASA}}\right)^{sc}_{aa} + \mathrm{SASA}^{bb}_{aa} \cdot \left(\frac{\mathrm{TFE}}{\mathrm{SASA}}\right)^{bb} \tag{5.3}$$

It differs from the Auton/Bolen model in the assumption that the protein backbone has the same TFE per SASA value in all amino acids (see the missing subscript *aa* in the last term of eq. 5.3). It is assumed that glycine is a good representation of the protein backbone. Therefore, the TFE of glycine is used as the backbone TFE. Experimentally, this is usually realized by measuring the TFE of a cyclic glycylglycine gly_2.[127] The side chain contribution per SASA is calculated from the total TFE of the amino acid in a fixed conformation with known SASA via eq. 5.3. The TFE of the side chain of an amino acid is the difference between the total TFE and the backbone contribution:

$$\left(\frac{\mathrm{TFE}}{\mathrm{SASA}}\right)^{sc}_{aa} = \frac{\mathrm{TFE}_{aa} - \left(\frac{\mathrm{TFE}}{\mathrm{SASA}}\right)^{bb} \cdot \mathrm{SASA}^{bb}_{aa}}{\mathrm{SASA}^{sc}_{aa}} \tag{5.4}$$

In the following section, we will explain in detail how to calculate the quantities necessary for applying the transfer model to predict TFEs of proteins in arbitrary conformations.

5.2 Calculating transfer free energies from simulations

First we need to calculate the surface area of a protein or peptide from a simulation trajectory. For this, we use the double cubic lattice algorithm[129] as implemented in `GROMACS` with the vdW radii by Lesser and Rose,[130] with a probe radius of 0.14 nm and a resolution of 500 points per sphere.

Next, the free energies of transfer have to be extracted from simulations. Shimizu derived a relation between the TFE and the excess (or depletion) of water and solute (i. e. osmolyte) around a macromolecule using Kirkwood-Buff theory.[131] Based on this relation, Horinek and Netz derived an exact expression that connects the derivative of the TFE with respect to the cosolvent (in our case TMAO, denoted as T) concentration to quantities accessible from simulations (eq. 5.5).[116]

$$\frac{\mathrm{d}\Delta G_{\mathrm{tr}}}{\mathrm{d}c_{\mathrm{T}}} = -\mathrm{k}T\bar{\Gamma}\left[1 + \left(\frac{\partial \ln y_{\mathrm{T}}}{\partial \ln c_{\mathrm{T}}}\right)_{p,T}\right] \tag{5.5}$$

As a model system for peptides, they used extended strands of periodic homopeptides that extend across the periodic boundary of the simulation box in one dimension. The difference between the relative excesses of TMAO (T) and water (W) around the peptide (P), denoted as $\bar{\Gamma}$, is then the integral over the difference between the peptide-water and peptide-cosolvent distribution functions $g(r)$, where r is the radial distance in cylindrical coordinates between the centers of mass of the molecules:

$$\bar{\Gamma} = \left[\frac{\Gamma_{\mathrm{T}}}{c_{\mathrm{T}}} - \frac{\Gamma_{\mathrm{W}}}{c_{\mathrm{W}}}\right] = \int_0^\infty \left(g_{\mathrm{PT}}(r) - g_{\mathrm{PW}}(r)\right) 2\pi r \mathrm{d}r \tag{5.6}$$

In practice, we calculate $\bar{\Gamma}$ by taking the average of the running integral in the range $1.5 - 1.6$ nm. The RDFs were divided by the average value in the same range to ensure that the integral can in principle converge to a constant value. The bracketed term in eq. 5.5 contains the molarity scale activity coefficient derivative of the cosolvent, which is obtained from bulk simulations of cosolvent/water mixtures using Kirkwood-Buff theory (see eq. 2.12). The quadratic fits of the activity coefficient derivative of TMAO (y_{TT} of the high pressure force field, see fig. 4.3) at 1 bar and 5 kbar have the following form:

$$y_{\mathrm{TT}}^{1\ \mathrm{bar}}(c_{\mathrm{T}}) = 0.097857 + 0.13418 \cdot c_{\mathrm{T}} + 0.12302 \cdot c_{\mathrm{T}}^2 \tag{5.7}$$

$$y_{\mathrm{TT}}^{5\ \mathrm{kbar}}(c_{\mathrm{T}}) = 0.20645 + -0.14404 \cdot c_{\mathrm{T}} + 0.18136 \cdot c_{\mathrm{T}}^2 \tag{5.8}$$

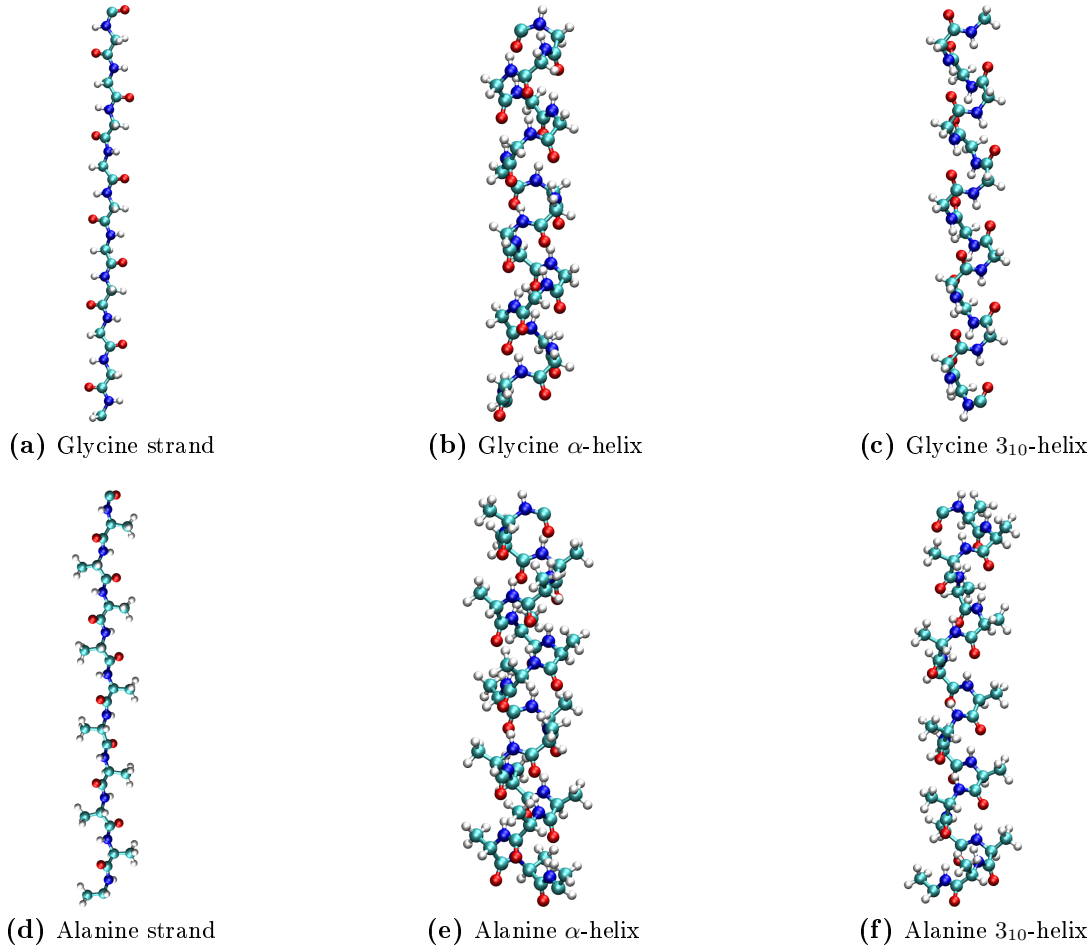
Finally, an integration of eq. 5.5 yields the TFE of the periodic homopeptides as a function of cosolvent concentration.

5.3 Simulation methods

We simulated periodic homopeptides of glycine and L-alanine at 1 bar and 5 kbar for a series of TMAO concentrations. The last amino acid in the peptide chain is bonded to the first amino acid across the periodic boundary of the simulation box in the z direction. Therefore, the box size in this direction was kept constant, and we used a semiisotropic barostat to scale the x and y dimensions, which were approximately 4.5 nm in size. Three different secondary structures were simulated (fig. 5.2): Extended strand, α-helix, and 3_{10}-helix. The extended strand is obtained by stretching the peptide along the z axis with a length of 3.8 nm with 10 amino acids in the system. The helical conformations of glycine and alanine, however, are generally not stable over the duration of our simulations. The helices were obtained by applying harmonic distance restraints between the carbonyl oxygen and the amide hydrogen in the backbone with an equilibrium distance of 0.202 nm and a force constant of $10^4\,\mathrm{kJ\,mol\,nm^{-2}}$. For the α-helix, the H-bond restraints are between amino acids i and $i+4$, which is approximately one turn in the helix (the exact value is 3.6 amino acids per turn). The α-helix is composed of 18 amino acids over a length

of 2.68 nm. In similar fashion, the 3_{10}-helix is stabilized by restraints between residues i and $i+3$. Each 3_{10}-helix contains 15 amino acids over 3.0 nm. For TMAO we use our force field parameters (see appendix B.2). Notably, we use the charge-scaled parameters for the 5 kbar simulations. The AMBER94[132] force field is used for the glycine and alanine strands without any modifications at high pressure. The exact version of AMBER is not important, since they differ only in the torsional potentials, which are irrelevant for fixed conformations. Depending on the conformation, between 5 and 17 concentrations for TMAO concentrations from 0.5 to 3.5 mol/L were simulated for 300 ns each. The time step was chosen as 1.5 fs, and all bonds to hydrogen atoms were constrained.

Additionally, we performed umbrella sampling simulations in order to determine the PMF between one TMAO molecule and the different amino acids and secondary structure types in water. Between 42 and 46 umbrella windows for TMAO-peptide distances from 0.20 to 2.0 nm at 1 bar and 0.20 to 1.84 nm at 5 kbar were simulated for 75 ns each. The force constant of the harmonic potential was 1500 kJ mol^{-1} nm^{-2}.



(a) Glycine strand **(b)** Glycine α-helix **(c)** Glycine 3_{10}-helix

(d) Alanine strand **(e)** Alanine α-helix **(f)** Alanine 3_{10}-helix

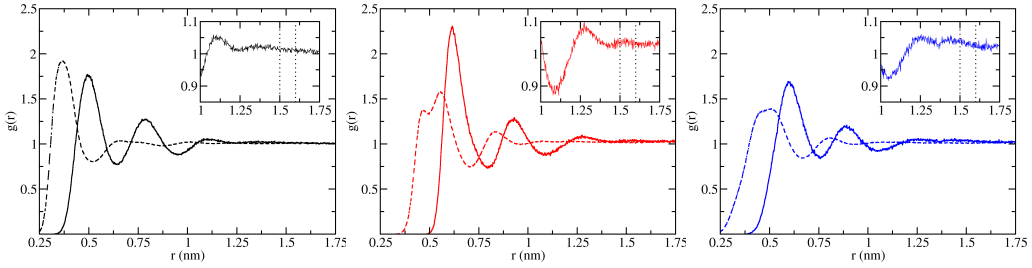
Figure 5.2.: Snapshots of the glycine and L-alanine periodic homopeptides in the different secondary structures.

5.4 Results

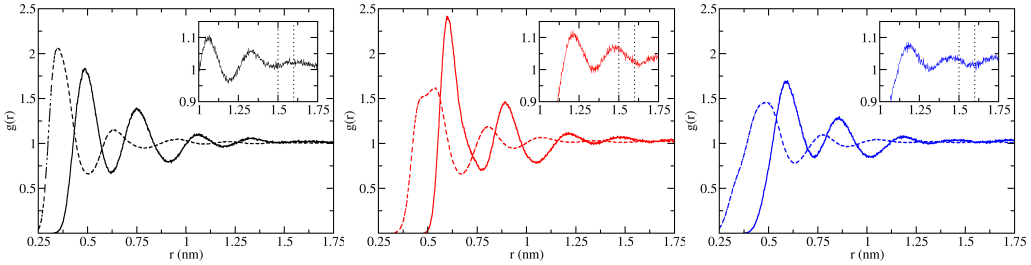
5.4.1 Transfer from water to TMAO solutions

The first step in the calculation of the transfer free energy of the periodic homopeptides to TMAO solutions is the determination of the relative excess of TMAO around the peptides. In fig. 5.3, we show the peptide-TMAO and peptide-water RDFs for the system with the highest TMAO concentration for the extended strands, α-helices, and 3_{10}-helices of glycine and alanine at 1 bar and 5 kbar. All RDFs show qualitatively that TMAO is excluded from glycine and alanine with a water layer between them. In the alanine systems, the RDFs have multiple peaks close to each other. The peak at shorter distances can be attributed to the distribution around the backbone, whereas the peak further away is the distribution around the side chain.

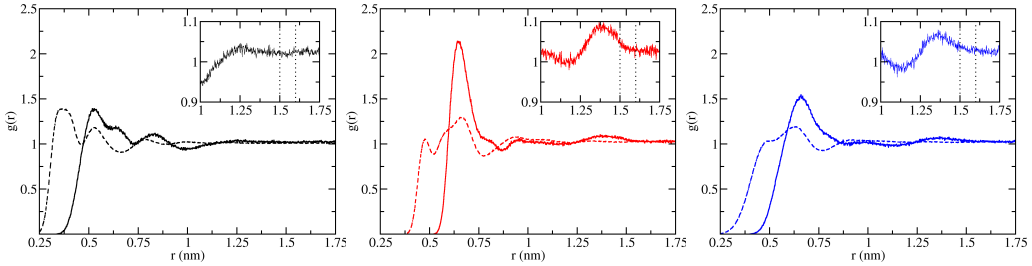
Next, the RDFs are divided by their bulk value, and we integrate over the difference between the TMAO and water RDFs (eq. 5.6) to obtain the relative excess coefficient $\bar{\Gamma}$ (fig. 5.4). We find that, with one exception, all systems exhibit a depletion of TMAO around the peptide within the uncertainty, which confirms the qualitative result obtained from the RDFs. Linear fits of $\bar{\Gamma}$ as a function of TMAO concentration are shown, even though the predictive power of linear regressions is limited due to the generally large error estimates of the individual values. Around glycine, an increasing excess with increasing TMAO content is found for the extended strand at both pressures. The glycine α-helix has an approximately constant $\bar{\Gamma}$ at 1 bar, and an increasing excess at 5 kbar. For the 3_{10}-helix, the trend of increasing TMAO excess at 1 bar inverts at 5 kbar. We find a more consistent picture for alanine peptides, where the α-helices have the largest TMAO excess, followed by 3_{10}-helices and the extended strands with the lowest excess. At 1 bar, the excess around all secondary structures increases with TMAO concentration and stays negative above 3 mol/L. The trend inverts when going to 5 kbar, where the excess is lowered with concentration and is even slightly positive for the α-helix at 1.2 mol/L. One would generally not expect that the relative amount of TMAO decreases with increasing TMAO concentration, but it already seems to be larger in dilute solutions than the maximum value at 1 bar. An important factor may be the activity coefficient, which is noticeably different at 5 kbar and which plays an important role in the calculation of the actual transfer thermodynamics. By multiplication of $\bar{\Gamma}$ with the term containing the activity coefficient derivative $(-1 - y_{\mathrm{TT}})$, the concentration derivative of the free energy of transfer is obtained (fig. 5.5). We find that the TFE derivative $\frac{\mathrm{d}\Delta G_{\mathrm{tr}}}{\mathrm{d}c_{\mathrm{T}}}$ increases with TMAO concentration for all systems and pressures and is on the order of one kJ per mole of amino acid divided by the TMAO concentration in mol/L. The highest values are found for the extended strand, followed by the 3_{10}-helix and the α-helix with the lowest values. In order to obtain the total TFEs, linear fits of $\frac{\mathrm{d}\Delta G_{\mathrm{tr}}}{\mathrm{d}c_{\mathrm{T}}}$ were integrated analytically (fig. 5.6). Now that we have values for the TFEs of the three different conformations, it is possible to check whether the transfer model holds for TMAO, i.e. whether the TFE is proportional to the SASA. For this, we require a value for the TFE per SASA of the peptide backbone, written as $\left(\frac{\mathrm{TFE}}{\mathrm{SASA}}\right)^{bb}$ in equations 5.3 and 5.4. We use the extended glycine strand for the protein backbone. The total TFE and SASA values of alanine, which are needed for the calculation of the side chain contribution in alanine, are likewise obtained from the extended strand system. The dashed lines in figures 5.5 and 5.6 are predictions of the derivative of the TFE and the TFE for the two helix types. All SASA values and fit parameters for the calculation of TFEs can be found in appendix G.



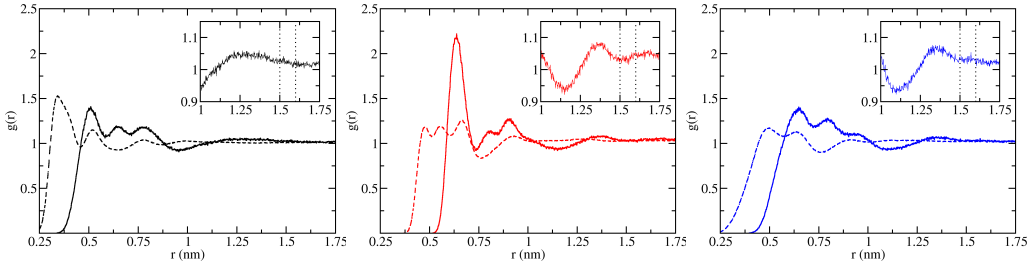
(a) Glycine 1 bar



(b) Glycine 5000 bar

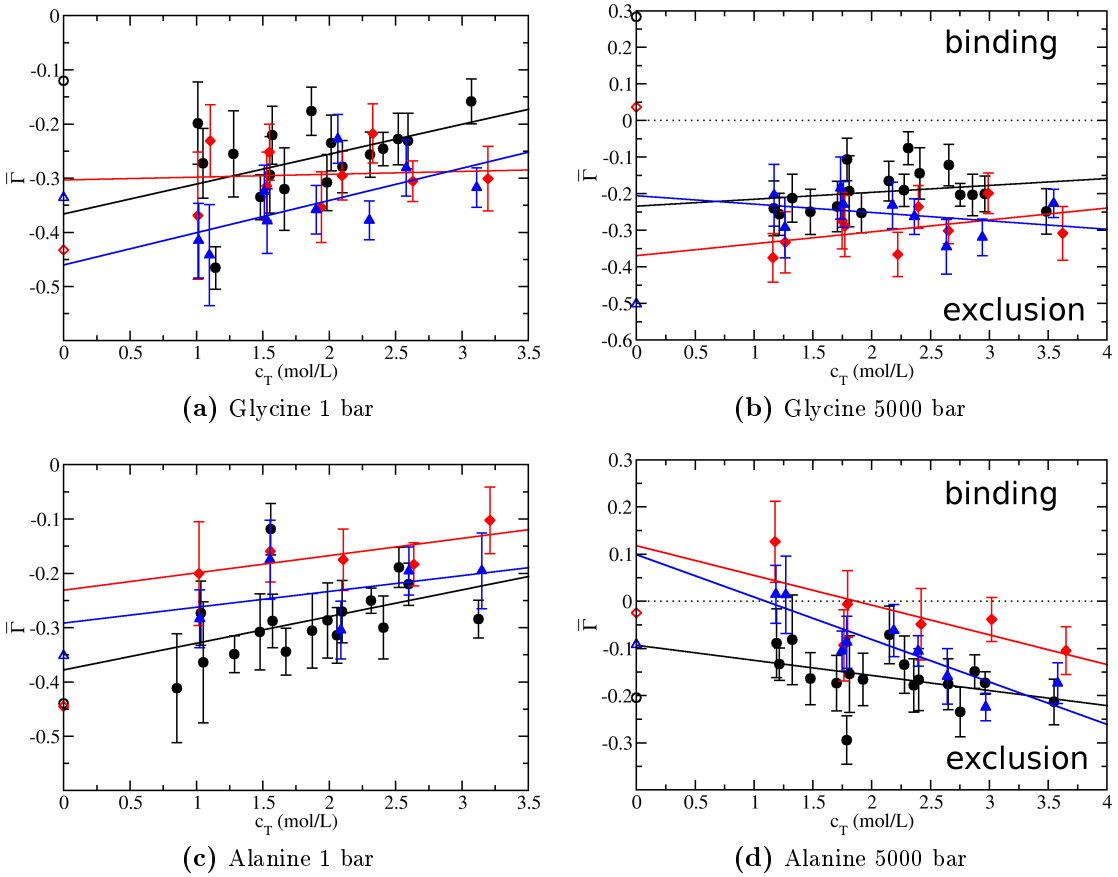


(c) Alanine 1 bar



(d) Alanine 5000 bar

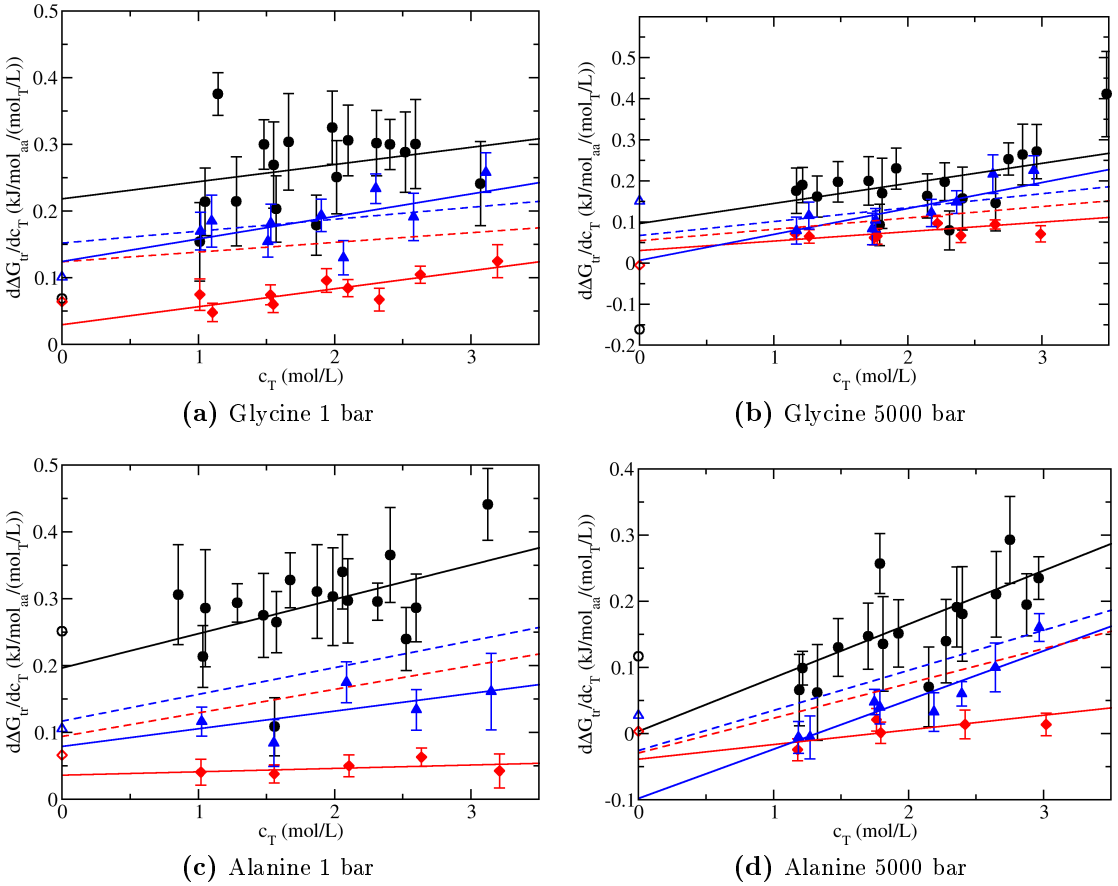
Figure 5.3.: Radial distribution functions between the periodic peptide and TMAO (full lines), and the peptide and water (dashed lines) at approximately $3-3.5$ mol/L. The colors mark the different secondary structures: Extended strand (black), α-helix (red), and 3_{10}-helix (blue). The insets show the tails of the peptide-TMAO RDFs and the range over which the bulk value and excess $\bar{\Gamma}$ are averaged (dotted lines).



(a) Glycine 1 bar **(b)** Glycine 5000 bar

(c) Alanine 1 bar **(d)** Alanine 5000 bar

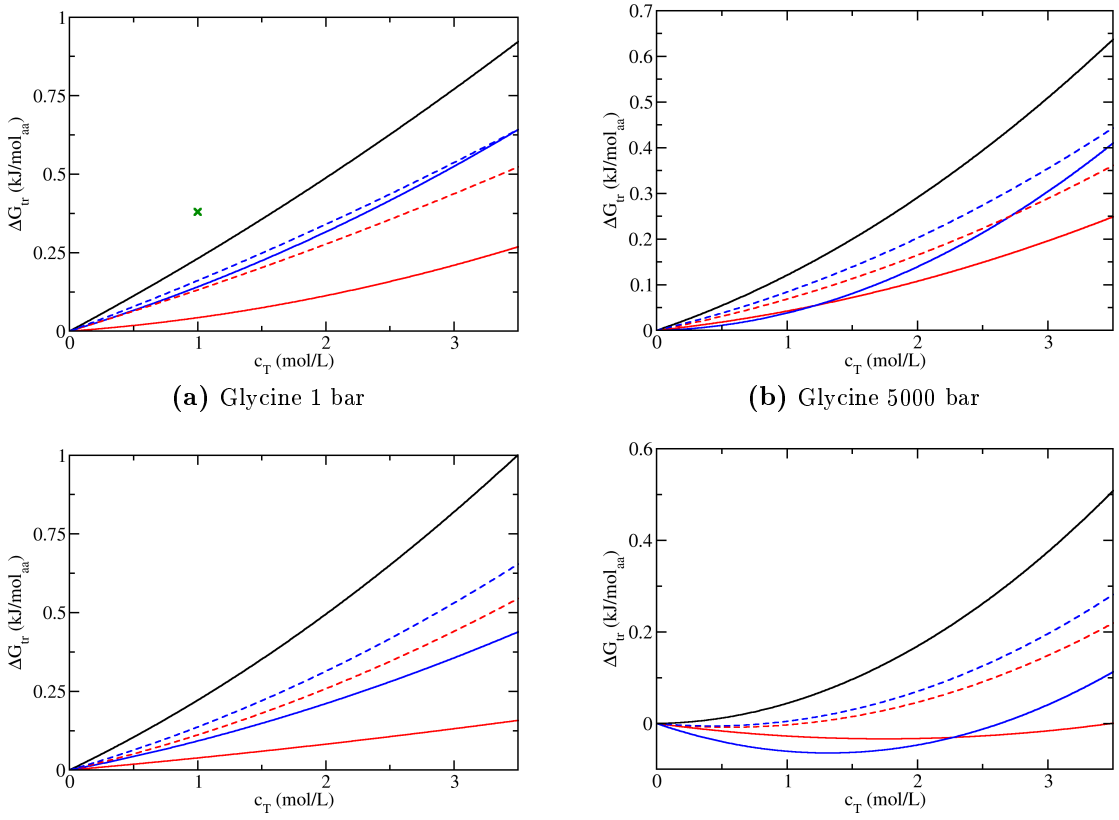
Figure 5.4.: The relative excess coefficient $\bar{\Gamma}$ of TMAO around the periodic peptides calculated from eq. 5.6. The values were obtained by taking the average over the running integral between $1.5 - 1.6$ nm. Secondary structures: Extended strand (black), α-helix (red), and 3_{10}-helix (blue). Error estimates are obtained by splitting each trajectory into 10 blocks of 30 ns and calculating the standard deviation of the mean of the blocks. Lines are linear fits. The empty symbols are values at infinite dilution from umbrella sampling (see below), which were not used in the fits.



(a) Glycine 1 bar **(b)** Glycine 5000 bar

(c) Alanine 1 bar **(d)** Alanine 5000 bar

Figure 5.5.: The derivative of the transfer free energy of the peptides to TMAO solutions for extended strands (black), α-helices (red), and 3_{10}-helices (blue). Full lines are linear fits, and dashed lines are predictions of the TFEs of the two helix types from using the TFE per SASA values of the extended strand (see below). Empty symbols are the values from umbrella sampling at infinite dilution (see below), which were not used in the fits.



(a) Glycine 1 bar

(b) Glycine 5000 bar

(c) Alanine 1 bar

(d) Alanine 5000 bar

Figure 5.6.: Transfer free energies of periodic glycine and alanine to TMAO solutions for extended strands (black), α-helices (red), and 3_{10}-helices (blue). Full lines are the values from analytic integration of the linear fits in fig. 5.5. Dashed lines are obtained from scaling the TFE per SASA of the extended strands with the SASA of the helices. The single experimental value for glycine at 1 bar (green) was taken from the supporting information of ref. 128.

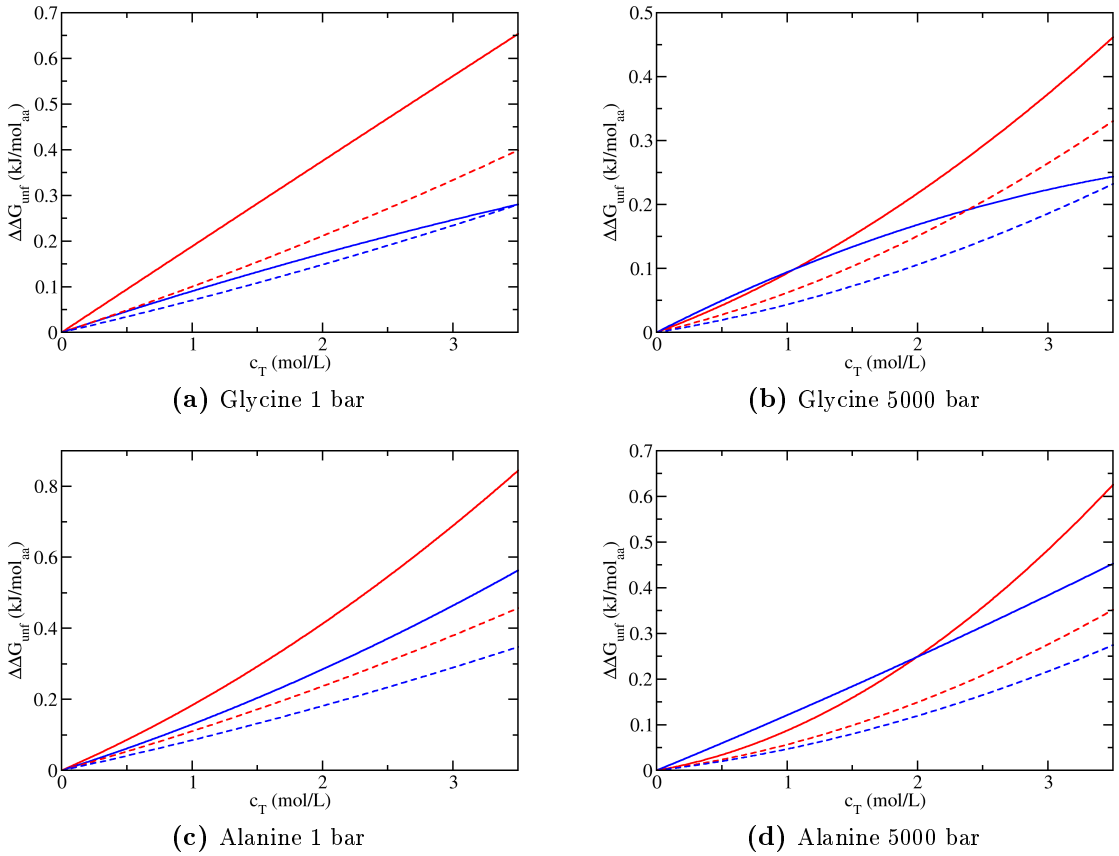
The TFE of the glycine systems at 1 bar is positive over the whole concentration range and increases up to about 1 kJ/mol for the extended strand. We show experimental data for the TFE of a cyclic gly_2 molecule to a 1 mol/L TMAO solution (SI of ref. 128). This is the closest we can get to a reference value for our periodic systems, because in short oligopeptides the zwitterionic character is problematic, and longer peptides assume conformations that are no longer a good approximation for an extended strand. This experiment gives a value of 0.38 kJ/mol, which is significantly above the value from simulations for the extended strand (0.23 kJ/mol) in relative terms, but acceptable considering the generally small energy scales we are dealing with. The TFEs of the 3_{10}-helix are lower than those of the extended strand, and the α-helix is least affected by TMAO. This is qualitatively in line with the transfer model, since the SASAs of the peptides decrease in the same order. However, the actual values obtained from scaling the TFE of the extended strand with the SASAs of the helices only quantitatively predict the TFEs of the 3_{10}-helix, and differ by up to 0.25 kJ/mol (50 %) in the case of the α-helix. At 5 kbar, the order of TFEs is different up to 1.25 mol/L, where the α-helix has a slightly higher TFE than the 3_{10}-helix. This is most likely not a significant result and would have to be confirmed in future work with additional sampling. The predictions perform reasonably well for the 3_{10}-helix, but are still too high for the TFE of the α-helix. We note that the TMAO concentrations are the actual concentrations at 5 kbar and do not correspond directly to the 1 bar systems. A comparison between the pressures with a common concentration scale is shown below.

We find the same qualitative behavior for alanine as for glycine at 1 bar. Again, the SASA-scaling method predicts TFEs that are too high with differences up to 0.2 kJ/mol for the 3_{10}-helix and 0.4 kJ/mol in the α-helix system at 3.5 mol/L TMAO. At 5 kbar, the extended strand follows the same trend and increases up to 0.5 kJ/mol at 3.5 mol/L. The helix simulations as well as the predictions have a local minimum at negative TFEs, although the SASA-scaled predictions are almost exactly zero. If this result could be verified quantitatively, this would mean that both helix types are actually more stable in TMAO solutions than in water at high pressures.

The goal of the transfer model is the prediction of the influence of cosolvents on unfolding free energies, $\Delta\Delta G_{\mathrm{unf}}(c_{\mathrm{T}})$. It can be expressed as the difference of TFEs by constructing a thermodynamic cycle (fig. 5.1, eq. 5.1). Then we can calculate the change in unfolding free energies of the helices by taking the difference between the TFEs of the extended strands and the helices (fig. 5.7):

$$\Delta\Delta G_{\mathrm{unf}} = \Delta G_{\mathrm{tr}}^{\mathrm{ext}} - \Delta G_{\mathrm{tr}}^{\mathrm{helix}} \tag{5.9}$$

Since we want to compare the values for the same system at 1 bar and 5 kbar, we transformed the molar concentration at 5 kbar, which depends on the system volume, to the concentration at 1 bar (only in fig. 5.7). The exact conversion procedure is shown in appendix H, which can also be used to transform the analytical expressions for the TFEs. Throughout all systems, $\Delta\Delta G_{\mathrm{unf}}$ is positive and increases with the TMAO concentration, which means that TMAO stabilizes the (folded) helical states compared to the extended conformation. It is a very important result that our TMAO model also stabilizes the helices at 5 kbar with $\Delta\Delta G_{\mathrm{unf}}$ values very similarly to 1 bar. However, we could not definitively confirm the proportionality of the transfer free energies to the solvent-accessible surface area of the peptides due to the sub-kJ/mol energy scale we are dealing with. A main issue is the choice of the integration range of the peptide-TMAO radial distributions, which can lead to quantitatively different TFEs due to statistical noise in the tail of the distributions. On a qualitative level, we were able to show that simulations of model peptides in conjunction with our TMAO model predict a stabilization of helical conformations against unfolding even at pressures of several kilobars.



(a) Glycine 1 bar

(b) Glycine 5000 bar

(c) Alanine 1 bar

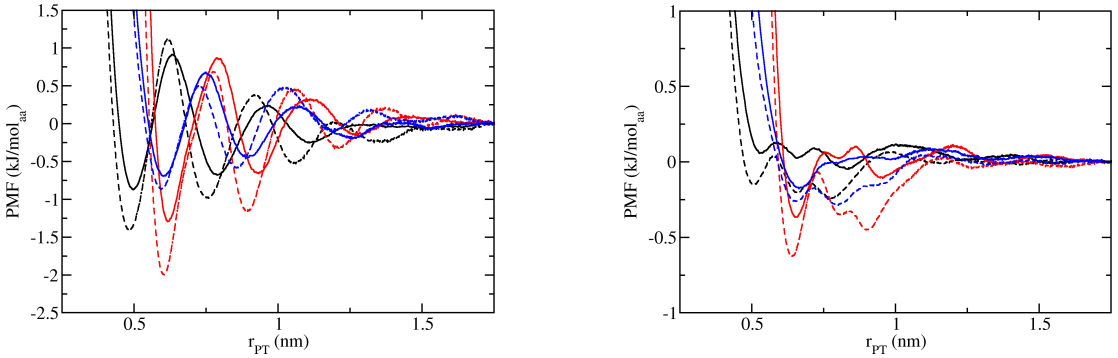
(d) Alanine 5000 bar

Figure 5.7.: Change in α-(red) and 3_{10}-helix (blue) unfolding free energies between water and TMAO solutions. Calculated by taking the difference between the extended strand and helix TFEs (eq. 5.9).

A comparison of our results for glycine at 1 bar with simulations by Schneck et al., who used the precursor of our TMAO model, gives good agreement for the exclusion from the protein and for the TFEs, but no data for the helix stability is available.[104] Our results also agree with recent work by Ganguly et al., in which simulations of multiple TMAO models, including this one, show that TMAO stabilizes compact conformations of an Ala_{15} oligopeptide.[36] Rodríguez-Ropero et al. have found that the unfolding free energy of a hydrophobic model polymer first increases with TMAO concentration, but that the effect inverts at very high concentrations.[42] We find no such dependence for our TMAO model in combination with our glycine and alanine helices. It may be necessary to test our force field for this model polymer to see if their results can be reproduced. Regarding experimental results at ambient pressure, our simulations slightly underestimate the TFE of glycine to 1 mol/L TMAO (SI of ref. 128). Although we only considered the amino acids glycine and alanine, our result that TMAO is excluded from the peptides is in qualitative agreement with vapor pressure osmometry data of bovine serum albumin by Courtenay et al.[106] At high pressures, we have no direct comparisons to experimental data available. However, the result that TMAO still stabilizes homopeptides can possibly be related to experiments which show that TMAO partially protects the activity of different enzymes against pressures of 1 kbar[27] and 2 kbar.[133] In SAXS experiments it was also shown that TMAO stabilizes the native state of SNase up to 3 kbar, which is over 1 kbar above its unfolding pressure in pure water.[37]

5.4.2 Peptide-TMAO interactions at infinite dilution

After treating the thermodynamics of periodic peptides at high TMAO concentrations, this section deals with TMAO-peptide interactions at infinite dilution. We calculated the potentials of mean force between the periodic glycine and alanine homopeptide extended strands, α-helices, and 3_{10}-helices using umbrella sampling. The PMFs at 1 bar and 5 kbar as a function of peptide-TMAO distance are shown in figure 5.8. First we see that



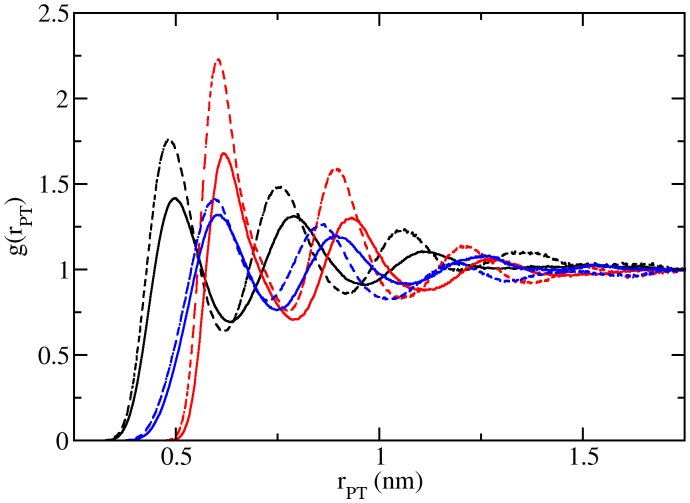
(a) Glycine **(b)** Alanine

Figure 5.8.: Potentials of mean force as a function of peptide-TMAO distance in cylindrical coordinates. Full lines are at 1 bar pressure, dashed lines are at 5 kbar. Extended strands are in black, α-helices are in red, and 3_{10}-helices are in blue.

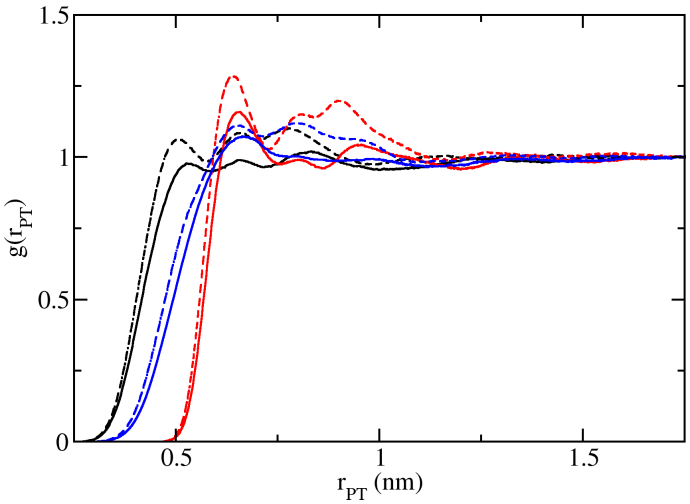
the PMFs do not become negative for distances shorter than 0.5 nm, which means that TMAO is separated from all peptides by a solvation layer of water. The position of the first minimum is not very sensitive to pressure, but the following minima are successively shifted towards shorter distances at 5 kbar. For glycine, the first minima are all at negative free energies and decrease further with pressure. The pressure effect is strongest for the α-helix, for which the PMF minimum at a separation of 0.6 nm decreases from -1.3 to -2.0 kJ/mol at 5 kbar. Around the extended strand, the distance with the highest probability is around 0.5 nm with a change in well depth from -0.9 to -1.4 kJ/mol. For the glycine 3_{10}-helix, it only changes from -0.7 to -0.85 kJ/mol at 0.6 nm. With increasing distance, the oscillations in the PMF decay towards zero. For alanine, the PMFs have multiple minima and maxima that overlap partially. The free energy of a TMAO molecule does not only depend on the distance from the peptide, but also on which side it approaches it. The second minimum is just the first layer of TMAO around the side chains, which is approximately 0.15 nm further away. For the extended strand, the PMF in the first layer is slightly positive at 1 bar. The two minima decrease from 0.05 to -0.15 and from 0.02 to -0.20 kJ/mol. At 0.64 nm, the first minimum of the alanine α-helix, the PMF decreases from -0.37 to -0.62 kJ/mol at 5 kbar. The PMF at the second minimum (0.80 nm) is initially positive at 1 bar and decreases from 0.02 to -0.35 kJ/mol. Finally, the 3_{10}-helix PMF goes from -0.17 to -0.26 kJ/mol at the first minimum upon compression. There is no discernible local minimum at 0.80 nm at 1 bar, but it exists at a free energy of -0.28 kJ/mol for 5 kbar. We generally observe a stronger exclusion of TMAO from the methyl sidechains around alanine, but the interactions become more favorable at higher pressures.

For easier comparison of these results with high TMAO concentrations, we transform the PMFs to the RDFs using eq. 4.11. These RDFs (fig. 5.9) show a very similar behavior to the concentrated systems shown previously in fig. 5.3. We note that excess coefficients $\bar{\Gamma}$ calculated from these RDFs (see fig. 5.4) do in some cases point towards binding of TMAO

to the peptide, which is unlikely to be the actual behavior. We assume it is statistical noise due to the large fluctuations of the integral in eq. 5.6 at large distances. For glycine, the peak positions are almost identical in both cases, but the increase of peak heights is more pronounced in the RDFs from umbrella sampling in pure water. We speculate that, at high concentrations, it is difficult for TMAO to be excluded from the peptide without causing a high TMAO concentration in the bulk, which is also unfavorable due to the hydrophobic frustration mechanism mentioned in the previous chapter. The general peak positions of the alanine RDFs in the dilute and concentrated systems are also similar, and the effect of pressure on the heights of the RDF maxima is qualitatively reproduced. The actual peak heights at infinite dilution are much lower than at multimolar TMAO content. The same argument used for glycine can be applied to explain the strong depletion of TMAO around the alanine conformations. Therefore, we find the expected result that in the limit of infinite dilution, TMAO is excluded more strongly from the polyalanines due to the unfavorable hydrophobic side chain interactions.



(a) Glycine



(b) Alanine

Figure 5.9.: Peptide-TMAO radial distributions at infinite dilution calculated from the PMFs (eq. 4.11) at 1 bar (full lines) and 5 kbar (dashed lines). Secondary structures: Extended strand (black), α-helix (red), and 3_{10}-helix (blue).

5.5 Conclusions and outlook

This chapter dealt with the application of our TMAO force field to simple model peptides, i. e. infinitely extended polyglycines and polyalanines in extended and helical conformations. We found that TMAO stabilizes α- and 3_{10}-helices compared to the extended strand even at high pressures up to 5 kbar through exclusion of TMAO from the peptides. The differences in unfolding free energies are on a scale of less than 1 kJ/mol of amino acid at very high TMAO concentrations up to 3.5 mol/L.

Our attempt to predict the effect of TMAO on the stability of the helices using the universal-backbone transfer model[34] was qualitatively successful but failed to resolve the sub-kilojoule differences in free energies. It is entirely possible that the class of stabilizing osmolytes is beyond the limits of the transfer model, or at least of this approach to compute TFEs. We have no reason to believe that the discrepancies are due to a systematic error in the treatment of the side chains in the transfer model, since we already fail to reproduce the TFEs of glycine.

Overall, our TMAO force field seems to be suited for the use with peptides under ambient and high pressure condition. Although it still remains to be seen whether protein force fields need to be adapted for high pressures similar to the charge scaling applied to TMAO, it is not likely that these modifications would significantly change the properties of homopeptides in fixed conformations. Therefore, the qualitative results in this chapter should remain valid.

Obviously, the content of this chapter is just the beginning of the characterization of TMAO-protein interactions. For an application of our TMAO model to a real peptide we refer to the next chapter. For future work it will also be important to investigate the interactions between these peptides and our urea force field, which we introduced in the previous chapter. A very detailed analysis of peptide-urea interactions for periodic peptides in combination with the KBFF urea model can be found in ref. 134. Since we have shown that the KBFF does not adequately reproduce the solvation structure in aqueous solutions, we expect to see some results to differ from the study in this reference.

Finally, we intend to combine the two osmolytes in water as a solvent for proteins. In nature, urea and TMAO are usually found in a 2:1 mixture. In a study by Lin and Timasheff it was shown that the effects of urea and TMAO on the transfer free energies of RNase T1 are additive.[35] It will be interesting to see if simulations can reproduce this result at ambient and under high pressure conditions.

Chapter 6

The Trp-cage miniprotein

In this chapter, we show results from REMD simulations of the Trp-cage miniprotein[47] (see chapter 1 and fig. 6.1) in water and TMAO solution at 1 bar and 10 kbar, where we make use of our high-pressure force field adaptation for TMAO. We calculate the stability diagram in temperature space as well as the effect of TMAO on the stability using the structural deviations from a NMR nuclear Overhauser effect reference structure as a reaction coordinate.

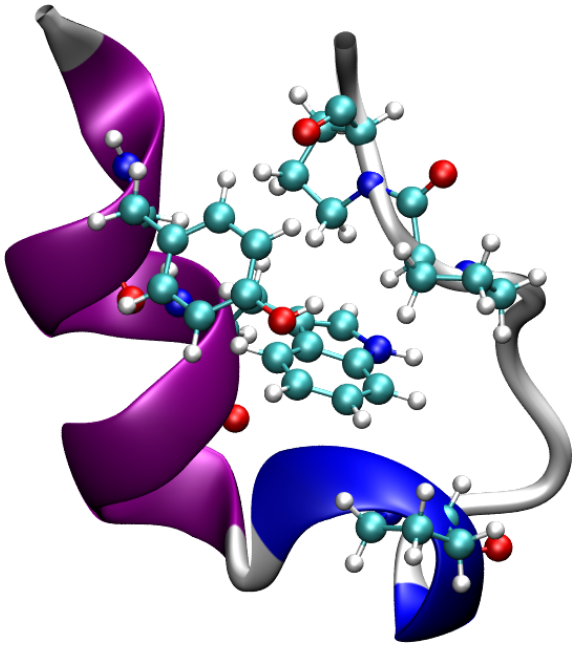


Figure 6.1.: Snapshot of the native structure of the Trp-cage miniprotein TC5B.[47] PDB structure 1L2Y. Sequence: NLYIQWLKDGGPSSGRPPPS. Only the tryptophane, proline, and tyrosine residues are shown explicitly. The colored bands signify the secondary structure as calculated from the STRIDE program.[135] Purple stands for α-helices, blue for 3_{10}-helices, and white signifies no discernible secondary structure. The polyproline II helix at the C terminus is not identified by STRIDE.

6.1 Methods

The Trp-cage miniprotein TC5B was simulated in water and 1 mol/L TMAO at 1 bar and 10 kbar using temperature replica exchange (see sec. 2.4.2) with 56 replicas. For Trp-cage, the Protein Data Bank (PDB)[136] structure 1L2Y was used with the AMBER03w[137] force field, which is a modification of AMBER03 to be used in combination with the TIP4P/2005 water model.[61] As in the previous chapter, we make the assumption that no scaling of charges or other corrections are necessary for using the protein force field under high pressure. Before simulating Trp-cage in solution, the PDB structure was denatured by simulation in a vacuum. The system composition in pure water is 1 Trp-cage, 2923 water molecules, and 1 chloride ion for charge neutrality. The 1 mol/L TMAO system contained 2642 water, 1 chloride, and 52 TMAO, where we used our parameters for TMAO with different partial charges for the 1 bar and 10 kbar cases (see appendix B.2). The edge length of the cubic box at 300 K and 1 bar was approximately 4.5 nm.

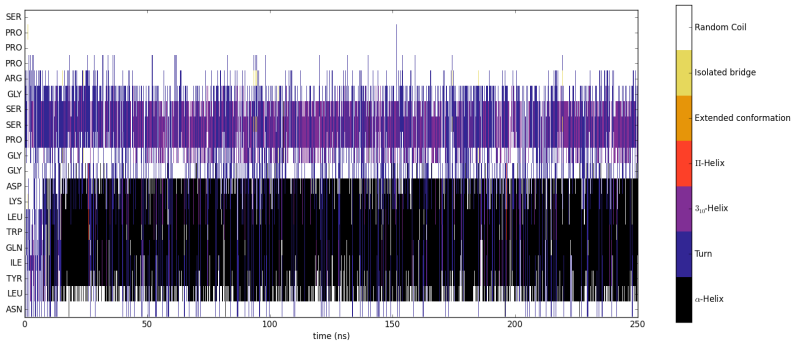
In order to obtain exchange rates between 0.1 and 0.3 during the replica exchange steps, the temperature range at 10 kbar was different from the range at 1 bar due to broadened potential energy distributions at high pressure. The temperatures of the 56 replicas were between 300 K and 592 K at 1 bar and between 300 K and 605 K at 10 kbar. After separate equilibration of the replicas at their respective temperatures, replica exchange simulations with a time step of 2 fs and an exchange frequency of 2 ps were performed for 250 ns, and the last 50 ns were used for analysis.

6.2 Results

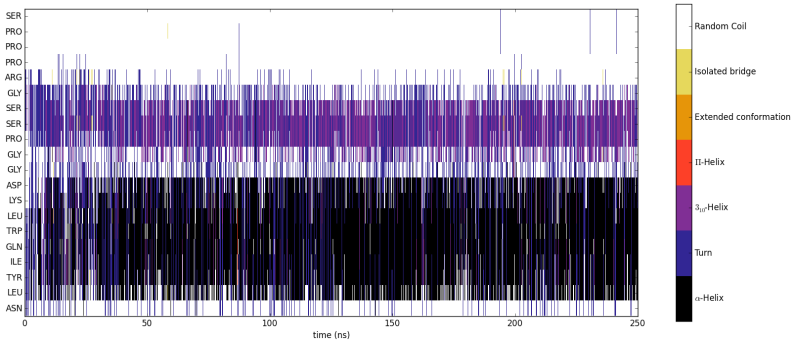
6.2.1 Secondary structure

The native secondary structure of Trp-cage from solution NMR (PDB 1L2Y, see fig. 6.1) is an α-helix (residues 2-8) followed by a single turn of a 3_{10}-helix (11-14), and ending in a polyproline II helix of three prolines close to the C-terminus. The aspartic acid and arginine residues (9 and 16) form a salt bridge. In the center of the protein, the tryptophane is surrounded by a tyrosine ring and multiple proline rings for favorable hydrophobic interactions, which gives Trp-cage its globular shape.

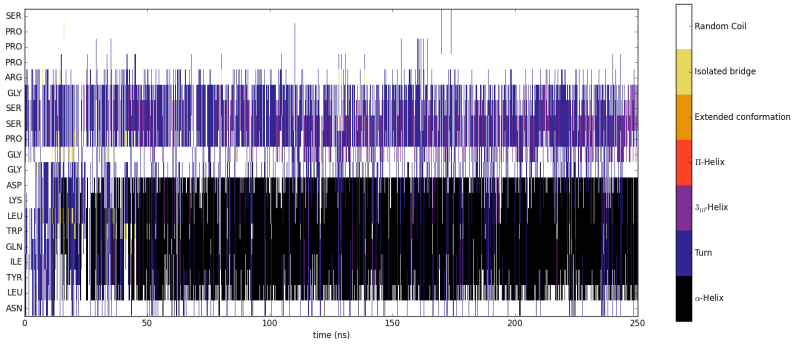
First we want to confirm whether the protein folds to a similar conformation in our simulations, which were started from a fully denatured structure. In figure 6.2 we show the secondary structure as a function of time for Trp-cage in water and in 1 mol/L TMAO solution at 1 bar and at 10 kbar. We used the STRIDE program for secondary structure determination, which uses a combination of H-bond energies and the ϕ and ψ backbone angles to assign a secondary structure to each amino acid.[135] We find that both α- and 3_{10}-helices form within the first 10 ns at 1 bar. At 10 kbar, the same secondary structure forms significantly slower within approximately 75 ns. The assigned structures of the residues in the 3_{10}-helix fluctuate between 'Turn' and '3_{10}-helix', which can be explained by their similarity for a short sequence.[135] We note that, since we are using replica exchange MD which significantly accelerates the folding kinetics, a simulated folding time of 10 ns does not correspond to the actual value. The experimental value for the folding time was determined as approximately $4\,\mu s$ using circular dichroism (CD) measurements.[48] We will not attempt to extract kinetic properties from our REMD simulations, although we note that some recent advances have been made in this area.[138]



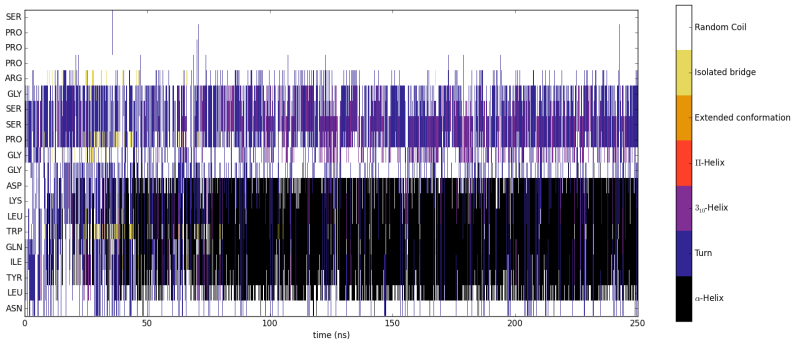
(a) Water at 1 bar and 300 K



(b) 1 mol/L TMAO at 1 bar and 300 K



(c) Water at 10 kbar and 300 K



(d) 1 mol/L TMAO at 10 kbar and 300 K

Figure 6.2.: Secondary structure of Trp-cage for the 300 K replicas as a function of time. The secondary structure was determined with the STRIDE program.[135]

In table 6.1 we show the fractions of secondary structures in the two sequences that form the α-helix (residues 2-8) and the single-turn 3_{10}-helix (residues 11-14) at 300 K. For the first sequence, we find that the fraction of α-helical conformations at 1 bar decreases from 79.5 % to 76.7 % with TMAO concentration. The amount of 3_{10}-helices increases by 8 %, which is almost balanced out by a decrease of turns. If we assume that 3_{10}-helices and turns are structurally similar enough to be considered identical in this case, the 3_{10}-helical sequence is not influenced significantly by TMAO. When going to 10 kbar pressure, the α-helix is stabilized with pressure, and we also find a 2 % increase in α-helical conformations with TMAO content. Around the 3_{10}-helix sequence, the fraction of turns is significantly higher than it is at normal pressure. The sum of turns and 3_{10}-helix conformations is about 7 % lower than at 1 bar and increases by 1 % from water to 1 mol/L TMAO. Interestingly, we find two opposing trends for the stability of the two main secondary structures. The alpha helix is stabilized with pressure, whereas the much shorter 3_{10}-helix is weakened. In order to confirm this result, further work at higher TMAO concentrations may be necessary, but our simulations at 10 kbar seem to confirm a weak protection of the secondary structure caused by TMAO.

	α-helix
Water, 1 bar	0.795
1 M TMAO, 1 bar	0.767
Water, 10 kbar	0.805
1 M TMAO, 10 kbar	0.824

(a) Residues 2-8 (LEU-LYS)

3_{10}-helix	Turn	3_{10}-helix + Turn
0.387	0.385	0.772
0.463	0.316	0.779
0.271	0.427	0.698
0.253	0.457	0.710

(b) Residues 11-14 (GLY-SER)

Table 6.1.: Fraction of secondary structures from the last 50 ns in the sequences that are α-helical (a) and 3_{10}-helical (b) in the native state.

6.2.2 Radius of gyration

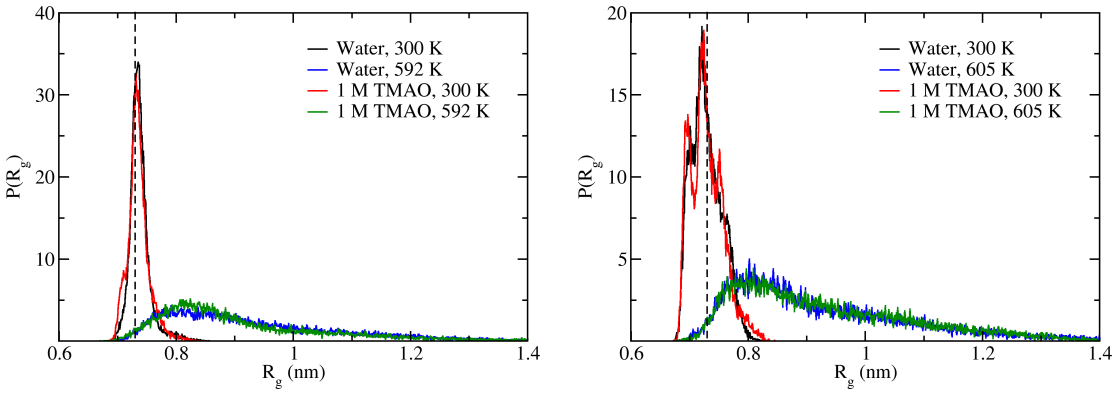
There are many possible coordinates to measure the degree of folding of a protein. As a very simple quantity, which is also accessible through scattering experiments, is the radius of gyration R_g:

$$R_g = \sqrt{\left(\sum_{i=1}^{N} m_i\right)^{-1} \sum_{i=1}^{N} m_i \left(\mathbf{r}_i - \mathbf{r}_{\text{com}}\right)^2} \tag{6.1}$$

Here, N is the number of atoms in the molecule, and m_i are the masses of atoms i. The radius of gyration is then the mass-weighted average root mean squared distance of atoms from the center of mass (com).

In fig. 6.3 we show the distributions of R_g for Trp-cage in water and 1 mol/L TMAO at 1 bar and 10 kbar for several temperatures. As a reference, the radius of gyration of the PDB structure (0.73 nm) is shown as a vertical line. At 1 bar, we find that R_g is distributed around this value at ambient conditions, but the distributions broaden and shift towards larger radii with temperature, which signifies unfolding. The addition of TMAO causes a weak shift towards smaller R_g, and a minor local minimum at 0.71 nm appears. This may point towards a slightly more compact state which becomes more favorable in TMAO solution. The temperature dependence of R_g at 10 kbar is similar to 1 bar, but the ambient temperature case is notably different. The broadened distributions in both water and TMAO solution have their maximum at 0.72 nm and a smaller peak around

0.70 nm, which has a higher probability and lower R_g in the TMAO system. When we compare the peak heights between the pressures, the more compact state, which is already distinguishable from the main peak in TMAO at 1 bar, is more populated at 10 kbar both in water and TMAO. Although only qualitative in nature, this result is another piece of evidence for the protecting effect of TMAO on proteins.



(a) 1 bar **(b)** 10 kbar

Figure 6.3.: Distributions of the radius of gyration R_g of Trp-cage in water and 1 mol/L TMAO at 1 bar (a) and 10 kbar (b). The vertical dashed lines mark the radius of gyration of the PDB structure 1L2Y.

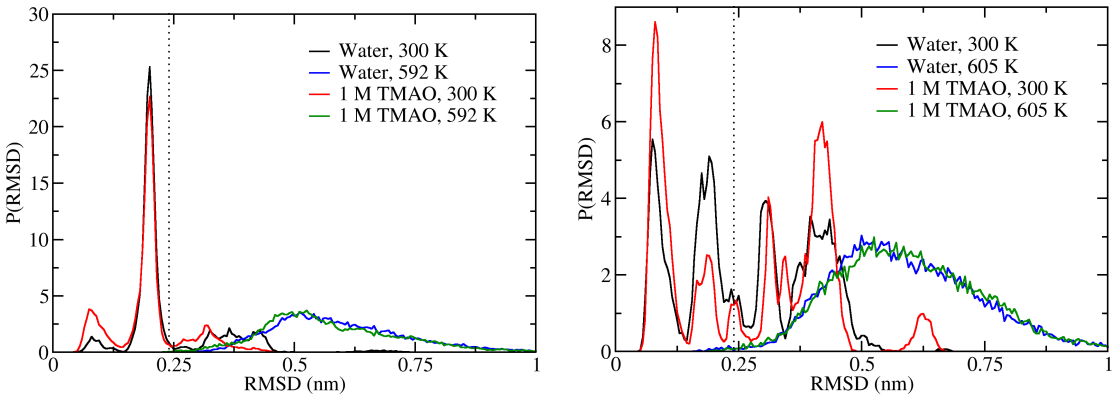
6.2.3 C_α-RMSD

Ideally, we would like to find a coordinate in which the distributions of the fully denatured state and the native conformation have no overlap, which would enable the application of a simple two-state model for the folded and unfolded states. A commonly used coordinate is the root mean squared deviation of the positions of the α-carbon atoms from the reference structure:

$$C_\alpha - \mathrm{RMSD} = \sqrt{\frac{1}{N_{aa}} \sum_{i=1}^{N_{aa}} (\mathbf{r}(C_\alpha^{aa})_{\mathrm{sim}} - \mathbf{r}(C_\alpha^{aa})_{\mathrm{exp}})^2} \tag{6.2}$$

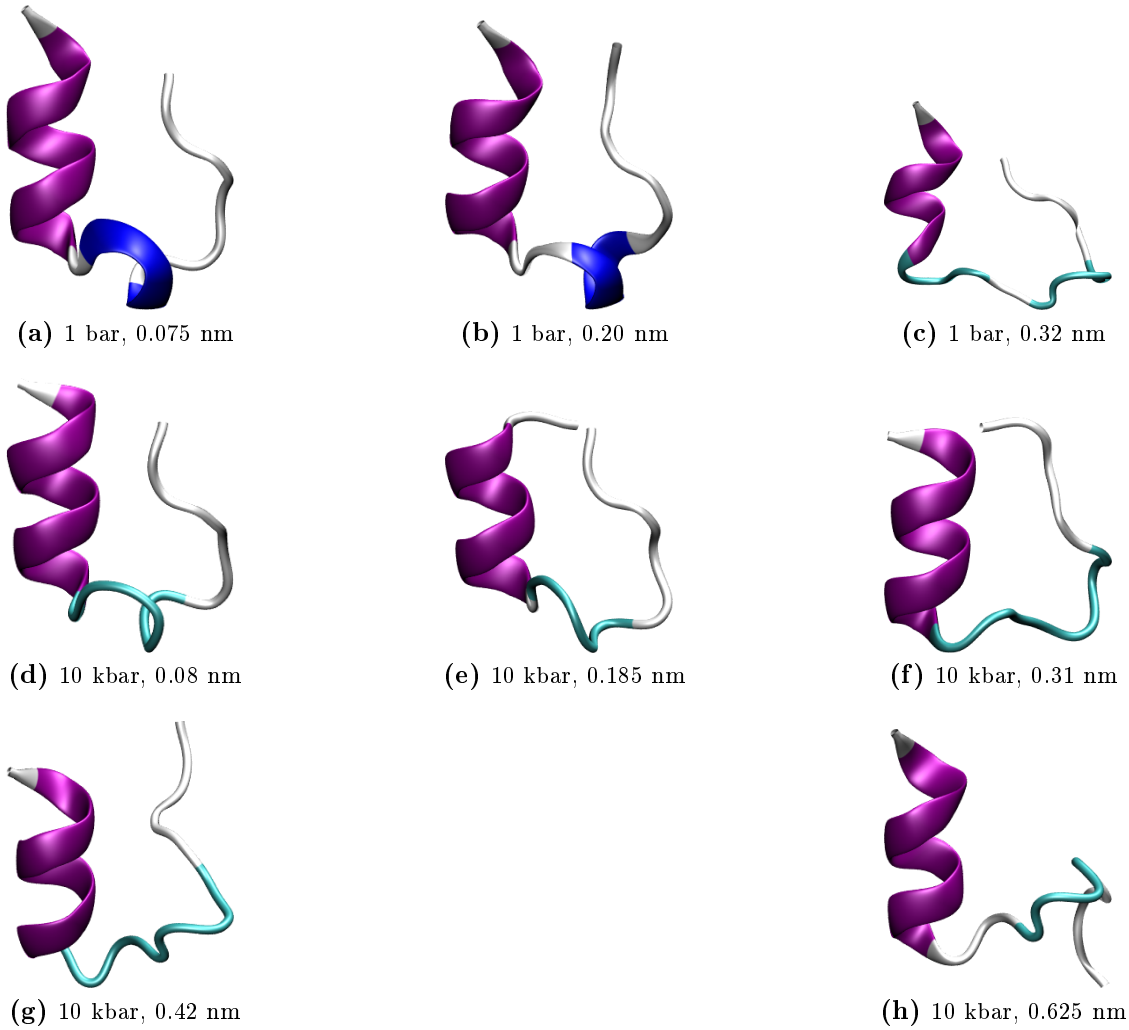
In order to explain why only the C_α atoms are used, it is useful to know how the reference PDB structure (1L2Y) of Trp-cage was determined. The distances between the C_α atoms were measured using NMR nuclear Overhauser effect spectroscopy and are listed in the initial Trp-cage publication.[47] The full structures are in fact obtained from MD simulations while restraining the C_α distances to the experimental values. This means that only the general backbone structure is a direct result of experiments, which is the reason for only comparing the C_α positions between structures. Using the C_α−RMSD raises a problem for our high-pressure simulations. To our knowledge, there exists no experimental structural data of Trp-cage at elevated pressures. Therefore, we make the assumption that the native states have similar structures at 1 bar and 10 kbar, and calculated the RMSD of our high-pressure simulations from the PDB structure at 1 bar. The RMSD distributions at both pressures for different temperature replicas are shown in fig. 6.4. Examples for structures at some RMSD values are shown in fig. 6.5. The distributions at 1 bar and ambient temperature have a maximum distributed around 0.2 nm as the main folded state and a local minimum at 0.08 nm. In 1 mol/L TMAO, the state with a lower RMSD has a

higher population compared to water, and the unstructured region between 0.25 nm and 0.5 nm also shifts towards lower RMSD values. In the replica with the highest temperature there is no significant change in the broad distributions at high deviations from the folded structure. The distributions of the native and fully unfolded states are well separated, which enables us to use a two-state model with a clear cutoff chosen as 0.24 nm. The same system at 10 kbar, however, differs significantly in its C_α−structure. Whereas the high-temperature distributions are comparable to 1 bar and are clearly above the cutoff value, we find multiple distinct peaks at 300 K distributed across a wide value range. We still find the two states around 0.08 and 0.2 nm with the equilibrium shifted towards the lower C_α deviation, which becomes more significant in TMAO solution. Although the 3_{10}-helix at 1 bar is often replaced by a turn at 10 kbar, as we mentioned earlier, the structures in figures 6.5 a) and d) are very similar, as are the structures b) and e). At RMSD values around 0.31 nm, the structures c) and f) show different examples of folding defects where the 3_{10}-helix or proper turn is not formed. Structure c) (1 bar) is an example structure in which the last turn of the α-helix is unfolded. This case is similar to the case at 10 kbar with a deviation of 0.42 nm (structure g). Only in TMAO at 10 kbar, the distribution has a small maximum at 0.625 nm, which corresponds to an intact α-helix, but completely unfolded 3_{10}-helix/turn, which prevents polyproline-tryptophane contacts (structure h). We used the last 50 of 250 ns for our analysis. As it turns out, this may not be sufficient, since structure h) only appears during the first 12 ns of this analysis. We therefore assume that this is not an effect of TMAO, but a consequence of insufficient sampling.



(a) 1 bar **(b)** 10 kbar

Figure 6.4.: Distributions of the C_α-RMSD between our REMD simulations and the PDB structure 1L2Y at 1 bar (a) and 10 kbar (b) for different temperatures in water and 1 mol/L TMAO. The dotted line marks the chosen cutoff at 0.24 nm between the folded and unfolded state in a two-state model.



(a) 1 bar, 0.075 nm **(b)** 1 bar, 0.20 nm **(c)** 1 bar, 0.32 nm

(d) 10 kbar, 0.08 nm **(e)** 10 kbar, 0.185 nm **(f)** 10 kbar, 0.31 nm

(g) 10 kbar, 0.42 nm **(h)** 10 kbar, 0.625 nm

Figure 6.5.: Stylized snapshots of the Trp-cage tertiary structure at different C_α−RMSD values (in nm) at 1 bar and 10 kbar and 1 mol/L TMAO. The colors correspond to the STRIDE secondary structure assignments: Purple to an α-helix, blue to a 3_{10}-helix, turquoise to a turn, and white to a random coil.

After noting these issues, we will continue working with the cutoff at 0.24 nm to calculate the unfolding free energy of Trp-cage from the ratio of unfolded to folded states:

$$\Delta G_{\text{unf}} = -\text{k}T \ln \left(\frac{P(\text{unfolded})}{P(\text{folded})} \right) \tag{6.3}$$

The probabilities of the folded and unfolded states are the numerical integrals over the distributions in fig. 6.4 up to and from the cutoff value. If we invert the folding curve at 1 bar in water (fig. 6.6 a), it is in qualitative agreement with the unfolding curve from CD experiments by Qiu et al.,[48] which has a similar sigmoidal shape but is shifted to lower temperatures. Compared to simulations by Canchi et al., the curves are also similar, but our choice of protein and water model unfolds at lower temperatures.[32] From $\Delta G_{\text{unf}} = 0$ in fig. 6.6 we can extract an approximate melting transition temperature of Trp-cage.

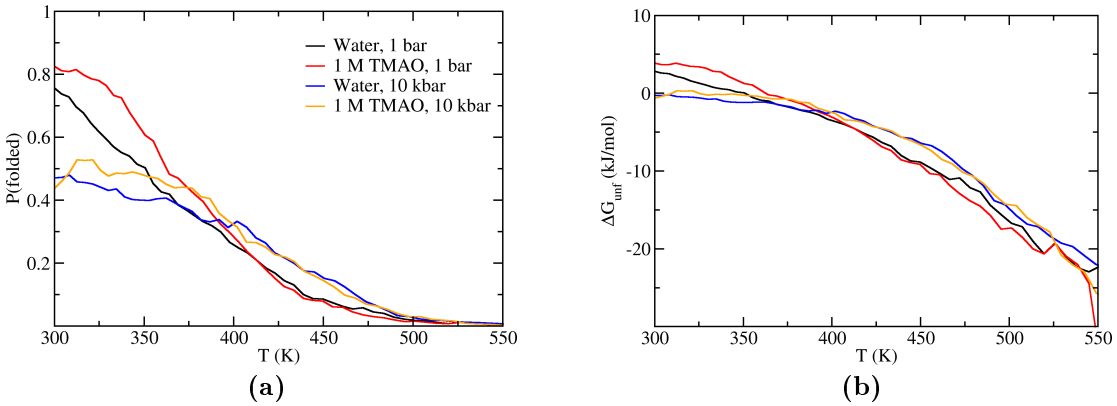


Figure 6.6.: Fraction of folded Trp-cage (a) and unfolding free energies (b) obtained from a two-state model of the C_α−RMSD as a function of temperature. The cutoff between the folded and unfolded states was 0.24 nm.

At 1 bar, the melting temperature is 350 K in water and 363 K in 1 molar TMAO, which is significantly above the experimental value of 314 K.[47,48] Simulations by Day et al.[53] were able to reproduce the experimental melting transition at 321 K. Although we cannot reach this accuracy, our results are significantly better than many previous simulation studies, which predict 400 K,[49] 440 K,[51] between 425 K and 445 K,[52] and 430 K.[32]

At 300 K, the protein is stabilized by only 1 kJ/mol in TMAO solution compared to water. The melting temperature at 10 kbar cannot be determined with reasonable accuracy, but it appears to be at or around 300 K. The free energy difference with the addition of osmolyte is smaller than it is at 1 bar, which is in qualitative agreement with the results of chapter 5.

If we do not want to limit ourselves to a two-state model, it is possible to calculate the surface of relative free energies by inverting the distributions for each temperature:

$$\Delta G = -\mathrm{k}T \ln P(\mathrm{RMSD}) \tag{6.4}$$

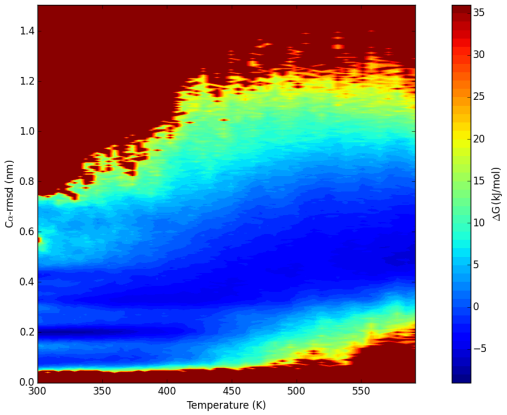
The resulting free energy surfaces from interpolating the regions between the temperature replicas are shown in fig. 6.7. At 1 bar, the main folded state around 0.2 nm can also be observed up to the previously determined melting temperature of 350 K. Around 425 K, the low-energy region shifts to RMSD values between 0.3 and 0.8 nm, where it remains up to 590 K. A comparison of this energy surface with results by Canchi et al.[32] yields good agreement, with one main difference: The main folded state in their work is at lower RMSD values around 0.14 nm, which is closer to the PDB structure than our results. We explain this with their choice of the AMBER94 force field for Trp-cage and the TIP3P water model, which are the parameters with which the PDB structure was determined.

We can quantify the influence of TMAO by subtracting the energy surface of water from the free energies of 1 mol/L TMAO (panel e). In TMAO solution, the region with a RMSD below 0.2 nm is stabilized by $2-4$ kJ/mol up to the melting point of Trp-cage in water (350 K). A region around 0.6 nm is stabilized by about 4 kJ/mol up to 320 K, whereas the energies of states with a RMSD of 0.4 nm are $2-4$ kJ/mol higher than in water. In the range from 400 to 450 K there is a region around 0.8 nm that is stabilized by about 4 kJ/mol in TMAO, although this finding might be a result of the large statistical noise at high temperatures.

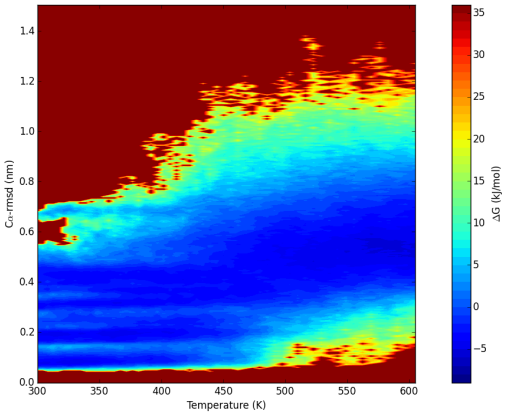
We have previously established that the reference structure of Trp-cage at 1 bar is most likely not ideal for analyzing the results at 10 kbar. However, it is nonetheless possible to

extract information from the resulting free energies. The first main difference to normal pressure is the fact that, in addition to the two states at 0.2 and 0.08 nm, there are at least two additional states at 0.3 and 0.45 nm. All these states have very similar free energies that remain relatively constant up to high temperatures well above the melting point of Trp-cage at 1 bar. The free energy at a RMSD around 0.4 nm is almost unchanged even in the highest temperature replica above 600 K. A notable difference between water and the TMAO solution is the high-energy region around 0.6 nm which is shifted towards 0.5 nm in TMAO. In the difference between the energies (panel f), we observe no significant stabilization of the low-RMSD region caused by the addition of TMAO. Instead, the regions around 0.2, 0.3, and 0.45 nm increase in energy by $1-4$ kJ/mol. At a RMSD between 0.6 and 0.9 nm, the TMAO solution favors lower RMSD values up to 450 K with free energy differences of ± 4 kJ/mol.

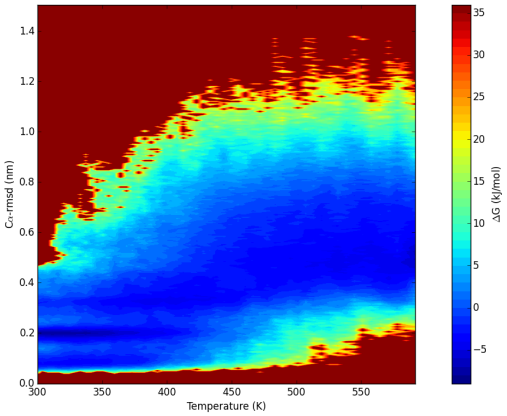
Whereas our results of the favored folded state at 1 bar are in agreement with our previous results, the stabilization of Trp-cage at 10 kbar is very minor and inconclusive.



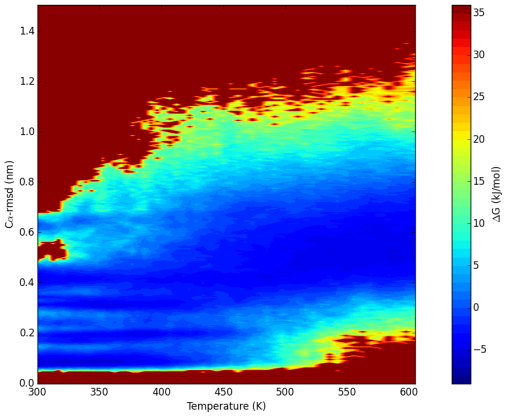
(a) Water at 1 bar



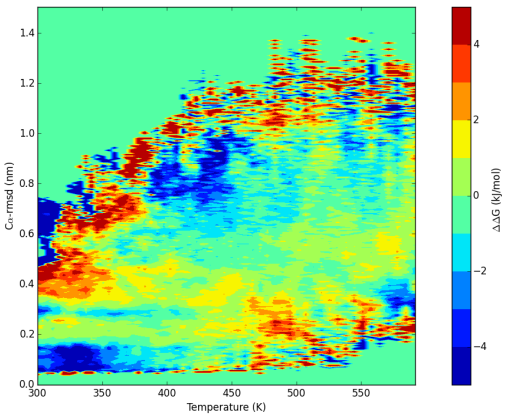
(b) Water at 10 kbar



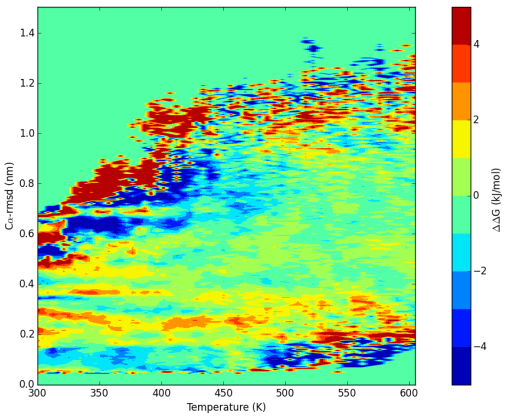
(c) 1 M TMAO at 1 bar



(d) 1 M TMAO at 10 kbar



(e) $\Delta\Delta G$ at 1 bar



(f) $\Delta\Delta G$ at 10 kbar

Figure 6.7.: (a-d) The Gibbs free energies of Trp-cage in water and 1 mol/L TMAO at 1 bar and 10 kbar were calculated from the distributions of the C_α−RMSD from the PDB reference structure. (e) Difference in free energy surfaces $\Delta\Delta G$ between 1 molar TMAO and water at 1 bar. (f) $\Delta\Delta G$ at 10 kbar. The 2D surfaces were calculated from the 1D distributions of each temperature replica. The visualization is a bicubic interpolation of these free energy surfaces.

6.3 Conclusions

From studying the folding behavior of the Trp-cage miniprotein in water and 1 mol/L TMAO solution at 1 bar and 10 kbar using the radius of gyration and the C_α−RMSD, we found the expected result that our TMAO model stabilizes the folded state and that the protein is less stable at high pressures. In the secondary structure, the α-helix remains stable at high pressures, but the 3_{10}-helix is often replaced by a structurally similar turn. We also found an increased probability for misfolding of the polyproline cage around tryptophane at 10 kbar, which shows that the formation of the hydrophobic core becomes more difficult under pressure.

Obviously, we did not exhaustively examine all structural properties. Further work is necessary to get an in-depth understanding of the workings of Trp-cage at high pressures. Possible avenues towards exploration of the structure that is native at high pressure include correlating different coordinates such as the radius of gyration and the C_α−RMSD simultaneously. Additional reaction coordinates could include e.g. eigenmodes of the folding process obtained from a principal component analysis of the full set of REMD trajectories.[49,52] Another option to identify new native structures is the application of clustering algorithms, e.g. on the set of backbone dihedral angles or the C_α−distances. If we know the equilibrium structures at different pressures, the next step is a qualitative comparison of the folding mechanisms and a quantitative study of the folding kinetics, which has some inherent difficulties when REMD is used.[138,139] There is still a wide variety of open questions on protein folding in general, which could be answered by using Trp-cage as a uniquely suited model system due to its extremely fast folding.

Chapter 7

Conclusions and outlook

The content of this thesis can be separated into two main topics. One is an in-depth treatment of the pressure-dependence of the hydrophobic/water solvation structure, and the second deals with method development for the simulation of osmolyte solutions and the application to peptide systems.

In an attempt to understand how hydrophobic solvation changes at high pressures, we have shown that the structure of water becomes more like that of hexagonal ice at elevated pressures on short and intermediate length scales, even though no such indications are found at ambient pressure. It will remain to be seen whether our results for a relatively even surface can be transferred to more complicated systems, e.g. to describe the solvation structure around the hydrophobic core of globular proteins. Since we studied this system at ambient temperature, no actual crystalline phase is observed. However, future work at or close to freezing temperatures could further our understanding of ice nucleation and how different types of interfaces regarding their chemical composition and surface geometry influence the freezing process.

Subsequently, we treated a different topic – we laid the groundwork for force field simulations of the osmolytes TMAO and urea in water at high pressure. For TMAO, we developed a general and systematic charge scaling approach which accounts for the increased molecular polarization caused by the elevated pressure. We showed that it was necessary to develop a completely new model for the nonbonded interactions of urea with water even for ambient pressure conditions, which required no further modifications for application at high pressures. The fact that a polarization correction was necessary only for TMAO is so far unexplained and remains a major question that needs to be answered in future research.

We then demonstrated that TMAO is excluded from glycine and alanine homopeptides and stabilizes their helical states, which confirms that the prevailing picture of the protection mechanism of TMAO remains valid at pressures of several kilobars. It turns out that the transfer model does not yield the correct unfolding free energies proportional to the solvent-accessible surface area, which could mean that it is not suited for the description of the influence of protecting osmolytes on protein stability. For the globular miniprotein Trp-cage, TMAO also stabilizes the native state, but the evidence at high pressures is too weak to be conclusive. This work marks just the beginning of the study of proteins and osmolytes at high pressures. So far we have no idea whether protein force fields require significant changes for a correct modeling of proteins under pressure. Our TMAO model has only been used with very few peptides and proteins, and our newly developed urea model is completely untested in this regard. However, the important result of this thesis

is the introduction of the tools necessary for high pressure force field simulations of these osmolytes, which gives access to the study of TMAO, urea, and their naturally occurring mixtures with biomolecules under these conditions.

Appendix A

Hydrogen bond network analysis using graph theory

In the field of graph theory, a graph is defined by a set of vertices and a set of edges. An edge is defined as the connection between two vertices, in which case the vertices are defined to be adjacent. The graphs we deal with in this work are simple, undirected, and unweighted. If a graph is simple, it has at most one edge between two vertices, and contains no loops from one vertex to itself. In undirected graphs, each edge is a path between two vertices in both directions. The edges in an unweighted graph have no weight, i.e. a numerical value associated with them, which is relevant e.g. when determining the shortest path between vertices. Figure A.1 shows a representation of an example graph with these properties.

We applied graph theory to the hydrogen bond network in water. Each water molecule corresponds to a vertex in the graph, and each hydrogen bond between water molecules is represented as an edge between these vertices. We implemented the algorithm using the `networkx` library,[140] which is written in python. Instead of finding all simple cycles, which is computationally very expensive, a subset of cycles is detected as follows: The graph is searched for all triplets of connected vertices, i.e. three vertices connected via two edges (or three if they already form a cycle). It is assumed that, due to the geometry of the water H-bond network, each triplet of connected vertices can be part of at most one cycle. A cycle is found if there exists a shortest path between the two non-adjacent vertices in the triplet, where the only the paths that do not contain the third vertex of the triplet are considered. If the path has length $n-2$, i.e. contains $n-2$ edges, the cycle is of length n. Finally, only cycles with a distinct set of vertices are relevant and all other equivalent cycles are discarded. Figure A.1 contains an example for this algorithm. We will refer to this subset of cycles as rings, even though this name is not defined in graph theory in this context.

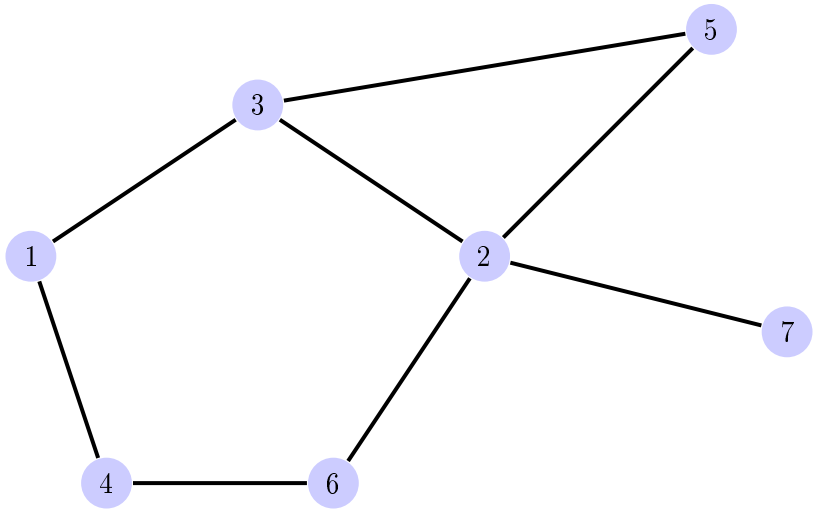


Figure A.1.: Representation of an example graph. The ring detection algorithm for this graph works as follows. The connected triplets are {(3, 5, 2), (5, 2, 3), (2, 3, 5), (6, 4, 1), (1, 3, 2), (3, 2, 6), (2, 6, 4), (4, 1, 3), (6, 2, 7), (5, 2, 7), (3, 2, 7)} (direction does not matter). The triplet (3, 5, 2) has a shortest path of length 1 between the vertices 3 and 2, so there exists a cycle of length 3. The next two triplets in the list lead to the same set of vertices and are not distinct cycles. Triplet (6, 4, 1) has multiple paths from 1 to 6. The shortest path that does not contain vertex 4 has length 3, which leads to a cycle with length 5. Four more triplets lead to the same set of vertices and are discarded. In this graph there are 2 distinct simple cycles without shortcuts. They contain the vertices {2, 3, 5} and {1, 2, 3, 4, 6}.

Appendix B

Force field parameters

B.1 Bonded potential energy functions

The TMAO and urea force fields in this work use the following potentials for bonded interactions:

$$V_{\mathrm{bond}}(r) = \frac{k_{\mathrm{b}}}{2}(r - r_0)^2 \tag{B.1}$$

$$V_{\mathrm{angle}}(\phi) = \frac{k_\phi}{2}(\phi - \phi_0)^2 \tag{B.2}$$

$$V_{\mathrm{dihedral}}(\theta) = k_\theta(1 + \cos(n\theta - \theta_0)) \tag{B.3}$$

V are the potential energies, k are force constants, r is a bond distance, ϕ is a bond angle, θ is a dihedral angle, and n is the multiplicity of the dihedral.

B.2 TMAO

Bond type	r_0 (nm)	k_{b} (kJ mol^{-1} nm^{-2})	
O-N	0.1407	143438	
N-C	0.1506	107258	
C-H	0.1082	constrained	
Angle type	ϕ_0 (deg)	k_ϕ (kJ mol^{-1} rad^{-2})	
O-N-C	109.99	255.16	
C-N-C	108.16	576.55	
N-C-H	108.07	209.10	
H-C-H	108.25	229.74	
Dihedral type	θ_0 (deg)	k_θ (kJ mol^{-1})	n
O-N-C-H	0	1.13	3
C-N-C-H	0	1.13	3

Table B.1.: Bond, bond angle, and dihedral parameters of the TMAO force field.

	O	N	C	H
σ (nm)	0.3266	0.2926	0.3707	0.2130
ϵ (kJ mol^{-1})	0.6389	0.8374	0.2830	0.07745

Table B.2.: Lennard-Jones parameters of the TMAO force field.

p (bar)	O	N	C	H
1	-0.815	0.605	-0.260	0.110
500	-0.8275	0.5812	-0.2557	0.1126
1000	-0.8368	0.5854	-0.2567	0.1135
2000	-0.8419	0.5842	-0.2570	0.1143
3000	-0.8467	0.5836	-0.2573	0.1150
4000	-0.8502	0.6024	-0.2603	0.1143
5000	-0.8598	0.5964	-0.2593	0.1157
7500	-0.8589	0.5958	-0.2606	0.1161
10000	-0.8599	0.5932	-0.2609	0.1166

Table B.3.: Partial charges in units of e of the TMAO force field for pressures up to 10 kbar.

B.3 Urea

Bond type	r_0 (nm)		
C-O	0.12290		
C-N	0.13350		
N-H	0.10100		
Angle type	ϕ_0 (deg)	k_ϕ (kJ mol^{-1} rad^{-2})	
C-N-H	120	390	
H-N-H	120	445	
O-C-N	121.4	730	
N-C-N	117.2	670	
Dihedral type	θ_0 (deg)	k_θ (kJ mol^{-1})	n
O-C-N-H	0.0	8.36800	1
O-C-N-H	180.0	10.46000	2
N-C-N-H	180.0	10.46000	2
N-N-C-O	180.0	43.932	2
C-H-N-H	180.0	4.184	2

Table B.4.: Bond, bond angle, and dihedral parameters of the urea force field.

	O	N	C	H^{cis}	H^{trans}
σ (nm)	0.31377	0.34452	0.36039	0.11333	0.11333
ϵ (kJ mol^{-1})	0.59432	0.51114	0.35982	0.065689	0.065689
q (e)	-0.6162	-0.8400	0.6068	0.4026	0.4421

Table B.5.: Lennard-Jones parameters and partial charges of the urea force field. H^{cis} (H^{trans}) refer to the hydrogen atoms on the same (opposite) side of the C-N bond as oxygen.

Appendix C

Density extrapolation

When the densities of aqueous solutions at extreme pressures beyond the experimentally accessible range are required, it is possible to extrapolate the measurements. Floriano and Nascimento developed an extrapolation technique for the density of water, which allows fitting the density vs pressure to the following function:[141]

$$\rho^{\mathrm{fit}}(p) = \rho(1\,\mathrm{bar}) + a_0 \ln \frac{a_1 + p}{a_1 + 1\,\mathrm{bar}} \tag{C.1}$$

Here the fit parameter a_0 is a density, and a_1 is a pressure. We use this function for the extrapolation of the densities of dilute solutions, thereby making the implicit assumption that this functional form is still reasonable outside of the pure water case.

Appendix D

High pressure force field adaptation through individual bond dipole scaling

In sec. 4.1.4 we presented a method to scale the partial charges in a force field to include the effect of polarization with increasing pressure. For a simple molecule like TMAO, the charges were scaled to reproduce the relative change in the molecular dipole moment from EC-RISM calculations. In the case of larger molecules with many significant contributions to the total dipole, a different approach may be necessary. We propose the the following method for scaling the charges to reproduce the relative change of the dipole moments of all individual chemical bonds in the molecule:
First the dipole moment μ_{ij} of each bond between atoms i and j is calculated with respect to the center of geometry of the two atoms.

$$\mu_{ij} = (q_j - q_i)\frac{r_{ij}}{2} \tag{D.1}$$

with the bond distance r_{ij} and charges q.
The force field dipoles for the target pressure p are scaled as follows:

$$\mu_{ij}^{\mathrm{ff}}(p) = \mu_{ij}^{\mathrm{ff}}(1\,\mathrm{bar})\frac{\mu_{ij}^{\text{EC-RISM}}(p)}{\mu_{ij}^{\text{EC-RISM}}(1\,\mathrm{bar})} \tag{D.2}$$

Next, we calculate the individual charges. A molecule with N atoms and no cycles has $N-1$ bonds. We then have the $N-1$ equations

$$\Delta q_{ij} = q_j - q_i \tag{D.3}$$

with the charge differences Δq_{ij}, which are known from eq. D.1 since the bond distances of the force field are fixed. We also know the total charge of the system:

$$q_{\mathrm{tot}} = \sum_{i=1}^{N} q_i \tag{D.4}$$

This leads to the system of N linear equations:

$$\mathbf{C}\vec{q} = \vec{\Delta q} \tag{D.5}$$

Here $\vec{q}$ are the charges of the N atoms. $\vec{\Delta q}$ contains the charge differences for each bond, and the last element is the total charge of the system q_{tot}. The elements of the coefficient matrix $\mathbf{C}$ can take the values -1,1, and 0. All elements in the last line of $\mathbf{C}$ are 1 from equation D.4 for the total charge. Solving this system of equations for $\vec{q}$ gives the force field charges that reproduce the scaled bond dipole moments.

Appendix E

Differential evolution algorithms (DEAs)

One of many strategies for solving global optimization problems of non-differentiable and nonlinear functions is called differential evolution. This family of algorithms, which was introduced by Storn and Price in 1996,[120] propagates multiple parameter vectors in the variable space, which helps in overcoming local extrema and finding the global optimum. In this work, we apply one specific variant of differential evolution to the problem of force field optimization by minimizing the cost function f. In the following, we elaborate on the technical details of this version of the algorithm as it is implemented in the python library `inspyred`,[142] which differs in its specifics from the initial paper[120] but conserves the general strategy. Differential evolution reuses terms from the field of genetics, which shows its proximity to the field of genetic algorithms. A set of D values of the variables, where D is the dimensionality of the parameter space, is called an individual x. The number of the current iteration is the generation G. We denote the element i of an individual as x_i^G. A set of individuals at any point in time is called the current population with size N_{pop}. At the beginning of the DEA optimization stands the generation of a starting population. A set of N_{pop} initial individuals x^0, which are D-dimensional vectors of real numbers, is generated randomly and uniformly in the parameter space of interest. It has been found that a good choice for the population size is $N_{\text{pop}} = 10 \cdot D$. During one iteration (generation), the following main steps are executed in order:

1. Selection
2. Crossover
3. Mutation
4. Evaluation
5. Replacement

In each generation, a number N_{sel} of individuals is selected from the population to be modified. A selected individual is called a candidate. Here we use the tournament selection procedure, in which multiple individuals are selected randomly to compete for one of the N_{sel} slots for selection. The number of competing individuals per candidate slot is called the tournament size. The individual with the lowest cost function value (since we are assuming a minimization problem) out of the competitors gets selected and becomes a candidate. After N_{sel} candidates are selected, they enter the crossover stage. In this step

we use the heuristic crossover method. Here, the candidates are randomly paired with each other, discarding one candidate if N_{sel} is odd. For each pair, crossover occurs with the crossover probability (crossover rate) $0 \leq P_{\text{cross}} \leq 1$, which is a constant parameter chosen beforehand. In crossover, the vector elements of each pair of two candidates a^G and b^G are randomly recombined to two new vectors u and v by applying the following scheme (when no superscript is given, the current generation G is implied):

$$u_i = \begin{cases} a_i + U([0,1]) \cdot (b_i - a_i) & \text{if } f(a) < f(b) \\ b_i + U([0,1]) \cdot (a_i - b_i) & \text{otherwise} \end{cases} \tag{E.1}$$

$U([0,1])$ is a random variable with a uniform distribution in the interval $[0,1]$. The new vector elements are created by generating a linear combination of the elements of the candidate pair scaled by a random factor between 0 and 1, and adding it to the candidate of the pair that has a lower cost function value. The second vector v is created in exactly the same way, and differs from u only in the random numbers. In the case where no crossover occurs, we simply use $u = a$ and $v = b$. After crossover, the candidates have a chance to be modified via Gaussian mutation. Whether mutation occurs is decided individually for each vector element. The probability $0 \leq P_{\text{mut}} \leq 1$ is called the mutation rate. For each mutating vector element, a number is randomly drawn from a normal distribution and added to the current value u_i to form the new vector z.

$$z_i = \begin{cases} u_i + N(\mu, \sigma) & \text{if } U([0,1]) < P_{\text{mut}} \\ u_i & \text{otherwise} \end{cases} \tag{E.2}$$

$N(\mu, \sigma)$ is a random variable with a normal distribution with mean μ and standard deviation σ. Finally, the cost functions $f(z)$ of the candidates have to be evaluated. In the replacement step we use the steady-state replacement rule: All z are accepted and replace the worst individuals in the population, even if they perform worse than the individuals they replace. In detail, this means that the next generation starts with the N_{sel} vectors z in their population vectors $x^{\text{G}+1}$. We manually terminated the optimization after a sufficient number of generations had passed and the best individuals were sufficiently converged.
Finally, we list our specific choices for all parameters in the optimization of our urea force field (sec. 4.2.4) in table E.1.

D	5
N_{pop}	50
N_{sel}	2
Tournament size	2
P_{cross}	1.0
P_{mut}	0.2
μ	0
σ	1.0

Table E.1.: Optimization parameters used in this work.

Appendix F

Isothermal and adiabatic compressibilities

We can use the heat capacity ratio to relate the adiabatic (isentropic) compressibility κ_S to the isothermal compressibility κ_T:

$$\frac{\kappa_T}{\kappa_S} = \frac{c_p}{c_V} \tag{F.1}$$

The heat capacities at constant volume and pressure c_V and c_p are calculated from the fluctuations of the total energy E of the system simulated in the NVT and NpT ensembles:[67]

$$c_V = \frac{\langle E^2 \rangle_{NVT} - \langle E \rangle^2_{NVT}}{\mathrm{k}T^2} \tag{F.2}$$

$$c_p = \frac{\langle E^2 \rangle_{NpT} - \langle E \rangle^2_{NpT}}{\mathrm{k}T^2} \tag{F.3}$$

The isentropic compressibility is then calculated as:

$$\kappa_S = \kappa_T \frac{\langle E^2 \rangle_{NVT} - \langle E \rangle^2_{NVT}}{\langle E^2 \rangle_{NpT} - \langle E \rangle^2_{NpT}} \tag{F.4}$$

Appendix G

Periodic peptide SASA and TFE data for TMAO solutions

This section contains the transfer free energies of polyglycine and polyalanine in different secondary structures to aqueous TMAO solutions (tab. G.1). Linear fits of the concentration derivative of the TFEs were integrated analytically to obtain the parameters in table G.2. Additionally, the parameters for the analytic form of the helix TFEs obtained from SASA-scaling of the extended strand TFEs are shown in table G.3.

Conformation	p (bar)	$\mathrm{SASA}^{bb}_{\mathrm{Gly}}(\mathrm{nm}^2)$	$\mathrm{SASA}^{sc}_{\mathrm{Gly}}(\mathrm{nm}^2)$
extended strand	1	0.874	0
extended strand	5000	0.874	0
α-helix	1	0.496	0
α-helix	5000	0.495	0
3_{10}-helix	1	0.609	0
3_{10}-helix	5000	0.608	0

(a) Glycine

Conformation	p (bar)	$\mathrm{SASA}^{bb}_{\mathrm{Ala}}(\mathrm{nm}^2)$	$\mathrm{SASA}^{sc}_{\mathrm{Ala}}(\mathrm{nm}^2)$
extended strand	1	0.538	0.516
extended strand	5000	0.538	0.515
α-helix	1	0.166	0.440
α-helix	5000	0.165	0.439
3_{10}-helix	1	0.239	0.478
3_{10}-helix	5000	0.239	0.477

(b) Alanine

Table G.1.: Solvent-accessible surface area in nm^2 per amino acid for glycine and alanine.

Conformation	p (bar)	a	b
extended strand	1	0.218420	0.012837
extended strand	5000	0.096426	0.024442
α-helix	1	0.029539	0.013478
α-helix	5000	0.030726	0.011516
3_{10}-helix	1	0.124180	0.016910
3_{10}-helix	5000	0.006269	0.031663

(a) Glycine

Conformation	p (bar)	a	b
extended strand	1	0.19609	0.025722
extended strand	5000	0.0034658	0.0404725
α-helix	1	0.035971	0.00259585
α-helix	5000	-0.038459	0.0110305
3_{10}-helix	1	0.079105	0.0132015
3_{10}-helix	5000	-0.09776	0.0371135

(b) Alanine

Table G.2.: Fit parameters of the analytically integrated linear fit to the TFE concentration derivatives. Form of the fit: $\Delta G_{\mathrm{tr}}(c_{\mathrm{T}}) = a \cdot c_{\mathrm{T}} + b \cdot c_{\mathrm{T}}^2$. The unit of the TFE is $\mathrm{kJ} \cdot \mathrm{mol}_{aa}^{-1}$, and all concentrations are in $\mathrm{mol}_{\mathrm{TMAO}} \cdot \mathrm{L}^{-1}$.

Conformation	p (bar)	a	b
α-helix	1	0.016764	0.007649
α-helix	5000	0.017402	0.006522
3_{10}-helix	1	0.086528	0.011783
3_{10}-helix	5000	0.004361	0.022026

(a) Glycine

Conformation	p (bar)	a	b
α-helix	1	0.093914	0.017624
α-helix	5000	-0.029463	0.026300
3_{10}-helix	1	0.116838	0.020007
3_{10}-helix	5000	-0.025392	0.030222

(b) Alanine

Table G.3.: Fit parameters of the α- and 3_{10}-helix TFEs from scaling the extended-strand TFEs with SASAs of the helices (eq. 5.3). Form of the fit: $\Delta G_{\mathrm{tr}}(c_{\mathrm{T}}) = a \cdot c_{\mathrm{T}} + b \cdot c_{\mathrm{T}}^2$. The unit of the TFE is $\mathrm{kJ} \cdot \mathrm{mol}_{aa}^{-1}$, and all concentrations are in $\mathrm{mol}_{\mathrm{TMAO}} \cdot \mathrm{L}^{-1}$.

Appendix H

Conversion between molar concentrations at 1 bar and 5 kbar

The problem with using the molar concentration scale is that it depends on the volume, which depends on the pressure and temperature. In order to convert the TMAO concentrations c_{T} at 5 kbar to the concentrations at 1 bar, we use the bulk TMAO solution densities of our force fields (using the high-pressure version at 5 kbar) as follows:

$$c_{\mathrm{T}}^{1\ \mathrm{bar}} = c_{\mathrm{T}}^{5\ \mathrm{kbar}} \frac{\rho^{1\ \mathrm{bar}}(c_{\mathrm{T}}^{5\ \mathrm{kbar}})}{\rho^{5\ \mathrm{kbar}}(c_{\mathrm{T}}^{5\ \mathrm{kbar}})} \tag{H.1}$$

For this, we fitted the bulk densities (in g/L) as a function of the known concentrations at 5 kbar:

$$\rho^{1\ \mathrm{bar}}(c_{\mathrm{T}}^{5\ \mathrm{kbar}}) = 996.39 + 1.8176 \cdot c_{\mathrm{T}}^{5\ \mathrm{kbar}} + 0.4031 \cdot (c_{\mathrm{T}}^{5\ \mathrm{kbar}})^2 \tag{H.2}$$

$$\rho^{5\ \mathrm{kbar}}(c_{\mathrm{T}}^{5\ \mathrm{kbar}}) = 1152.8 - 4.6678 \cdot c_{\mathrm{T}}^{5\ \mathrm{kbar}} + 0.36887 \cdot (c_{\mathrm{T}}^{5\ \mathrm{kbar}})^2 \tag{H.3}$$

Publication list

1. Christoph Hölzl, Patrick Kibies, Sho Imoto, Roland Frach, Saba Suladze, Roland Winter, Dominik Marx, Dominik Horinek, and Stefan M. Kast. Design principles for high-pressure force fields: Aqueous TMAO solutions from ambient to kilobar pressures. *J. Chem. Phys.*, **144**, 144104, 2016.

2. Florian J. Wirkert, Christoph Hölzl, Michael Paulus, Paul Salmen, Metin Tolan, Dominik Horinek, and Julia Nase. The Hydrophobic Gap at High Hydrostatic Pressures. *Angew. Chem. Int. Ed.*, **56**, 12958, 2017.

3. Christoph Hölzl and Dominik Horinek. Pressure Increases the Ice-like Order of Water at Hydrophobic Interfaces. *Phys. Chem. Chem. Phys.*, **20**, 21257, 2018.

4. Vira Agieienko, Christoph Hölzl, Dominik Horinek, and Richard Buchner. The Interplay of Methyl-Group Distribution and Hydration Pattern of Isomeric Amphiphilic Osmolytes. *J. Phys. Chem. B*, **122**, 5972, 2018.

Bibliography

[1] E. G. Nisbet and N. H. Sleep. The habitat and nature of early life. *Nature*, 409(6823):1083–1091, feb 2001.

[2] Y. Levy and J. N. Onuchic. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.*, 35(1):389–415, 2006.

[3] P. Yancey, M. Clark, S. Hand, R. Bowlus, and G. Somero. Living with water stress: evolution of osmolyte systems. *Science*, 217(4566):1214–1222, 1982.

[4] K. Akasaka. Probing conformational fluctuation of proteins by pressure perturbation. *Chem. Rev.*, 106(5):1814–35, 2006.

[5] R. Ravindra and R. Winter. On the temperature - Pressure free-energy landscape of proteins. *ChemPhysChem*, 4(4):359–365, 2003.

[6] H. R. Kalbitzer, M. Spoerner, P. Ganser, C. Hozsa, and W. Kremer. Fundamental link between folding states and functional states of proteins. *J. Am. Chem. Soc.*, 131(9):16714–16719, 2009.

[7] E. E. Meyer, K. J. Rosenberg, and J. Israelachvili. Recent progress in understanding hydrophobic interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 103(43):15739–46, 2006.

[8] J. A. Long, B. M. Rankin, and D. Ben-Amotz. Micelle Structure and Hydrophobic Hydration. *J. Am. Chem. Soc.*, 137(33):10809–10815, 2015.

[9] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640, 2005.

[10] H. S. Frank and M. W. Evans. Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; Partial molal entropy in dilute solutions; Structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.*, 13(11):507, 1945.

[11] A. A. Bakulin, M. S. Pshenichnikov, H. J. Bakker, and C. Petersen. Hydrophobic molecules slow down the hydrogen-bond dynamics of water. *J. Phys. Chem. A*, 115(10):1821–1829, 2011.

[12] D. Laage, G. Stirnemann, and J. T. Hynes. Why water reorientation slows without iceberg formation around hydrophobic solutes. *J. Phys. Chem. B*, 113:2428–2435, 2009.

[13] K. Meister, S. Strazdaite, A. L. DeVries, S. Lotze, L. L. C. Olijve, I. K. Voets, and H. J. Bakker. Observation of ice-like water layers at an aqueous protein surface. *Proc. Natl. Acad. Sci. U. S. A.*, 111(50):17732–17736, 2014.

[14] C.-Y. Lee, J. A. McCammon, and P. J. Rossky. The structure of liquid water at an extended hydrophobic surface. *J. Chem. Phys.*, 80(9):4448, 1984.

[15] F. Sedlmeier, J. Janecek, C. Sendner, L. Bocquet, R. R. Netz, and D. Horinek. Water at polar and nonpolar solid walls (Review). *Biointerphases*, 3(3):FC23, 2008.

[16] C. S. Tian and Y. R. Shen. Structure and charging of hydrophobic material/water interfaces studied by phase-sensitive sum-frequency vibrational spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 106(36):15148, 2009.

[17] P.-L. Chau and A. J. Hardwick. A new order parameter for tetrahedral configurations. *Mol. Phys.*, 93(3):511, 1998.

[18] J. R. Errington and P. G. Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409(6818):318, 2001.

[19] J. Fidler and P. M. Rodger. Solvation Structure around Aqueous Alcohols. *J. Phys. Chem. B*, 103(36):7695–7703, 1999.

[20] A. H. Nguyen and V. Molinero. Identification of Clathrate Hydrates, Hexagonal Ice, Cubic Ice, and Liquid Water in Simulations: The CHILL+ Algorithm. *J. Phys. Chem. B*, 119(29):9369–9376, 2015.

[21] M. Mezger, H. Reichert, S. Schöder, J. Okasinski, H. Schröder, H. Dosch, D. Palms, J. Ralston, and V. Honkimäki. High-resolution in situ x-ray study of the hydrophobic gap at the water-octadecyl-trichlorosilane interface. *Proc. Natl. Acad. Sci. U.S.A.*, 103(49):18401, 2006.

[22] M. Mezger, F. Sedlmeier, D. Horinek, H. Reichert, D. Pontoni, and H. Dosch. On the origin of the hydrophobic water gap: An X-ray reflectivity and MD simulation study. *J. Am. Chem. Soc.*, 132(19):6735–41, 2010.

[23] F. J. Wirkert, C. Hölzl, M. Paulus, P. Salmen, M. Tolan, D. Horinek, and J. Nase. The Hydrophobic Gap at High Hydrostatic Pressures. *Angew. Chem. Int. Ed.*, 56:12958, 2017.

[24] P. H. Yancey and J. F. Siebenaller. Co-evolution of proteins and solutions: protein adaptation versus cytoprotective micromolecules and their roles in marine organisms. *J. Exp. Biol.*, 218(12):1880–1896, 2015.

[25] T. Arakawa and S. N. Timasheff. The stabilization of proteins by osmolytes. *Biophys. J.*, 47(3):411–4, 1985.

[26] A. 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):9101–8, 1997.

[27] P. H. Yancey and J. F. Siebenaller. Trimethylamine oxide stabilizes teleost and mammalian lactate dehydrogenases against inactivation by hydrostatic pressure and trypsinolysis. *J. Exp. Biol.*, 202(Pt 24):3597–603, 1999.

[28] Q. Zou, B. J. Bennion, V. Daggett, and K. 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.

[29] C. Y. Hu, G. C. Lynch, H. Kokubo, and B. M. Pettitt. Trimethylamine N-oxide influence on the backbone of proteins: an oligoglycine model. *Proteins*, 78(3):695–704, 2010.

[30] D. R. Canchi and A. E. García. Cosolvent effects on protein stability. *Annu. Rev. Phys. Chem.*, 64(December 2012):273–93, 2013.

[31] P. H. Yancey, M. E. Gerringer, J. C. Drazen, A. a Rowden, and A. Jamieson. Marine fish may be biochemically constrained from inhabiting the deepest ocean depths. *Proc. Natl. Acad. Sci. U. S. A.*, 111(12):4461–5, 2014.

[32] D. R. Canchi, D. Paschek, and A. E. García. Equilibrium study of protein denaturation by urea. *J. Am. Chem. Soc.*, 132(7):2338–2344, 2010.

[33] E. J. Guinn, L. M. Pegram, M. W. Capp, M. N. Pollock, and M. T. Record. Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. *Proc. Natl. Acad. Sci. U. S. A.*, 108(41):16932–16937, 2011.

[34] B. Moeser and D. Horinek. Unified Description of Urea Denaturation: Backbone and Side Chains Contribute Equally in the Transfer Model. *J. Phys. Chem. B*, 118(1):107–114, 2014.

[35] T.-Y. Lin and S. N. Timasheff. Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry*, 33(42):12695–12701, 1994.

[36] P. Ganguly, P. Boserman, N. F. A. van der Vegt, and J.-E. Shea. Trimethylamine-N-oxide Counteracts Urea Denaturation by Inhibiting Protein-Urea Preferential Interaction. *J. Am. Chem. Soc.*, 140(1):483–492, 2018.

[37] C. Krywka, C. Sternemann, M. Paulus, M. Tolan, C. Royer, and R. Winter. Effect of osmolytes on pressure-induced unfolding of proteins: a high-pressure SAXS study. *ChemPhysChem*, 9(18):2809–15, 2008.

[38] M. W. Jaworek, V. Schuabb, and R. Winter. The effects of glycine, TMAO and osmolyte mixtures on the pressure dependent enzymatic activity of α-chymotrypsin. *Phys. Chem. Chem. Phys.*, 20(3):1347–1354, 2018.

[39] M. A. Schroer, Y. Zhai, D. C. F. Wieland, C. J. Sahle, J. Nase, M. Paulus, M. Tolan, and R. Winter. Exploring the piezophilic behavior of natural cosolvent mixtures. *Angew. Chem. Int. Ed.*, 50:11413–11416, 2011.

[40] T. Q. Luong and R. Winter. Combined pressure and cosolvent effects on enzyme activity - a high-pressure stopped-flow kinetic study on α-chymotrypsin. *Phys. Chem. Chem. Phys.*, 2015.

[41] M. Gao, L. Arns, and R. Winter. Modulation of the Thermodynamic Signatures of an RNA Thermometer by Osmolytes and Salts. *Angew. Chem.*, 129(9):2342–2346, 2017.

[42] F. Rodríguez-Ropero, P. Rötzscher, and N. F. A. van der Vegt. Comparison of Different TMAO Force Fields and Their Impact on the Folding Equilibrium of a Hydrophobic Polymer. *J. Phys. Chem. B*, 120(34):8757–8767, 2016.

[43] S. C. Hyde, P. Emsley, M. J. Hartshorn, M. M. Mimmack, U. Gileadi, S. R. Pearce, M. P. Gallagher, D. R. Gill, R. E. Hubbard, and C. F. Higgins. Structural model of ATP-binding proteing associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature*, 346(6282):362–365, 1990.

[44] S. B. Prusiner. Chemistry and biology of prions. *Biochemistry*, 31(49):12277–12288, 1992.

[45] C. R. Brown, L. Q. Hong-Brown, J. Biwersi, A. S. Verkman, and W. J. Welch. Chemical chaperones correct the mutant phenotype of the delta F508 cystic fibrosis transmembrane conductance regulator protein. *Cell Stress Chap.*, 1(2):117–25, 1996.

[46] C. R. Brown, L. Q. Hong-Brown, and W. J. Welch. Correcting temperature-sensitive protein folding defects. *J. Clin. Investig.*, 99(6):1432–44, 1997.

[47] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen. Designing a 20-residue protein. *Nat. Struct. Mol. Biol.*, 9(6):425–430, 2002.

[48] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.*, 124(44):12952–12953, 2002.

[49] R. Zhou. Trp-cage: folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. U. S. A.*, 100:13280–13285, 2003.

[50] J. Juraszek and P. G. Bolhuis. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. U. S. A.*, 103(43):15859–15864, 2006.

[51] D. Paschek, H. Nymeyer, and A. E. García. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. *J. Struct. Biol.*, 157(3):524–533, 2007.

[52] D. Paschek, S. Hempel, and A. E. García. Computing the stability diagram of the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. U. S. A.*, 105:17754–17759, 2008.

[53] R. Day, D. Paschek, and A. E. Garcia. Microsecond simulations of the folding/ unfolding thermodynamics of the Trp-cage miniprotein. *Proteins*, 78(8):1889–1899, 2010.

[54] D. R. Canchi, P. Jayasimha, D. C. Rau, G. I. Makhatadze, and A. E. Garcia. Molecular mechanism for the preferential exclusion of TMAO from protein surfaces. *J. Phys. Chem. B*, 116(40):12095–104, 2012.

[55] Z. Su, F. Mahmoudinobar, and C. L. Dias. Effects of Trimethylamine-N-oxide on the Conformation of Peptides and its Implications for Proteins. *Phys. Rev. Lett.*, 119(10):108102, 2017.

[56] S. Kitazawa, M. J. Fossat, S. A. McCallum, A. E. Garcia, and C. A. Royer. NMR and Computation Reveal a Pressure-Sensitive Folded Conformation of Trp-Cage. *J. Phys. Chem. B*, 121(6):1258–1267, 2017.

[57] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577, 1995.

[58] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[59] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684, 1984.

[60] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182, 1981.

[61] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.*, 123(23):234505, 2005.

[62] J. L. Aragones, M. M. Conde, E. G. Noya, and C. Vega. The phase diagram of water at high pressures as obtained by computer simulations of the TIP4P/2005 model: The appearance of a plastic crystal phase. *Phys. Chem. Chem. Phys.*, 11(3):543–555, 2009.

[63] R. Kumar, J. R. Schmidt, and J. L. Skinner. Hydrogen bonding definitions and dynamics in liquid water. *J. Chem. Phys.*, 126, 2007.

[64] S. Imoto, H. Forbert, and D. Marx. Water structure and solvation of osmolytes at high hydrostatic pressure: pure water and TMAO solutions at 10 kbar versus 1 bar. *Phys. Chem. Chem. Phys.*, 17(37):24224, 2015.

[65] J. G. Kirkwood and F. P. Buff. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.*, 19(6):774, 1951.

[66] A. Ben-Naim. *Molecular theory of solutions.* Oxford University Press, 2006.

[67] D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, 2002.

[68] P. Ganguly and N. F. A. van der Vegt. Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 9(3):1347–1355, 2013.

[69] J. Milzetti, D. Nayar, and N. F. A. van der Vegt. Convergence of Kirkwood-Buff Integrals of Ideal and Non-Ideal Aqueous Solutions Using Molecular Dynamics Simulations. *J. Phys. Chem. B*, 122(21):5515–5526, 2018.

[70] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *J. Chem. Phys.*, 145(14):141103, 2016.

[71] J. Kästner. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1:932–942, 2011.

[72] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[73] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The Weighted Histogram Analysis Method for Free Energy Calculations on Biomolecules: I. The Method. *J. Comp. Chem.*, 13(8):1011–1021, 1992.

[74] J. S. Hub, B. L. De Groot, and D. Van Der Spoel. G-wham-a free Weighted Histogram Analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.*, 6:3713–3720, 2010.

[75] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, 1999.

[76] N. Metropolis, M. Rosenbluth, A. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(1953):1087–1092, 1957.

[77] A. Patriksson and D. van der Spoel. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.*, 10(15):2073, 2008.

[78] C. Hölzl, P. Kibies, S. Imoto, R. Frach, S. Suladze, R. Winter, D. Marx, D. Horinek, and S. M. Kast. Design principles for high-pressure force fields: Aqueous TMAO solutions from ambient to kilobar pressures. *J. Chem. Phys.*, 144(14):144104, 2016.

[79] S. Imoto, J. Noetzel, and D. Marx. personal communication.

[80] J. Vandevondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103–128, 2005.

[81] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. Vandevondele. Cp2k: Atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 4:15–25, 2014.

[82] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Phys. Rev. B*, 59:7413–7421, 1999.

[83] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15), 2010.

[84] P. Kibies and S. M. Kast. personal communication.

[85] T. Kloss, J. Heil, and S. M. Kast. Quantum chemistry in solution by combining 3D integral equation theory with a cluster embedding approach. *J. Phys. Chem. B*, 112(14):4337–43, 2008.

[86] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[87] C. M. Breneman and K. B. Wiberg. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. 11(3):361–373, 1990.

[88] C. Hölzl and D. Horinek. Pressure Increases the Ice-like Order of Water at Hydrophobic Interfaces. *Phys. Chem. Chem. Phys.*, pages 21257–21261, 2018.

[89] C. Hölzl. Master's thesis (unpublished). 2014.

[90] E. Papirer. *Adsorption on silica surfaces.* Marcel D., New York, USA, 1 edition, 2000.

[91] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.*, 24(16):1999, 2003.

[92] W. L. Jorgensen, D. S. Maxwell, and J. 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(45):11225–11236, 1996.

[93] M. Matsumoto, T. Yagasaki, and H. Tanaka. GenIce: Hydrogen-Disordered Ice Generator. *J. Comp. Chem.*, 39(1):61, 2018.

[94] William H., Andrew D., and Klaus S. VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[95] J. P. R. B. Walton, D. J. Tildesley, J. S. Rowlinson, and J. R. Henderson. The pressure tensor at the planar surface of a liquid. *Mol. Phys.*, 48(6):1357–1368, 1983.

[96] S. R. Wasserman, G. M. Whitesides, I. M. Tidswell, B. M. Ocko, P. S. Pershan, and J. D. Axe. The Structure of Self-Assembled Monolayers of Alkylsiloxanes on Silicon: A Comparison of Results from Ellipsometry and Low-Angle X-ray Reflectivity. *J. Am. Chem. Soc.*, 111(15):5852–5861, 1989.

[97] A. Rahman and F. H. Stillinger. Hydrogen-Bond Patterns in Liquid Water. *J. Am. Chem. Soc.*, 95(24):7943–7948, 1973.

[98] M. Matsumoto, T. Yagasaki, and H. Tanaka. Chiral Ordering in Supercooled Liquid Water and Amorphous Ice. *Phys. Rev. Lett.*, 115(19):197801, 2015.

[99] M. Matsumoto, A. Baba, and I. Ohmine. Topological building blocks of hydrogen bond network in water. *J. Chem. Phys.*, 127(13):134504, 2007.

[100] F. Sedlmeier, D. Horinek, and R. Netz. Nanoroughness, Intrinsic Density Profile, and Rigidity of the Air-Water Interface. *Phys. Rev. Lett.*, 103(13):136102, 2009.

[101] G. J. MacDonald. Role of methane clathrates in past and future climates. *Climatic Change*, 16(3):247–281, 1990.

[102] R. Noto, V. Martorana, A. Emanuele, and S. L. Fornili. Comparison of the water perturbations induced by two small organic solutes: ab initio calculations and molecular dynamics simulation. *J. Chem. Soc. Faraday Trans.*, 91(21):3803, 1995.

[103] K. M. Kast, J. Brickmann, S. M. Kast, and R. S. Berry. Binary Phases of Aliphatic N -Oxides and Water: Force Field Development and Molecular Dynamics Simulation. *J. Phys. Chem. A*, 107(27):5342–5351, 2003.

[104] E. Schneck, D. Horinek, and R. R. Netz. Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations. *J. Phys. Chem. B*, 117(28):8310–21, 2013.

[105] K. Usui, Y. Nagata, J. Hunger, M. Bonn, and M. Sulpizi. A new force field including charge directionality for TMAO in aqueous solution. *J. Chem. Phys.*, 145(6):064103, 2016.

[106] E. S. Courtenay, M. W. Capp, C. F. Anderson, and M. T. Record. Vapor Pressure Osmometry Studies of Osmolyte-Protein Interactions: Implications for the Action of Osmoprotectants in Vivo and for the Interpretation of "Osmotic Stress" Experiments in Vitro. *Biochemistry*, 39(15):4455–4471, 2000.

[107] I. Leontyev and A. Stuchebrukhov. Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys.*, 13(7):2613–2626, 2011.

[108] D. Markthaler, J. Zeman, J. Baz, J. Smiatek, and N. Hansen. Validation of Trimethylamine-N-oxide (TMAO) Force Fields Based on Thermophysical Properties of Aqueous TMAO Solutions. *J. Phys. Chem. B*, 121(47):10674–10688, 2017.

[109] R. A. Kuharski and P. J. Rossky. Molecular Dynamics Study of Solvation in Urea-Water Solution. *J. Am. Chem. Soc.*, 106(20):5786–5793, 1984.

[110] E. M. Duffy, D. L. Severance, and W. L. Jorgensen. Urea: Potential Functions, log P, and Free Energy of Hydration. *Isr. J. Chem.*, 33(3):323–330, 1993.

[111] J. Tsai, M. Gerstein, and M. 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.

[112] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, 91(1):1–41, 1995.

[113] S. Weerasinghe and P. E. Smith. A Kirkwood-Buff Derived Force Field for Mixtures of Urea and Water. *J. Phys. Chem. B*, 107(16):3891–3898, 2003.

[114] M. Knierbein, C. Held, and G. Sadowski. personal communication.

[115] M. Brehm and B. Kirchner. TRAVIS - A free analyzer and visualizer for monte carlo and molecular dynamics trajectories. *J. Chem. Inf. Model.*, 51(8):2007–2023, 2011.

[116] D. Horinek and R. R. Netz. Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations. *J. Phys. Chem. A*, 115(23):6125–36, 2011.

[117] F. Sokolić, A. Idrissi, and A. Perera. Concentrated aqueous urea solutions: A molecular dynamics study of different models. *J. Chem. Phys.*, 116(4):1636, 2002.

[118] G. Scatchard, W. J. Hamer, and S. E. Wood. Isotonic Solutions. I. The Chemical Potential of Water in Aqueous Solutions of Sodium Chloride, Potassium Chloride, Sulfuric Acid, Sucrose, Urea and Glycerol at 25°. *J. Am. Chem. Soc.*, 60(12):3061–3070, 1938.

[119] F. T. Jr. Gucker, F. W. Gage, and C. E. Moser. The Densities of Aqueous Solutions of Urea at 25 and 30 degree and the Apparent Molal Volume of Urea. *J. Am. Chem. Soc.*, 60:2582–2588, 1938.

[120] R. Storn and K. Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *J. Glob. Optim.*, 11(4):341–359, 1997.

[121] T. M. Nymand and P. Linse. Molecular dynamics simulations of polarizable water at different boundary conditions. *J. Chem. Phys.*, 112(14):6386–6395, 2000.

[122] P. J. Dunlop and L. J. Gosting. A Study of the Diffusion of Glycolamide in Water at 25° with the Gouy Interference Method. *J. Am. Chem. Soc.*, 75(20):5073–5075, 1953.

[123] H. Endo. The Adiabatic Compressibility of Nonelectrolyte Aqueous Solutions in Relation to the Structures of Water and Solutions. *Bull. Chem. Soc. Jpn.*, 46(4):1106–1111, 1973.

[124] P. Salmen and J. Nase. personal communication.

[125] J. Wyman. Dielectric Constants: Ethanol-Diethyl Ether and Urea-Water Solutions between 0 and 50°. *J. Am. Chem. Soc.*, 55(10):4116–4121, 1933.

[126] C. Tanford. Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *J. Am. Chem. Soc.*, 86(10):2050–2059, 1964.

[127] M. Auton and D. W. Bolen. Additive Transfer Free Energies of the Peptide Backbone Unit That Are Independent of the Model Compound and the Choice of Concentration Scale. *Biochemistry*, 43:1329–1342, 2004.

[128] M. Auton and D. W. Bolen. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. U. S. A.*, 102(42):15065–8, 2005.

[129] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. 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. Comput. Chem.*, 16(3):273–284, 1995.

[130] G. J. Lesser and G. D. Rose. Hydrophobicity of amino acid subgroups in proteins. *Proteins*, 8(1):6–13, 1990.

[131] S. Shimizu. Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments. *Proc. Natl. Acad. Sci. U. S. A.*, 101(5):1195–1199, 2004.

[132] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.

[133] C. Rosin, K. Estel, J. Hälker, and R. Winter. Combined Effects of Temperature, Pressure, and Co-Solvents on the Polymerization Kinetics of Actin. *ChemPhysChem*, 16(7):1379–1385, 2015.

[134] B. Moeser. *Protein Stability in Mixed Solvents: From Transfer Thermodynamics to the Denaturation by Urea and Back.* PhD thesis, Universität Regensburg, 2015.

[135] D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579, 1995.

[136] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–242, 2000.

[137] R. B. Best and J. Mittal. Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B*, 114(46):14916–14923, 2010.

[138] L. S. Stelzl and G. Hummer. Kinetics from Replica Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 13(8):3927–3935, 2017.

[139] D. Van Der Spoel and M. M. Seibert. Protein folding kinetics and thermodynamics from atomistic simulations. *Phys. Rev. Lett.*, 96(23):1–4, 2006.

[140] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11–15, Pasadena, CA USA, 2008.

[141] W. B. Floriano and M. A. C. Nascimento. Dielectric constant and density of water as a function of pressure at constant temperature. *Braz. J. Phys.*, 34(1):38–41, 2004.

[142] A. Garrett. inspyred: Bio-inspired algorithms in Python, 2015. [Online; accessed 2016-12].

Acknowledgments

In the end, i would like to thank the many people that directly and indirectly contributed to the existence of this work:

Most importantly, i thank my advisor Prof. Dr. Dominik Horinek, who gave me the freedom to follow my own interests while at the same time giving advice and inspiration when needed. His approach enabled me to get the most out of my time in his group.

I also owe thanks to Prof. Dr. Stefan M. Kast, who not only took the time to be a referee for this thesis but also welcomed me as a guest in his group in Dortmund.

I am very grateful to Dr. Beate Moeser, who mentored me during the early days of my thesis and who diligently answered my countless questions. She and my colleagues Sebastian, Heida, Minh, Eva, and Maximilian (among others) made my time in the office, during breaks, and on occasional evenings interesting, entertaining, and pleasant.

During my brief stay at the TU Dortmund the members of the group of Prof. Kast immediately made me feel at home. Special mentions go to Tim and Lukas, as well as my host Patrick, who also granted me permission to use some of his results in this thesis.

Within the DFG research group FOR1979 i had the pleasure to collaborate with too many people to list them all here. I am especially thankful to Dr. Julia Nase, Dr. Paul Salmen, Michael Knierbein, Dr. Christoph Held, and Prof. Dr. Gabriele Sadowski from the TU Dortmund and Dr. Sho Imoto, Jan Noetzel, and Prof. Dr. Dominik Marx from the Ruhr-University Bochum for allowing me to use their research as reference data for my own.

Furthermore, i am grateful to Prof. Dr. Matubayasi for giving me the opportunity to spend a couple of weeks at his group at Osaka University. My stay there was both educational and entertaining also thanks to Drs. Yamamori, Ishii, and Yamada.

My parents' significant contribution to this document is their ceaseless support, for which i will always be grateful.

Lastly, i thank my friends that simple were there whenever they were needed. You know who you are.