universität wien

Dissertation

Titel der Dissertation

Solvation of Biomolecules in Ionic Liquid - Water Mixtures

Verfasser
Mag. Michael Haberler

angestrebter akademischer Grad
Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2012

Studienkennzahl laut Studienblatt:	A 091 490
Dissertationsgebiet laut Studienblatt:	Molekulare Biologie
Betreuer:	Prof. Othmar Steinhauser

Danksagung

Ich möchte mich bei allen bedanken, die mich beim Erstellen dieser Arbeit begleitet und unterstützt haben. In dieser Zeit habe ich am Institut für Computer-gestützte biologische Chemie eine inspirierende Atmosphäre erlebt, in der man Rückschläge mit Humor verkraftet, Leistungen anerkannt werden und wo man sich mit Leidenschaft Herausforderungen widmen kann.

Mein Dank gilt vor allem meinem Betreuer Othmar Steinhauser, der mit Geduld und Vorausblick meine Ideen förderte, mir stets mit Begeisterung Hilfestellung leistete und mich in eine wissenschaftliche Arbeitsweise einführte, die einfach Spaß macht. Auch Christian Schröder war immer ein wichtiger Ansprechpunkt, der mit seiner Kompetenz und Erfahrung sowie mit seinem Humor zu kritisieren und zu motivieren wusste. Danken möchte ich auch meinen Kollegen Sonja Maurer und Gerhard König, die beide viele Jahre neben mir sitzen mussten und immer für Fröhlichkeit und gute Stimmung sorgen konnten. Auch Gregor Neumayr und Thomas Taylor möchte ich besonders danken. Die gemeinsame Arbeit am Projekt "Gepetto"war spannend und bereichernd. Schließlich gilt mein Dank auch Stefan Bruckner und Stefan Boresch, mit denen ich mich stets austauschen konnte und die mich so an den unterschiedlichsten Problemen der Computersimulation teilnehmen ließen.

An dieser Stelle möchte ich auch noch meiner Freundin Bettina, meinen Freunden und meiner Familie danken. Sie haben mich immer unterstützt und mir den so wichtigen Ausgleich verschafft.

Zusammenfassung

Die Solvatisierung von Biomolekülen in hydratisierten ionischen Flüssigkeiten wurde durch Molekulardynamik-Simulation untersucht. Das betrifft vor allem Protein-Stabilität, Solvens-Dynamik und -Struktur und sogar makroskopische Eigenschaften wie die dielektrische Konstante und Viskosität, die so auf atomarer Ebene erklärt werden. Drei Serien von hydratisierten ionischen Flüssigkeiten (1-Ethyl-3-methylimidazolium als Kation, Trifluoromethansulfonat als Anion und Wasser) mit Wasser-Molenbrüchen zwischen 1.0 und 0.5 wurden simuliert: Eine Serie mit dem reinen Solvens, eine Serie mit dem Solvens und dem Zink-Finger Protein (PDB-Kode 5ZNF), und eine dritte Serie mit dem Solvens und Ubiquitin (PDB-Kode 1UBQ).

Es werden neue Methoden für die Auswertung von Molekulardynamik Trajektorien präsentiert, die eine quantitative Beschreibung der Solvatisierung auf atomarer Ebene erlauben: Die Parameter-freie Voronoi-Tesselierung zur Aufteilung des Raums ermöglicht es benachbarte Moleküle eindeutig zu identifizieren und so einer Solvatationshülle zuzuordnen. Dadurch zeigt sich, dass sich die Kationen überproportional in der Solvatationshülle der Biomoleküle aufhalten, aber keine Präferenz für bestimmte Typen von Aminosäuren haben. Im Gegenzug bleiben die Anionen in der Nähe von positiv geladenen Aminosäuren. Die Dynamik der gesamten Solvatationshülle verlangsamt sich exponentiell im Verhältnis zum Wasser-Molenbruch. Die Translationsdiffusion verhält sich proportional zur Viskosität, entsprechend der Stokes-Einstein Formel. In Bezug auf Rotation zeigt einzig die Dynamik der molekularen Dipolmomente ein ähnliches Verhalten. Translations- und Rotationsbewegungen der Moleküle werden auch im jeweiligen Körper-fixierten Bezugssystem untersucht. Dadurch wird es möglich die molekulare Bewegung entlang und um die Achsen des Körper-fixierten Bezugsystems zu analysieren. Darüber hinaus wird eine Theorie der Rotationsdiffusion präsentiert, die Winkelgeschwindigkeits-Korrelationsfunktionen in Beziehung zu mittleren quadratischen Winkeldifferenz setzt.

Abstract

The effects of hydrated ionic liquids on biomolecular solvation have been investigated by means of molecular dynamics simulation. This includes phenomena such as protein stability, solvent dynamics and structure, and even macroscopic properties like the dielectric constant and viscosity, which can be explained at atomic resolution. Three series of hydrated ionic liquids (1-ethyl-3-methylimidazolium as cation, trifluoromethanesulfonate as anion, and water) with the water mole fraction varying between 1.0 and 0.5 were simulated: one series containing only the solvent, one the solvent and the zinc finger protein as solute (PDB-code 5ZNF), and one the solvent and ubiquitin as solute (PDB-code 1UBQ).

Novel methods for the analysis of molecular dynamics trajectories are presented to describe the solvation effect in a quantitative way: The parameter-free Voronoi decomposition of space permits to unambiguously identify neighbouring molecules and solvation shells. Thus it is found that the cation predominantly populates the protein solvation shell. However, it shows no preference for any types of amino acids. In contrast, the location of the anion with its higher charge density is restricted to positively charged amino acids. The dynamics of the protein solvation shells scales exponentially with the water mole fraction. The translational diffusion scales with viscosity as expected from the Stokes-Einstein relation. From the rotational diffusion, only the dipole moment axis relaxes at times that scale with the viscosity. The translational and rotational motion is transformed into the body-fixed frame for each molecule. This allows to decompose the motion of each molecule along and about body-fixed axes. A theory of rotational diffusion is presented linking the rotational velocity autocorrelation functions to the angular displacement, analogous to the Einstein relation of translational diffusion.

Contents

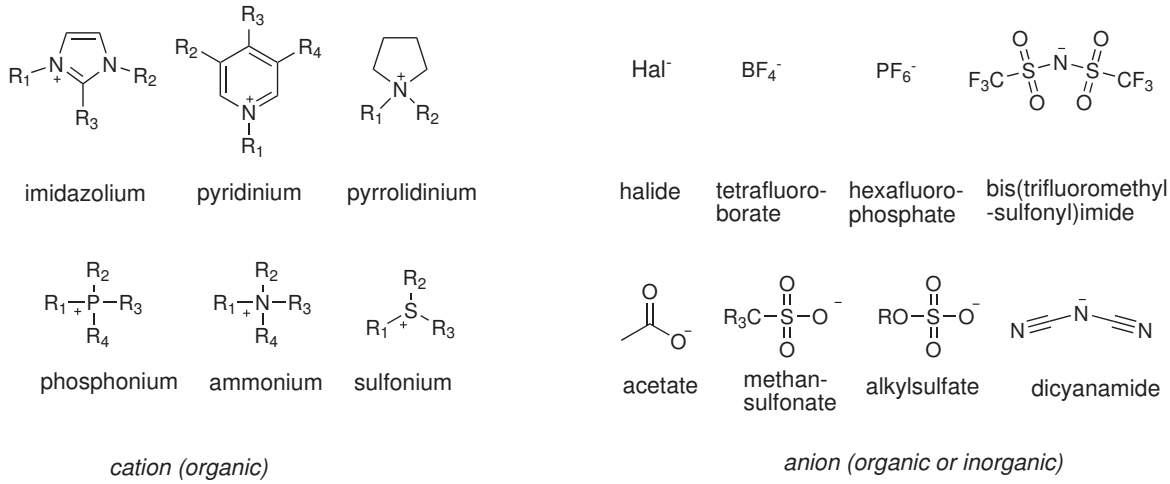


Figure 1: Typical cations and anions of ionic liquids (from Ref. [2]).

1 General Introduction

1.1 Motivation for Biomolecules in Hydrated Ionic Liquids

Ionic liquids (IL) have been discovered about 100 years ago and have gained increased attention in the past ten years with the development of innovative applications. Roughly speaking, ILs are salts that are liquid at room temperature. This is achieved by using large asymmetric (organic) molecules as ions which have their charge distributed over all their atoms. In other words, the ions do not arrange in such an ordered way as their commonly known inorganic salts do because of their anisotropic shape [1]. Figure 1 displays cations and anions commonly encountered in science and industry. Obviously, the physicochemical properties of ionic liquids vary depending on the cation-anion combination as well as on the choice of substituents on the ions.

ILs have two important properties: First, as a class of substances, they can dissolve a wide range of materials. Second, being a salt, they have negligible vapour pressure. Both characteristics are highly desirable in industrial applications and this is why research in this field is growing. Additionally, some ILs are miscible with water or organic solvents and in this way it is also possible to alter the physicochemical properties of ILs. In particular IL-water mixtures, so-called hydrated ILs, show promise for biotechnological purposes. It might well be possible that hydrated ILs can mimic the complex environment within which biochemical processes occur similar to molecular crowding agents.

For example, the enzyme *Candida antarctica lipase B* (CALB) was shown to be more stable in the IL BMIM-PF_6 (1-butyl-3-methylimidazolium hexafluorophosphate) as compared to t-butanol which allowed the temperature to be increased for the synthesis of a pharmaceutical intermediate (see left side of Figure 2). Thus the catalysed hydrolysis reaction was four times faster [3].

Another example is the dissolution of cellulose and even wood in ionic liquids. In a high-throughput screening EMIM-OAc (1-ethyl-3-methylimidazolium acetate) was found to be the most efficient IL for this purpose (see right side of Figure 2). Consequently, the dissolved cellulose is accessible to further downstream processing, such as

Figure 2: Left side: The reaction catalysed by CALB in different solvent media (from Ref. [6]). Right side: The ionic liquid EMIM-OAc (yellow), wood chips of the common birch, and completely dissolved wood chips in EMIM-OAc (brown) (from Ref. [4]).

hydrolysis and fermentation to produce biofuel [4].

Further examples of applications for ILs can be found in the extensive reviews of Plechkova and Seddon [5], Gorke *et al.* [6], and Moniruzzaman *et al.* [7].

So far however, ILs suitable for specific purposes have been discovered arguably by trial and error. It would be highly desirable to know how the molecular structure of an ionic liquid and its mixture with additional components affect its physical properties so that it is possible to rationally design ILs to suit specific purposes. In turn, this could lead to more efficient biotechnological processes that reduce economic and environmental expenses.

1.2 Solvation of Biomolecules

Although there are a couple of highly probable models on the origin of life most include the liquid phase, water, as the indispensable environment [8]. The aqueous solvent environment has been shown in numerous studies to not only stabilise biomolecules such as proteins and nucleic acids but actively engage in biological processes involving the former [9]. Examples include protein folding, water acting as acid or base in biochemical reactions, and ligand-receptor recognition and binding. Moreover, in biochemical systems one can rarely find a solute and solvent in isolated pure form. Additional biomolecules, such as lipid membranes, and inorganic molecules populate this environment and further modulate the physicochemical properties. Therefore solvent chemistry and in particular the process of solvation are important branches of research, both for industrial as well as medical purposes.

The IUPAC (International Union Of Pure and Applied Chemistry) defines in its "Gold Book" [10] "solvation" as "any stabilising interaction of a solute and the solvent. Such interactions generally involve electrostatic forces and Van der Waals forces, as well as chemically more specific effects such as hydrogen bond formation." In general the solvation free energy, or solvation free energy difference, serves as an excellent measure to describe the amount of work that a thermodynamic system can perform, i.e. the extent of the solvation process. Moreover, protein-ligand binding and protein folding, two very important research branches in biochemistry, can be quantified in terms of solvation free energy differences as well. Also, the recent development of powerful

computers helped in the application of theoretic models of interactions in complex biomolecular systems in order to explain biological processes in terms of solvation free energy.

With regard to ionic liquids Deschamps *et al.* have conducted computer simulation studies to calculate the solvation free energy differences for small molecules in imidazolium based ionic liquids and water (e.g. argon, oxygen, nitrogen, methane, carbon dioxide, water) [11].

However, the solvation free energy is determined by the number of possible states of a molecular system at equilibrium. This number quickly rises to virtually infinity for reasonably sized systems comprising biomolecular solutes and a solvent environment and the correct calculation of free energies depends decisively on the representative sampling of this immense state space. Gerhard König conducted his research in this area and revealed deficiencies in coarse models that try to circumvent this difficulty inherent to this thermodynamic quantity [12]. In general it is hardly possible to decompose the solvation free energy difference into contributions from groups of atoms and even individual atoms, their arrangement and motions.

The problem remains: How do the molecular structure and dynamics of a molecular system (such as solute and solvent) affect its macroscopically readily observable physicochemical properties? An answer could provide routes to choose a suitable cation-anion combination (for instance from Figure 1) in such a away to obtain the desired properties, such as improved stabilisation of an enzyme solute and increased velocities of biochemical reactions.

Besides the calculation of the solvation free energy there are numerous other techniques to study the liquid phase on the molecular level, in particular solvation. Theoretical approaches, such as molecular dynamics (MD) and Monte Carlo simulations, permit to investigate molecular systems at spatial resolutions and time scales inaccessible to experimental procedures. They rely on a model of atomic interactions based on classical mechanics that can explain a wide array of phenomena that go in hand with solvation. Section 1.3 will provide a brief overview of the theory of MD computer simulation which was the method to investigate the topic of this thesis. Quantum mechanical methods that account for interactions at subatomic level, and coarse grained MD simulation that only consider groups of multiple atoms, as well as hybrids thereof present complementary theoretical approaches for spatial resolutions and time scales outside the scope of MD.

As an alternative, experimental methods can be employed to study solvation from a different aspect. Among the most frequently used techniques are X-ray crystallography to locate crystal water sites, neutron diffraction to study liquid structure [13], NMR spectroscopy, in particular the magnetic relaxation dispersion technique and the pulsed-field-gradient spin-echo method that permit insight into single molecule dynamics [14, 15], time-resolved fluorescence spectroscopy to study solvent dynamics near solutes [16], and dielectric dispersion spectroscopy for collective dipole moment reorientation and displacement [17]. However, these experimental methods usually profit crucially from models and refinements based on computational techniques.

1.3 Theory of Molecular Dynamics

Molecular dynamics (MD) computer simulations have been applied in this thesis to study the effects of biomolecular solvation in ionic liquid water mixtures. Here, a brief overview of this method will be given with a special focus on those aspects that form the basis of the following publications. For this purpose many intricate details of MD must be dropped and for which the reader is referred to the comprehensive textbooks "Computer simulations of liquids" [18] and "Molecular Modelling: Principles and Applications" [19].

MD uses classical mechanics, i.e. Newton's laws of motion, to simulate the physical movements of atoms and molecules. This distinguishes it from quantum mechanical methods that also consider the wave like properties of atoms and subatomic particles, electrons and atomic nuclei. In fact, a sound theoretical justification for MD to ignore quantum mechanics is given by the Ehrenfest theorem. In brief, it states that the mean value of an observable, for instance the momentum $\mathbf{p}(t)$ of an atomic nucleus, follows classical mechanics. Therefore, it is sufficient to solve the differential equation below expressing Newton's second law to perform a simulation of atoms.

$$\frac{d^2\mathbf{x}_i(t)}{dt^2} = \frac{\mathbf{F}_i(t)}{m_i} \tag{1}$$

The force $\mathbf{F}_i(t)$ acting on the atom i with mass m_i determines the acceleration $\mathbf{a}_i(t) = d^2\mathbf{x}_i(t)/dt^2$ which in turn can be integrated numerically for consecutive timesteps $t = \Delta t \cdot n$ in intervals of Δt ($n = 0, 1, 2, 3, ...$). The molecular systems considered in this thesis included from 12000 up to 27000 atoms for each of which the force was calculated for up to 100 million times with time steps of $\Delta t = 0.002$ picoseconds which amounts to 200 nanoseconds.

In the present study the leap-frog integration algorithm has been used to calculate the positions $\mathbf{x}_i(t)$ and velocities $\mathbf{v}_i(t)$. The name accounts for the asynchronous calculation of the two variables which requires a third step (Equ. 4) to obtain them for the same time t.

$$\mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \Delta t \cdot \mathbf{v}_i(t + \frac{1}{2}\Delta t) \tag{2}$$

$$\mathbf{v}_i(t + \frac{1}{2}\Delta t) = \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \Delta t \cdot \mathbf{a}_i(t) \tag{3}$$

$$\mathbf{v}_i(t) = \frac{1}{2}\big(\mathbf{v}_i(t + \frac{1}{2}\Delta t) + \mathbf{v}_i(t - \frac{1}{2}\Delta t)\big) \tag{4}$$

The series of consecutive positions and velocities for all simulated atoms is called *trajectory* and serves as the raw data for the present study. Typically, the trajectory was calculated in consecutive slices of 50000 steps (100 ps) each on single nodes of the Vienna Scientific Cluster using eight processor cores. Such a slice was calculated in about 30 minutes for the small systems and in up to 90 minutes for the large systems. The position trajectory slices were stored to files for the entire 200 ns simulation yielding 2000 files in the size of hundreds of megabytes each if only every 50th step was actually recorded. Such a coarse graining however eliminates most of the information in velocity trajectories where the individual atomic velocities frequently change. Thus, for some 2 ns of each system's simulation time the position and velocity trajectory was recorded at each timestep resulting in over 600 gigabytes of data for each system.

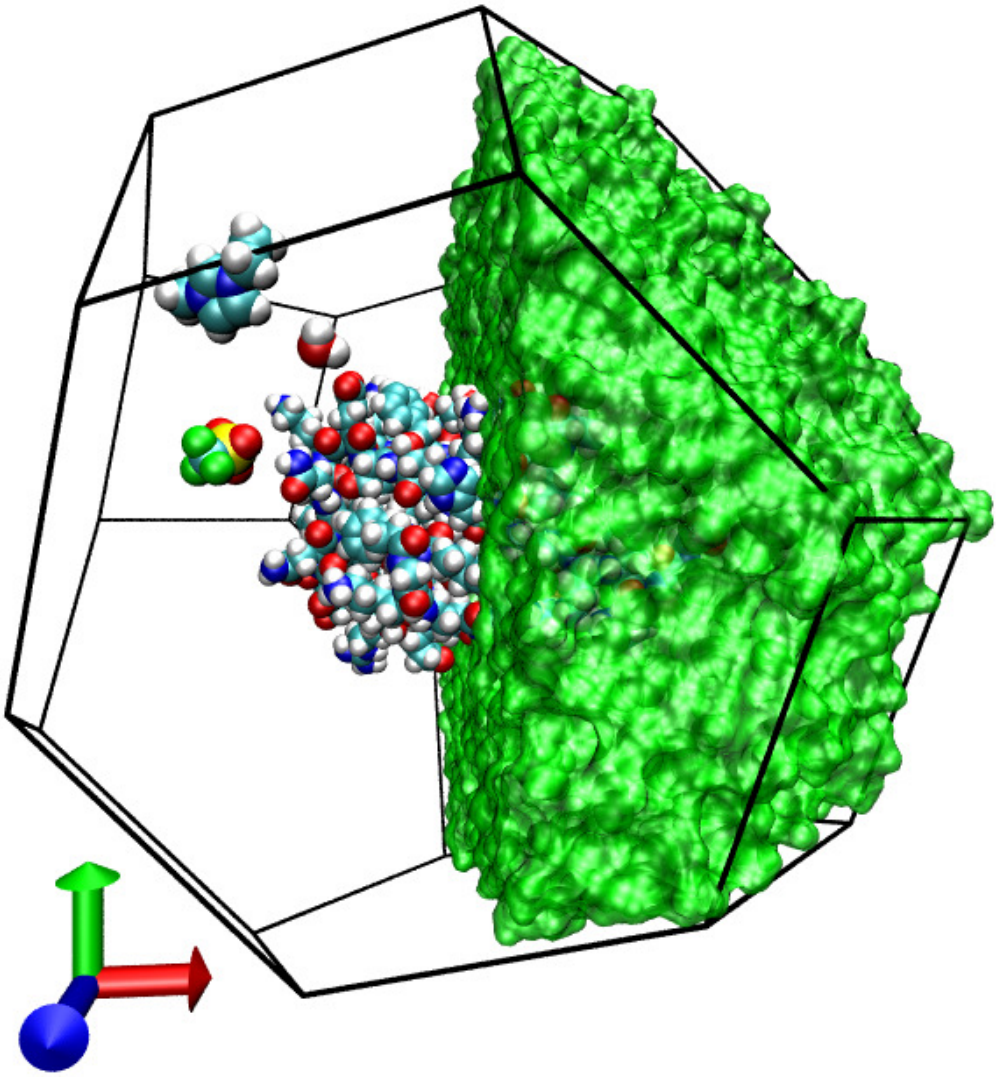


Figure 3: Snapshot of the trajectory of ubiquitin in the hydrated ionic liquid EMIM-TRIF-water. The simulation cell, a truncated octahedron, is outlined by black lines. In the right half of the cell the surface of all atoms is depicted in green. In the left half the only four molecules are visible: parts of the protein ubiquitin and the three solvent molecules EMIM, TRIF, and water (white spheres=hydrogen, turquoise=carbon, blue=nitrogen, red=oxygen, yellow=sulfur, green=fluoride).

One of the major resource drawing parts of MD simulations is the calculation of the force $\mathbf{F}_i(t)$ on each atom i which is ultimately responsible for the atom's behaviour. The forces are determined by the interactions between the atoms which are usually modelled by an empirical potential V, also called *force field*, which is treated pair-wise additive as an approximation.

$$\mathbf{F}_i = -\nabla V(\mathbf{x}_i) = -\nabla \sum_j V(\mathbf{x}_i, \mathbf{x}_j) \tag{5}$$

For the purposes of MD simulations the Born-Oppenheimer approximation is valid, i.e. electrons move much faster than atomic nuclei. In consequence, the Hellmann-Feynman theorem can be used to calculate covalent as well as electrostatic and Van der Waals forces among the atoms, considering only the positions of the atom nuclei. Thus, the empirical force field from CHARMM [20] that has been used in the present thesis can be given in compact form describing the total potential energy V of atom i

$$\begin{aligned} V(\mathbf{x}_i) = & \sum_{bonds} K_{ij}(r_{ij} - r_{ij,0})^2 \\ & + \sum_{angles} K_{ijk}(\theta_{ijk} - \theta_{ijk,0})^2 \\ & + \sum_{dihedrals} K_{ijkl}\big(1 - \cos(n_{ijkl}\phi_{il} - \delta_{ijkl})\big) \\ & + \sum_i \sum_{i<j} \varepsilon_{ij}\Big[\big(\frac{r_{\min,ij}}{r_{ij}}\big)^{12} - 2\big(\frac{r_{\min,ij}}{r_{ij}}\big)^{6}\Big] \\ & + \sum_i \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \end{aligned} \tag{6}$$

Here, five different types of interactions are included, expressed conveniently in so called internal coordinates (r_{ij}, θ_{ijk}, and ϕ_{il}) which can be derived from the cartesian coordinates $(\mathbf{x}_i, \mathbf{x}_j)$:

1. the covalent bond to an atom j with the force constant K_{ij} and the equilibrium distance $r_{ij,0}$,

2. the angle between the bonds of atoms i and k to an atom j with the force constant K_{ijk} and the equilibrium value $\theta_{ijk,0}$,

3. the dihedral angle between the plane constructed by the three atoms i, j, and k, and the plane from the atoms j, k, and l with the force constant K_{ijkl}, the multiplicity of dihedral angle n_{ijkl} (1 to 3 in biomolecules), and a shift constant δ_{ijkl} (0° or 180° degrees),

4. the Lennard-Jones potential to model the Van der Waals interaction between the atoms i and j with the parameters $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ and $r_{\min,ij} = (r_{\min,i} + r_{\min,j})/2$,

5. and the Coulomb potential for electrostatic interactions with the dielectric permittivity constant ε_0, partial atomic charges q_i and q_j, as well as the interatomic distance r_{ij}.

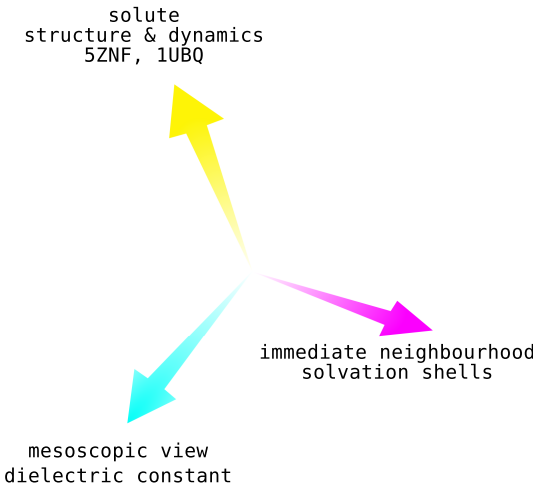
It has to be noted that this is a qualitative picture that conceals the numerous additional considerations of which the MD computer simulation program has to take care of. For instance, it is crucial to construct neighbour lists for Van der Waals interactions and to use the particle mesh Ewald summation for electrostatic interactions in order to cope with the long-range nature of Coulomb forces and to avoid the double summation over all atoms and thus massively reduce the computational effort.

Furthermore, the choice of the parameters, like the force constants, the equilibrium coordinates, and the atomic partial charges, will eventually be reflected in the results to a greater or lesser extent. These parameters are carefully selected to reproduce experimentally measured quantities, such as density, enthalpy of vaporisation, self-diffusion rate, and the dielectric constant. Unfortunately, it is quite impossible to develop force fields for complex biomolecular systems that correctly describe all experimentally measurable properties and at the same time provide a compact model that can be efficiently simulated. Therefore, slight divergences have to be accepted in favour of a reduced computational effort. For example, the water in the hydrated ionic liquids systems studied in this thesis is represented by the TIP3P model [21] which matches experimental values for the enthalpy of vaporisation and the dielectric constant at the expense of a reduced density and an enhanced self diffusion rate. This is of course a much bigger issue in the case of the ionic liquid molecules for which empirical models and experimentally measured quantities rarely exist.

Now molecular dynamics simulations per se provide invaluable insights into molecular structure and motion based on used model of molecular interactions. However, the concepts of statistical mechanics provide the link between the molecular level of simulation and the experimentally measurable macroscopic properties of materials built up by these molecules, such as pressure, viscosity, or free energies.

One of these concepts is the *statistical ensemble* that represents a theoretical collection of all possible configurations ($\mathbf{x}_i$ and $\mathbf{v}_i$ for all atoms i) of a molecular system and that enables to extract thermodynamic properties, i.e. macroscopic properties. The molecular systems simulated in this thesis were members of *canonical ensembles*, also referred to as NVT-ensembles, which have in common that the number of particles N, the volume V, and the temperature T remain constant. This implies that they are in contact to a heat bath, a thermostat, which is another detail of which the MD simulation program has to take care of.

In fact, the notion of statistical ensembles provides a framework to calculate macroscopic properties, such as the electrical conductivity from the time average of the fluctuations of the charge current which is derived in detail in the Theory Section of Ref. [22].

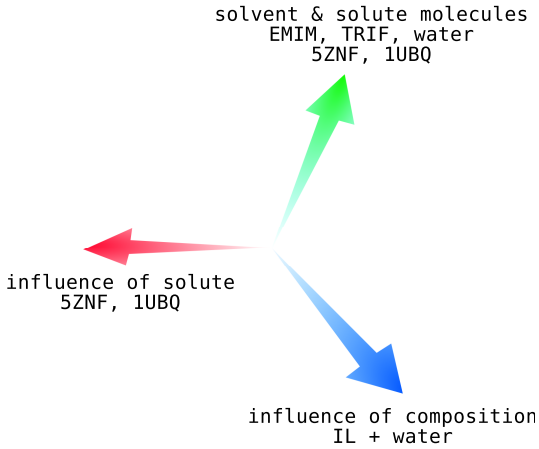


1.4 Summary/Roadmap of the Following Publications

In this thesis I will explain the influence of the composition of a hydrated IL on its structure and dynamics as well as on that of a solvated protein. Vice versa, I will also point out the effects of the solvated protein on the hydrated IL. To this end I have conducted molecular dynamics simulations of a series of hydrated ILs at various mole fractions without any solute, as well as a series containing a small zinc finger protein and one containing the larger ubiquitin. I have developed novel methods for the investigation of protein and solvent structure based on the work on spatial Voronoi decomposition by Gregor Neumayr [23]. This allowed to unambiguously identify neighbouring molecules and solvation shells without need for parameters. Furthermore I have studied molecular motion in the body-fixed frame as opposed to the laboratory frame. This enabled us to devise a theory of anisotropic translational and rotational diffusion analogous to the existing theory of isotropic translational diffusion [18].

The first article, *Solvation Studies of a Zinc Finger Protein in Hydrated Ionic Liquids* (Section 3), focuses on the protein and its solvation shells. The variation of the water mole fraction revealed a solvent composition where the structural properties of the protein are extremal. In aqueous solutions the electrostatic interactions between ions are screened by the effect of the collective dynamics of the water dipoles. Molecular ions do have dipoles but it is their net charge that largely determines their electrostatic potential. Therefore, in an ionic liquid, the interaction between two ions is reduced by the ions in between, i.e. by their charges. We describe this as charge screening and the extremal protein characteristics at a relatively high water mole fraction is interpreted as the transition point from dipolar screening to charge screening of electrostatic interactions. This is also confirmed by the initial sharp drop of the static dielectric constant when reducing the water content in the hydrated IL.

The second article, *On the influence of hydrated ionic liquids on the dynamical structure of model proteins: A computational study* (Section 4), expands the first in several ways: A larger protein, ubiquitin, is used, a broader range of water mole fractions is considered, and alternative ways to study the mutual influences between the protein and the solvent are developed. Specific regions of ubiquitin having no secondary structure are tracked down to be responsible for the characteristic features displayed



by the protein as a whole. The composition of the solvent in the first solvation shell around proteins is different from the bulk solvent composition: cations are enriched at the protein surface whereas the bulk solvent establishes a strong anion-water network. At the lowest water mole fractions we find that any water dipole dynamics is controlled by anion dipole dynamics and thus the dielectric effect is completely determined by the ionic liquid and stays almost constant irrespective of further reduction of water content.

Finally, the article *Hydrated Ionic Liquids With and Without Solute: The Influence of Water Content and Protein Solutes* (Section 5) presents an extensive study of the structure and – in particular – of the dynamics of the solvent. This time, the influence of the protein on the solvent is of special interest. We incorporate radial distribution functions, generalised to averaged angular functions and decomposed by Voronoi shells, to illustrate the liquid structure. We present that viscosity scaling applies to the average translation diffusion of the centre-of-mass as well as to the reorientation of the dipole moment but not to the reorientation of perpendicular axes. We decompose the translational and rotational motion in the body-fixed frame and observe the necessity to incorporate the metric, the off-diagonal elements of the diffusion tensor in the body-fixed frame. Furthermore, we establish a relation between single particle and collective dynamics which encourages to estimate the dielectric spectra of hydrated IL - protein solutions.

2 Declaration by the Author

In this thesis I have investigated the influence of the composition of a hydrated IL on its structure and dynamics as well as on that of a solvated biomolecule. The results have been summarised in three articles of which two have been published (Section 3 and Section 4) and one is in revision (Section 5).

Additionally, I have added a documentation of the software that I have developed in cooperation with Gregor Neumayr and Thomas Taylor (Section 7).

Section 3
: M. Haberler, C. Schröder, and O. Steinhauser. Solvation studies of a zinc finger protein in hydrated ionic liquids. *Phys. Chem. Chem. Phys.*, 13:6955–6969, 2011

Section 4
: Michael Haberler and Othmar Steinhauser. On the influence of hydrated ionic liquids on the dynamical structure of model proteins: a computational study. *Phys. Chem. Chem. Phys.*, 13:17994–18004, 2011

Section 5
: Michael Haberler, Christian Schröder, and Othmar Steinhauser. Hydrated ionic liquids with and without solute: The influence of water content and protein solutes. 2012. in revision

I have conducted all the MD simulations and the entire analysis all on my own for the articles listed. O. Steinhauser provided assistance in the design of the studies, in the derivation and development of the underlying theory, and interpretation of results. C. Schröder helped setting up the MD simulations, in the derivation of the underlying theory, and in the interpretation of results.

Moreover, the studies profited from the weekly reviews and discussions with the entire team: O. Steinhauser, C. Schröder, G. Neumayr, T. Taylor, and S. Maurer.

3 Solvation Studies of a Zinc Finger Protein in Hydrated Ionic Liquids

PCCP

View Online
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 6955–6969

www.rsc.org/pccp

PAPER

Solvation studies of a zinc finger protein in hydrated ionic liquids

Michael Haberler, Christian Schröder and Othmar Steinhauser

Received 9th September 2010, Accepted 1st February 2011
DOI: 10.1039/c0cp02487b

The solvation of the zinc finger protein with the PDB-ID "5ZNF" in hydrated ionic liquids was studied at varying water content. 1-Ethyl-3-methylimidazolium and trifluoromethanesulfonate were the cation and anion, respectively. The protein stability as well as the solvation structure, the shell dynamics and the shell resolved dielectric properties were investigated by means of molecular dynamics simulations. The lengths of the respective trajectories extended up to 200 nanoseconds in order to cover the complete solvent dynamics. Considering the above mentioned properties as a function of the water content they all exhibit a maximum or minimum at the very same mole fraction. While the exact value $x_{H_2O} = 0.927$ depends on the underlying force field, its origin may be traced back to the competition between the van der Waals and the electrostatic energy of the protein as well as to the transition from aqueous dielectric screening to ionic charge screening with decreasing water content. The parameter-free Voronoi decomposition of space served as a basis for the analysis of most results. In particular, solvation shells were naturally inferred from this concept. In addition to the molecular analysis a mesoscopic view is given in terms of dielectric properties. Thereby, the net dielectric constant is decomposed into contributions from the protein, the first and second solvation shells as well as the bulk. Cross-terms between these components are given, too.

I. Introduction

Molecular ionic liquids (MIL) are an environmentally friendly alternative to traditionally used volatile organic compounds. This role as green chemistry solvents becomes most obvious in biomolecular solvation. In fact, MILs may enhance the stability and/or the catalytic activity of certain enzymes.[1,2] Furthermore, the range of enzymes can be extended using mixtures of water and MILs, so-called hydrated ionic liquids. Thereby, water as well as MILs may act as a co-solvent.[3–13]

While macroscopic, thermodynamic or kinetic data are available for many enzyme- or protein-MIL (-water) solutions, *e.g.* ref. 14 and citations therein, their explanation at the molecular level lags behind. In other words, experimentally and computationally, *e.g.* ref. 15, molecular solvation studies so far focus on small molecules like, CO_2, benzene, chromophores, *etc.* Experimentally, a summary of structural data from neutron diffraction may be found in ref. 16. Corresponding dynamic results obtained by fluorescence spectroscopy and similar techniques are reported in ref. 17–19. A comprehensive summary of simulation studies is given in ref. 20.

Solvation studies of proteins at molecular resolution focus mainly on the aqueous environment.[21–23] Rare examples where ionic liquids act as a solvent or co-solvent are discussed ref. 24 and 25.

This study presents extensive simulations and detailed analyses of a zinc finger protein in hydrated ILs. We have chosen this model protein for several methodological reasons: first, the solute should exhibit characteristic secondary structure elements of a protein, α-helix and β-sheet. Second, it should be small enough to permit the inclusion of a sufficient number of solvent shells. This is important for several reasons: (a) the molecular volume of the cations and anions of MILs exceeds that of water by a factor of three to seven. Therefore, even a single solvation shell needs considerable space. (b) Due to the strong electrostatic coupling of charged species one needs several solvation shells in order to achieve the undisturbed formation of the first and second solvation layers. If the number of shells would be restricted to these first layers boundary artifacts would emerge. (c) In order to avoid surface artifacts computer simulations of solutions are performed under periodic boundary conditions. In an aqueous environment the high dielectric constant of water perfectly screens the protein solute and thus suppresses spurious self-interactions between solute replicas. As the dielectric constant of MILs is 10 to 15[26–28] the screening effect is modest and has to be compensated by the inclusion of more solvation shells. Third, the complete system, protein plus solvent, must be simulated for a time period of 100 nanoseconds and more in order to permit appropriate sampling that covers slow relaxation

University of Vienna, Department of Computational Biological Chemistry, Währingerstr. 17, 1090 Vienna, Austria. E-mail: os@mdy.univie@ac.at

Table 1 Basic properties of the simulated systems: water mole fraction x_{H_2O}, ion concentration c_{IL}, number of molecules #, mass density ρ, edge length of the truncated cuboctahedron d, MD-system volume V. The water mole fraction x_{H_2O} and the ionic strength, or equivalently, the ion concentration c_{IL}, are used to label the MD-systems.

x_{H_2O}	c_{IL} /mol l^{-1}	#EMIM$^+$	#$CF_3SO_3^-$	#H_2O	ρ/g l^{-1}	d/Å	V/Å^3
1.000	0	0	4	5061	1022.29	58.6	155 035
0.927	2.36	200	204	2557	1206.82	56.8	140 873
0.854	3.38	300	304	1766	1286.62	57.6	147 330
0.751	4.09	312	316	947	1343.99	54.8	126 670
0.675	4.41	345	349	721	1369.44	55.2	129 792

processes. We think that the zinc finger protein with the Protein Data Bank (PDB) ID "5ZNF" fulfils all these requirements to a high extent. As compared to the standard zinc finger motif PDB-ID "1ZNF" the number of charged amino acids is enhanced thus offering additional target sites for strong electrostatic interaction with solvent ions.

As a solvent we use the hydrated ionic liquid with 1-ethyl-3-methylimidazolium (EMIM$^+$) as the cation and trifluoromethanesulfonate ($CF_3SO_3^-$) as the anion. EMIM$^+$ was chosen because—together with its butyl-analogue—it is the most frequently used cation in experimental studies. $CF_3SO_3^-$ was used because of its resemblance to the frequently used bis(trifluoromethane)sulfonimide[9] and the absence of conformational flexibility which would complicate interpretation. Furthermore, its high polarity enables the necessary dielectric screening. In order to elucidate the mutual interplay of the three solvent species, EMIM$^+$, $CF_3SO_3^-$ and water, both in the solvent layers of the protein and in the bulk, we have studied five systems of varying molar water fractions in Table 1, three of which cover 200 nanoseconds.

Solvation studies are intrinsically connected with the concept of solvation shells. Traditionally, distance-based methods are used to define these shells. However, this involves a plethora of parameters which must be predefined. As a major methodological innovation our analysis of solvation shells is exclusively based on the Voronoi method which is a parameter-free technique to decompose space. Although computationally very expensive Voronoi decomposition of space is prerequisite for most results presented in this study which comprise: protein stability in terms of volume, surface and contact matrices, solvation structure in terms of coordination numbers, solvent contacts and distribution as well as shell dynamics characterised by mean residence times. Finally, the dielectric behaviour of the whole solution is decomposed into contributions from the protein, the first and second solvation shells and bulk.

II. Theory

A Dielectric permittivity and mesoscopic electrostatic interaction energy

If a sample containing a set of charges $\{q_i\}$ located at positions $\{\vec{r}_i\}$ is exposed to a spatially homogeneous external electric field $\vec{E}_0$, each charge experiences a potential $\phi_i = -\vec{r}_i \cdot \vec{E}_0$. Therefore, the interaction energy of the charge set with the external field is given by

$$U_{int} = \sum_i q_i(-\vec{r}_i \cdot \vec{E}_0) = -\left(\sum_i q_i \vec{r}_i\right) \cdot \vec{E}_0 = -\vec{M} \cdot \vec{E}_0 \quad (1)$$

Here, $\vec{M} = \sum_i q_i \vec{r}_i$ is the total dipole moment and is a measure of the electric anisotropy of the sample. It should be noticed that $\vec{M}$ is uniquely defined for a neutral charge set only ($\sum_i q_i = 0$). Otherwise a point of reference $\vec{r}_{ref}$ has to be specified

$$\vec{M} = \sum_i q_i(\vec{r}_i - \vec{r}_{ref}) = \sum_i q_i \vec{r}_i - \left(\sum_i q_i\right)\vec{r}_{ref} \quad (2)$$

In equilibrium, *i.e.* in the absence of an external field, the average value of $\langle\vec{M}\rangle_{eq}$ is zero. In the presence of an external field a residual value $\langle\vec{M}\rangle_{\vec{E}_0}$ remains. It can be shown that for not too strong external fields this residual value is given by

$$\langle\vec{M}\rangle_{\vec{E}_0} = \frac{\langle\vec{M}^2\rangle_{eq}}{3k_B T}\vec{E}_0 \quad (3)$$

Therefore, the average interaction energy of the charge set with the external field is proportional to the mean square equilibrium dipole moment[22]

$$\langle U_{int}\rangle = -\langle\vec{M}\rangle_{\vec{E}_0} \cdot \vec{E}_0 = -\frac{\langle\vec{M}^2\rangle_{eq}}{3k_B T}\vec{E}_0^2 \quad (4)$$

These theoretical relations are linked to experiment by the so-called constitutive relation

$$\frac{\langle\vec{M}\rangle_{\vec{E}_0}}{V} = \chi\vec{E} \quad (5)$$

involving the Maxwell field $\vec{E}$ and the susceptibility χ. The Maxwell field is the sum of the external field $\vec{E}_0$ and the average net field exerted by the molecules of the sample. The difference between both fields, external and Maxwell fields, may be quite important. Only in the case of so-called conducting boundary conditions[29,30] they are equal. For this special case the combination of eqn (3) and (5) leads to

$$\varepsilon - 1 = 4\pi\chi = \frac{4\pi}{3Vk_B T}\langle\vec{M}^2\rangle_{eq} \quad (6)$$

Here, we have expressed the susceptibility in terms of the experimentally accessible dielectric permittivity ε. Furthermore, T and V are the temperature and volume of the sample.

Strictly speaking, this relation is restricted to samples composed of neutral molecules. In the case of molecular ions, one has to add the so-called dielectric conductivity[31,32] which can be shown to be of little influence in the present case (data not shown).

B Dipolar decomposition

While dielectric experiments measure the total mean square dipole moment $\langle \vec{M}^2 \rangle_{eq}$ the detailed information provided by computer simulations offers a decomposition of the total dipole moment $\vec{M} = \sum_k \vec{M}_k$. As both, the dielectric permittivity as well as the mesoscopic interaction energy involve the square of the dipole moment $\vec{M}^2$, a linear decomposition automatically creates self-terms, $\vec{M}_k^2$, and cross terms $\vec{M}_k \cdot \vec{M}_l$.

$$\langle \vec{M}^2 \rangle = \sum_k \langle \vec{M}_k^2 \rangle + 2 \sum_{k \neq l} \langle \vec{M}_k \cdot \vec{M}_l \rangle \tag{7}$$

In the result section we will use three different types of decompositions: (1) species-specific ($\vec{\mu}_{5ZNF}$ for the protein 5ZNF, $\vec{M}_{\{0\}}$ for H_2O, $\vec{M}_{\{+\}}$ for $EMIM^+$, $\vec{M}_{\{-\}}$ for $CF_3SO_3^-$), (2) shell-specific ($\vec{M}_{\{1\}}$ for the first shell, $\vec{M}_{\{2\}}$ for the second, and $\vec{M}_{\{b\}}$ for the bulk), and (3) per residue ($\vec{M}_{\{aa\},k}$ for amino acid k).

$$\vec{M} = \vec{\mu}_{5ZNF} + \vec{M}_{\{0\}} + \vec{M}_{\{-\}} + \vec{M}_{\{+\}} \tag{8}$$

$$\vec{M} = \vec{\mu}_{5ZNF} + \vec{M}_{\{1\}} + \vec{M}_{\{2\}} + \vec{M}_{\{b\}} \tag{9}$$

$$\vec{\mu}_{5ZNF} = \sum_k \vec{\mu}_{\{aa\},k} \tag{10}$$

C The Poisson–Boltzmann equation

When we presented the relations for the dielectric properties of a solution we started from the electrostatic potential $\phi(\vec{r})$. This quantity is also important for ionic solutions, *i.e.* systems where cations and anions are dissolved in a solvent, usually water. Under the assumption that the solvent can be simplified to a dielectric continuum the electrostatic potential acting in such solutions can be computed from the Poisson–Boltzmann equation:

$$\nabla \cdot \varepsilon \nabla \phi(\vec{r}) - \kappa^2 \phi(\vec{r}) = 0 \tag{11}$$

Here, ε and κ are material constants. ε is the dielectric constant of the solvent, for pure water $\varepsilon = 78$. The inverse of κ is called Debye screening length and measures the thickness of the solvent layer around the ions. It is related to the ionic strength I

$$I = \frac{1}{2} \sum_i z_i^2 c_i \tag{12}$$

by the relation

$$\kappa = \frac{2Ie^2}{\varepsilon k_B T} \tag{13}$$

with e being the electronic charge and T the temperature. Since in our systems the concentrations, c, of cations and anions are almost equal and the charge number $z_i = \pm 1$, the ionic strength I is identical to the concentration c_{IL}. Eqn (11) may be generalised to heterogeneous systems by introducing spatially varying material constants $\varepsilon(\vec{r})$ and $\kappa(\vec{r})$.

$$\nabla \cdot \varepsilon(\vec{r}) \nabla \phi(\vec{r}) - \kappa(\vec{r})^2 \phi(\vec{r}) + 4\pi\rho(\vec{r}) = 0 \tag{14}$$

In this form it is frequently used for the qualitative description of biomolecular solvation. Practically, the biomolecule is represented by a cavity of $\varepsilon = 1$ and $\kappa = 0$, while the ionic solution outside is represented by its material constants. Since the ion distribution is implicitly included *via* κ in the Poisson–Boltzmann equation, one only has to add the explicit charge distribution $\rho(\vec{r})$ of the biomolecule. While the simple Poisson–Boltzmann equation (eqn (11)) can be solved analytically its extended version (eqn (14)) must be solved numerically using techniques from finite element methods.[33]

D Residence times

As solvation shells are not static but evolve in time we have introduced a binary residence function for a single solvent particle i

$$n_i(t) = \begin{cases} 1: & i \in \text{shell} \\ 0: & i \notin \text{shell} \end{cases} \tag{15}$$

depending on whether particle i is a member of a solvation shell or not at time t. We emphasise that the molecular identity is conserved, *i.e.* recurrence of a particle is not recognised as the entrance of a new particle into the solvation shell. $\{n_i(t)\}$ is a time series, simply a sequence of 1 s or 0 s. In order to get a concise measure of such a time series one traditionally uses time correlation functions

$$C_n(t) = \sum_{i=1}^{N} \langle n_i(0) n_i(t) \rangle \tag{16}$$

One term in the above sum represents the memory of a specific solvent particle to be still a member of the solvent shell after some time interval t. The initial value $\langle n_i(0) n_i(0) \rangle$ represents the probability of a particle i to be a member of the first shell. The sum over all particles N gives the coordination number $C_n(t = 0) = CN$. Particles which are not members of the first shell are automatically ruled out by a zero residence function $n_i(t) = 0$. In other words, $C_n(t)$ describes the memory of the average particle.

As opposed to many correlation functions which decay to zero in the long-time limit $C_n(t)$ approaches a steady state

$$\lim_{t \to \infty} C_n(t) = \frac{CN^2}{N} \tag{17}$$

The characteristic time to reach this steady state may be interpreted as a mean residence time (MRT). The easiest way to get the MRT is to fit $C_n(t)$ to the analytical expression[34]

$$C_n(t) = \frac{CN^2}{N} + \frac{(N - CN)CN}{N} e^{(-\frac{t}{\tau})^\beta} \tag{18}$$

Here, N denotes the number of solvent molecules. Those molecules with a long residence time, *i.e.* "bound" molecules, have to be subtracted from N. This type of Kohlrausch–William–Watts functions (KWW) is an extension of a mono-exponential relaxation to a more complex shell dynamics. This transition is modelled by a single parameter β. The smaller the β the larger the deviation from monoexponential behaviour, or equivalently, the larger the spread of relaxation times. The average over this distribution of these relaxation times is given by the analytic expression

$$\langle \tau \rangle = \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right) \tag{19}$$

where Γ denotes the Gamma function representing the generalisation of the factorial to real numbers.

III. Methods

A System setup procedures

Five molecular dynamics simulations of 5ZNF in hydrated MIL, $EMIM^+$ $CF_3SO_3^-$, with varying water mole fractions x_{H_2O} between 1.0 and 0.675 were performed with the software package CHARMM (ref. 35). To set up the simulation system the coordinates from the 5ZNF PDB-entry were used for the protein. For the solvent, first, $EMIM^+$ and $CF_3SO_3^-$ molecules were placed randomly around the protein like in a Monte-Carlo simulation where each addition of a molecule is a move which, depending on the energy, can be accepted or not. Then equilibrated water boxes with TIP3P[36] as the water model were overlaid and overlapping water molecules removed. Finally, a box with the geometry of a truncated cuboctahedron was cut out which is more spherical than a cube and has exactly one half of the volume of a circumscribed cube. Subsequently, the energy of this MD-system was minimised and an equilibration run in the constant pressure, constant temperature (T = 300 K) (CPT) ensemble was started with a time step of Δt = 2 fs. This value of the time step was used for all further simulations. After one nanosecond the average edge length of the truncated cuboctahedron was fixed (Table 1) and the equilibration was continued in the constant volume, constant temperature (NVT) ensemble for another nanosecond. The simulation was continued in the NVT ensemble for up to 200 nanoseconds for the five systems displayed in Table 1 with a water mole fraction $1.000 < x_{H_2O} < 0.675$. Additionally, three simulations of roughly 65 nanoseconds were made for x_{H_2O} = 0.921, 0.886 and 0.833 to confirm the systematic trend of several properties analysed in the result section.

The forces acting between the components of the system were computed in the following way: the force field parameters of the protein were taken from CHARMM.[37] For the solvent molecules, the parameters were identical to those already used in ref. 38. The mutual interaction between the protein and the solvent was modelled by electrostatic interactions and Lennard-Jones terms. In the latter case mixing rules between the solute and solvent force fields were used. The procedure to handle the long range electrostatic forces was along the lines described in the Method section of ref. 38 with a κ of 0.41 $Å^{-1}$ and a real space cut-off of 10 Å.

All bond lengths were kept fixed by the shake algorithm,[39] whereas bond angles and torsions are left flexible.

B Spatial decomposition and definition of solvation shells

Once a solute is immersed in a solvent it structures its neighbourhood. This automatically poses the question how to classify the degree of neighbourhood. Intuitively, one thinks of solvation shells. For a small solute composed of a few atoms the concept of concentric spherical shells is appropriate. This goes along with the traditional concept of radial distribution functions (RDF).

For a large, anisotropic solute, however, this concept has to be extended as visible in Fig. 2 of ref. 40. There, we have shown that even for a solvent composed of a single species the structure of the RDF originates from the overlap of at least two Voronoi shells. This superposition stems from the anisotropy of the solute which mixes different radial shells. Therefore, alternative routes to decompose the space around a large anisotropic solute have to be followed.

In addition, the heterogeneity of the solvent consisting of rather different molecular species has to be accounted for. Therefore, a distance based shell concept would require a distance parameter for each solvent species. While this may be manageable for the first shell around the large anisotropic solute, the classification of second and further shells becomes extremely difficult in the case of a heterogeneous solvent since this would necessitate shell specific mutual distance parameters. However, the introduction of this plethora of parameters can be avoided because the method of the Voronoi tessellation offers a parameter free decomposition.

Given a set of points representing atomic coordinates, a Voronoi decomposition creates an ensemble of space-filling disjunct polyhedra, each containing all space closer to its associated point than to any other point of the given set.

The faces of each of these Voronoi polyhedra are constructed by planes perpendicular to the vectors between the associated point and its neighbour points.

As the direct construction of Voronoi polyhedra is computationally demanding one exploits the duality between Voronoi decomposition and Delaunay tessellation. We have implemented a very efficient Delaunay algorithm the details of which are described in ref. 41. The complete tessellation of one coordinate frame of our systems and the assignment of solvation shells take roughly one second on an Intel DualCore 3.00 GHz CPU.

As a final result there is a one-to-one correspondence between each atom and its Voronoi polyhedron. The volume of a molecule is obtained by summing up all Voronoi polyhedra of the constituting atoms. This method can be used to compute the fluctuations of the molecular volume. In the context of molecular ionic liquids this was already used in ref. 42 to compute the molecular volume of the small solute CO_2.

Of course, the sum of all Voronoi polyhedra gives the volume of the whole sample.

With this result at hand the definition of neighbourhood and solvation shells is straightforward. If the polyhedron of a solute atom shares a common face with the polyhedron of a solvent atom, the respective solvent molecule is called a first neighbour. The set of all these first neighbours constitutes the first solvation shell. Solvent molecules outside the first shell but sharing a polyhedral face with members of the first shell belong to the second shell and so on. In this way the solvent is organised in successive layers. In order to get a visual impression we present Fig. 1 which shows the union of the atomic Voronoi polyhedra of the solute, the protein 5ZNF. Each face of a Voronoi polyhedron can be classified as an inner or an outer face depending on whether it has already contact to another polyhedron of the same solute or not. The union of all outer faces constitutes the Voronoi surface of the solute which is offered to the polyhedra of the solvent molecules. Those which have contact create the first solvation shell. As a symbolic example we have also given in Fig. 1 single water,

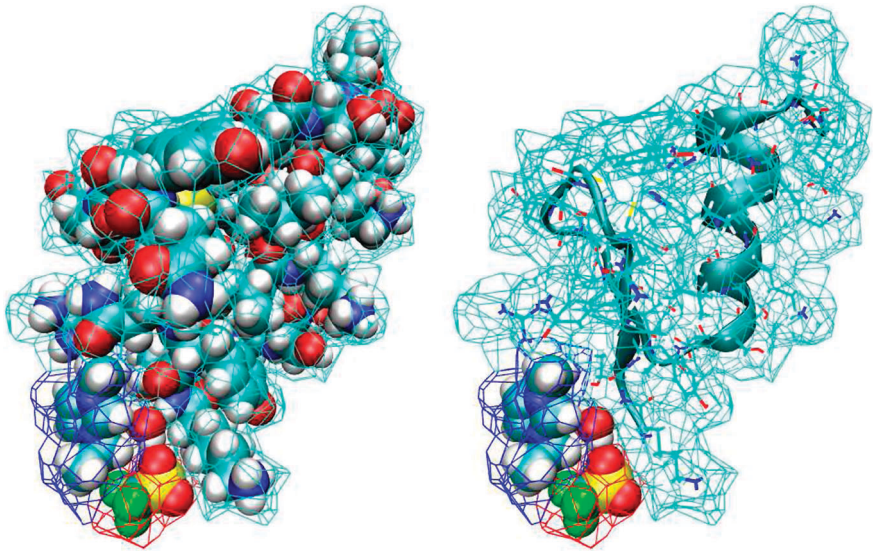


Fig. 1 Union of the atomic Voronoi polyhedra of 5ZNF and 3 solvent molecules ($EMIM^+$, $CF_3SO_3^-$ and H_2O) in the first shell. (a) All atoms displayed by their vdW-spheres. (b) 5ZNF displayed as structure with ribbon diagram.

$EMIM^+$ and $CF_3SO_3^-$ molecules with their respective polyhedra as members of the first shell.

C Occupancy

Mole fraction, concentration and ionic strength are traditional measures to specify the composition of a solution. However, they tell nothing about the space covered by a species. This occupancy can be easily derived from a Voronoi decomposition. The sum of all polyhedral volumes of all atoms of a species is the space occupied by that species. In Fig. 2 we depict the relative occupancy of the four species, protein, water, $EMIM^+$ and $CF_3SO_3^-$ for each simulated system.

One can see that at a mole fraction of $x_{H_2O} = 0.927$ water covers slightly more than only one half of the entire volume. Alternatively, at $x_{H_2O} = 0.675$ $EMIM^+$ already covers more than one half of the volume. This behaviour comes from the difference in molecular Voronoi volumes which are—on average—30 $Å^3$ for water, 200 $Å^3$ for $EMIM^+$ and 100 $Å^3$ for $CF_3SO_3^-$.

IV. Results and discussion

Here, we will present and discuss results for the protein itself and its immediate neighbourhood, the so-called first solvation shell. This analysis will be completed by a mesoscopic view based on the whole system's static dielectric properties. All properties were computed by averaging over the complete lengths of the trajectories with a graining appropriate to the statistical variation of the respective property.

A The protein itself

1 Protein volume and surface. As a starting point we have calculated the protein's Voronoi volume and surface. For each mole fraction of water or equivalently, each ionic strength, we found a Gaussian distribution, slightly skewed to high values. The mean values of these distributions are given in Table 2. Only a slight variation of these mean values as a function of ionic strength is observed. Yet there is an indication that the behaviour at $x_{H_2O} = 0.927$ is somewhat exceptional. In order to clear up this point we have also computed the mean extension of the protein along its three principal axes of inertia. The three average values $a = \langle x_{max} - x_{min} \rangle/2$, $b = \langle y_{max} - y_{min} \rangle/2$, $c = \langle z_{max} - z_{min} \rangle/2$ are given in Table 2, too. Considering a, b and c as the axes of a general ellipsoid it is possible to compute alternative values for the volume and surface using the formulas $V = 4\pi/3 \cdot abc$ and $S = ((ab)^{1.6} + (ac)^{1.6} + (bc)^{1.6})^{1/1.6}/3$. We have used the formulas for a general ellipsoid, although the close proximity of b and c as compared to a favours the picture of a prolate ellipsoid. Here, the exceptionality at $x_{H_2O} = 0.927$ becomes

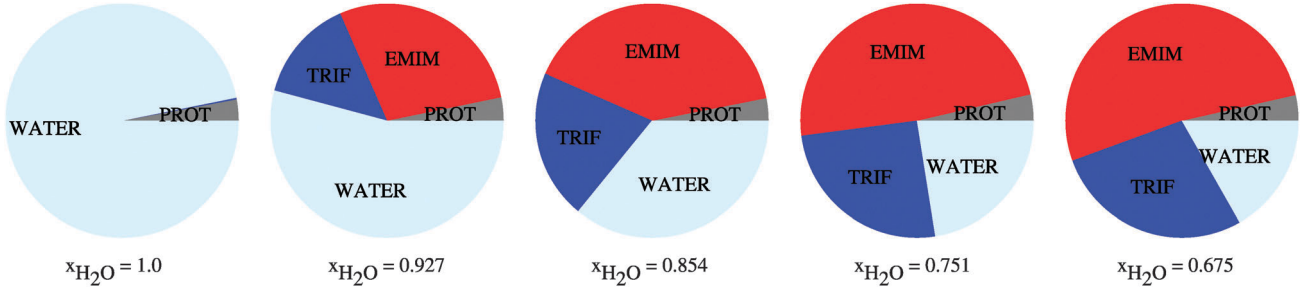


Fig. 2 Piecharts illustrating the volume–volume percentage, the occupancy, of 5ZNF, $EMIM^+$, $CF_3SO_3^-$ and H_2O for each system.

Table 2 Volumes and surfaces of the protein in the simulated systems calculated by means of Voronoi decomposition of the system and ellipsoid approximation of the protein

x_{H_2O}	c_{IL}/mol l^{-1}	$V_{Voronoi}$/Å^3	$S_{Voronoi}$/Å^2	a/Å	b/Å	c/Å	V/Å^3	S/Å^2
1.0	0	4574	2928	18.4	11.5	9.1	8066	2071
0.927	2.36	4570	2942	17.5	12.35	9.0	8148	2073
0.854	3.38	4602	2953	17.3	11.75	9.3	7919	2019
0.751	4.09	4609	2918	18.35	11.75	9.0	8128	2085
0.675	4.41	4621	2934	17.35	11.3	8.9	7309	1928

obvious for the first time as a maximum of these elliptic volumes. (The corresponding surfaces, however, show no clear trend.) Upon closer inspection the maximum of the elliptic volume comes from the increase of b, *i.e.* the ellipsoid is slightly shifted towards an oblate shape. For higher ionic strengths the protein again seems to relax to a prolate shape. The elliptic volumes almost double the Voronoi volumes. This might be explained by the fact that the protein is not a strict convex body but is characterised by several concave regions which contribute a lot to the surface area but not to the volume.

2 Structural behaviour of the backbone. The crystal structure of the protein as given in the Protein Data Bank entry "5ZNF" shows two principal secondary elements, an anti-parallel β-sheet around amino acids 3, 4, 11, and 12 and an α-helix extending from amino acid 15 to 25. We have analysed the behaviour of these structural elements in a three-fold manner: by the DSSP and the STRIDE method[43,44] as well as by a self-designed method based on contact matrices derived from Voronoi tessellation.

DSSP. Discarding the initial 5 ns, a time series of snapshots taken at every 100 ps was analysed with the DSSP program. The resulting evaluation was compared to the standard secondary structure given in the PDB. The respective amino acids 3, 4, 11 and 12 constituting the β-sheet and residues 15 to 25 building up the α-helix were assigned a value of 1 or 0 depending on whether the DSSP evaluation agrees with the PDB standard or not. This secondary structure indicator was subsequently averaged over all frames considered as well as over all residues within a secondary structure element, β-sheet or α-helix. The dependence of the resulting indicators on the ionic strength is displayed in Fig. 3. Now the exceptionality at the point $x_{H_2O} = 0.927$ becomes pretty obvious as a minimum of the secondary structure indicator.

STRIDE. Applying an analogous procedure of the secondary structure indicator assignment to the evaluation results of the STRIDE method the respective curves given also in Fig. 3 were obtained. The pronounced minimum is again visible but the additional criteria and features typical for STRIDE somewhat flatten the curve.

Contact matrix. Alternatively to the standard methods DSSP and STRIDE we have used a simple criterion to define turns and bridges on the basis of Voronoi tessellation: if two amino acids i and $i + 4$ have a common Voronoi face they are termed a 1–4-turn. More that two subsequent 1–4-turns define an α-helix. An analogous criterion based on Voronoi contacts

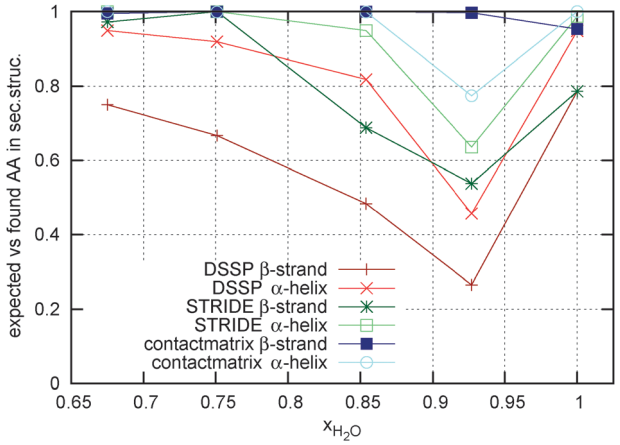


Fig. 3 The secondary structure indicators from the results of DSSP, STRIDE and contact matrix are given for the two secondary structures of the protein 5ZNF.

is used to define a β-sheet. The resulting curves are shown in Fig. 3, too. It turns out that this mild criterion produces a minimum for the α-helix. The respective curve for the β-sheet is almost constant with a slight decline at zero ionic strength.

3 Structural behaviour of the side chains. With the secondary structure indicator we have characterised the backbone in a global way. As a global indicator of the side chains we use the protein's net dipole moment μ_{5ZNF} (Fig. 4). One might argue that the α-helical part of the backbone contributes to μ_{5ZNF} as well. This is certainly true but we have found that in our case its contribution is around 20%. In other words, μ_{5ZNF} is essentially determined by the dipole moments of the side chains. In order to ascertain that the "fraying ends" (lys1, lys28, glu29, lys30) do not play a major role we have computed their contribution separately. Of course, they make a contribution at the quantitative level but do not affect the qualitative interpretation.

Again we find a maximum at $x_{H_2O} = 0.927$ in accordance with the minima of the secondary structure indicator of the backbone. Therefore the whole structure of the protein, backbone and side chains, behaves in a synchronous manner. In order to elucidate the maximum we have decomposed μ_{5ZNF} into contributions from the individual amino acids. Thereby, all charge positions ($\vec{r}_{\alpha,i}$) were referred to the center of mass of the protein ($\vec{r}_{COM,5ZNF}$). This ensures the additivity of all dipole moments referring to residues or any other arbitrary division into atom groups. It is interesting to analyse lengths of the residual dipole moments $\vec{\mu}_{\{aa\},i}$ and the angles they make with $\vec{\mu}_{5ZNF}$ separately. While the lengths of the residual

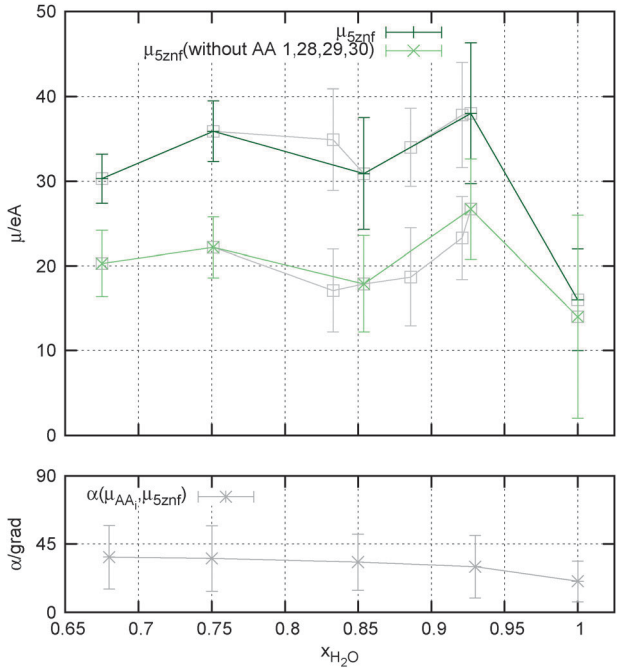


Fig. 4 Upper part: the dipole moment of the protein $|\vec{\mu}_{5ZNF}|$ is given, from the entire protein (dark-green), and the protein without "fraying ends"—amino acids (lys1, lys28, glu29, lys30) (light-green). The respective values from the additionally simulated systems are given in grey. Lower part: $\Delta\alpha$.

Table 3 The absolute deviation from 90° of the angle between the dipole vector for an amino acid i, $\vec{\mu}_{\{aa\},i}$, and the dipole vector for the entire protein $\vec{\mu}_{5ZNF}$ was averaged over all timesteps and all amino acids ($\Delta\alpha$)

x_{H_2O}	$\Delta\alpha(\vec{\mu}_{5ZNF},\vec{\mu}_{\{aa\},i})/^\circ$
1	20.4
0.927	30.1
0.854	33.1
0.751	35.5
0.675	36.4

moments do not vary significantly with ionic strength, the angles are strongly affected. In order to have a globular measure of angular distribution we have computed the average absolute deviation of all angles from 90°, *i.e.* from random orientation (Table 3). As can be seen from the table this value increases monotonically with ionic strength. In other words, residual dipoles orient more and more in the direction of the net dipole moment $\vec{\mu}_{5ZNF}$. By the way, $\vec{\mu}_{5ZNF}$ approximately coincides with the *b*-axis of the inertia ellipsoid. Again we have paid attention to the "fraying ends" lys1, lys28, glu29, and lys30 as well as to the zn31 center. They give quantitative contribution but they do not change the qualitative interpretation.

4 Energetic analysis. So far we have found that at the mole fraction $x_{H_2O} = 0.927$ the volume of the inertia ellipsoid has a maximum, the secondary structure indicator of the backbone displays a minimum—for all three methods: DSSP, STRIDE and contact matrices—and the net dipole moment of the protein shows a maximum again. In order to search for a common root of this synchronous behaviour of a plethora of properties we have performed an energy analysis. Starting from the total energy of the protein, *i.e.* neglecting any interaction with the solvent, we first observed that the so-called bonded interactions (bond lengths, bond angles, dihedral angles, *etc.*) show almost no variation with ionic strength and thus can be ignored for a qualitative interpretation. Thus the energetic analysis reduces to the interplay of van der Waals (E_{vdW}) and electrostatic interactions (E_{elec}) (Fig. 5). In fact, both are a monotonic function of ionic strength. The first one rises, whereas the latter decays monotonically. As the slope is different the sum of both exhibits a maximum, right at the point $x_{H_2O} = 0.927$ where we have shown that all properties discussed so far exhibit an exceptional behaviour. In order to secure that this magic point $x_{H_2O} = 0.927$ is not an artifact we have also simulated systems with nearby mole fractions $x_{H_2O} = 0.921$, $x_{H_2O} = 0.886$ and $x_{H_2O} = 0.833$ but of shorter trajectory length. The results are in close correspondence to that of the magic point.

On the level of interpretation the interplay of van der Waals and electrostatic energy may be viewed from the contributions of the individual amino acids. The change from solvation in pure water to solvation in hydrated ionic liquids corresponds

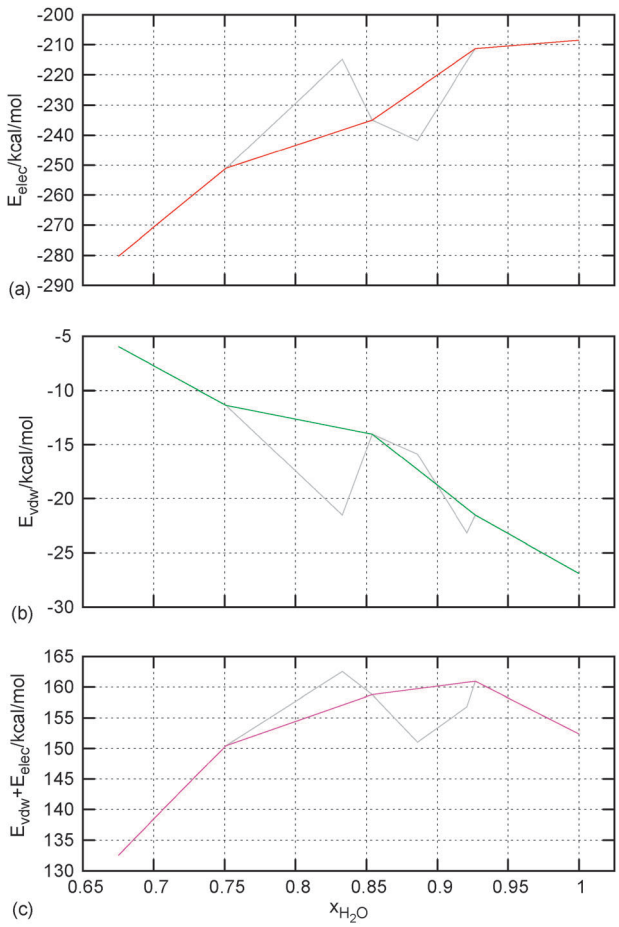


Fig. 5 The total intramolecular non-bonded interaction energies (E_{elec} and E_{vdW}) only between the atoms of 5ZNF are given for the simulated systems, (a), (b) separately and (c) as a sum. The respective values for the additionally simulated systems are given in grey.

to a transition from dipolar screening to charge screening of electrostatic forces. Dipolar screening means that the small water molecules penetrating the space between amino acids screen the electrostatic interactions and reduce their mutual attraction or repulsion. In contrast, their size prevents molecular ions from migration between residues. Therefore they propagate or transmit their electrostatic potential in order to screen the solute's charges. Once this charged screening has been established the approaching amino acids are no longer hindered by interstitial water molecules and can attract each other without any steric hindrance by the solvent. However, the enhanced electrostatic attraction of residues is accompanied by an increase of their van der Waals interactions.

This also explains the enhanced dipolar correlations: as the interaction of a pair of dipoles is proportional to their negative scalar product divided by the cube of the mutual distance, $-\vec{\mu}_i\vec{\mu}_j/r_{ij}^3$, residue dipoles pointing in the same direction as the total dipole moment $\vec{\mu}_{5ZNF}$ are energetically favoured. Visualising the residue dipoles as pick-up sticks as shown in Fig. 6 one may speak of a "Mikado"-effect: with increasing water content the intruding water molecules open the cone of pick-up sticks. The mole fraction $x_{H_2O} = 0.927$ seems to be the transition state where dipolar screening changes to charge screening. At this point dipolar screening is considerably weakened but charge screening isn't fully active already. This generates a non-optimal situation for the solvation of the zinc finger protein since regions with dipolar screening are intercalated with regions of charge screening. Beyond the transition point charge screening dominates. This can be also seen from the decreasing influence of the van der Waals energy and the dominance of the electrostatic energy. In this way the transition point corresponds to the highest total energy which explains the lowest level of ordering of the zinc finger protein in terms of secondary structure elements.

B The protein's immediate neighbourhood

1 3D Solvent densities. In order to get an intuitive expression of the distribution of cations and anions as well as water molecules around the protein we start with Fig. 6–9 which refer to the highest and lowest non-zero ionic strengths. In the middle of these figures we give the 3-dimensional histogram of the distribution of solvent molecule centers around the protein where we used the center of mass as a point representation. These histograms were constructed by subdividing the whole space into little cubes (bins) of edge length 0.5 Å. If the center of a solvent molecule falls within such a cube the bin entry is incremented by one. Having run through all frames the histogram is normalised to the average density. All entries

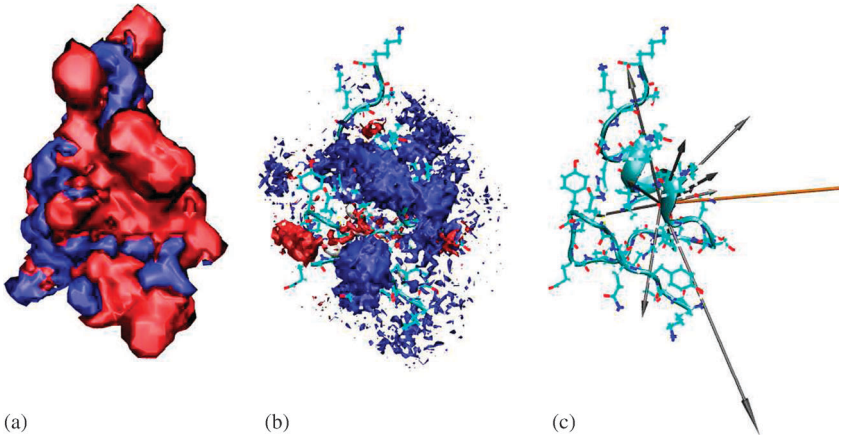


Fig. 6 Front view for $x_{H_2O} = 0.927$. (a) The blue isosurface represents -0.5 kcal mol^{-1}, the red one $+0.5$ kcal mol^{-1}. (b) Cations are represented by red, anions by blue and water molecules by gray clouds. (c) Structure of 5ZNF and secondary structure as a ribbon diagram, as well as the dipole vectors for charged amino acids (in grey) and for the total protein 5ZNF (in orange). The 3D density in figure (b) is sampled from the entire trajectory, figures (a) and (c) represent snapshots of the protein.

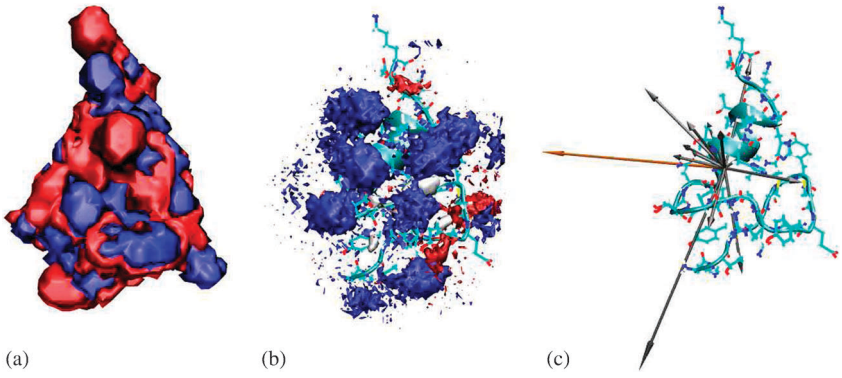


Fig. 7 Back view for $x_{H_2O} = 0.927$. Parameters and keys are the same as in Fig. 6.

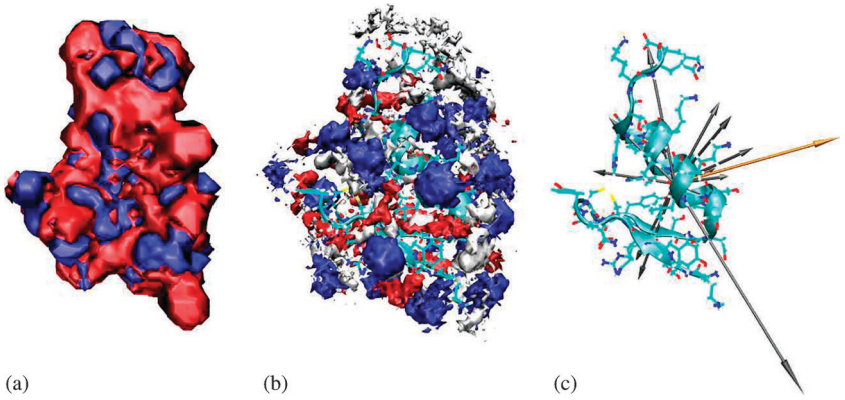


Fig. 8 Front view for $x_{H_2O} = 0.675$. Parameters and keys are the same as in Fig. 6.

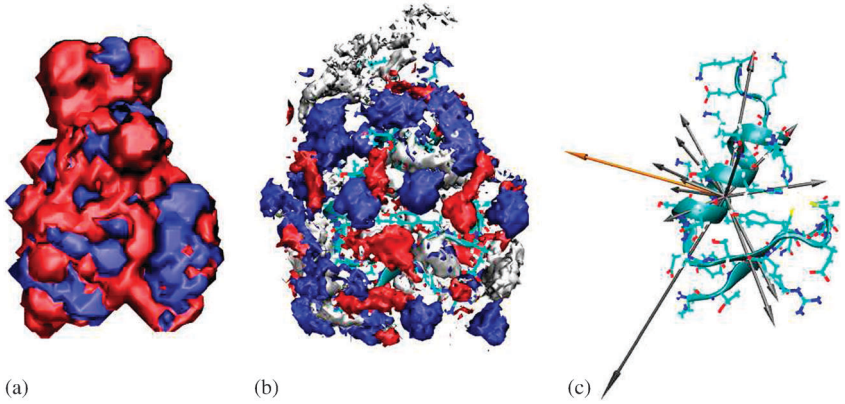


Fig. 9 Back view for $x_{H_2O} = 0.675$. Parameters and keys are the same as in Fig. 6.

below 2.5 times the average density were subsequently eliminated. The so-obtained importance histogram was displayed as a set of clouds, both as a front and a back view. For a better understanding of these cloud histograms we have also computed the contour isosurfaces of the electrostatic potential as a solution of the Poisson–Boltzmann equation.[33] Thereby, the actual ionic strength as well as the simulated dielectric constant to be discussed in the next section were used as input parameters.

For low ionic strengths the histogram of ion centers follows—at least qualitatively—the solution of the Poisson–Boltzmann equation. In other words, anions cluster preferentially in regions of positive electrostatic potential. The preference of the cations for the negative regions is not so pronounced. As we shall see later in the section discussing mean residence times this has to do with the higher mobility of the cations. The only region with a pronounced density of cations is a part of the protein surface with no preferred sign of the electrostatic potential because of the absence of charged amino acids. This region of polar amino acids is also concave from a steric point of view. So it might be called "fossa neutralis". The meaning of this special region will become clearer on a more quantitative level when discussing the residue-resolved contact matrices in the next section. For high ionic strengths the distribution of ions departs from the solution of the Poisson–Boltzmann equation: anions and cations organize in an alternating patchwork of charge clouds. The regions of ion densities with uniform sign are broken up into islands of alternating sign. This might be explained by the mutual repulsion of like charges which becomes more and more important when the ionic strength and thus the number of charge carriers increases. In addition the increase of ionic strengths goes along with the reduction of the amount of water and a subsequent loss of dielectric screening. One might get the impression that water molecules reside preferentially at the C-terminus. One has to bear in mind, however, that the heavy ions—in contrast to the mobile water molecules—cannot really follow the motions of the highly flexible C-terminal end ("fraying ends"). Therefore, they are ruled out by a density threshold of 2.5. Furthermore, the small water molecules can fill out gaps as already discussed in the energetic analysis above.

2 Coordination number of the protein. Within the Voronoi method coordination numbers can be easily computed: every solvent molecule which has a common Voronoi face with the protein is a first neighbour or belongs to the first solvation shell. Table 4 shows the coordination numbers for all three solvent species at all ionic strengths studied. What is striking at first sight is the higher coordination number of cations as compared to that of anions despite the fact that the protein is positively charged ($+4e$). This excess of cations becomes even clearer when normalising the coordination numbers to the number of solvent molecules of that species. As Table 4 shows,

Table 4 Absolute (CN_{Species}) and relative coordination numbers $\left(\frac{CN_{\mathrm{Species}}}{N_{\mathrm{Species}}}\right)$ for 5ZNF and the solvent molecules

x_{H_2O}	c_{IL}/mol l^{-1}	$CN_{\{0\}}$	$CN_{\{+\}}$	$CN_{\{-\}}$	$\frac{CN_{\{0\}}}{N_{\{0\}}}$	$\frac{CN_{\{+\}}}{N_{\{+\}}}$	$\frac{CN_{\{-\}}}{N_{\{-\}}}$
1	0	319.6		1.41	0.06		0
0.927	2.36	196.28	25.05	23.65	0.08	0.13	0.12
0.854	3.38	138.78	38.88	32.5	0.08	0.13	0.11
0.751	4.09	89.69	48.59	39.58	0.09	0.16	0.13
0.675	4.41	64.86	51.39	40.72	0.09	0.15	0.12

this normalised coordination numbers are constant for water and anions, but increase with ionic strength for cations. In some sense one might argue that there is a tendency to exclude cations from the water–anion system thus forcing some cations to approach the protein surface. The alternative view is that the protein extracts the cations from the solvent. However, an intensive search based on atom–atom pair correlation functions favours the former view (results too bulky to be shown).

In addition we have individually spotted excess cations on the protein surface. We have found hot spots in the negatively charged region of the β-sheet (glu9, cys5 and cys8) as well as in rather apolar regions at amino acids leu18 and ile22. This spotting is in accordance with suggestions from the Poisson–Boltzmann method and will be supported by the residue-resolved contact matrices below.

The 3D density plots also suggest an excess of anions in the first solvation shell which seems to be contrary to the greater coordination number of cations. This can be resolved by the fact that cations form regions of lower but contiguous density and thus quite often fall below the threshold. In contrast, anions are found preferably at high density but with distinct large gaps between hot spots.

3 Residue-resolved contact matrices. As a quantitative counterpart to the intuitive 3D histograms we have computed residue-resolved contact matrices between the protein and the solvent species cation, anion and water. Each row of a contact matrix corresponds to a specific amino acid. Each entry in a row stands for the probability that a certain solvent molecule has a Voronoi contact with the protein. A Voronoi contact means that at least one Voronoi polyhedron of a protein atom has a common face with the Voronoi polyhedron of a solvent molecule. As a typical example a contact matrix of the protein with 50 $EMIM^+$ molecules is shown in Fig. 10. The sum of each row represents the probability that the solvent species as such has contact with the respective amino acid. Such a contact vector is also shown in the figure. If there were no multiple contacts between a solvent molecule and several amino acids the sum over the contact vector would give the coordination number. This additivity could be enforced by introducing a criterion for unique assignment of every protein–solvent contact to a single amino acid. With the concept of a contact vector the inherent 3-dimensionality of neighbourhood may be transformed to a 1-dimensional measure of neighbourhood statistics. Subtracting the mean value from all entries of a contact vector one gets a quantitative function of the enhanced presence or absence of a solvent molecule in contact with an amino acid.

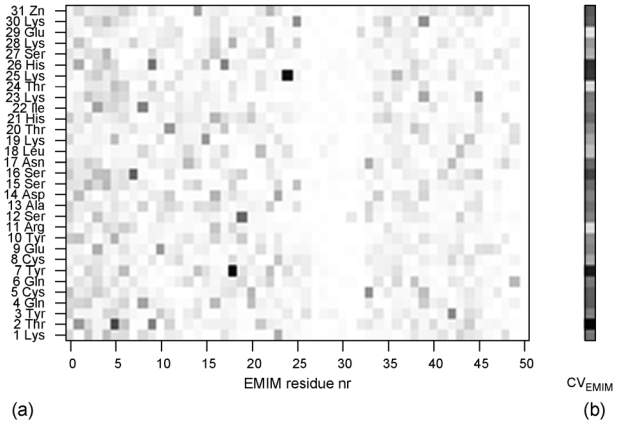


Fig. 10 (a) The entries of the contact matrix lie within the interval [0,1] and are mapped onto a greyscale. They are given for the amino acids of the protein and 50 $EMIM^+$ molecules (residues). (b) The entries of each row of the contact matrix are summed, mapped onto a greyscale and given as a column, a contact vector. The intensity of each component displays the frequency of Voronoi contacts between an amino acid-residue and a solvent-species.

As a first application of the concept of the contact vector Fig. 11 compares the mean value corrected contact vectors for the highest and lowest non-zero ionic strengths. Both, anions (Fig. 11(a)) and cations (Fig. 11(b)) show a quite different variation as a function of the amino acids. While at low ionic strengths the contact vector components change gradually, rather sudden changes are observed at high ionic strengths. This correlates with the observations found in the 3-dimensional ion densities: regions preferably or exclusively occupied by charges of the same sign are shrinking in favour of a patchwork of small charge islands. A striking feature hardly discernible in the 3D density shows up in the contact vector component analysis: at the positively charged or polar residues 17 and 21 as well as 24 and 28 a characteristically enhanced probability of anions is found. The relative distance in sequence of 4 is typical for the amino acid contact pattern in the α-helix. The cations show an analogous pattern at the rather apolar residues 18, 22 and 26. In other words, the anionic and cationic patterns are shifted by one residue. This systematic intercalation of cations and anions maybe induced by the protein but cannot be solely due to the protein. Rather the ionic network at the surface of the protein organises in a specific way in order to optimize the balance between repulsive and attractive electrostatic interactions between ions. Consequently, the charge network organises in small regions of alternating sign because a large region covered by ions of the same charge sign would result in a too high repulsive energy. Therefore, it is not surprising that all these findings are observed at high ionic strength. For lower ion concentrations the enhanced screening by water molecules blurs the distinct ion pattern.

The concept of contact vectors may also give some hints where the excess cations are located on the protein surface: while the anions prefer special regions as already inferred from the 3D densities and clearly visible in the relative contact vectors given in Fig. 11, the cations tend to a more ubiquitous distribution despite some special spots like the amino acid

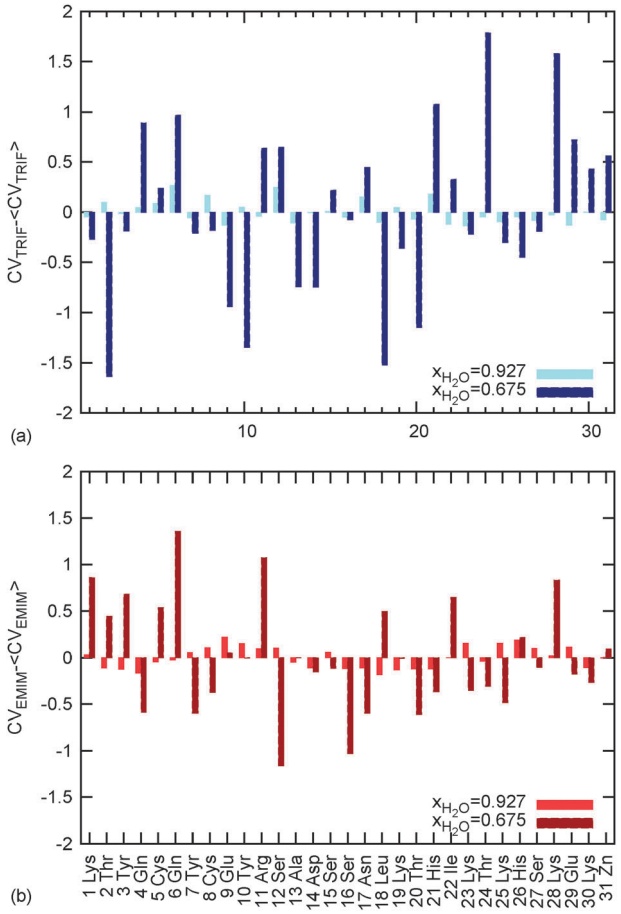


Fig. 11 The average of all components of the contact vector $CV_{Species}$ was subtracted from each component to display above and below average frequencies of Voronoi contacts, (a) for the anion $CF_3SO_3^-$ and (b) for the cation $EMIM^+$. The systems with lowest and highest ionic strengths are compared. Note the peaks for the anion at residues 17 and 21 as well as 24 and 28, and for the cation at 18, 22 and 26 at high ionic strength.

triad 18, 22 and 26. Even in this case the approach of cations towards rather apolar residues is not obvious at first sight. However, space on a protein surface is limited and the high number of cations (as compared to the anions) forces them to occupy unfavourable regions. A further reason might be that the charge density of a cation is rather flat: indeed, if summing within a cation or anion the absolute charges—neglecting the sign—one gets 2.141 for the cation and 3.74 for the anion. This shows the stronger variation of charge density within the anion and explains its tendency to approach charged and polar residues. Furthermore, the lower sum of absolute charges of the cation goes along with a volume of double size as compared to the anion.

In order to analyse the approach of cations to the protein surface in more detail we have decomposed the cation into three parts: head, ring and tail (methyl group, aromatic ring and ethyl group) and subsequently computed the residue-resolved contact vectors for all three parts. Taking the pairwise difference of contact vector components a positive overshoot of the ring and tail over the head is clearly visible. For the pair ring and tail the situation is not so simple: at low

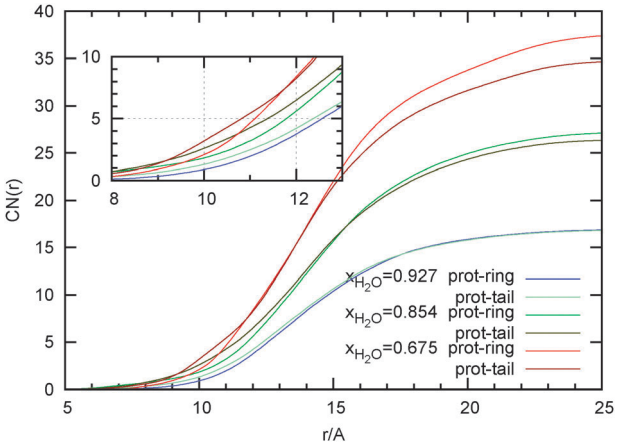


Fig. 12 The coordination number as a function of distance to the center of mass of the protein ($CN_{Species}(r)$). The ethyl-part and the ring-part of $EMIM^+$ were treated as individual species and are compared here.

ionic strengths ring and tail contacts are fairly equal. With increasing ionic strength, however, ring contacts become more important. The average value of the difference contact vector starting at almost zero increases with ion concentration. For a final evidence we have computed the running coordination number from the radial distribution functions between the protein and the ring/tail. Strictly speaking, we refer to the distance between respective center of mass, may be protein or head/ring. We have only considered the ring/tail centers which make a direct Voronoi contact with the protein. As can be seen from the insert in Fig. 12, coordination by the tail is slightly dominant at short distances to the center. However, it is finally overruled by ring coordination, as shown in the full figure. A possible explanation for this prevalence of ring coordination at higher ionic strengths and longer distances might be the specific interaction of charged ends of certain amino acids with the ring-charges. In this sense one would speak of a preferred ring-charge group interaction as opposed to the bare steric interaction of the tail.

4 Mean residence times. So far we have focused on the static, *i.e.* time averaged, view of the solvation structure. However, solvation shells are dynamic. They are subject to permanent restructuring as its members enter and leave their shell at a certain rate. In order to have a measure of this rate of migration we have computed the residence functions already outlined in the Theory section. A systematic discussion is facilitated considerably by fitting these residence functions to analytic expressions like the KWW functions (eqn (18)). The so-obtained pair of fit parameters τ and β can be converted to a mean residence time according to relation (19). The mean residence times for the three species, water, $EMIM^+$ and $CF_3SO_3^-$ as a function of mole fraction x_{H_2O} are shown in Table 5 and Fig. 13. One sees immediately that the anion MRTs are about twice as long as for cations. This goes along with the findings from the 3D densities where cations exhibit low but contiguous densities while anions populate selected sites with large gaps in between. These selected sites point to a strong interaction between a single anion and the respective region on the protein surface. The probability to find a second

Table 5 Mean residence times of each of the three solvent-molecule species within the first Voronoi shell of the protein

x_{H_2O}	$\langle\tau\rangle_{\{0\}}$/ps	$\langle\tau\rangle_{\{+\}}$/ps	$\langle\tau\rangle_{\{-\}}$/ps
1.0	49.45		
0.927	176	1067	2146
0.854	551	3090	4156
0.751	1872	13 903	28 909
0.675	4064	28 098	66 422

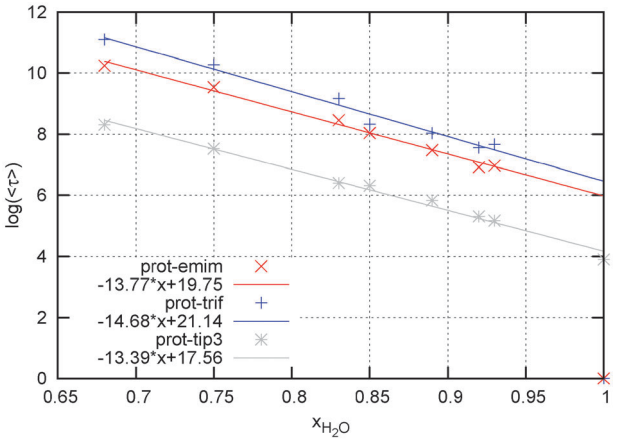


Fig. 13 Logarithmic plot of the mean residence times given in Table 5.

adjacent anion is rather low because the charge density on the anions is very high. In contrast, the hydrophobic parts, *e.g.* the ethyl chain, of the cations facilitate direct contacts between cations.[45] On the one hand this permits a contiguous distribution of cations on the protein surface. On the other hand the more diffusive charge distribution impedes strong interaction between the protein surface and a single cation. In fact, we have plotted the logarithm of $\langle\tau\rangle$ and obtained almost perfect linear relations. In other words the mean residence times increase exponentially with decreasing mole fraction. As the logarithmic plots for the three species are almost parallel the rate of increase is very similar among the species. This points to a collective solvent phenomenon which is most probably the viscosity. If one wants to discuss the slight differences in the slope one finds that the retardation of the mean residence times increases in the order water to $EMIM^+$ to $CF_3SO_3^-$. This follows the same order as the absolute values and shows that the larger cation resides for a shorter time than the smaller anion. This is in agreement with the general observation of the higher mobility of the cation as compared to the anion already observed for the pure ionic liquid without any protein.[20] This significant difference in mobility also persists at the surface of the protein.

Fig. 14, however, gives a more subtle explanation. Therein, the residue-resolved contact vector of the $EMIM^+$ molecules is confronted with the residue-specific mean residence times. One sees a clear anti-correlation. The contact vector measures the average number of solvent molecules which are simultaneous first neighbours of an amino acid. There seem to be two different patterns of residence: an amino acid may bind loosely to a couple of solvent molecules which reside for a short time, or the amino acid fixes a single molecule for a long time. The cation $EMIM^+$ seems to prefer the first pattern while the anion $CF_3SO_3^-$ the second one.

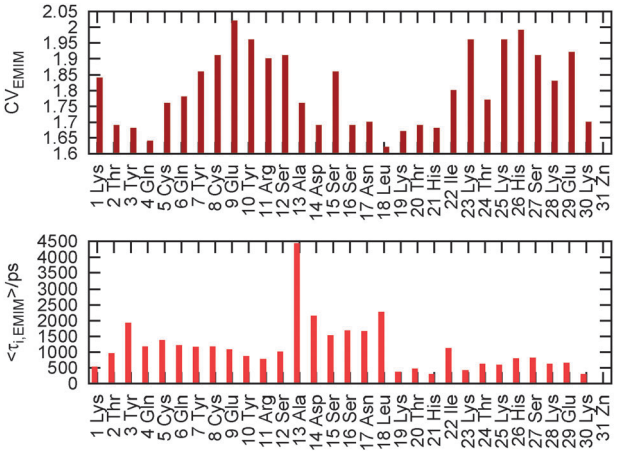


Fig. 14 Contact vector (CV_{EMIM}) and mean residence times ($\langle\tau_{EMIM}\rangle$) of $EMIM^+$ within the first Voronoi shell of each amino acid for x_{H_2O} = 0.927.

In order to get an impression of the actual numbers Table 5 lists MRTs in picoseconds for the three species as a function of the mole fraction. The spread is enormous. From the magic point to the highest ionic strength MRTs rise by a factor 25 to 30, *i.e.* from the picosecond to the nanosecond regime.

C Mesoscopic view

So far we have analysed the systems at the molecular level. The Voronoi method enables a more coarse grained resolution considering a group or set of Voronoi polyhedra as one moiety. A natural choice for such mesoscopic moieties would be the protein, its solvation shells and the remaining bulk. This coarse graining of space goes along with the dipolar decomposition already outlined in the Theory section. Therefore, the moieties or regions are characterized by their dipolar self-terms $\langle\vec{M}_k^2\rangle$ and cross terms $\langle\vec{M}_k\cdot\vec{M}_l\rangle$. For x_{H_2O} = 1 five complete Voronoi shells may be constructed for the given system size. With increasing ionic strength the thickness of the Voronoi shells increases such that the number of complete shells finally reduces to three. As a common number we therefore choose three solvation shells. A more detailed analysis, however, showed that the third solvation shell already resembles the properties of the bulk for all ionic strengths. Therefore, a four component description with the protein, the first and second solvation shells and the bulk are sufficient.

In the following tables the self and cross terms are given as entries to a 4 × 4-matrix for each mole fraction (Table 6). The prefactor $4\pi/(3Vk_BT)$ is already included. Therefore, one can directly read off their respective contributions to the dielectric permittivity. The behaviour of the dielectric self-terms as a function of ionic strength is quite diverse. The self-terms of the bulk decrease monotonically. Except for the system x_{H_2O} = 0.927 this is also true for the self-terms of the first and second shells. The self-term of the protein rises from the almost pure water system x_{H_2O} = 0.927 but then remains confined to an interval between 15 and 25.

This set of matrices gives a very clear answer concerning the cross terms. At the lowest ionic strengths, or equivalently, at x_{H_2O} = 0.927, even a pattern of chain coupling, *i.e.* 5ZNF to 1st shell, 1st shell to 2nd shell and 2nd shell to bulk, can be

Table 6 Spatially resolved contributions to the dielectric constant: self and cross terms. Above each table the sum of the self-terms ε_{diag} and the dielectric constant ε are given for comparison

$x_{H_2O} = 1.0$, $\varepsilon_{diag} + 1 = 89.6$, $\varepsilon = 108.8$

	5ZNF	1st shell	2nd shell	Bulk
5ZNF	4.4	0.4	0.2	1.6
1st shell	0.4	2.8	1.1	1.7
2nd shell	0.2	1.1	4.6	4.6
Bulk	1.6	1.7	4.6	76.8

$x_{H_2O} = 0.927$, $\varepsilon_{diag} + 1 = 64$, $\varepsilon = 72.8$

	5ZNF	1st shell	2nd shell	Bulk
5ZNF	25.0	1.4	0.8	0.1
1st shell	1.4	2.5	0.6	0.1
2nd shell	0.8	0.6	5.6	1.4
Bulk	0.1	0.1	1.4	29.9

$x_{H_2O} = 0.854$, $\varepsilon_{diag} + 1 = 41.1$, $\varepsilon = 41.7$

	5ZNF	1st shell	2nd shell	Bulk
5ZNF	15.8	−0.5	−0.1	0.3
1st shell	−0.5	1.7	0.2	−0.1
2nd shell	−0.1	0.2	4.5	0.5
Bulk	0.3	−0.1	0.5	18.1

$x_{H_2O} = 0.751$, $\varepsilon_{diag} + 1 = 40.5$, $\varepsilon = 39.1$

	5ZNF	1st shell	2nd shell	Bulk
5ZNF	24.0	−0.4	0.2	−0.4
1st shell	−0.4	1.5	0.1	−0.1
2nd shell	−0.2	0.1	4.1	−0.1
Bulk	−0.4	−0.1	−0.1	9.9

$x_{H_2O} = 0.675$, $\varepsilon_{diag} + 1 = 30.9$, $\varepsilon = 30.5$

	5ZNF	1st shell	2nd shell	Bulk
5ZNF	16.6	0.5	−0.1	−0.3
1st shell	0.5	1.5	−0.1	0.0
2nd shell	−0.1	−0.1	3.5	−0.2
Bulk	−0.3	0.0	−0.2	8.3

detected. For high ionic strengths this chain is more and more broken. In other words, once the ionic strength has exceeded the magic value $x_{H_2O} = 0.927$ the importance of cross terms is drastically reduced. In fact, the sum of the self-terms comes very close to the sum over all terms, *i.e.* to the total value of the dielectric permittivity.

The case $x_{H_2O} = 0.927$ is once more pronounced: the contribution from the protein is at the maximum and at the same time the coupling to its environment, in particular to the first shell, is maximal, too. A similar but weaker trend is observed for the almost pure water system, in particular the protein seems to have a longer range of influence on the shells and even on the bulk. In fact, two times the cross term protein–bulk rivals already the protein self-term.

Besides the marginalisation of the cross terms at higher ionic strengths the dominant contribution of the protein is the most striking feature. As the solvent's contribution decreases drastically with ionic strength the protein's contribution becomes more and more important. For the highest ionic strength it already makes up one half of the total dielectric permittivity. Thus, the dipole density in the solvent is reduced more and more and in the end the protein remains the only zone of high dipolar density.

From the combination of eqn (4), (6) and (7) one can learn that the off-diagonal elements of the dielectric matrices presented above may be also interpreted as a measure of the dipolar coupling between moieties or regions.

For the almost pure water solvent the mutual coupling is strongest: this concerns not only the coupling among the water moieties, first shell, second shell and bulk, but also the coupling between protein and bulk. The sum of all these couplings makes up 20% of the total dielectric effect.

At the magic point the coupling between the protein and its immediate neighbourhood, *i.e.* first shell, is most pronounced. All other contributions are slightly reduced by the change from dipolar screening to charge screening. In particular, the protein–bulk interaction becomes very small. Nevertheless all couplings remain positive, *i.e.* corresponding regional dipoles are aligned parallel on average. This picture changes for higher ionic strengths where antiparallel dipolar alignment dominates.

As a consequence the sum of the diagonal elements seems to completely determine the total dielectric effect. However, this dominance of the diagonal terms comes from the fact that the already small off-diagonal elements cancel each other to a large extent when summed up.

Reduction of the mutual coupling of shell dipole moments indicates that the ordering influence of the protein on the solvent decreases. Usually, such a screening effect goes along with a high dielectric constant. Therefore, this finding seems contradictory at first sight. However, we have already learned from the discussion of the 3-dimensional ion densities and contact vectors that systems at high ionic strengths do not follow a continuum description, because cations and anions establish a highly ordered charge pattern.

In order to elucidate the vanishing dielectric influence of the solvent we have also performed a solvent-species specific decomposition of the total dielectric permittivity. The results are given in Table 7. Although the contribution of the protein is confined to an interval an overall dielectric decrement is observed. This is so for three reasons: first, the water components show a decrement as a consequence of a reduced mole fraction x_{H_2O}, *i.e.* it is more and more substituted by cations and anions. Second, a dielectric increment is observed for the anion. However, the slope is modest as compared to decrement of water. Third, the contribution of the cations is so marginal that is increment is negligible.

Quite generally, dielectric contributions are closely related to dipolar densities, *i.e.* to the square of the dipole moment of a voxel divided by its volume. Identifying the voxel with the Voronoi polyhedron around a single molecule gives its isolated contribution to the dielectric constant. The dipolar densities of water, $CF_3SO_3^-$ and $EMIM^+$ follow the ratio 18 : 31 : 1.8. Thereby we have already included the prefactor $4\pi/(3Vk_BT)$ which converts the dipolar density to a dielectric contribution (eqn (6)). For the ideal case of non-interacting molecules one could calculate the dielectric constant as a linear combination of the three values weighting each with its respective mole fraction. This simplified consideration shows already that the replacement of the water component by $EMIM^+$ has a mere "thinning out"-effect and is the major reason for the strong dielectric decrement. The enrichment of cations in the first coordination shell found already in the discussion of the

Table 7 Contributions to the dielectric constant, resolved by species: self- and cross-terms

x_{H_2O} = 1.0, ε = 108.8				
	5ZNF	H_2O	$CF_3SO_3^-$	$EMIM^+$
5ZNF	4.4	2.2	0.0	0.0
H_2O	2.2	98.7	0.1	0.0
$CF_3SO_3^-$	0.0	0.1	0.1	0.0
$EMIM^+$	0.0	0.0	0.0	0
x_{H_2O} = 0.927, ε = 72.8				
	5ZNF	H_2O	$CF_3SO_3^-$	$EMIM^+$
5ZNF	24.1	1.0	0.0	0.0
H_2O	1.0	29.2	1.9	0.1
$CF_3SO_3^-$	0.0	1.9	3.8	0.1
$EMIM^+$	0.0	0.1	0.1	0.5
x_{H_2O} = 0.854, ε = 41.7				
	5ZNF	H_2O	$CF_3SO_3^-$	$EMIM^+$
5ZNF	19.7	0.5	−0.1	−0.3
H_2O	0.5	12.9	1.9	0.1
$CF_3SO_3^-$	−0.1	1.9	4.3	0.1
$EMIM^+$	−0.3	0.1	0.1	0.7
x_{H_2O} = 0.751, ε = 39.1				
	5ZNF	H_2O	$CF_3SO_3^-$	$EMIM^+$
5ZNF	24.0	0.2	−0.5	−0.3
H_2O	0.2	7.2	1.0	−0.1
$CF_3SO_3^-$	−0.5	1.0	4.8	0.2
$EMIM^+$	−0.3	−0.1	0.2	1.0
x_{H_2O} = 0.675, ε = 30.5				
	5ZNF	H_2O	$CF_3SO_3^-$	$EMIM^+$
5ZNF	16.6	−0.1	0.1	0.1
H_2O	−0.1	4.7	1.0	0.0
$CF_3SO_3^-$	0.1	1.0	4.9	0.2
$EMIM^+$	0.1	0.0	0.2	0.9

coordination number above enhances this effect. Replacement by $CF_3SO_3^-$ produces an increment and thus counteracts this effect. However, interdipolar coupling neglected so far reduces the contribution of the isolated $CF_3SO_3^-$ molecules because of their essentially antiparallel alignment. This was already found in previous studies: collectivity reduces the contribution of the isolated $CF_3SO_3^-$ by a typical factor of one half. This is quite opposite to water where collectivity enhances the single water molecule's contribution by a factor of two to three. Therefore, the $CF_3SO_3^-$ -component cannot really compensate for the reduction of the water component. So it is finally left to the protein to maintain the dielectric constant of the complete solution.

So far we have discussed the species-specific self-terms. Table 7 shows an interesting behaviour of the cross terms, too. From the above discussion it is not surprising that the cross terms with $EMIM^+$ are marginal. The only cross terms of importance for all non-zero ionic strengths are those between water and $CF_3SO_3^-$. Cross correlations between the protein and water are only observed in the water-rich systems. Therefore, the water–$CF_3SO_3^-$ network already observed at the molecular level in systems without the solute seems to persist at the mesoscopic dielectric level, too.

V. Summary and conclusion

In this study we have analysed the solvation of a zinc finger protein "5ZNF" in a hydrated ionic liquid composed of $EMIM^+$, $CF_3SO_3^-$ and water. The mole fraction of water x_{H_2O}, or equivalently the ionic strength I, were varied systematically from pure water to x_{H_2O} = 0.675.

This very inhomogeneous system was divided up into three principal regions: the protein itself, its immediate neighbourhood and the remaining bulk. For such intuitive decomposition, however, we need a rational concept. While the protein may be easily defined as the union of amino acids, the characterisation of the immediate neighbourhood is usually done by distance based methods. For a homogeneous solvent this would already require the introduction of parameters like shell thickness, *etc*. Therefore, the heterogeneous solvent used here would require an extended set of parameters. It is a definite goal of this study to describe solvation shells introducing no parameters at all. Therefore we have used the computationally more demanding Voronoi method in order to spatially decompose the complete system into atomic polyhedra. In this way neighbourhood may be easily defined by common or shared polyhedral faces and the union of atomic polyhedra gives moieties, solvation shells, *etc*.

Volume and surface were the first simple properties of the protein itself which were investigated. Later on the structural behavior of the backbone was analysed using DSSP, STRIDE, and contact matrices. The behaviour of the side chains was characterised by their dipole moments, *i.e*. by their contribution to the protein's net dipole moment. All these properties exhibit a maximum or minimum at the magic point x_{H_2O} = 0.927. This corresponds to the transition from a pure dipolar screening by water molecules to a beginning charge screening by ions. The total energy of the protein as the common root for all these structural properties exhibits a maximum at the magic point. This was brought about by the monotonic increase of the van der Waals energy and the monotonic decrease of the electrostatic energy as a function of the mole fraction. It should be pointed out that the inherent force field inaccuracies would make a quantitative prediction of the exact water concentration for this magic point very difficult. In any event, the main result is that such a point exists.

The immediate neighbourhood of the protein was first characterised by 3D density plots which were compared to the electrostatic potential computed as a solution of the Poisson–Boltzmann equation. For low ionic strength larger regions of the electrostatic potential of uniform sign are found. The distribution of ions follows this uniform sign of the potential. For higher ionic strengths however, a checkerboard pattern with changing sign of the ions emerges. This cannot be described by the Poisson–Boltzmann equation anymore. Obviously, the higher number of ions has to arrange in such a way to minimize the mutual repulsion of like charges.

This intuitive picture of the first solvation shell can be rationalised by a series of properties: already the simplest of them, the coordination number, shows that cations are enriched at the protein surface although the protein itself is positively charged. This is an indirect effect of the water–anion network which expels cations. A residue specific resolution of

solute–solvent contacts can be achieved by Voronoi based contact matrices which in their turn may be contracted to contact vectors. In this way, 3D densities can be converted to a quantitative one-dimensional information. Thereby, amino acid residues are only ordered according to the sequence. Specific neighbourhood relations are not considered. Nevertheless, specific statements can be made. A typical example is the change of the charge pattern as a function of the ionic strength which shows up as slower or faster oscillations in the components of the contact vector.

As a measure for solvent shell dynamics we have computed mean residence times. Defining shell specific mean residence times characteristic for the complete set of molecules of a respective species we find a linear relationship between the logarithms of these mean residence times, or life times, as a function of the mole fraction. In other words, the species-specific MRTs are an exponential function of the mole fraction. This exponential rise makes clear why simulations at high ionic strengths are computationally very demanding. MRTs up to 66 nanoseconds were found.

In order to characterise the system as a whole we have also introduced a mesoscopic view. Thereby each region or moiety is characterised by its contribution to the total dielectric constant of the system. Such a decomposition involves both self-terms and cross terms. The latter are a measure for the mutual coupling of regions. At the magic point $x_{H_2O} = 0.927$, the self-terms are high, but the cross terms are remarkable, too. In particular, the protein is strongly coupled to its environment, especially to the first solvation shell. One might think that cross terms increase together with the ionic strength. Quite the opposite is true, however: with increasing ionic strength the total dielectric constant is more and more determined by the sum of self-terms and the cross terms loose their importance. $EMIM^+$ more or less acts as a dilution because of its marginal dipole density. The higher dipole density of $CF_3SO_3^-$—as compared to water—is of modest influence because of the marginal cooperativity between the $CF_3SO_3^-$ ions in terms of dipolar coupling. The water component with its pronounced dipolar cooperativity is quenched simply by its decreasing mole fraction.

Acknowledgements

This work was performed on the "Vienna Scientific Cluster" [www.zid.tuwien.ac.at/vsc] of the University of Vienna, the Vienna University of Technology, and the University of Natural Resources and Applied Life Science Vienna. We thank for generous allocation of computer time. This work was supported by Project No. P19807 of the FWF Austrian Science Fund.

References

1 H. Olivier-Bourbigou, L. Magna and D. Morvan, *Appl. Catal., A*, 2010, **373**, 1.
2 M. Moniruzzaman, K. Nakashima, N. Kamiya and M. Goto, *Biochem. Eng. J.*, 2010, **48**, 295.
3 U. Kragl, M. Eckstein and N. Kaftzik, *Curr. Opin. Biotechnol.*, 2002, **13**, 565.
4 S. Park and R. J. Kazlauskas, *Curr. Opin. Biotechnol.*, 2003, **14**, 432.
5 H. Zhao, *J. Mol. Catal. B: Enzym.*, 2005, **37**, 16.
6 E. Fehér, B. Major, K. Bélafi-Bakó and L. Gubicza, *Biochem. Soc. Trans.*, 2007, **35**, 1624.
7 F. van Rantwijk and R. A. Sheldon, *Chem. Rev.*, 2007, **107**, 2757.
8 K. Fujita, D. R. MacFarlane, M. Forsyth, M. Yoshizawa-Fujita, K. Murata, N. Nakamura and H. Ohno, *Biomacromolecules*, 2007, **8**, 2080.
9 C. Roosen, P. Müller and L. Greiner, *Appl. Microbiol. Biotechnol.*, 2008, **81**, 607.
10 Z. Yang, Y. Yue, W. Huang, X. Zhuang, Z. Chen and M. Xing, *J. Biochem. (Tokyo)*, 2009, **145**, 355.
11 K. Fujita, N. Nakamura, K. Igarashi, M. Samejima and H. Ohno, *Green Chem.*, 2009, **11**, 351.
12 R. J. Lichtenecker and W. Schmid, *Monatsh. Chem.*, 2009, **140**, 509.
13 J. Gorke, F. Srienc and R. Kazlauskas, *Biotechnol. Bioprocess Eng.*, 2010, **15**, 40.
14 D. Constatinescu, C. Herrmann and H. Weingärtner, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1756.
15 T. G. A. Youngs, J. D. Holbrey, M. Deetlefs, M. Nieuwenhuyzen, M. F. C. Gomes and C. Hardacre, *ChemPhysChem*, 2006, **7**, 2279.
16 C. Hardacre, J. D. Holbrey, M. Nieuwenhuyzen and T. G. A. Youngs, *Acc. Chem. Res.*, 2007, **40**, 1146.
17 M. N. Kobrak, *J. Chem. Phys.*, 2007, **127**, 184507.
18 S. Arzhantsev, H. Jin, G. A. Baker and M. Maroncelli, *J. Phys. Chem. B*, 2007, **111**, 4978.
19 A. Samanta, *J. Phys. Chem. Lett.*, 2010, **1**, 1557.
20 E. J. Maginn, *J. Phys.: Condens. Matter*, 2009, **21**, 373101.
21 C. Schröder, T. Rudas, S. Boresch and O. Steinhauser, *J. Chem. Phys.*, 2006, **124**, 234907.
22 T. Rudas, C. Schröder, S. Boresch and O. Steinhauser, *J. Chem. Phys.*, 2006, **124**, 234908.
23 M. Feig, *Modeling Solvent Environments: Applications to Simulations of Biomolecules*, Wiley-VCH, 2010.
24 N. M. Micaêlo and C. M. Soares, *J. Phys. Chem. B*, 2008, **112**, 2566.
25 T. A. Page, N. D. Kraut, P. M. Page, G. A. Baker and F. V. Bright, *J. Phys. Chem. B*, 2009, **113**, 12825.
26 C. Schröder, T. Rudas and O. Steinhauser, *J. Chem. Phys.*, 2006, **125**, 244506.
27 C. Schröder, C. Wakai, H. Weingärtner and O. Steinhauser, *J. Chem. Phys.*, 2007, **126**, 084511.
28 C. Schröder, J. Hunger, A. Stoppa, R. Buchner and O. Steinhauser, *J. Chem. Phys.*, 2008, **129**, 184501.
29 G. Löffler, H. Schreiber and O. Steinhauser, *J. Mol. Biol.*, 1997, **270**, 520.
30 S. Boresch, S. Ringhofer, P. Höchtl and O. Steinhauser, *Biophys. Chem.*, 1999, **78**, 43.
31 C. Schröder, M. Haberler and O. Steinhauser, *J. Chem. Phys.*, 2008, **128**, 134501.
32 C. Schröder and O. Steinhauser, *J. Chem. Phys.*, 2009, **131**, 114504.
33 I. Klapper, R. Hagstrom, R. Fine, K. Sharp and B. Honig, *Proteins: Struct., Funct., Bioinf.*, 1986, **1**, 47.
34 G. Neumayr, C. Schröder and O. Steinhauser, *J. Chem. Phys.*, 2009, **131**, 174509.
35 B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels and S. Boresch, *et al.*, *J. Comput. Chem.*, 2009, **30**, 1545.
36 W. L. Jorgensen, *J. Am. Chem. Soc.*, 1981, **103**, 335.
37 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo and S. Ha, *et al.*, *J. Phys. Chem. B*, 1998, **102**, 3586.
38 C. Schröder and O. Steinhauser, *J. Chem. Phys.*, 2010, **132**, 244109.
39 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327.
40 G. Neumayr, T. Rudas and O. Steinhauser, *J. Chem. Phys.*, 2010, **133**, 084108.
41 C. Schröder, G. Neumayr and O. Steinhauser, *J. Chem. Phys.*, 2009, **130**, 194503.
42 X. Huang, C. J. Margulis, Y. Li and B. J. Berne, *J. Am. Chem. Soc.*, 2005, **127**, 17842.
43 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577.
44 D. Frishman and P. Argos, *Proteins: Struct., Funct., Bioinf.*, 1995, **23**, 566.
45 J. N. A. C. Lopes and A. A. H. Padua, *J. Phys. Chem. B*, 2006, **110**, 3330.

4 On the Influence of Hydrated Ionic Iiquids on the Dynamical Structure of Model Proteins: A Computational Study

PCCP

View Online
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 17994–18004

www.rsc.org/pccp

PAPER

On the influence of hydrated ionic liquids on the dynamical structure of model proteins: a computational study

Michael Haberler and Othmar Steinhauser*

Received 12th July 2011, Accepted 23rd August 2011
DOI: 10.1039/c1cp22266j

The solvation of the protein ubiquitin (PDB entry "1UBQ") in hydrated molecular ionic liquids was studied for varying water content or, equivalently, a diversity of ionic strengths. The cations and anions were 1-ethyl-3-methylimidazolium and trifluoromethanesulfonate, respectively. The protein's shape and stability as well as the solvation structure, the shell dynamics and the shell resolved dielectric properties were investigated by means of molecular dynamics simulations. The respective simulation trajectories covered 200 nanoseconds. Besides the characteristic point already found for the zinc finger motif at the transition from the pure aqueous environment to the ionic solution an even more pronounced state is found where several properties show extremal behaviour (maximum or minimum). This second characteristic point occurs at the transition from the ionic solution to the hydrated ionic melt where water changes its role from a solvent to a co-solvent. Most of the data analysis presented here is based on the Voronoi decomposition of space.

1 Motivation

As numerously stated, molecular ionic liquids (MIL) are a benign alternative to traditionally used volatile organic compounds. In biomolecular solvation their role as green chemistry solvents becomes most obvious. In fact, the stability and/or catalytic activity of certain enzymes may be enhanced by MILs.[1,2] The range of biomolecules may be further extended using mixtures of water and MILs, so-called hydrated molecular ionic liquids. Thereby water as well as MILs can act as a co-solvent as reviewed in ref. 3 and 4.

Of course, the overwhelming part of publications on biomolecular solvation refers to the aqueous environment.[5] A smaller part is devoted to solvation studies in nonaqueous solvents, mainly classic organic compounds.[6] Biomolecular solvation studies in (hydrated) MILs are rare but exist.[7] Irrespective of the solvent environment most experimental studies provide macroscopic data, mainly thermodynamic and kinetic. Few exceptions for a molecular resolution are the fluorescence measurements of Page *et al.*[8] and the EPR experiments of Akdogan *et al.*[9] Alternatively, simulation studies provide a tool for a systematic investigation of molecular structure and dynamics. To our best knowledge we are aware of the study of glucose by ref. 10, of cutinase by ref. 11 and of a lipase by ref. 12. These studies are based on analysis tools adapted from the traditional studies on aqueous biomolecular systems. In recent papers, however, we could show that novel solvents need novel tools:[13–15] water and diluted ionic solutions either consist of small polar molecules or atomic charges. MILs, however, are composed of sterically anisotropic, electrically polar, and charged molecules. It is important to note that all these properties are combined in the very same molecule. This simultaneous occurrence of molecular features requires novel approaches and techniques of analysis. In particular, the traditional way to characterise solvation shells of proteins on the basis of radial distribution functions becomes cumbersome involving a plethora of parameters. Therefore our complete analysis is based on the parameter-free Voronoi decomposition of space. We have demonstrated the power of this approach in a simulation study of the zinc finger model protein (PDB entry "5ZNF") immersed in a mixture of MIL and water where we systematically varied the mole fraction or concentration of the solvent constituents.[16] The underlying paper deals with the solvation of ubiquitin (PDB entry "1UBQ"). The aim of this study is twofold: first, we will investigate whether the features found for a model system like 5ZNF are also found for a classic NMR-benchmark protein of more than twice the size. Second and even more important, we want to extend the diversity of mole fractions or ionic strengths. For 5ZNF we focused on the transition from the pure aqueous environment to an ionic solution and found a characteristic point where the dielectric screening of solvent water changes to the charge screening by MILs. Here we want to see how the further reduction of the water component affects the structure and dynamics of the protein and its immediate neighbourhood. In particular we want to figure out the transition from a situation where the MIL acts as a co-solvent (ionic solution) to the case where water plays this role (hydrated ionic melt).

University of Vienna, Department of Computational Biological Chemistry, Währingerstr. 17, 1090 Vienna, Austria. E-mail: os@mdy.univie.ac.at

Table 1 Decreasing mole fraction is synonymous with increasing ionic strength. The relation is monotonic but nonlinear. We give the parameters for the 1UBQ systems only, for 5ZNF we refer to Table 1 in ref. 16

x_{H_2O}	c_{IL}/mol l^{-1}	c_{H_2O}/mol l^{-1}	#EMIM$^+$	#$CF_3SO_3^-$	#H_2O	d/Å	ρ/g l^{-1}
1	0	53.6	0	0	8523	70.01	1019.06
0.88	3.1	22.4	461	461	3366	68.64	1261.55
0.8	3.8	15.2	561	561	2244	68.35	1317.45
0.75	4.1	12.1	592	592	1763	67.90	1339.46
0.66	4.4	8.5	649	649	1244	68.10	1364.94
0.52	4.7	5.2	698	698	768	68.22	1386.42

2 Methods

2.1 Details of simulation

The core data for the subsequent analysis are based on five molecular dynamics (MD) simulations performed with the CHARMM package. The protein ubiquitin (PDB entry "1UBQ") was immersed in a mixture of water, EMIM$^+$ and $CF_3SO_3^-$ (TRIF). Results for the pure aqueous system were taken over from previous papers.[17,18] Table 1 shows a collection of the respective water mole fractions, ionic strengths, water concentrations, number of EMIM$^+$, $CF_3SO_3^-$, and water molecules, the edge length of the truncated octahedron and the mass density of the simulation box. (The edge length of the cube to which the truncated octahedron is inscribed is longer by the factor $\sqrt{4/3}$.)

The starting configurations were generated with the program PACKMOL[19] filling a cube of edge length 82 Å with the appropriate number of solvent molecules. Subsequently the cube corners were cut in order to get the truncated octahedron. If necessary the system was neutralised by replacing excess ions by counter ions or water. Then the system was minimised with the steepest descent method for 200 steps. A first equilibration in the NPT-ensemble covered one nanosecond. A second one in the NVT-ensemble lasted half a nanosecond. The average edge lengths of the second half of the initial equilibration was used as the constant value for the NVT-ensemble. The second equilibration was followed by the production run in the NVT-ensemble of about 200 nanoseconds at a temperature of 300 K. The respective time step was 2 femtoseconds.

The atomic forces were calculated in the same way as described at the end of the Method section A in ref. 16. The force field parameters of the protein were taken from CHARMM.[20] For the solvent molecules, the parameters were identical to those already used in ref. 21. The mutual interaction between the protein and the solvent was modelled by electrostatic interactions and Lennard-Jones terms. In the latter case mixing rules between the solute and solvent force field were used. The procedure to handle the long range electrostatic forces was along the lines described in the Method section of ref. 21 with a κ of 0.41 Å^{-1} and a real space cut-off of 10 Å.

All bond lengths were kept fixed by the shake algorithm,[22] whereas bond angles and torsions are left flexible.

For analysis, comparison, and interpretation the results of the five longer (mole fractions 1.0, 0.93, 0.85, 0.75, 0.68) and three shorter simulations (mole fractions 0.92, 0.89, 0.83) reported in ref. 16 were also included in those cases where analogy is helpful.

2.2 Voronoi decomposition as a basic tool

Except for the energy, the RMSF, RMSD, and the species-specific dipole moments, the complete analysis was based on the Voronoi decomposition of space. In other words, every frame (trajectory snapshot) considered was subjected to a Voronoi tessellation which uniquely assigns a Voronoi polyhedron to each atom of the system. This polyhedron comprises all space being closer to this atom than to any other one. As atoms constitute molecules these atomic polyhedra are the building blocks of larger moieties, *e.g.* molecules, solvation shells, *etc.* The detailed description of the construction and analysis of these Voronoi polyhedra is given in ref. 14. Except for a preliminary study of spherical holes[23] our application of the Voronoi method for the analysis of molecular ionic liquids is quite novel. We also use the Voronoi concept for a parameter-free definition of neighbourhood: moieties that share at least one Voronoi face are called neighbours. The sum of all shared surfaces is called the Voronoi interface surface area (VISA).

The Voronoi method is a coarser decomposition of space than the usual description in terms of atom pair correlation functions resolving the distance between atom pairs into radial bins. For solvents composed of molecules of varying size and anisotropic shape it enables a parameter free and rational characterisation of neighbourhood. It is even superior to the popular 3D density plots giving only a visual and approximate impression of neighbourhood and depending intrinsically on the threshold chosen. Fig. 1 shows how the surface elements of the protein's VISA can be uniquely mapped to the nearest neighbour solvent molecule. This done per residue: that solvent species which—on average—shares the largest VISA with the respective residue is considered as the dominant species and the complete VISA of the residue is coloured according to this species. We consider this sequence of coloured protein VISA for different mole fractions as a seed for the subsequent discussion where try to reveal step by step the underlying phenomena of protein solvation as a function of ionic strength.

3 Results and discussion

3.1 The protein itself: a global view

Fig. 1 contains a twofold information. On the one hand, the Voronoi interface surface area (VISA) can be regarded as a measure of the compactness of the protein. On the other hand, the colouring of each residue shows which solvent species makes the dominant contribution to the total VISA. From this graphical representation one gets the impression that the VISA

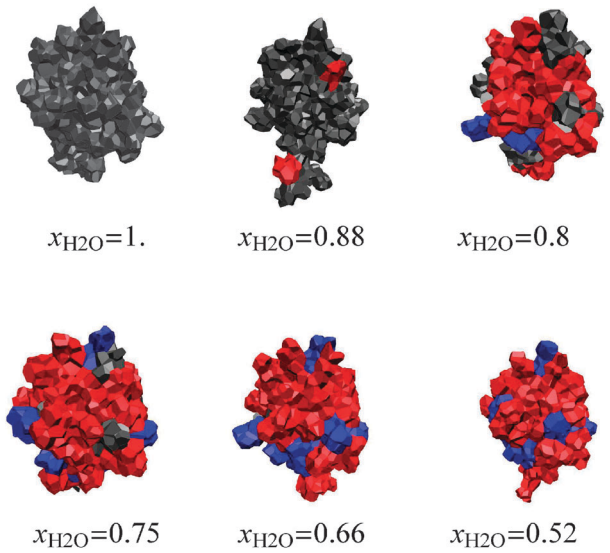


Fig. 1 Representative Voronoi polyhedra for 1UBQ at the different mole fractions. The faces of the polyhedron are coloured according to the dominant solvent species (gray for H_2O, red for $EMIM^+$, and blue for $CF_3SO_3^-$).

is not a monotonic function of the ionic strength. Rather the protein seems to adopt a maximum compactness at a certain mole fraction. Of course the pictures given refer to snapshots but we have ascertained that this high compactness is preserved—on the average—over the whole trajectory. This is documented in Fig. 2(a) giving the average VISA for ubiquitin and its two subunits, comprising amino acids 1–38 and 39–76, respectively. The average VISA of 5ZNF is also given. This subdivision of ubiquitin is motivated by the fact that the first half of the molecule resembles the structure of a zinc finger motif: a combination of a β-hairpin and an α-helix despite of the lack of a zinc ion. The second half of the molecule contains remnant features of secondary structure elements connected by rather unstructured regions.

As can be seen in Fig. 2(a) subunit 1 of 1UBQ and 5ZNF are both characterised by a point of maximum compactness which seems to be a typical feature of a zinc finger motif. On further inspection, the increase of compactness seems to go along with an increase of sphericity. Therefore, Fig. 2(b) gives the average sphericity, $\Psi = \pi^{1/3}(6\text{Volume})^{2/3}/\text{VISA}$, which in fact displays a maximum at the mole fraction $x_{H_2O} = 0.75$ for ubiquitin. A local maximum can be also found for 5ZNF at $x_{H_2O} = 0.88$.

The major player in this measure of sphericity, however, is the VISA as the Voronoi volume is rather independent of the mole fraction.

Since the VISA is a geometrical measure we next considered the total energy of the protein and its decomposition into van der Waals and electrostatic contributions. Comparing Fig. 3(a) and (b), the bonded interactions cause a more or less constant shift of the total energy but do not change the behaviour of the nonbonded energy, *i.e.* the sum of van der Waals and electrostatic contributions. These latter two are shown separately in Fig. 3(c) and (d), respectively.

The van der Waals energy is a monotonically increasing function of the ionic strength. The electrostatic energy is essentially decreasing with a local minimum at the point of maximum compactness where the interatomic distances are shortest and the electrostatic attraction is most favourable. This behaviour of the total electrostatic energy again comes from subunit 1 because subunit 2 displays no fine structure among the nonaqueous systems. Summing up, the van der Waals and electrostatic contributions lead to a flattening of the total nonbonded energy. The fine structure, however, is preserved due to the dominating influence of the electrostatic energy.

When changing from the pure aqueous environment to an aqueous solution of ILs all curves are characterised by an initial rise. This behaviour was already found for 5ZNF in ref. 16 and is again confirmed for 1UBQ in this study. The possible mechanism for this change in solvation was already discussed in ref. 16 (see for example the discussion preceding and following Fig. 5).

It is interesting to compare the internal energy of the protein discussed so far to the interaction energy between the protein and the solvent. This is shown in Fig. 4. In both cases, zinc finger and ubiquitin, the local minimum of the internal energy correlates with a local maximum of the interaction energy between protein and solvent. In other words, the protein compensates for the reduced interaction strength with the solvent by enhancing internal interactions which in their turn increase compactness. As for the internal energy, the fine structure is again caused by the interplay of electrostatic and van der Waals energy. While the latter decays monotonically the local enhancement is already visible in the former.

So far our discussion was based on geometrical (VISA, sphericity) and energetic properties. In order to elucidate the influence of ionic strength on the protein's structure and

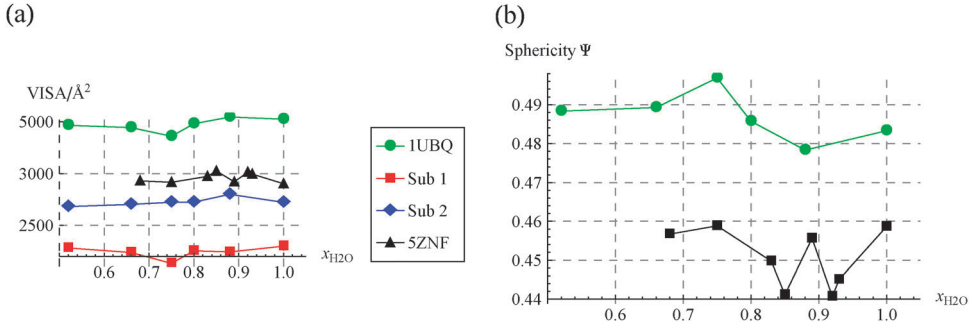


Fig. 2 VISA (a) and sphericity (b) of 1UBQ and 5ZNF ("Sub 1" and "Sub 2" stand for subunit 1 and subunit 2, respectively).

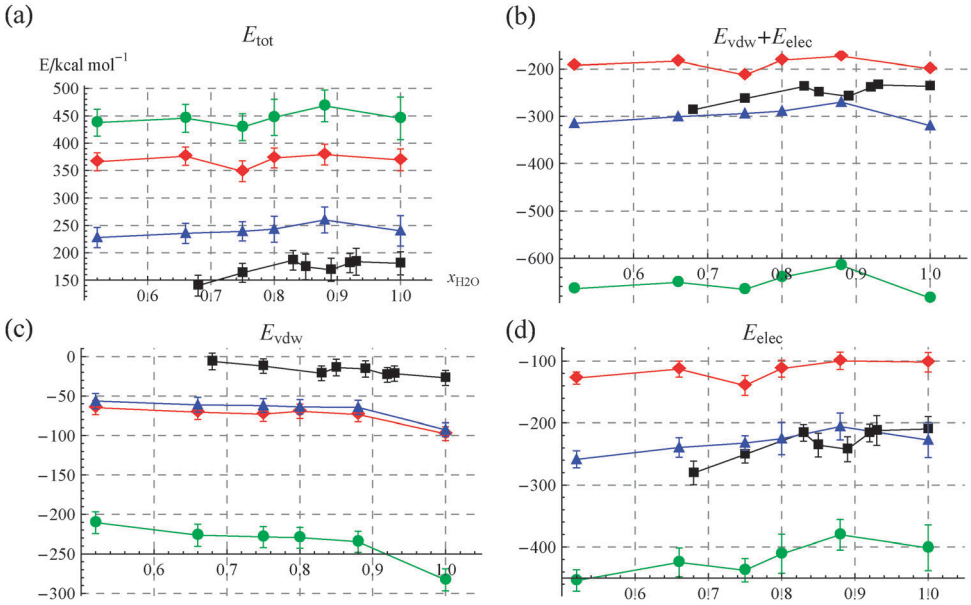


Fig. 3 Total energy of 1UBQ and 5ZNF (a), sum of van der Waals and electrostatic energy (b), van der Waals energy (c), and electrostatic energy (d). The colour code is the same as in Fig. 2(a).

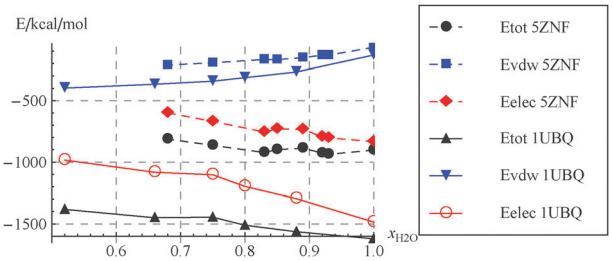


Fig. 4 Total interaction energy between protein, zinc finger or ubiquitin, and solvent and its decomposition into electrostatic and van der Waals contributions.

dynamics Fig. 5 displays the root mean square deviation of all C_α-atoms (RMSD, with respect to the crystal structure) and the root mean square fluctuations (RMSF) of all backbone atoms being proportional to the B-factors used in crystallography. For ubiquitin, the RMSD curve for the whole protein is again split into the contributions from the subunits. One might argue that the contribution of the fraying ends (residues 72–76 for 1UBQ, and residues 1, 29, 30 for 5ZNF) should be omitted in RMSD and RMSF curves. When doing this the absolute values go down but the characteristic features remain unaffected. Therefore we have included the fraying ends. As a consequence, the curve for the whole protein is the mean of the curves for the two subunits. The characteristic features of all properties analysed globally in this section will be further resolved from a regional viewpoint in Section 3.3.

3.2 The proteins immediate neighbourhood

3.2.1 Coordination numbers for the total protein surface. In the previous section we have found a consistent behaviour of the geometry, energetics, structure, and dynamics of the protein itself. Of course, these properties are not isolated but influenced by the surrounding solvent. Therefore, in a first step, we analyse the protein's immediate neighbourhood, *i.e.* the first solvation shell. As outlined in the Methods section the first solvation shell was deduced from the Voronoi tessellation: those solvent molecules which have at least one Voronoi face in common with the protein are called first neighbours. The respective total number of neighbours of a species, *i.e.* the

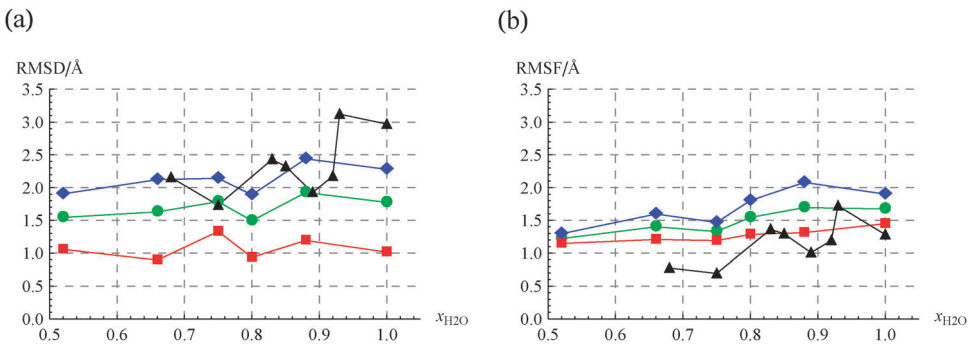


Fig. 5 (a) RMSD of C_α-atoms of 1UBQ (green), C_α-atoms of subunit 1 (red), C_α-atoms of subunit 2 (blue), C_α-atoms of 5ZNF (black). (b) RMSF of backbone atoms of 1UBQ (green), backbone atoms of subunit 1 (red), subunit 2 (blue) and backbone atoms of 5ZNF (black).

Table 2 Species specific coordination numbers of 1UBQ

x_{H_2O}	c_{IL}/mol l^{-1}	c_{H_2O}/mol l^{-1}	CN_+	CN_-	CN_0
1	0	53.6	0	0	516.2
0.88	3.1	22.4	52.7	40.3	256.2
0.8	3.8	15.2	65.0	48.8	178.0
0.75	4.1	12.1	70.6	52.4	138.7
0.66	4.4	8.5	75.9	57.2	106.1
0.52	4.7	5.2	82.8	62.7	73.3

coordination number, is collected in Table 2 for all ionic strengths or mole fractions studied.

3.2.2 Coordination numbers for partial surfaces. In order to show up the composition of solvent molecules in the first shell as compared to the overall stoichiometric composition we have calculated the local mole fraction

$$x_{i,\mathrm{local}} = \frac{CN_i}{(CN_+ + CN_-)/2 + CN_0}$$
$$i = \{+ \text{ for } EMIM^+, - \text{ for } CF_3SO_3^-, 0 \text{ for } H_2O\} \quad (1)$$

The respective curves are shown in Fig. 6 together with the straight lines x_{H_2O} and $x_{IL} = 1 - x_{H_2O}$. For ubiquitin, the local mole fraction for $CF_3SO_3^-$ sticks closely to the stoichiometric mole fraction. $EMIM^+$ cations are enriched while water molecules are slightly underrepresented in the first shell. In the case of 5ZNF this enrichment of cations and the lack of water molecules are slightly enhanced. These observations go along with our findings for pure hydrated 1-butyl-3-methylimidazolium tetrafluoroborate where we stated a strong anion–water network.[24,25] A similar mechanism seems to be in action here: the cations are expelled from the anion–water network and are enriched at the protein surface. Fig. 1 gives a visual expression of this enrichment of cations. It is important to note that the anion density in the vicinity of the zinc finger also exceeds the bulk density. In this case however, the positive net charge of the zinc finger—as compared to the zero net charge of ubiquitin—seems to be the obvious reason.

Considering mole fractions, either global or local, is not exclusive to one molecular species but always includes the influence of the other species. Therefore we now consider the relation between the coordination number in the first shell CN, and the respective global concentration, c_{IL}. In order to get a deeper insight into the coordination numbers of cations CN_+^S and anions CN_-^S we have computed these quantities with respect to different VISA as indicated by the superscript S: the surface of the zinc finger, S(5ZNF), the surface of subunits 1 and 2, S(Sub1) and S(Sub2), of ubiquitin. The mean values of these surfaces over all mole fractions considered are $\langle S(5ZNF)\rangle = 2968$ Å^2, $\langle S(Sub1)\rangle = 2231$ Å^2, $\langle S(Sub2)\rangle = 2726$ Å^2, $\langle S(1UBQ)\rangle = 4957$ Å^2. In order to emphasise the relation between the surface specific coordination numbers, CN_i^S, and the surface itself, we have computed the ionic surface density $CN_i^S/\langle S\rangle$. These densities are shown in Fig. 7. To a first approximation these normalised coordination numbers seem to be a linear function of ionic strength

$$\frac{CN_i^S}{\langle S\rangle} = k_i c_{IL} \quad (2)$$

with a species-specific universal constant k_i. We find $k_+ = 0.004$ and $k_- = 0.0031$.

Alternatively, the surface specific coordination number is given by $CN_i^S = k_i c_{IL}\langle S\rangle$ showing that the product of the ionic strength and the surface determines the absolute value of CN_i^S. From the previous section we know that the point of maximum compactness occurs at $c_{IL}^{5ZNF} = 3.1$ M for 5ZNF and at $c_{IL}^{Sub1} = 4.1$ M for the zinc finger motif in ubiquitin, referred to as subunit 1. (The respective water mole fractions are $x_{H_2O} = 0.88$ and $x_{H_2O} = 0.75$, see Table 2.) The ratio of these concentrations agrees with the inverse ratio of the respective surfaces

$$\frac{c_{IL}^{5ZNF}}{c_{IL}^{Sub1}} = \frac{\langle S(Sub1)\rangle}{\langle S(5ZNF)\rangle} = 0.75 \quad (3)$$

(The surface of the zinc finger motif in ubiquitin is smaller than that of the isolated zinc finger because of the second half of the ubiquitin molecule, subunit 2.) In other words the point of minimum VISA occurs when both the zinc finger and subunit 1 of ubiquitin have the same absolute value of the ionic coordination number. This offers the possibility to match the position of the characteristic minimum or maximum—observed for most properties presented in the previous section—between the zinc finger and ubiquitin. For example, Fig. 8 again presents the sphericity but now the ionic strength is used instead of the mole fraction as abscissa which was rescaled by the factor 0.75 found above in case of the zinc finger. Of course, this procedure can be done for all properties presented in the previous section to make the characteristic minimum (or maximum) coincide at the same position for the zinc finger and ubiquitin.

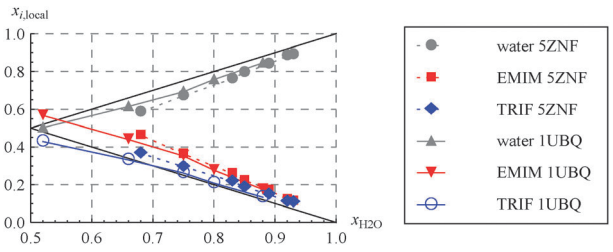


Fig. 6 As given in the legend the symbols represent the species specific local mole fractions within the first solvation shells. The straight lines in black stand for the stoichimetric global mole fractions, x_{H_2O} and x_{IL}.

3.2.3 Mean residence times. So far we have considered the average static coordination of the proteins by the solvent. Now we will analyse the dynamic reorganisation of the first solvation shell. In the Theory section of our previous paper[16] the definition, computation and fitting of the so-called residence function $\langle n(0)n(t)\rangle$ are described where $n(t)$ is a binary function with values 1 or 0 depending on whether a solvent molecule is a member of the first shell or not. For a minimum set of parameters the KWW-function $A_0 + Ae^{-(t/\tau)\beta}$ has proved as the most appropriate fit function. By the simple relation $\langle\tau\rangle = (\tau/\beta)\Gamma(1/\beta)$ a mean residence time (MRT) can be easily derived.

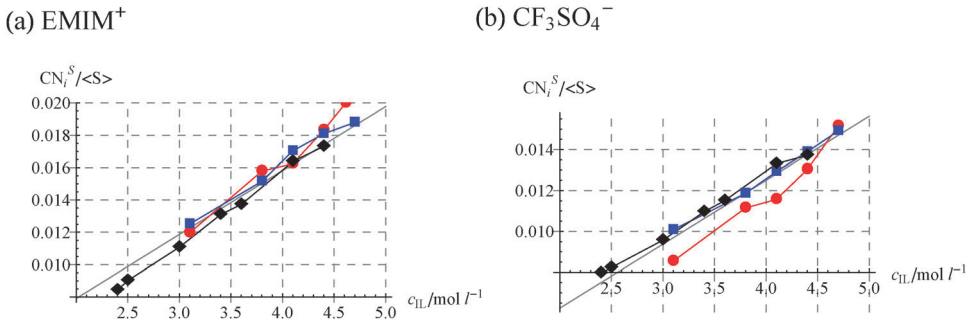


Fig. 7 Solvent specific surface density $CN_i^S/\langle S\rangle$ as a function of ionic concentration. The color code is the same as in Fig. 2. Straight lines with gradients $k_+ = 0.004$ (a) and $k_- = 0.0031$ (b) are shown in black.

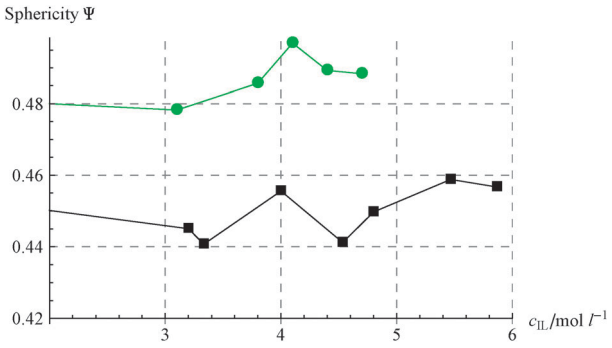


Fig. 8 Matching the characteristic points of 1UBQ and 5ZNF by rescaling the abscissa according to the inverse ratio of the zinc finger motif surfaces (green: 1UBQ, black: 5ZNF).

In order to get an impression of the original residence functions Fig. 9 shows $\langle n(0)n(t)\rangle$ for the cations and anions for the diversity of mole fractions. The retardation with increasing ionic strength is directly visible as well as the longer residence of the anions as compared to the cations. The respective coordination numbers can be read directly from the initial values of these correlation functions. The dominance of cations over anions in the first solvation shell can be seen again.

This intuitive picture can be rationalised by computing the mean residence times, $\langle\tau\rangle$, which are collected in Table 3 for the cations, anions, and water. We are aware of the fact that MRTs exceeding the simulation period occur at high ionic strengths. But as we see later on these extrapolated values follow the general trend already observed at low ionic strengths. This general trend can be found plotting logarithm of the MRT *versus* the mole fraction. As can be seen from Fig. 10 the MRT are exponential

Table 3 Species specific mean residence times as a function of mole fraction or ionic strength

x_{H_2O}	c_{IL}/mol l^{-1}	$\langle\tau\rangle_0$/ps	$\langle\tau\rangle_+$/ps	$\langle\tau\rangle_-$/ps
1	0	65.7	0	0
0.88	3.1	523	2998	4446
0.8	3.8	1579	9890	20 791
0.75	4.1	2754	19 457	35 432
0.66	4.4	7449	52 127	195 317
0.52	4.7	17 030	213 264	467 131

functions of the mole fraction as their logarithm follows a straight line. The corresponding linear fit functions are also shown and their analytic expressions are given in the figure caption. The respective figure and table for the zinc finger were already given in ref. 16 (see Table 5 and Fig. 13 of ref. 16). As the surface of ubiquitin is larger than that of the zinc finger migration of the solvent molecules on the surface and simultaneously staying in the first shell is facilitated. Therefore, one would expect that ubiquitin has a stronger retarding influence for the solvent molecules as compared to the zinc finger. This is confirmed by the comparison of Table 3, here, and Table 5 of ref. 16. Thus the relaxation of the first Voronoi shell of the zinc finger and the ubiquitin occurs on a similar timescale but the retardation increases with the size of the protein surface. In other words, the MRTs reflect—at first sight—the solvent's composition. A subtle influence of the protein solute can be observed however.

3.3 The protein itself: A regional view

So far we have documented that several properties are not a monotonic function of the ionic strength: first, one observes an initial rise when going from the pure aqueous environment to

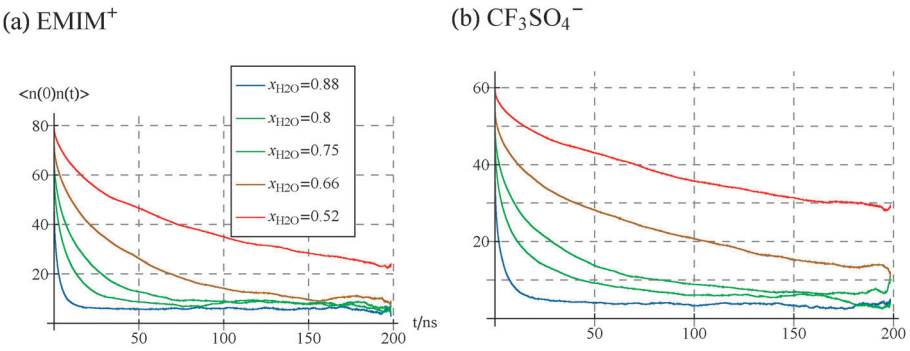


Fig. 9 Residence functions for $EMIM^+$ (a) and $CF_3SO_3^-$ (b) around 1UBQ for varying mole fractions.

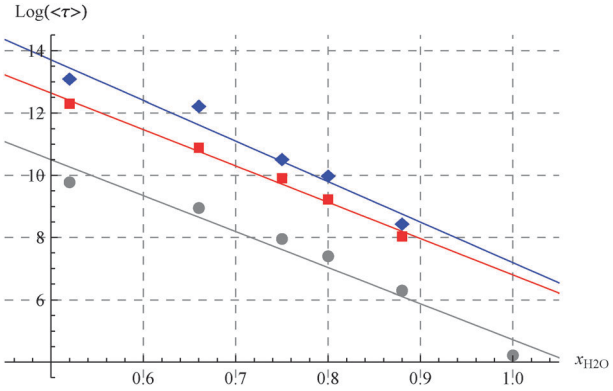


Fig. 10 Logarithm of MRTs as a function of mole fractions for 1UBQ. Straight lines have been fitted to the data points: $16.3 - 11.6x$ for H_2O in gray, $18.5 - 11.7x$ for $EMIM^+$ in red, and $20.2 - 13.0x$ for $CF_3SO_3^-$ in blue. For the respective curves and linear fits for 5ZNF we refer to Fig. 13 of ref. 16.

the diluted ionic solution. With increasing ionic strength a second characteristic point is found where most properties show a local minimum or maximum. For ubiquitin this characteristic point occurs at $c_{IL} = 4.1$ M and for the zinc finger at $c_{IL} = 3.1$ M. In a first analysis we could show that the relation of these two concentrations correlates with the respective Voronoi interactions surface areas (VISA) of the zinc finger motif in ubiquitin and the isolated zinc finger. An alternative view why this point appears at these concentrations will be presented in the subsequent section. Here, we want to clarify to what extent different regions of the proteins contribute to the phenomena at the characteristic point. To carry out this analysis we use a measure, the VISA, which is strictly additive from an atomistic level over individual amino acids, secondary structure elements, protein subunits, *etc*. *Vice versa*, the measure enables a decomposition of the protein into these building blocks. Since the Voronoi analysis is done by atomic tessellation it fulfils this criterion automatically. As an alternative to this static measure we will also use the root mean square fluctuations (RMSF) which reflect the dynamics of the protein. Of course, they are not additive but can be averaged over the building bocks.

As we are interested in the formation of the characteristic points and not in the absolute VISA, we subtracted the average value of each building block from its actual VISA. In this case averaging refers to the mean over all mole fractions.

In a first step of decomposition we calculated the total VISA of the amino acids in secondary and non-secondary structure elements. As expected, the amino acids in secondary structures are not responsible for the characteristic point. Therefore we further investigated the regions between the secondary structure elements as shown in Fig. 11. Here, the major deviations come from the loop of the β-hairpin (β-loop), the region linking β-strand 2 with the α-helix (β2-α), and the fraying ends (amino acids 72 to 76 for 1UBQ and 1, 28, 29, and 30 for 5ZNF). In the case of 5ZNF, the β2-α region consists of three amino acids only and therefore has a marginal influence. In both cases, however, the loop of the β-hairpin is the dominant player for the characteristic point (Fig. 11). The sum of the three major players in ubiquitin is also given and already displays the shape of total VISA. (The constant shift at low ionic strength stems from subunit 2). In the case of 5ZNF, the overall shape of the total VISA is already determined by the β-loop. In order to facilitate the identification of these important regions they are specifically marked in the two cartoons of Fig. 12.

Our regional analysis of the VISA performed so far can be extended in two directions. First, the VISA is a bilateral quantity involving both protein and solvent atoms sharing a common Voronoi surface. Therefore, it can be used to study the solvent specificity of selected amino acids. Amino acids may be grouped into positively and negatively charged, as well as polar and apolar residues and one may ask which of these groups is preferred by a solvent species $EMIM^+$, $CF_3SO_3^-$ or H_2O. Upon measuring the common surface between a residue out of a specific group of residues and a solvent species the following preferences are found: not unexpectedly, the anions share the largest VISA with positively charged residues. On the contrary, the cations do not follow the counter ion pattern. In fact, they seem to be rather promiscuous having high values of common VISA with polar, apolar and even positively charged amino acids. This behaviour can already be deduced from Fig. 1 where the anions reside at specific positions while the

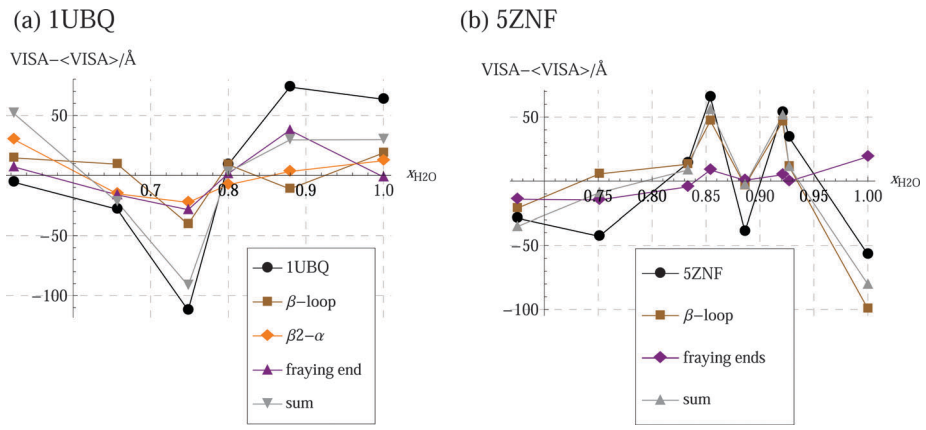


Fig. 11 The VISA for the three (two) most important regions in 1UBQ (5ZNF). Instead of absolute values the deviations from the mean value are given for a better comparison. By mean value the averaging over mole fractions is meant.

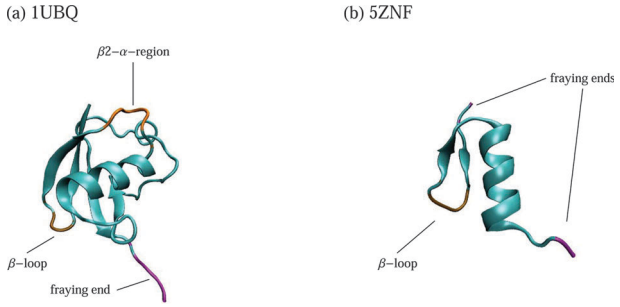


Fig. 12 Backbone structure of 1UBQ and 5ZNF.

cations roam over large areas of the protein surface. It further correlates with the higher mean residence time of anions as compared to cations shown in Fig. 9.

Second, the regional VISA may be finer grained by studying individual amino acids and even atoms. For example, we calculated the distribution of atomic VISA. Extracting that part of the distribution which refers to very low VISA (below 1 Å^2) one finds a pattern reciprocal to the total VISA (Fig. 2(a)): at the characteristic point the number of atoms with a marginal VISA is enhanced. Synchronously, the number of atoms with a medium VISA (20 to 30 Å^2) is reduced.

Finally, we want to contrast the static quantity, the VISA, with a dynamical one, the RMSF. Fig. 13 gives the RMSF averaged over the same regions as discussed above for the VISA. In agreement with our findings for the VISA, the average RMSF of the β-loop already resembles the shape of the overall RMSF. The β2-α region has a shallow minimum at the characteristic point. Even the fraying ends follow the general trend, although not unexpectedly, the respective amplitudes are high. However, upon averaging the RMSF over all three regions, β-loop, β2-α region, and fraying ends, their contribution is suppressed as they comprise only five amino acids. For the zinc finger analogous arguments hold. As for the VISA, the β-loop and the fraying ends together constitute the characteristic pattern of the total RMSF.

3.4 Mesoscopic view

Up to now the description was entirely on the atomistic or molecular level. The Voronoi method provides a natural way for coarse graining: a group or set of Voronoi polyhedra may be considered as one moiety. The straightforward choice for moieties would be the protein, its solvation shells and the remaining bulk. Along with this coarse graining of space charge distributions may be coarse grained as well by representing the electric anisotropy of a moiety by its dipole moment. Thus, the spatial decomposition into moieties implies a decomposition of the total dipole moment

$$M_{\mathrm{tot}} = \sum_i M_i \text{ with } i = \{1\mathrm{UBQ},\ \mathrm{shell},\ \mathrm{bulk}\} \qquad (4)$$

Its mean square, $\langle M_{\mathrm{tot}}^2 \rangle$ is related to the experimentally accessible static dielectric constant

$$\varepsilon - 1 = \frac{4\pi}{3Vk_{\mathrm{B}}T}\langle M_{\mathrm{tot}}^2 \rangle_{\mathrm{eq}} \qquad (5)$$

According to the decomposition eqn (4) the dielectric constant is made up of self-terms $\langle M_i^2 \rangle$ and cross-terms $\langle M_i \cdot M_j \rangle$

$$\langle M_{\mathrm{tot}}^2 \rangle = \sum_i \langle M_i^2 \rangle + 2\sum_{i \neq j} \langle M_i \cdot M_j \rangle \qquad (6)$$

which—at the mesoscopic level—represent the interaction within or between moieties.

From the Voronoi analysis we could see that the size of the systems studied here permits the undisturbed formation of at least three solvation shells. The dielectric contributions from the protein, the three solvation shells and the bulk together with the respective cross-terms are given in Table 4 for the two mole fractions $x_{\mathrm{H_2O}} = 1.0$ and $x_{\mathrm{H_2O}} = 0.88$. As the cross terms are below one and most of them even below one half we give only the self-terms for the higher mole fractions (see Table 5). The maximum change in dielectric properties is to be expected for the transition from the pure aqueous system to the ionic solution. Table 4 shows this transition in a threefold respect: the anti-correlation between the first shell and the protein changes to slightly positive values. The cross-term between the protein and the bulk almost vanishes. The correlation between neighbouring shells, in particular between the third shell and the bulk is strongly reduced. For the higher mole fractions the suppression of cross-terms increases. This is also shown by the last two columns of Table 5, trace (sum of self-terms) and total, where the trace approaches the dielectric constant with

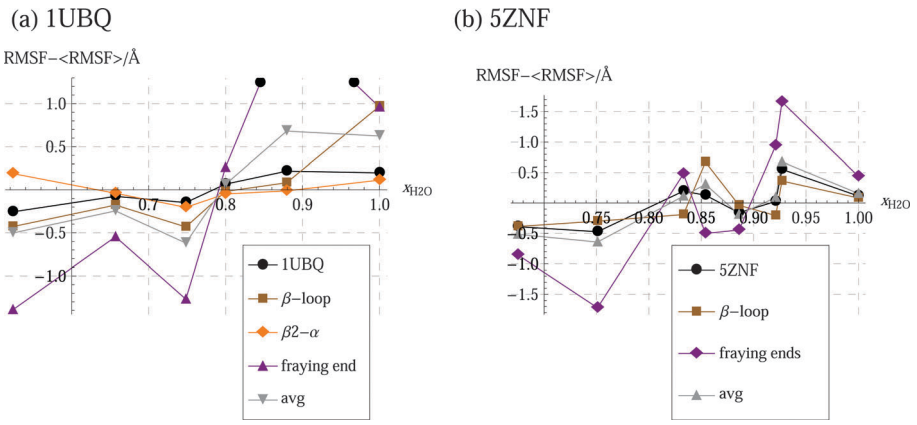


Fig. 13 The RMSF for the three (two) most important regions in 1UBQ (5ZNF). Instead of absolute values the deviations from the mean value are given for a better comparison. By mean value the averaging over mole fractions is meant.

Table 4 Spatially resolved contributions to the dielectric constant: self- and cross-terms. Above each table the sum of the self-terms ε_{diag} and the dielectric constant ε are given for comparison. ε_{diag} is incremented by one to permit a direct comparison with ε

$x_{H_2O} = 1.0$, $\varepsilon_{diag} + 1 = 105.8$, $\varepsilon = 143.9$

	1UBQ	1st shell	2nd shell	3rd shell	bulk
1UBQ	28.4	−2.	0	1.1	9.5
1st shell	−2.	2.6	0.9	0.2	0.5
2nd shell	0	0.9	3.7	1.5	2.3
3rd shell	1.1	0.2	1.5	5.5	4.9
Bulk	9.5	0.5	2.3	4.9	64.6

$x_{H_2O} = 0.88$, $\varepsilon_{diag} + 1 = 62.3$, $\varepsilon = 68.5$

	1UBQ	1st shell	2nd shell	3rd shell	bulk
1UBQ	35.3	0.5	0.6	−0.1	0.3
1st shell	0.5	1.7	0.4	0	0
2nd shell	0.6	0.4	4.	0.6	0
3rd shell	−0.1	0	0.6	6.5	0.9
Bulk	0.3	0	0	0.9	13.8

Table 5 Contributions of the dielectric self-terms for the protein, its three solvation shells and the bulk. Their sum is given in the previous to last column while the last column contains the total dielectric constant

H_2O	1UBQ	1st shell	2nd shell	3rd shell	Bulk	$\varepsilon_{diag} + 1$	ε
1.0	28.4	2.6	3.7	5.5	64.6	105.8	143.9
0.88	35.3	1.7	4.	6.5	13.8	62.3	68.5
0.80	32.	1.4	3.2	5.3	8.8	51.7	53.1
0.75	20.9	1.2	3.	4.8	6.9	37.7	38.
0.66	33.7	1.	2.5	4.2	5.4	47.8	47.7
0.52	24.	1.	2.1	3.5	3.9	35.5	34.1

increasing ionic strength. As Table 5 further shows, these self-terms of the shells and the bulk decrease monotonically with decreasing mole fraction. The protein is the only exception. At the characteristic point its dielectric contribution reaches a minimum. In other words, its higher sphericity goes along with a smaller electric anisotropy, *i.e.* with a lower mean squared dipole moment. As the protein makes the largest contribution to the dielectric constant for the nonaqueous systems this minimum becomes a characteristic feature of the dielectric constant as such (see Fig. 14).

Beyond this characteristic point the variation of the dielectric constant is small. In order to elucidate this behaviour we have decomposed the dielectric constant in an alternative way. Instead of discerning the protein, its solvation shells, and the bulk we have calculated separately, the mean square dipole moment of the protein, the molecular ions $EMIM^+$ and $CF_3SO_3^-$, and the water component. The corresponding dielectric contributions are summarised in Table 6 for mole fractions studied. This table shows: first, the contribution from the cations is marginal, in fact, it never exceeds one. Second, the contribution from the anions remains fairly constant for higher mole fractions. This can be only explained by the fact that the increasing number of ionic dipoles is compensated by their mutual anti-correlation. This compensation seems to be almost perfect. Third, of course the water component decreases with increasing ionic strength but the coupling within the anion–water network decreases remarkably slower. In fact, if one normalises the cross-terms between the anions and water to the water self-term, this ratio increases

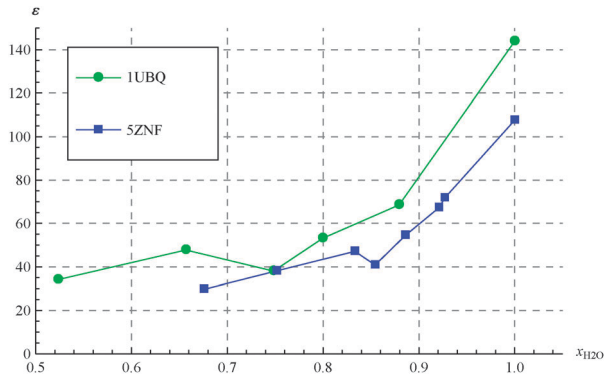


Fig. 14 Total dielectric constant as a function of mole fractions.

Table 6 Species specific contributions to the dielectric constant: self- and cross-terms. Above each table the sum of the self-terms $\varepsilon_{diag} + 1$ and the dielectric constant ε are given for comparison

$x_{H_2O} = 1.0$, $\varepsilon_{diag} + 1 = 126.5$, $\varepsilon = 143.9$

	1UBQ	$EMIM^+$	$CF_3SO_3^-$	H_2O
1UBQ	28.4	0	0	8.7
$EMIM^+$	0	0	0	0
$CF_3SO_3^-$	0	0	0	0
H_2O	8.7	0	0	97.1

$x_{H_2O} = 0.88$, $\varepsilon_{diag} + 1 = 61.6$, $\varepsilon = 68.5$

	1UBQ	$EMIM^+$	$CF_3SO_3^-$	H_2O
1UBQ	35.3	0.1	−0.1	1.3
$EMIM^+$	0.1	0.7	0.1	0.2
$CF_3SO_3^-$	−0.1	0.1	4.5	2.
H_2O	1.3	0.2	2.	20.1

$x_{H_2O} = 0.8$, $\varepsilon_{diag} + 1 = 49.2$, $\varepsilon = 53.1$

	1UBQ	$EMIM^+$	$CF_3SO_3^-$	H_2O
1UBQ	32	0	−0.8	0.9
$EMIM^+$	0	0.8	0.2	0.1
$CF_3SO_3^-$	−0.8	0.2	4.7	1.6
H_2O	0.9	0.1	1.6	10.8

$x_{H_2O} = 0.75$, $\varepsilon_{diag} + 1 = 35.0$, $\varepsilon = 38.0$

	1UBQ	$EMIM^+$	$CF_3SO_3^-$	H_2O
1UBQ	20.9	0	−0.3	0.2
$EMIM^+$	0	0.9	0.2	0.1
$CF_3SO_3^-$	−0.3	0.2	4.6	1.3
H_2O	0.2	0.1	1.3	7.6

$x_{H_2O} = 0.66$, $\varepsilon_{diag} + 1 = 44.8$, $\varepsilon = 47.7$

	1UBQ	$EMIM^+$	$CF_3SO_3^-$	H_2O
1UBQ	33.7	0.2	−1.1	1.2
$EMIM^+$	0.2	1.	0.2	0.1
$CF_3SO_3^-$	−1.1	0.2	4.7	0.9
H_2O	1.2	0.1	0.9	4.4

$x_{H_2O} = 0.52$, $\varepsilon_{diag} + 1 = 33.2$, $\varepsilon = 34.1$

	1UBQ	$EMIM^+$	$CF_3SO_3^-$	H_2O
1UBQ	24.	−0.1	−0.6	0.3
$EMIM^+$	−0.1	1.	0.4	0
$CF_3SO_3^-$	−0.6	0.4	5.	0.5
H_2O	0.3	0	0.5	2.2

monotonically from 0.1 to 0.23. This demonstrates the orienting influence of the anions on the water dipoles. As opposed to an ionic solution where the diluted ions also have some directing influence but the majority of water molecules is still subjected to the rules of the water hydrogen bonded network, all water molecules are now under the complete influence of the ions. In other words, the solvent has changed from an ionic solution to a hydrated ionic melt. Therefore, in dielectric terms, the characteristic point found in this paper marks the change from an ionic solution to a hydrated ionic melt. The characteristic point already found in our previous paper (ref. 16) is typical for a change from a pure aqueous system to an ionic solution. Of course this turning point is also found for the ubiquitin systems studied here, although at higher ionic strength. Therefore, its appearance is less pronounced.

4 Summary and conclusion

The focus of this study is on the solvation of the protein ubiquitin by hydrated ionic liquids for a diversity of mole fractions or ionic strengths. For comparison results of an analogous previous study of the solvated zinc finger (5ZNF) are also discussed here in order to find general principles. Our analyses are based on simulation trajectories covering typically 200 nanoseconds. Therefore, our findings refer to this time window. The system was analysed at four levels of resolution: first, properties representative for the entire protein like the protein–solvent interface surface (VISA), the energy, or the root mean squared deviation and fluctuation, were interpreted. Second, the first solvation shell of the protein was characterised in terms of coordination numbers and mean residence times. Third, certain regions of the protein were tracked down to be responsible for the characteristic features found at the two previous levels. Fourth, the characterisation of the whole system and its components was done at the mesoscopic level: the atomistic charge distributions were replaced by a dielectric polarisation, *i.e.* by the net dipole moments of a component divided by the system's volume. Two alternative ways of decomposition were done: protein, several solvation shells and bulk, or, species specific, protein, cations, anions and water.

Considering the above-mentioned properties as a function of the mole fraction or ionic strength two characteristic transition points can be figured out. The first one refers to the transition from the pure aqueous system to the ionic solution. It was already found for the zinc finger in our previous study[16] and was also confirmed for the ubiquitin systems studied here. The second characteristic point corresponds to the transition from the ionic solution to the hydrated ionic melt. These transitions can be best elucidated at the mesoscopic dielectric level discussed below.

Starting from the visual behaviour of the protein's Voronoi surface a couple of properties was investigated thereby gradually unfolding the level of resolution. It turned out that the first half of ubiquitin—called subunit 1 in the paper—does not only look like a zinc finger motif but also behaves this way, analogously to the isolated zinc finger. As part of the Voronoi surface of this zinc finger unit in ubiquitin is blocked by the second half of the molecule the phenomena observed for the isolated zinc finger are observed at a higher ionic strength. In fact, the ratio between the ionic strengths where the characteristic point occurs scales with the inverse ratio of the respective Voronoi surfaces.

When considering several properties as a function of the ionic strength the second characteristic point appears as a minimum or maximum. Our regional analysis shows that only a small group of amino acids is responsible for this phenomenon. They are all located in random coils or turns and not in ordered secondary structures which seem to be stable enough to resist the attack of the molecular ions. In fact, the loop linking the two first β-strands in the zinc finger motif, either isolated or as part of ubiquitin, is the major player. It is assisted by the region linking the end of the second β-strand with the α-helix. To a minor degree the fraying ends have an influence, too.

The relation between the appearance of the two characteristic points and the composition of the system as described by the mole fraction or ionic strength becomes obvious from the dielectric behaviour of the system. From the pure aqueous system to the ionic solution (first transition point) the dielectric constant decays steeply with decreasing water content. In other words, the amount of water is reduced but the water network can still exert its orientating influence on water dipoles. Once the amount of water has fallen below some certain threshold (second transition point) the remaining water component is completely under the control of the molecular ions, in particular the anions. Thus the orientation of the water molecules is "frozen" and their dielectric contribution is almost constant. One might think that the $CF_3SO_3^-$ dipoles should take over the role of dipolar orientation. However, our study shows that the contribution of the $CF_3SO_3^-$ dipoles stays constant as well. In other words, the increasing number of ionic dipoles is almost perfectly compensated by their antiparallel alignment. In summary, we state that two transition points are caused by the competition of the water hydrogen network and the ionic network. First there is a change from a pure aqueous system to an ionic solution which then changes to a hydrated ionic melt which completely overrules water dipolar correlations.

Acknowledgements

This work was performed on the "Vienna Scientific Cluster" [www.zid.tuwien.ac.at/vsc] of the University of Vienna, the Vienna University of Technology, and the University of Natural Resources and Applied Life Science Vienna. We thank for generous allocation of computer time. This work was supported by Project No. P19807 of the FWF Austrian Science Fund.

References

1 H. Olivier-Bourbigou, L. Magna and D. Morvan, Ionic liquids and catalysis: Recent progress from knowledge to applications, *Appl. Catal., A*, 2010, **373**(1–2), 1.
2 M. Moniruzzaman, K. Nakashima, N. Kamiya and M. Goto, Recent advances of enzymatic reactions in ionic liquids, *Biochem. Eng. J.*, 2010, **48**, 295–314.

3 C. Roosen, P. Müller and L. Greiner, Ionic liquids in biotechnology: applications and perspectives for biotransformations, *Appl. Microbiol. Biotechnol.*, 2008, **81**, 607–614.
4 J. Gorke, F. Srienc and R. Kazlauskas, Toward advanced ionic liquids, polar, enzyme-friendly solvents for biocatalysis, *Biotechnol. Bioprocess Eng.*, 2010, **15**(1), 40–53.
5 B. Halle, Protein hydration dynamics in solution: a critical survey, *Philos. Trans. R. Soc. London, Ser. B*, 2004, **359**, 1207–1223.
6 A. M. Klibanov, Improving enzymes by using them in organic solvents, *Nature*, 2001, **409**, 241–246.
7 D. Constatinescu, C. Herrmann and H. Weingärtner, Patterns of protein unfolding and protein aggregation in ionic liquids, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1756–1763.
8 T. A. Page, N. D. Kraut, P. M. Page, G. A. Baker and F. V. Bright, Dynamics of loop 1 of domain I in human serum albumin when dissolved in ionic liquids, *J. Phys. Chem. B*, 2009, **113**, 12825–12830.
9 Y. Akdogan, M. J. N. Junk and D. Hinderberger, Effect of ionic liquids on the solution structure of human serum albumin, *Biomacromolecules*, 2011, **12**(4), 1072.
10 T. G. A. Youngs, J. D. Holbrey, M. Deetlefs, M. Nieuwenhuyzen, M. F. C. Gomes and C. Hardacre, A molecular dynamics study of glucose solvation in the ionic liquid 1,3-dimethylimidazolium chloride, *ChemPhysChem*, 2006, **7**, 2279–2281.
11 N. M. Micaêlo and C. M. Soares, Protein structure and dynamics in ionic liquids. Insights from molecular dynamics simulation studies, *J. Phys. Chem. B*, 2008, **112**, 2566–2572.
12 M. Klähn, G. S. Lim, A. Seduraman and P. Wu, On the different roles of anions and cations in the solvation of enzymes in ionic liquids, *Phys. Chem. Chem. Phys.*, 2011, **13**, 1649–1662.
13 G. Neumayr, C. Schröder and O. Steinhauser, Relaxation of Voronoi shells in hydrated molecular ionic liquids, *J. Chem. Phys.*, 2009, **131**, 174509.
14 C. Schröder, G. Neumayr and O. Steinhauser, On the collective network of ionic liquid/water mixtures. III. Structural analysis of ionic liquids on the basis of Voronoi decomposition, *J. Chem. Phys.*, 2009, **130**, 194503.
15 G. Neumayr, T. Rudas and O. Steinhauser, Global and local Voronoi analysis of solvation shells of proteins, *J. Chem. Phys.*, 2010, **133**, 084108.
16 M. Haberler, C. Schröder and O. Steinhauser, Solvation studies of a zinc finger protein in hydrated ionic liquids, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6955–6969.
17 C. Schröder, T. Rudas, S. Boresch and O. Steinhauser, Simulation studies of the protein–water interface. I. Properties at the molecular resolution, *J. Chem. Phys.*, 2006, **124**, 234907.
18 T. Rudas, C. Schröder, S. Boresch and O. Steinhauser, Simulation studies of the protein–water interface. II. Properties at the mesoscopic resolution, *J. Chem. Phys.*, 2006, **124**, 234908.
19 L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, Packmol: A package for building initial configurations for molecular dynamics simulations, *J. Comput. Chem.*, 2009, **30**(13), 2157–2164.
20 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B*, 1998, **102**, 3586.
21 C. Schröder and O. Steinhauser, Using fit functions in computational dielectric spectroscopy, *J. Chem. Phys.*, 2010, **132**, 244109.
22 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.*, 1977, **23**, 327–341.
23 X. Huang, C. J. Margulis, Y. Li and B. J. Berne, Why is the partial molar volume of CO_2 so small when dissolved in a room temperature ionic liquid? Structure and dynamics of CO_2 dissolved in [bmim+] [pf6(−)], *J. Am. Chem. Soc.*, 2005, **127**, 17842–17851.
24 C. Schröder, T. Rudas, G. Neumayr, S. Benkner and O. Steinhauser, On the collective network of ionic liquid/water mixtures: I. Orientational structure, *J. Chem. Phys.*, 2007, **127**, 234503.
25 C. Schröder, J. Hunger, A. Stoppa, R. Buchner and O. Steinhauser, On the collective network of ionic liquid/water mixtures. II. Decomposition and interpretation of dielectric spectra, *J. Chem. Phys.*, 2008, **129**, 184501.

5 Hydrated Ionic Liquids With and Without Solute: The Influence of Water Content and Protein Solutes

Hydrated Ionic Liquids With and Without Solute: The Influence of Water Content and Protein Solutes

Michael Haberler, Christian Schröder and Othmar Steinhauser

University of Vienna, Department of Computational Biological Chemistry, Austria

(Dated: April 10, 2012)

Abstract

Computational studies of the ternary hydrated ionic liquid 1-ethyl-3-methylimidazolium, trifluoromethanesulfonate, and water, are presented for a diversity of water mole fractions. For the analysis novel tools are applied:

Radial distribution functions are decomposed into contributions from different Voronoi shells and generalised to include averaged angular functions.

The applicability of hydrodynamic relations for single particle motion are tested. Viscosity scaling is found to be valid for the translation of molecular species and the reorientation of their dipolar axis.

Using the body-fixed frame as an alternative reference system translation and rotation can be decomposed into contributions along and about the axes of a well defined orthogonal trihedron. When decomposing the overall translational and rotational diffusion coefficient the inclusion of the metric is essential.

A relation between single particle and collective dynamics is established: For pure systems average relaxation times can be scaled by the static collective Kirkwood order parameter. In case of systems containing additional solutes collective dynamics has to be corrected excluding solvent molecules bound to the solute.

The scaling of collective dynamics by a static order parameter encouraged us to present dielectric spectra even in a regime exceeding the time window covered by our simulation runs.

I. INTRODUCTION

Since the pioneering work of Hanke and Lynden-Bell[8] a large body of computer simulations of molecular ionic liquids (IL) has established[3,4,7,17,21,23,34,36] For ionic liquids mixed with water, so-called hydrated ionic liquids, simulation studies are rare[11,25,29,32,35]. Computational studies of solutes dissolved in hydrated ILs are extremely rare[5,13,16,22]. In this paper we collect a plethora of simulation studies of hydrated IL which span from almost aqueous systems to those with upcoming features of pure ILs without water. Thereby we will not follow the classical route of presenting thermodynamic data, densities, radial distribution functions (RDF), single particle dynamics, and maybe viscosity, but we will focus on a couple of major issues:

First, anisotropy in shape and charge distribution of the underlying molecular species will play a major role in our analysis. For example, we will present angular and dipolar RDFs in addition to simple RDFs. Furthermore, we will decompose translational and rotational single particle dynamics in the body-fixed frame. A second major issue will be collectivity. We will test the Kievelson-Madden relation linking single particle and collective reorientational dynamics. We will also interpret the rotational part of the dielectric spectra. Finding out the limits of hydrodynamic rules for the interpretation of molecular motion will be a third issue. Last but not least, we will analyse the impact and influence of protein solutes on the above mentioned properties.

II. THEORY

A. Characterisation of Molecular Motion

The motion of an atom α in a molecule i, may be water (TIP3), 1-ethyl-3-methylimidazolium (EMIM), trifluoromethanesulfonate (TRIF), or protein, can be decomposed into translation, rotation and internal motion. With respect to its velocity $\vec{v}_{i,\alpha} = \mathrm{d}/\mathrm{d}t\, \vec{r}_{i,\alpha}$ this means

$$\vec{v}_{i,\alpha} = \vec{v}_i^{\,\mathrm{com}} + \vec{\omega}_i \times (\vec{r}_{i,\alpha} - \vec{r}_i^{\,\mathrm{com}}) + \vec{v}_{i,\alpha}^{\,\mathrm{int}} \tag{1}$$

where we have introduced the centre of mass velocity, $\vec{v}_i^{\,\mathrm{com}} = \mathrm{d}/\mathrm{d}t\, \vec{r}_i^{\,\mathrm{com}}$, and the angular velocity, $\vec{\omega}_i$, which characterise the motion of the molecule as a whole. The supplementary internal velocity, $\vec{v}_{i,\alpha}^{\,\mathrm{int}}$, stands for the non-rigidity of the molecule. In order to facilitate

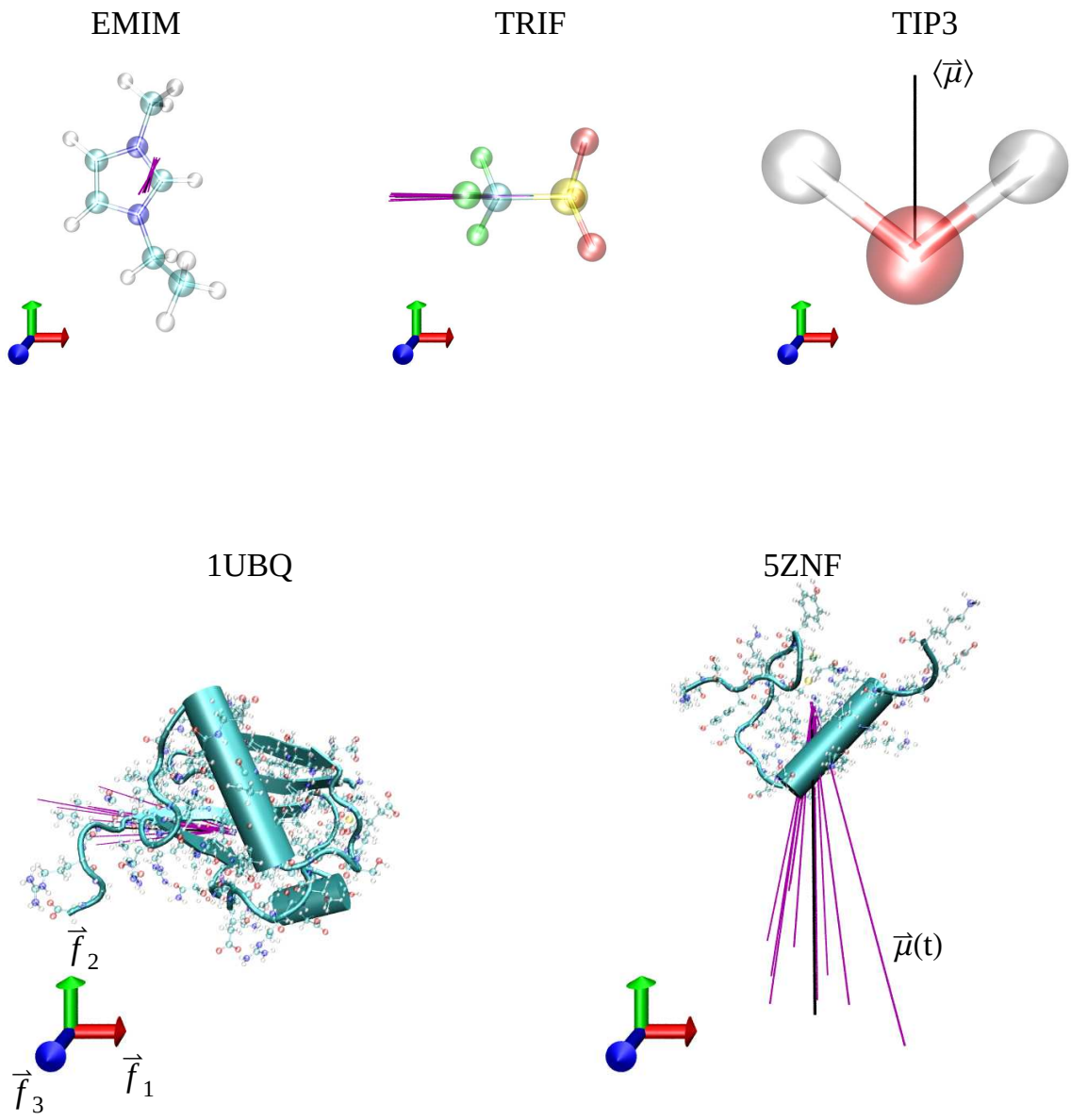


FIG. 1. Reference conformations of the three solvent molecules (EMIM, TRIF, and TIP3) as well as the two solutes (5ZNF and 1UBQ). The trihedron defining the body-fixed frame ($\vec{f}_1$ red, $\vec{f}_2$ green, $\vec{f}_3$ blue) is included; the origin of each body-fixed frame is located at the centre of mass of the molecule. Additionally, the mean dipole moment ($\langle\vec{\mu}\rangle$) for each molecule is given as a black line, and the dipole moment at ten random time steps ($\vec{\mu}(t)$) is given in purple. For the solvent molecules the dipole moment is scaled by three and the unit is $e\mathring{A}$.

our motional analysis we consider the motion of pseudo-rigid molecules. Their rotation is obtained by the least-squares fit[10] of the actual molecular geometry to a reference molecule whose axes $\vec{f}_1, \vec{f}_2, \vec{f}_3$ (see Figure 1) are aligned along the axes $\vec{e}_1, \vec{e}_2, \vec{e}_3$ of the laboratory frame. These body-fixed axes are the principal axes of inertia. The respective operator $\hat{R}(t)$ which performs the rotation

$$\hat{R}(t)\vec{f}_k(t) = \vec{e}_k \tag{2}$$

corresponds to a representation matrix, $R(t)$, transforming the coordinates (v_1, v_2, v_3) of an arbitrary vector $\vec{v}$ in the laboratory frame to the coordinates (u_1, u_2, u_3) in the body-fixed frame:

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \tag{3}$$

Since $R(t)$ is a rotation matrix its transpose, $R^t(t)$, is also its inverse. Both matrices are built up by the coordinates of the body-fixed axes $\vec{f}_1, \vec{f}_2, \vec{f}_3$ with respect to the laboratory frame, i.e. $\vec{f}_1 = (R_{11}, R_{12}, R_{13})$, $\vec{f}_2 = (R_{21}, R_{22}, R_{23})$, and $\vec{f}_3 = (R_{31}, R_{32}, R_{33})$. This can be shown by inserting the coordinates into the general relation Equ. 3.

The concept of a transformation matrix, $T(t)$ greatly facilitates the computation of the angular velocity circumventing the usual problems associated with angles. Quite generally, the time evolution of an arbitrary vector $\vec{r}(t)$ is given by

$$\vec{r}(t) = T(t)\,\vec{r}(0) \tag{4}$$

. In case of a pure rotation the velocity can be expressed by

$$\vec{v}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\,\vec{r}(t) = \frac{\mathrm{d}T(t)}{\mathrm{d}t} \cdot \vec{r}(0) = \vec{\omega}(t) \times \vec{r}(t) \tag{5}$$

Rewriting the cross product as

$$\vec{\omega}(t) \times \vec{r}(t) = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} \cdot \vec{r}(t) = \Omega(t) \cdot \vec{r}(t) \tag{6}$$

and $\vec{r}(0) = T^t(t)\,\vec{r}(t)$, it follows from Equ. 5 that the matrix, $\Omega(t)$, containing the coordinates of the angular velocity with the laboratory frame, is given by

$$\Omega(t) = \frac{\mathrm{d}T(t)}{\mathrm{d}t} \cdot T^t(t) \tag{7}$$

. Multiplying the identity

$$R(0)\,\vec{f}_k(0) = R(t)\,\vec{f}_k(t) = e_k \tag{8}$$

by $R^t(t)$ from the left side

$$\vec{f}_k(t) = \underbrace{R^t(t)\,R(0)}_{T(t)}\,\vec{f}_k(0) \tag{9}$$

the matrix, $\Omega(t)$, is finally given by

$$\Omega(t) = \frac{\mathrm{d}R^t(t)}{\mathrm{d}t}\,R(0)\,R^t(0)\,R(t) = \frac{\mathrm{d}R^t(t)}{\mathrm{d}t}\,R(t) \tag{10}$$

In this derivation we have used the fact that for a pure rotation an arbitrary vector $\vec{r}(t)$ has constant coordinates, s_k, with respect to the body-fixed axes $\vec{f}_k(t)$, i.e. $\vec{r}(t) = \sum_k s_k \vec{f}_k(t)$. Therefore, the transformation matrix of Equ. 9 applies to an arbitrary vector as well.

The off-diagonal elements of the matrix $\Omega(t)$ yield the coordinates $(\omega_1, \omega_2, \omega_3)$ of the angular velocity $\vec{\omega}(t)$ in the laboratory frame. By means of Equ. 3 it may be readily converted to body-fixed coordinates $(\gamma_1 = d\phi_1/dt, \gamma_2 = d\phi_2/dt, \gamma_3 = d\phi_3/dt)$ which characterise the angular velocity of rotation about the body-fixed axes $\vec{f}_1, \vec{f}_2, \vec{f}_3$, respectively.

For ionic liquids molecular anisotropy of shape and charge distribution is essential. In order to figure out the impact of this anisotropy on the dynamics we project the linear and angular velocity onto the body-fixed axes:

$$\vec{v}(t) = \sum_k u_k(t)\vec{f}_k(t) \tag{11}$$

$$\vec{\omega}(t) = \sum_k \gamma_k(t)\vec{f}_k(t) \tag{12}$$

This implies for the respective time auto-correlation functions

$$\langle \vec{v}(0) \cdot \vec{v}(t) \rangle = \sum_k \sum_l \langle u_k(0) u_l(t) M_{kl}(t) \rangle \tag{13}$$

$$\langle \vec{\omega}(0) \cdot \vec{\omega}(t) \rangle = \sum_k \sum_l \langle \gamma_k(0) \gamma_l(t) M_{kl}(t) \rangle \tag{14}$$

not only the auto- and cross-correlation of the coordinates u_k and γ_k, but also the relaxation of the "metric"

$$M_{kl}(t) = \vec{f}_k(0) \cdot \vec{f}_l(t) \tag{15}$$

At time t=0 the metric equals the unit matrix. For intermediate times off-diagonal elements appear and diagonal elements decrease. At sufficiently long times the metric approaches the zero matrix.

For experimental reasons, the dipole moment vector, $\vec{\mu}(t)$, is an important probe of molecular rotation. It can be also represented as a linear combination of body-fixed axes:

$$\vec{\mu}(t) = \sum_k \mu_k \vec{f}_k(t) \tag{16}$$

The coordinates μ_k are independent of time for a pseudo-rigid molecule. Therefore we have for the dipolar autocorrelation function

$$\langle \vec{\mu}(0)\vec{\mu}(t)\rangle = \sum_k \sum_l \mu_k \mu_l \langle M_{kl}(t)\rangle \tag{17}$$

. In principle, the matrix of the metric contains nine elements, but when taking the time correlation functions $\langle M_{kl}(t)\rangle$ pairs of off-diagonal elements become statistically equivalent. We have accounted for this by taking the sum $\langle M_{kl}(t)\rangle + \langle M_{lk}(t)\rangle = \langle \bar{M}_{kl}(t)\rangle$.

Integrating Equ. 13 and Equ. 14 over time yields the translational

$$3D^{\mathrm{trans}} = \int_0^\infty \langle \vec{v}(0)\vec{v}(t)\rangle \mathrm{d}t = \sum_k \sum_l D_{kl}^{\mathrm{trans}} \tag{18}$$

and rotational

$$3D^{\mathrm{rot}} = \int_0^\infty \langle \vec{\omega}(0)\vec{\omega}(t)\rangle \mathrm{d}t = \sum_k \sum_l D_{kl}^{\mathrm{rot}} \tag{19}$$

diffusion coefficients as a superposition of the elements of the "diffusion tensors",

$$D_{kl}^{\mathrm{trans}} = \int_0^\infty \langle u_k(0)u_l(t)M_{kl}(t)\rangle \mathrm{d}t \tag{20}$$

and

$$D_{kl}^{\mathrm{rot}} = \int_0^\infty \langle \gamma_k(0)\gamma_l(t)M_{kl}(t)\rangle \mathrm{d}t \tag{21}$$

. As a first approximation the complete correlation function may be replaced by the product of the coordinate correlation function and that of the metric

$$\bar{D}_{kl}^{\mathrm{trans}} = \int_0^\infty \langle u_k(0)u_l(t)\rangle \langle M_{kl}(t)\rangle \mathrm{d}t \tag{22}$$

$$\bar{D}_{kl}^{\mathrm{rot}} = \int_0^\infty \langle \gamma_k(0)\gamma_l(t)\rangle \langle M_{kl}(t)\rangle \mathrm{d}t \tag{23}$$

. Under the assumption that the coordinates, u_k, γ_k, and the metric, M_{kl} relax on completely different timescales the expression for the diffusion tensors may be simplified to

$$\tilde{D}_{kl}^{\mathrm{trans}} = \int_0^\infty \langle u_k(0)u_l(t)\rangle \mathrm{d}t \tag{24}$$

and

$$\tilde{D}_{kl}^{\mathrm{rot}} = \int_0^\infty \langle \gamma_k(0)\gamma_l(t)\rangle \mathrm{d}t \tag{25}$$

. A further simplification often used is the neglect of the off-diagonal elements

$$\tilde{D}_{kl} = \tilde{D}_{kk}\delta_{kl} \tag{26}$$

. While the relaxation of the metric characterises the relaxation of the body-fixed axes the angular velocity autocorrelation function in the body-fixed frame describes rotation about the body-fixed axes. The cumulant expansion developed by Kubo[15] provides a relationship between these two types of rotational motion

$$ln\langle M_{kk}(t)\rangle = -\sum_{l\neq k}\int_0^t \int_0^t \langle \gamma_l(t')\gamma_l(t'')\rangle \mathrm{d}t'\mathrm{d}t'' \tag{27}$$

(cf. Equ. 8 of Ref. 19). Using the translational symmetry in time this relation may be rewritten

$$ln\langle M_{kk}(t)\rangle = -\sum_{l\neq k}\int_0^t \mathrm{d}t' \int_0^t \langle \gamma_l(0)\gamma_l(s)\rangle \mathrm{d}s = -t\sum_{l\neq k}\tilde{D}_{ll}^{\mathrm{rot}}(t) \tag{28}$$

. For times t exceeding the relaxation time τ_γ of the angular velocity correlation function, $\tilde{D}_{ll}^{\mathrm{rot}}(t)$ approaches a constant value such that $\langle M_{kk}(t)\rangle$ decays exponentially with a relaxation time τ_k. In this limit we then have

$$\tau_k = \frac{1}{\sum_{l\neq k}\tilde{D}_{ll}^{\mathrm{rot}}} \tag{29}$$

. Integrating the angular velocity and disregarding periodicity

$$\Delta\phi_k(t) = \int_0^t \gamma_k(s)\mathrm{d}s \tag{30}$$

we get the angle swept through during time t[26]. The corresponding mean square displacement (MSD) $\langle\Delta\phi_k^2(t)\rangle$ is the rotational analogue of the translational MSD. According to this analogy the angular MSD may be also related to the time autocorrelation functions of the respective velocities[2]

$$\langle\Delta\phi_k^2(t)\rangle = 2t\int_0^t \langle\gamma_k(0)\gamma_k(s)\rangle\,\mathrm{d}s - 2\int_0^t s\cdot\langle\gamma_k(0)\gamma_k(s)\rangle\,\mathrm{d}s \tag{31}$$

. For sufficiently long times when the correlation function and, even more important, its product with time have levelled off the MSD fulfils a linear relationship such that a diffusion coefficient may be derived from the linear slope.

$$\tilde{D}_{kk}^{\mathrm{rot}} = \frac{1}{2}\lim_{t\to\infty}\frac{\mathrm{d}}{\mathrm{d}t}\langle\Delta\phi_k^2(t)\rangle = \int_0^t \langle\gamma_k(0)\gamma_k(s)\rangle\,\mathrm{d}s \tag{32}$$

In the last equation we have explicitly indicated the axis about which rotation occurs.

Since the pioneering work of Alder *et al.*[1] we are familiar with the interpretation of molecular diffusion coefficients by hydrodynamic relations

$$\frac{1}{D^{\mathrm{trans}}} = \frac{6\pi r}{k_B T}\,\eta \tag{33}$$

$$\frac{1}{D^{\mathrm{rot}}} = \frac{8\pi r^3}{k_B T}\,\eta \tag{34}$$

which involve the shear viscosity η and the hydrodynamic radius r. Combining the result of the cumulant expansion (Equ. 29) with the respective hydrodynamic formula (Equ. 34) "viscosity scaling" is also found for the relaxation time

$$\tau \sim \frac{8\pi r^3}{k_B T}\,\eta \tag{35}$$

For liquids composed of neutral molecules application of these hydrodynamic formulae provides realistic hydrodynamic radii. For ionic liquids, however, the so obtained values are to small as shown for example in Ref. 33 but viscosity scaling is conserved[31]. In other words, dynamics of one system may be converted to that of another system by simple viscosity scaling. We will follow this approach in Section IV B.

Thereby, the viscosity will be computed either from the Green-Kubo relation[14]

$$\eta = \frac{V}{k_B T}\int_0^\infty \langle P_{\alpha\beta}(0) P_{\alpha\beta}(t)\rangle \mathrm{d}t \tag{36}$$

where $P_{\alpha\beta}$ are the off-diagonal element of the stress tensor, or from the Helfand moment[9]

$$\eta = \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t}\langle [G_{\alpha\beta}(0) - G_{\alpha\beta}(t)]^2\rangle \tag{37}$$

where $G_{\alpha\beta}$ is the time integral of $P_{\alpha\beta}$.

III. METHODS

The raw data for the subsequent analysis are based on eight molecular dynamics (MD) simulations of a mixture of water (TIP3), 1-ethyl-3-methylimidazolium (EMIM), and trifluoromethanesulfonate (TRIF) (without solute, subsequently referred as pure system). We further analyse 7 (5) simulations of the zinc finger protein (5ZNF) as well as 6 of the ubiquitin protein (1UBQ) immersed in the same hydrated ionic liquid.

x_{H_2O}	$c_{H_2O}/\mathrm{mol\,l^{-1}}$	$c_{IL}/\mathrm{mol\,l^{-1}}$	#EMIM	#TRIF	#TIP3	$L/\mathring{A}$	$\rho_{mass}/\mathrm{g\,l^{-1}}$
0.97	43.87	1.18	200	200	7428	65.51	1097.8
0.93	32.15	2.34	400	400	5489	65.69	1189.06
0.9	26.2	2.93	500	500	4467	65.66	1235.34
0.85	20.24	3.53	600	600	3444	65.62	1282.22
0.79	14.87	4.06	700	700	2561	65.89	1325.25
0.71	10.66	4.46	770	770	1841	65.95	1352
0.65	8.76	4.66	800	800	1504	65.81	1370.76
0.43	3.86	5.12	900	900	678	66.34	1401.54

TABLE I. Note: Decreasing mole fraction, x_{H_2O}, is synonymous with increasing ionic strength, c_{IL}. The relation is monotonic but nonlinear. ρ_{mass} is the respective density and L the edge length of the cubic box. Parameters are given for the composition of the pure systems. The parameters for the systems with protein solutes, 5ZNF and 1UBQ, are collected in Table 1 in Ref. 5 and Table 1 in Ref. 6, respectively.

The simulation details for the protein systems have been already given in previous studies[5,6], while those of the pure systems are described below.

Table I collects the respective water mole fractions, ionic strengths, water concentrations, number of EMIM, TRIF, and water molecules, the edge length of the cube and the mass density of the simulation box.

The starting configurations of the system were minimised with the steepest descent method for 200 steps. A first equilibration in the NPT-ensemble covered one nanosecond. A second one in the NVT-ensemble lasted half a nanosecond. The average edge lengths of the second half of the initial equilibration was used as the constant value for the NVT-ensemble. The second equilibration was followed by the production run of about 50 nanoseconds at a temperature of 300K. The respective time step was 2 femtoseconds. The atomic forces were calculated in the same way as described at the end of the Method section A in Ref. 5.

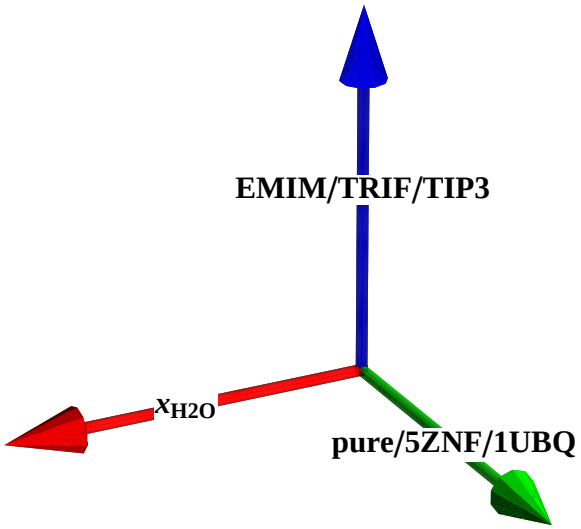


FIG. 2. Three dimensional analysis: (1) water mole fraction x_{H_2O}/ionic strength, (2) pure/5ZNF/1UBQ hydrated-IL systems, (3) solvent molecules, EMIM, TRIF, and TIP3.

IV. RESULTS AND DISCUSSION

As depicted in Figure 2 the presentation and discussion of our results will follow a "three dimensional" route depending on whether the mole fraction of water (or equivalently, the ionic strength) varies, or the type of the solute (none, 5ZNF, or 1UBQ) is changed, or a specific solvent species (TIP3, TRIF, or EMIM) is considered. Thus we have a 8x3x3-manifold of system parameters. Of course, we will not discuss the complete manifold but rather selected combinations of system parameters.

This section is organised as follows: First, we will discuss the packing of solvent molecules as a function of ionic strength and solute type. Next, we thoroughly investigate the translational and rotational single particle dynamics with respect to two reference frames: the laboratory frame (LF) and the body-fixed frame (BF). Fourth, we will analyse the collective orientational structure and dynamics probing the relation between single particle and collective motion. Finally, we will present the rotational dielectric spectra.

A. Packing

If the systems were composed of atomic ions the packing would be described by the cation-cation, cation-anion, anion-anion radial distribution function (RDF) only. Considering molecular ions the RDF between the respective centres of mass (COM)

$$g^{000}(r) = \frac{1}{4\pi\rho r^2 N} \sum_i \sum_{j \neq i} \langle 1 \cdot \delta(\vec{r} - \vec{r}_{ij}) \rangle \tag{38}$$

is the appropriate analogue (shown in the first row of Figure 3). The alternating layers of positively and negatively charged ions is usually described by the charge ordering function

$$Q(r) = g^{000}_{\mathrm{EMIM-EMIM}}(r) + g^{000}_{\mathrm{TRIF-TRIF}}(r) - 2\, g^{000}_{\mathrm{EMIM-TRIF}}(r) \tag{39}$$

depicted in Figure 4. Following intuition increasing water content should lower the peak heights and extend the structural oscillations. In order to rationalise this we have fitted the charge ordering function to the expression[26]

$$Q_{\mathrm{fit}}(r) = \frac{A}{r}\, e^{-\frac{r}{\sigma}}\, \sin\left(\frac{2\pi r}{\lambda} + \phi\right) \tag{40}$$

in the distance range from $r_{\min} = 4\mathring{A}$ to $r_{\max} = 30\mathring{A}$. For the phase shift a constant value of $\phi = 24.82$ was used. The amplitudes, A, the damping factor, $1/\sigma$, and the "wave length", λ, are listed in Table II. Without any exception the parameter sets follow intuition: Damping factor and "wave length" decrease monotonically as a function of ionic strength. In physical terms, reduced damping means that the ordering of charge layers extends over longer range. The direct influence of electrostatic forces becomes more important as the screening effect of water dipoles is weakened. The decreasing "wave length" may be interpreted as the contraction of charge layers reducing their thickness. Of course the behaviour of the charge ordering function can be traced back to the individual RDFs. The screening effect of water molecules is also clearly visible in $g^{000}_{\mathrm{EMIM-TRIF}}$ and $g^{000}_{\mathrm{TRIF-TRIF}}$. Simultaneously, weaker attraction between cation and anion results in a slight shift towards longer distances. The weaker repulsion between anions results in a slight shift to shorter COM distances. As these two RDFs enter the charge ordering function, $Q(r)$, with a total weight of three as compared to $g^{000}_{\mathrm{EMIM-EMIM}}$ the observed screening effects are conserved despite of the counterintuitive behaviour of $g^{000}_{\mathrm{EMIM-EMIM}}$. In order to elucidate this unusual behaviour we inspect the RDFs TRIF-TIP3 and TIP3-TIP3 given in the second row of Figure 3. Here, we find extremely

high first peaks with values of more than 3.0 and even more than 7.0, respectively. In other words, a strong anion-water network forms at high ionic strength. As the rather bulky cations do not fit into this network they are expelled to separate regions of space. Thus, they approach each other more closely explaining the characteristic features. This may be also indicative of nanostructured regions in ILs already reported in literature[18].

The screening influence of water dipoles on the anion-anion RDF discussed above is only one feature. Even more striking is the formation of a peak at around $7.5\mathring{A}$ with increasing ionic strength. In fact, the shape of this RDF changes from a double peak structure to a single peak with shoulder. However, a radial analysis is not capable to resolve its origin: As already discussed[28] earlier a description solely in terms of radial distributions is not appropriate for anisotropic molecules because molecules in different solvation shells may contribute to the same radial bin of the RDF. The method of Voronoi tesselation offers the possibility to decompose RDFs into contributions from different shells. This is done in Figure 5 for $g^{000}_{\mathrm{TRIF-TRIF}}$ and $g^{000}_{\mathrm{TRIF-TIP3}}$. We learn that the emerging peak around $7.5\mathring{A}$ mainly stems from the second solvation shell which changes massively as a function of ionic strength while the first peak at around $6.0\mathring{A}$ preserves its shape. This different behaviour is caused by the anisotropy of the triflate molecule: The angular RDF

$$g^{101}(r) = \frac{1}{4\pi\rho r^2 N} \sum_i \sum_{j \neq i} \langle \hat{\mu}_i \cdot \hat{r}_{ij} \cdot \delta(\vec{r} - \vec{r}_{ij}) \rangle \tag{41}$$

is a generalisation of the simple RDF (Equ. 38) measuring the angular distribution of neighbouring molecules around a reference TRIF. This is achieved replacing unity in the simple RDF by the scalar product $\hat{\mu}_i \cdot \hat{r}_{ij}$ representing the cosine between the dipole axis of TRIF, $\hat{\mu}_i$, and the unit vector $\hat{r}_{ij}$ joining the reference TRIF with its neighbours. In this way one gets information about the average cosine and thus the latitude of the angular distribution. Decomposing again $g^{101}_{\mathrm{TRIF-TIP3}}$ into contributions from different shells we find that the average cosine is of opposite sign in these two shells. A positive sign tells us that the neighbouring molecule is in the northern hemisphere of the reference TRIF, in other words closer to the CF_3 unit. Analogously, a negative sign is indicative of a closer proximity to the sulfonyl group (SO_3^-). With the additional information that the cosine between dipolar axis $\hat{\mu}_i \cdot \hat{\mu}_j$ discussed in more detail in Section IV D 1 is strictly negative in both shells one can conclude that the first peak in $g^{000}_{\mathrm{TRIF-TIP3}}$ corresponds to a closer proximity CF_3 units while the second peak represents a closer proximity of SO_3^- moieties. A fine tuning can be

x_{H_2O}	A	$1/\sigma$	λ
0.97	294.35	0.48	7.22
0.93	81.71	0.28	6.97
0.9	53.96	0.22	6.8
0.85	41.23	0.17	6.64
0.79	35.53	0.15	6.49
0.71	32.11	0.13	6.38
0.65	31.35	0.13	6.34
0.43	28.08	0.11	6.23

TABLE II. Parameters specifying the fit function of the charge ordering function $Q(r)$.

achieved by inspecting

$$g^{202}(r) = \frac{1}{4\pi\rho r^2 N} \sum_i \sum_{j \neq i} \langle (\frac{3}{2}(\hat{\mu}_i \cdot \hat{r}_{ij})^2 - \frac{1}{2}) \cdot \delta(\vec{r} - \vec{r}_{ij}) \rangle \tag{42}$$

which also shows a positive negative sign pattern at distances around 7.5$\mathring{A}$. (On a simple radial scale this fine pattern cannot be distinguished as the two solvation shells annihilating each other.) Figure 2 of Ref. 30 tells us that the combination of two positive signs for $g^{101}_{\mathrm{TRIF-TRIF}}$ and $g^{202}_{\mathrm{TRIF-TRIF}}$ stands for a preference of the neighbouring molecule for the northern calotte starting at 54.7° latitude, i.e. a very close proximity of the "hydrophobic" CF_3 groups. The combination of two negative signs represents the region from 0° (equator) to 54.7° southern latitude, i.e. the proximity of the hydrophilic SO_3^- units. (SO_3^- units seem to avoid the southern calotte for spacing filling arguments.) Not to forget, however, that Voronoi analysis classifies SO_3^- units as secondary neighbours. In other words, there must be a bridging molecule. The extremely high peak found in $g^{000}_{\mathrm{TRIF-TIP3}}$ clearly point to a bridging water molecule which is the only of the three solvent species, EMIM, TRIF, and TIP3, which can fill such a small gap. In fact, the sharp peak is at 4.5$\mathring{A}$. If the COM of the reference TRIF, the oxygen of TIP3, and the COM of the neighbour TRIF would be collinear the two TRIF molecules would be separated by a distance of 9$\mathring{A}$. Taking the actual distance of 7.5$\mathring{A}$ leads to a triangular geometry with an angle of 113°.

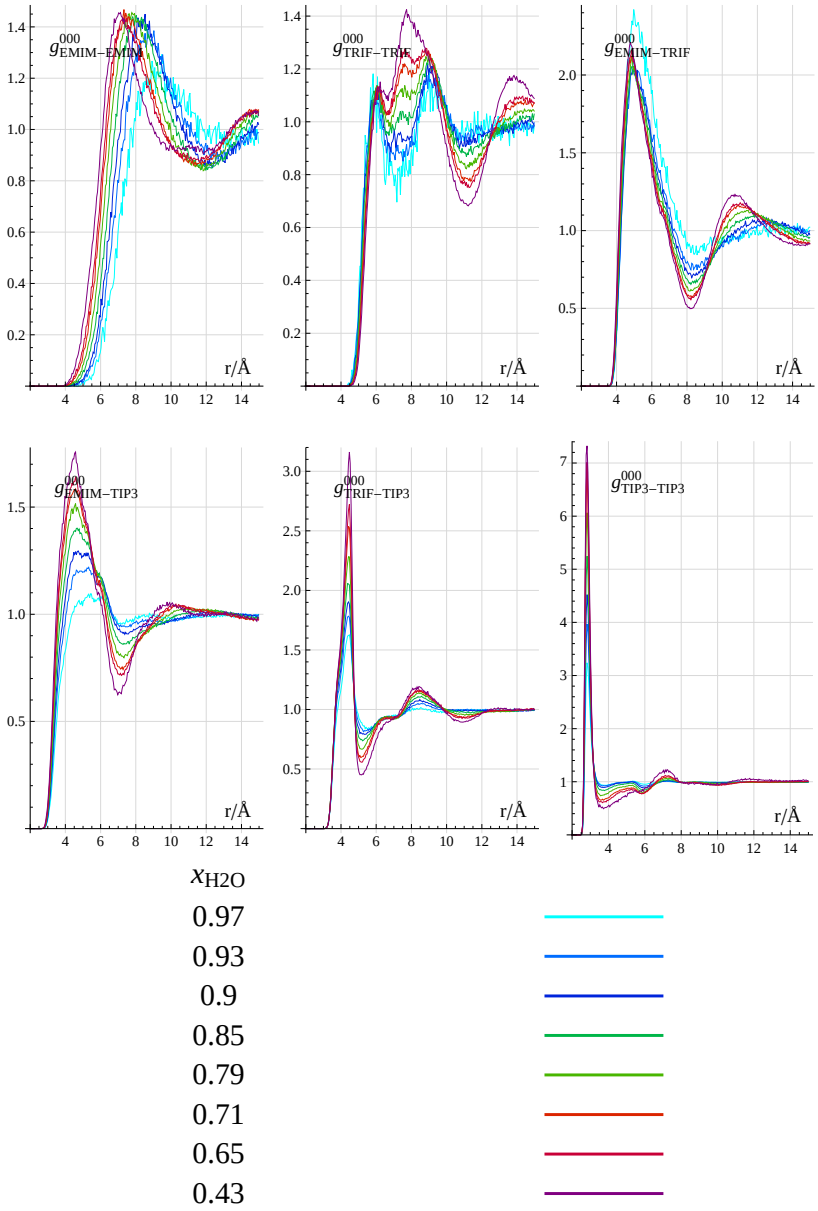


FIG. 3. $g^{000}(r)$ (simple COM-RDF) for all solvent species combinations in the pure water-IL systems.

B. Single Particle Dynamics in the Laboratory Frame

As representative examples we present the translational diffusion coefficient (Equ. 18) and for rotation the average relaxation time, $\langle\tau\rangle$, obtained from a multi exponential fit of the dipolar correlation function (Equ. 17). For TRIF and TIP3 the dipole moment vector exactly coincides with one body-fixed axis: $\vec{\mu}(t) = \mu\vec{f}_1(t)$ for TRIF, and $\vec{\mu}(t) = \mu\vec{f}_2(t)$ for TIP3. The dipole moment vector of EMIM is very close to body-fixed y-axis, $\vec{f}_2(t)$.

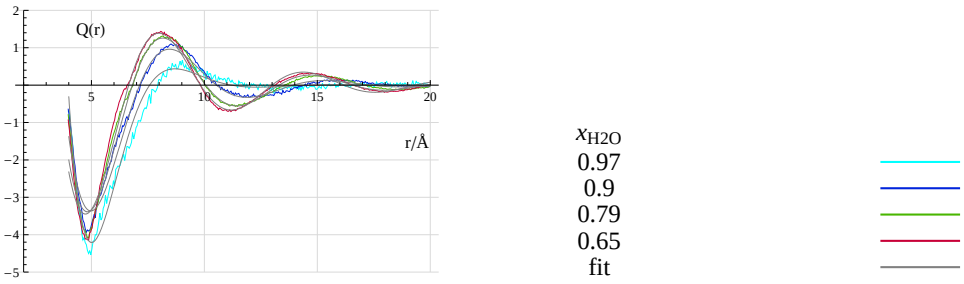


FIG. 4. Charge ordering function $Q(r)$ and fits (parameters in Table II) for selected water mole fractions of the pure water-IL systems.

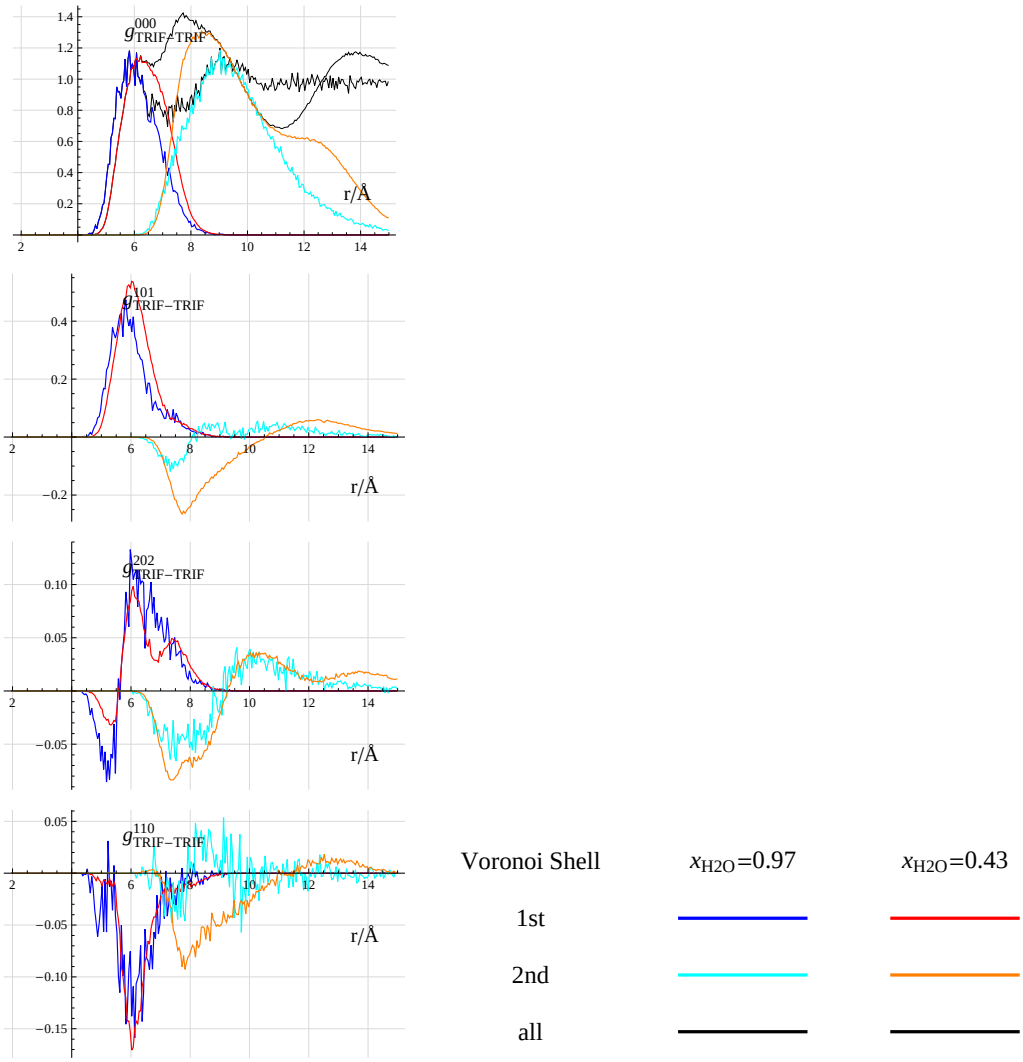


FIG. 5. $g^{000}(r)$ (simple COM-RDF), $g^{101}(r)$ and $g^{202}(r)$ (angular RDFs), and $g^{110}(r)$ (dipole-dipole spatial correlation function) for TRIF around TRIF. In addition, the correlation functions are resolved into Voronoi shells.

In order to present the plethora of data in a compact way, we have fitted their dependence on the water molar fraction by logistic functions

$$f(x) = a_0 + \frac{a}{1 + e^{k(x-x_0)}} \tag{43}$$

. We emphasise that our fit is restricted to the range of actual values and should not be used for extrapolation. The obtained parameters are gathered in Table III. The corresponding curves are presented in Figure 6, although multiplied by a constant, c. For each curve the respective constant was obtained by a least-squares fit of $c \cdot f(x)$ to the viscosity values. Bearing in mind the diversity of molecular species comprising water, TRIF, EMIM, and even proteins, the almost perfect coincidence illustrates the "viscosity scaling" (cf. Equ. 33 and Equ. 35). The constants may be converted to hydrodynamic radii but the so-obtained values are far from realistic values, as expected from previous studies[29]. Taking one half of the maximum intra molecular distance as a reference for realistic radii, solvent molecules are one half and proteins are twice as large.

The parameters of the logistic functions can be interpreted as follows: The function rises monotonically from a_0 to $a_0 + a$ with an inflexion point at x_0. The slope is given by the parameter k which is a measure of the influence of the ionic strength. For the systems with protein the influence is larger as in the pure system; among the proteins the impact is stronger in the zinc-finger system. Correspondingly, the inflexion point of the zinc-finger system is at a higher mole fraction as compared to the ubiquitin system. It reaches the lowest value for the pure systems. In other words, the proteins seem to act as an effective enhancement of the ionic strength. Interestingly, the viscosity of the systems follows a slightly different trend. It is high for the ubiquitin system, medium for the zinc finger, and low for the pure systems.

So far, reorientation in the laboratory frame was restricted to a single body-fixed axis, the dipole moment vector. The inherent anisotropy of molecular ions, however, requires a description of reorientation in terms of all three body-fixed axes, in other words, the relaxation of the metric $M_{kl}(t)$ (Equ. 15). Fitting the diagonal terms of the metric, $\langle M_{kk}(t) \rangle$, again multi exponentially, yields the average relaxation times, $\langle \tau_k \rangle$, as a function of mole fraction as shown in Figure 7. For the pure systems, the three solvent species exhibit the following behaviour: TRIF and EMIM appear – in a dynamical sense – as rotational prolate ellipsoids with the axis coinciding with the dipole moment vector being the major axis.

PURE $1/D\,[6\cdot10^4 cm^{-2}s]$

	a0	a	k	x0
TIP3	-1.24	44.68	8.63	0.58
TRIF	-6.71	336.37	9.34	0.55
EMIM	-7.75	275.56	7.98	0.54

PURE $\langle\tau\rangle\,[10^{-12}s]$

	a0	a	k	x0
TIP3	0.	682.82	9.75	0.54
TRIF	0.	1729.46	9.51	0.58
EMIM	0.	2980.72	9.92	0.55

5ZNF $1/D\,[6\cdot10^4 cm^{-2}s]$

	a0	a	k	x0
TIP3	-0.41	51.4	11.27	0.61
TRIF	0.16	207.79	17.63	0.7
EMIM	0.	124.76	18.5	0.73
5ZNF	4.65	3377.04	20.57	0.72

5ZNF $\langle\tau\rangle\,[10^{-12}s]$

	a0	a	k	x0
TIP3	0.	1500.	13.	0.64
TRIF	27.38	2002.18	16.43	0.72
EMIM	0.1	2100.	15.	0.68
5ZNF	1000.	$1.\times10^6$	20.	0.73

1UBQ $1/D\,[6\cdot10^4 cm^{-2}s]$

	a0	a	k	x0
TIP3	0.	51.38	11.17	0.6
TRIF	0.	629.34	12.72	0.56
EMIM	0.	316.15	12.29	0.6
1UBQ	40.12	9177.61	17.14	0.6

1UBQ $\langle\tau\rangle\,[10^{-12}s]$

	a0	a	k	x0
TIP3	-4.48	3364.71	11.05	0.46
TRIF	0.	5498.54	11.55	0.6
EMIM	4.87	4471.95	12.96	0.59
1UBQ	2400	2.1×10^6	15	0.6

TABLE III. Collection of parameters specifying the logistic fit Equ. 43 of the reciprocal diffusion coefficient, $1/D$, and the average reorientation times $\langle\tau\rangle$ for the three solvent species, EMIM, TRIF, and TIP3, and the two solutes, 5ZNF and 1UBQ.

The rotational dynamics of TIP3 resembles that of a general ellipsoid with three different relaxation times. It is interesting to note that for all three solvent molecules only the major axis, i.e. dipole axis, obeys viscosity scaling as shown above in Figure 6. In fact, the two minor axes depend on the ionic strength to a lesser extent as compared to the viscosity. The rotational dynamics of the two proteins, 5ZNF and 1UBQ, seems to be that of a spherical

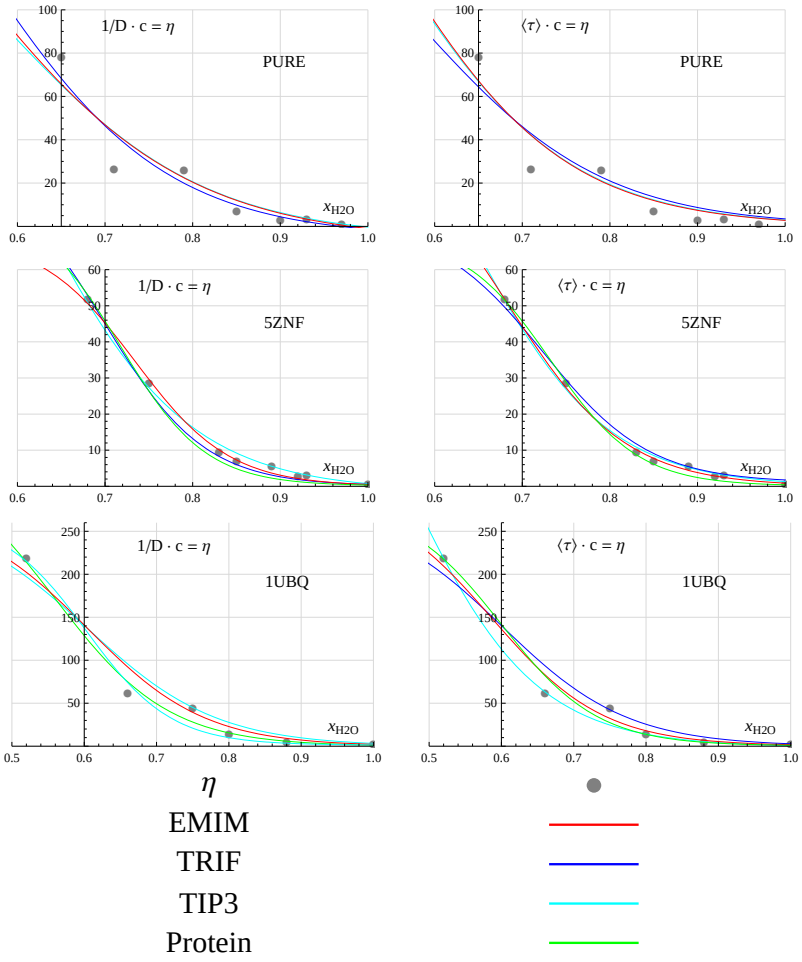


FIG. 6. Logistic fits of the reciprocal diffusion coefficient, $1/D$, and the average reorientation times $\langle\tau\rangle$ for the three solvent species, EMIM, TRIF, and TIP3, and the two solutes, 5ZNF and 1UBQ, scaled by a factor c to coincide with the viscosities.

top (Figure 7) within the statistical accuracy achievable in a 200ns simulation.

C. Single Particle Dynamics in the Body Fixed Frame

Anisotropy of shape and charge distribution is a genuine feature of molecular ionic liquids and the ultimate reason for their liquidity at room temperature. Of course this anisotropy is also reflected in the static structure but dynamics is usually much more sensitive to the interactions between molecules. Therefore, we will resolve the anisotropic character of translational and rotational dynamics of molecular species in full detail. We switch from a description in the laboratory frame to the projection of dynamics to body-fixed axes.

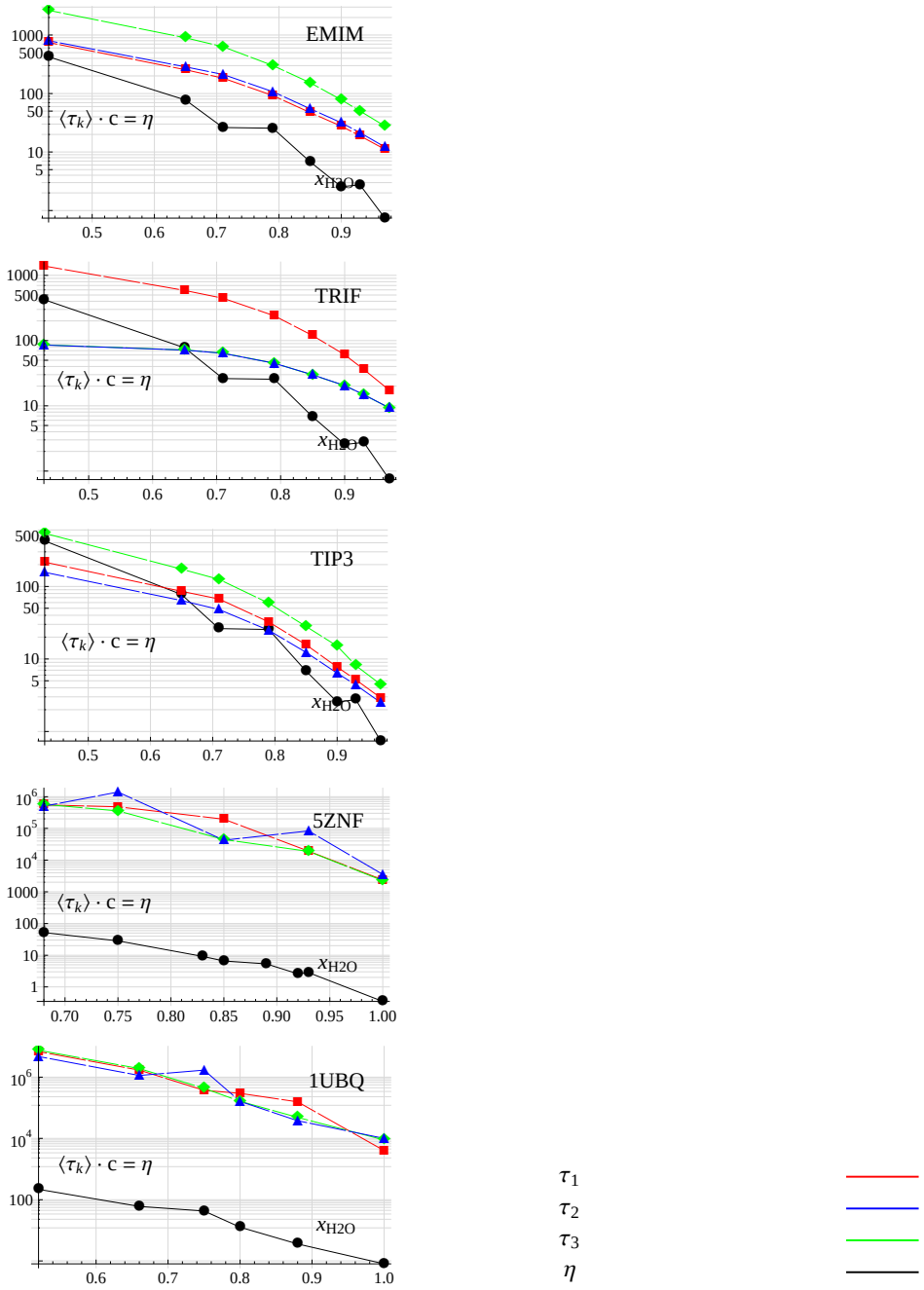


FIG. 7. Average relaxation times, $\langle \tau_k \rangle$ (in ps), of the diagonal terms of the metric, $\langle M_{kk}(t) \rangle$, and viscosities, η (in cPs). The values for the solvent species (EMIM, TRIF, and TIP3) are from the pure IL-water systems.

Thereby we will always analyse the dependence on water mole fraction, or equivalently, on ionic strength, in order to characterise the transition from hydrated ionic liquids to pure ionic liquids. Although in our systems, studied at high ionic strengths, water is still present

the results found here are already indicative or even representative of pure ionic liquids.

We start with the decomposition of translational motion and present the rotational one subsequently.

1. Decomposition of Translational Motion

Figure 8 and Figure 9 display the decomposition of the translational velocity autocorrelation function, $\langle \vec{v}(0)\vec{v}(t) \rangle$ (left column), and the respective diffusion coefficient, D^{trans} (middle and right column), according to Equ. 18 and Equ. 20.

For the highest and lowest mole fraction a set of four curves is given including the three diagonal terms coloured in accordance with the definition of body-fixed axes in Figure 1. The six off-diagonal elements are pairwise identical for statistical reasons. Thus we could also give the three off diagonal elements individually. However, in order to avoid overcrowding the figure we give their sum only.

Comparing the lowest and highest ionic strength a dramatic change of the decomposed translational motion (first column in Figure 8 and Figure 9) can be observed. The behaviour of ubiquitin is most striking: The set of non negative correlation functions, typical for a diffusive process, changes to a oscillatory pattern with pronounced negative regions. This change can be observed for all other molecular species but to a lesser extent. For all components of the velocity correlation function, their first minimum steps down at high ionic strength and shifts to shorter times. This shift can be clearly seen in the respective power spectra (data not shown). It can be interpreted as an enhanced backwards motion, or rebound, of a molecule trapped in a cage built up by its neighbours following the initial forward movement along a preferred body-fixed axis. The elevated frequency of this "rattling the cage" is caused by the stronger electrostatic forces at high ionic strength. Of course, this has a dramatic impact on the diffusion coefficients. Taking again the extreme example of ubiquitin, the running integral of the diffusion coefficient, $D_{kl}^{\text{trans}}(t)$ (middle column in Figure 9), increases at low ionic strength while it decreases – after an initial rise – at high ionic strength. This decrease after an initial rise can be observed for all other molecular species. The amount of decay, however, is rather different depending on ionic strength (middle column, Figure 8 and Figure 9). In some cases this leads to counterintuitive results and subtle differences between molecular species:

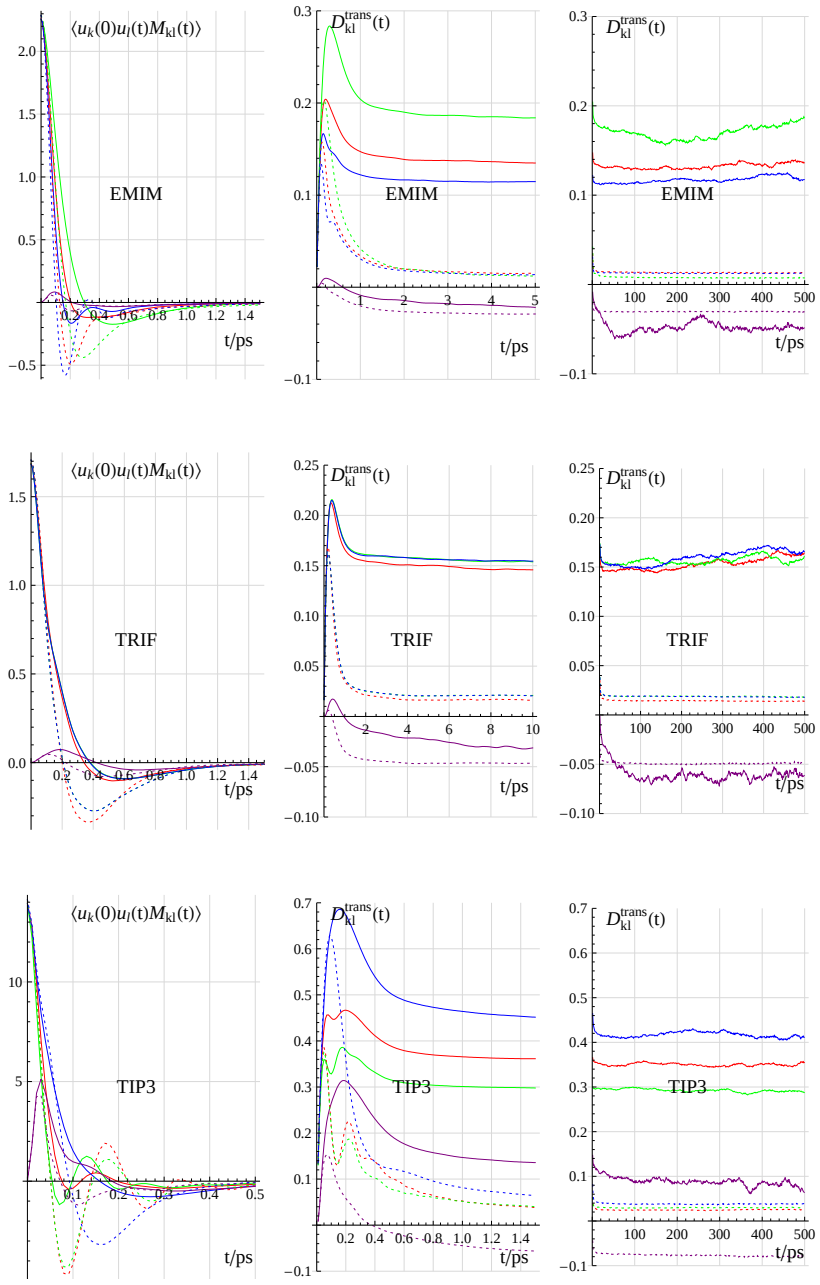


FIG. 8. Left column: translational velocity autocorrelation function, decomposed into diagonal terms ($k = l = 1, 2, 3$) and the sum of off-diagonal terms ($k \neq l$). Middle and right column: Integral of the translational velocity autocorrelation function, short-term and long-term behaviour. The legend in Figure 9 applies.

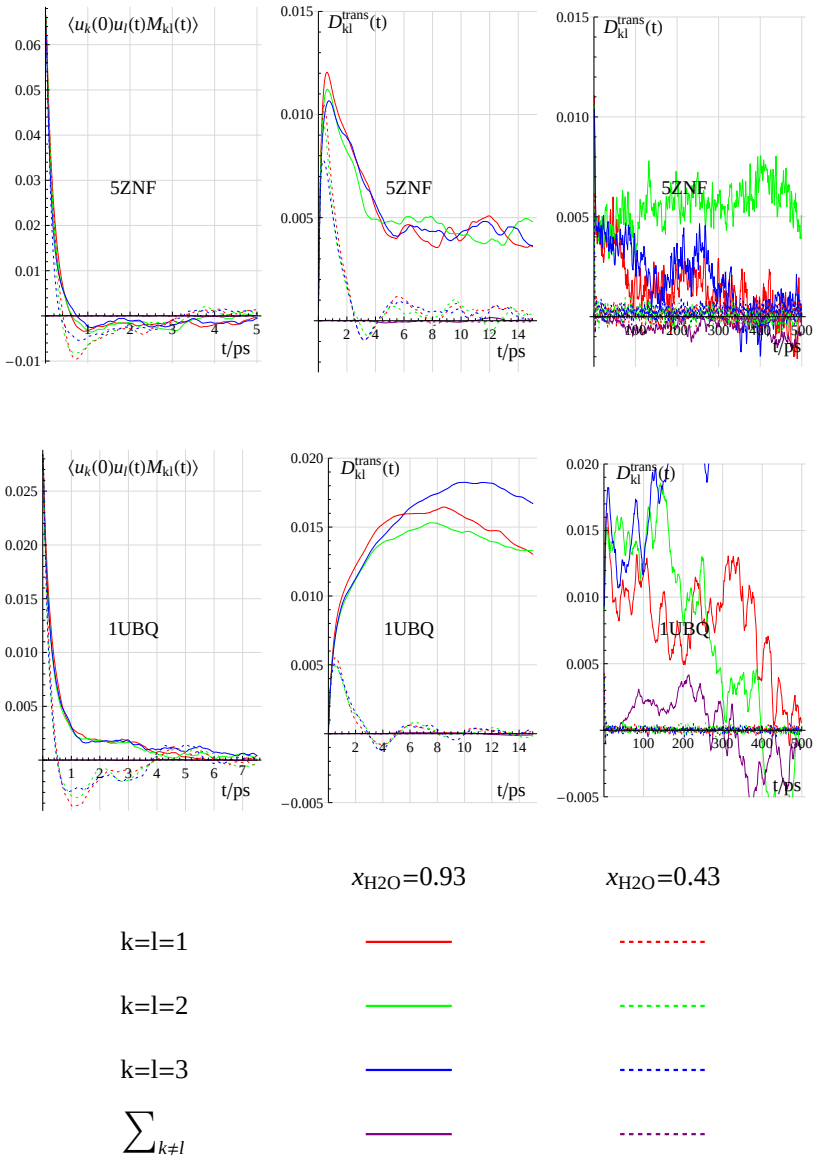


FIG. 9. Left column: translational velocity autocorrelation function, decomposed into diagonal terms ($k = l = 1, 2, 3$) and the sum of off-diagonal terms ($k \neq l$). Middle and right column: Integral of the translational velocity autocorrelation function, short-term and long-term behaviour.

The molecular structure of triflate suggests that diffusion along the C-S axis (dipole moment axis and axis of minimum moment of inertia) is accelerated as compared to the two (equivalent) perpendicular axes. The higher oscillatory character, i.e. the deeper negative region in the respective velocity component correlation function, however, leads to a smaller partial diffusion coefficient (cf. right column in Figure 8, D_{11}^{trans} is smaller than D_{22}^{trans} and D_{33}^{trans}).

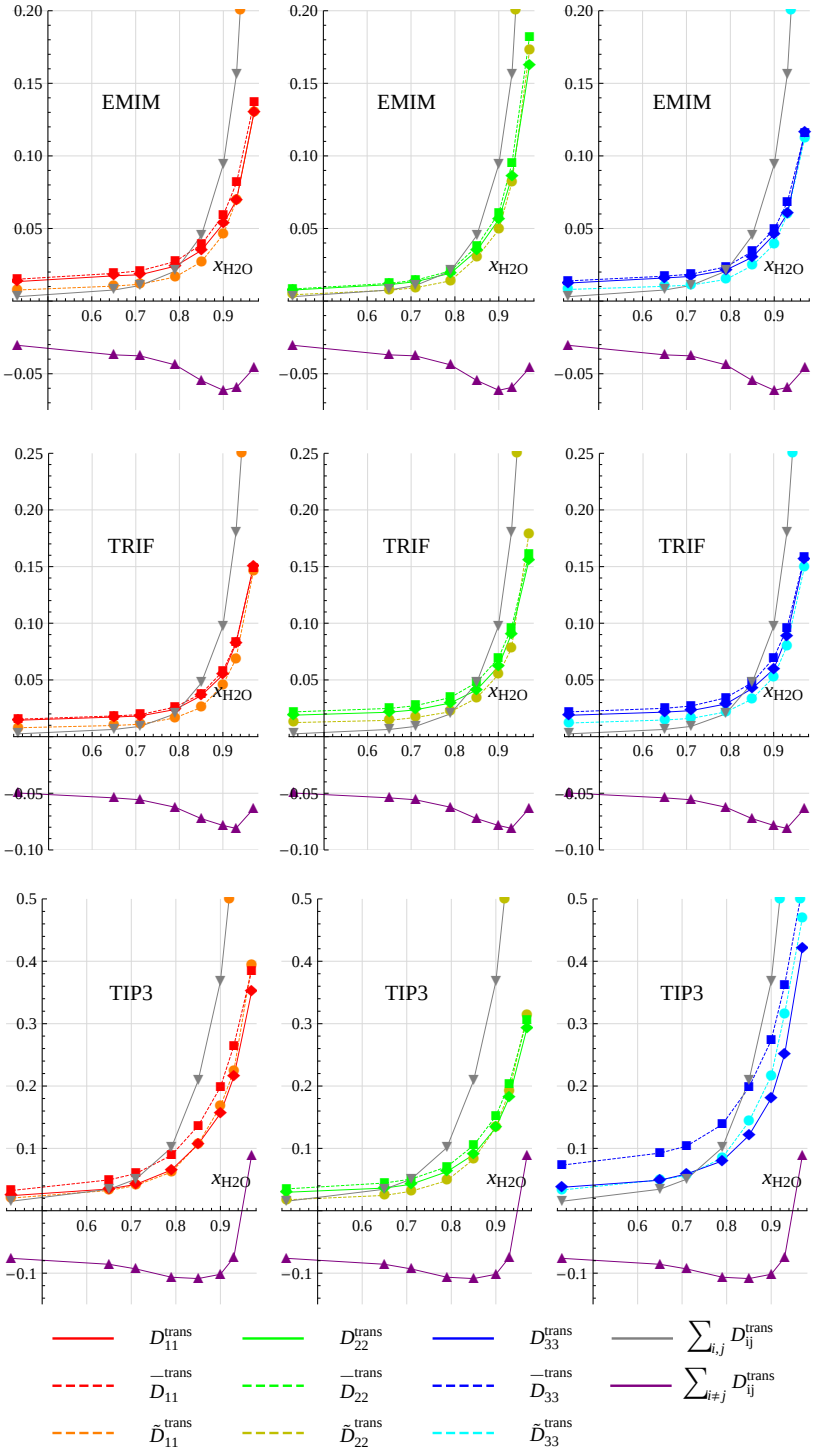


FIG. 10. Partial translational diffusion coefficients (diffusion tensor elements), D_{kk}^{trans}, and approximations, $\bar{D}_{kk}^{\mathrm{trans}}$, and $\tilde{D}_{kk}^{\mathrm{trans}}$, as well as the sum of the off-diagonal elements and the total sum, the translational diffusion coefficient (times three).

For EMIM a more subtle behaviour is observed: At low ionic strength it is intuitive, i.e. diffusion is faster along the axis of minimum moment of inertia (close to the dipole moment axis), but it is counterintuitive, i.e. slower, at high ionic strength (right column in Figure 8).

At first sight the presentation of the third column in Figure 8 and Figure 9 seems superfluous: The integrand in the first column dies off quite rapidly and is confined to short time interval. For an intermediate time interval the running integral (middle column) seems to converge. However, in order to get the final value, an asymptotic time regime up to 0.5ns is necessary. In fact, we took the final values of the partial diffusion coefficients (Figure 10) from the average in the time window 0.1ns to 0.3ns.

This demonstrates once more the high compensatory character of correlations in ionic liquids which almost annihilate each other upon integration leading to a very small final value. A most prominent example is the low conductivity of ionic liquids[27].

So far we have found an almost perfect compensation of positive and negative contributions upon integration. However, this is not limited to contributions within one component, may be diagonal or off-diagonal, but applies to a mutual compensation between diagonal and off-diagonal components, as well. Quite generally, the sum of off-diagonal elements makes a negative contribution with the single exception of TIP3 at high water mole fraction (Figure 10).

Beyond this initial region of composition the sum of off-diagonal terms achieves a steady state. The diagonal terms, however, show a steady decrease finally reaching values comparable to the sum of the off-diagonal terms. As a result the total diffusion coefficient, representing the sum of diagonal and off-diagonal elements, is further reduced according to the increase of the relative importance of the off-diagonal terms. As they describe the cross correlation between different axes, increase of their relative importance implies an increased hampering of translation along preferred axes. While at low ionic strength translation along different axes is almost free or slightly coupled, at high ionic strength translation along one axis immediately induces translation along other axes such that the net or overall displacement is quite modest.

Regarding a pure ionic liquid as the lower limit of mole fraction it is to be expected that these compensatory features observed above reach their maximum value.

The previous analysis is based on the full correlation functions simultaneously considering the body-fixed coordinates and the metric Equ. 13. Comparing the timescales in Figure 7

with that of Figure 8 and Figure 9 which differ by at least two orders of magnitude it might be tempting to consider the correlation functions of the body-fixed coordinates alone leaving aside the metric. Figure 10 gives the diagonal diffusion coefficients $\tilde{D}_{kk}^{\mathrm{trans}}$ computed from the body-fixed velocity coordinates (Equ. 24). There is a clear tendency to underestimate the diagonal diffusion coefficients and the deviation increases from a few percent at low ionic strength up to forty percent at the highest ionic strength. This behaviour can be explained by the partial suppression of the negative subregion in the correlation functions of the coordinates by the decay of the metric from the initial of unity. This makes it reasonable to include the metric at least in the product approximation (Equ. 22). Indeed, for TRIF and EMIM this creates a mere overestimation of typically ten percent. For TIP3, however, the deviations are larger – especially for $\tilde{D}_{33}^{\mathrm{trans}}$ – because of the faster relaxation of the metric (Figure 7). For the off-diagonal elements the product approximation fails because the metric stays close to zero for too long time. As the above analysis has revealed the importance of the off-diagonal terms at higher ionic strength, inclusion of the metric is essential.

2. Decomposition of Rotational Motion

This section presents the rotational analogue of the decomposition of the velocity correlation function involving the angular velocity $\vec{\omega}(t)$. Traditionally this quantity is formulated in terms of Eulerian angles, quaternions, etc. . As our definition of the rotation of a pseudo-rigid body is based on the least-squares fit algorithm of Horn[10] we infer $\vec{\omega}(t)$ directly from the rotation matrix (Equ. 6 and Equ. 10).

Most of the findings, e.g. viscosity scaling, compensatory behaviour within the diagonal elements of the diffusion tensor and between diagonal and off-diagonal elements, and essentiality of metric relaxation, of the previous section referring to the decomposition of the translational motion are also valid for the decomposition of the rotational motion presented here. Therefore we restrict our discussion to additional features and possible differences.

In Figure 11 and Figure 12 the high frequency undulations are a striking new feature in addition to the already observed typically oscillatory correlation functions found in ionic liquids (first column). The reason or origin of both types of oscillations can be demonstrated in the respective power spectra (Figure 13) for the lowest and highest ionic strength.

For EMIM, the high frequency peak appears in the region of 160THz for $k = 2$ and

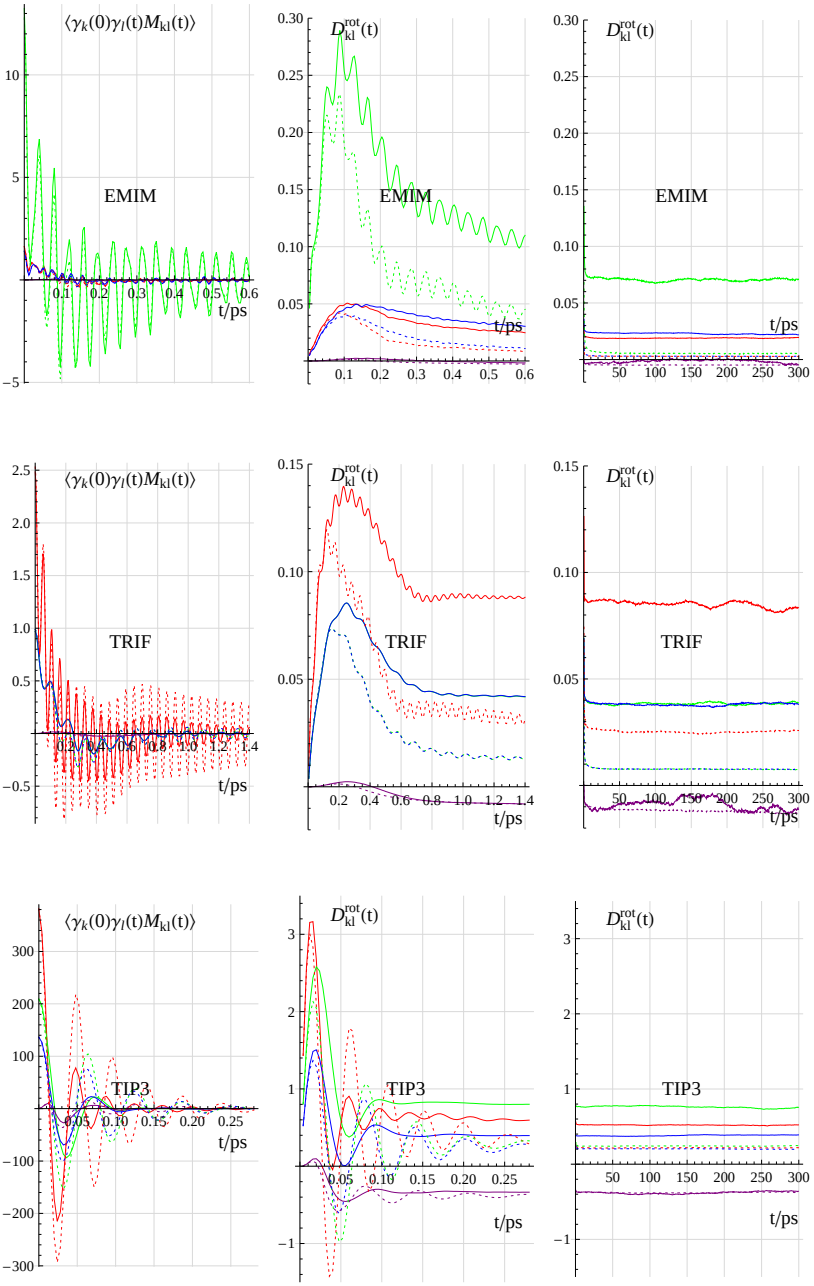


FIG. 11. Left column: angular velocity autocorrelation function, decomposed into diagonal terms ($k = l = 1, 2, 3$) and the sum of off-diagonal terms ($k \neq l$). Middle and right column: Integral of the angular velocity autocorrelation function, short-term and long-term behaviour. The legend in Figure 12 applies.

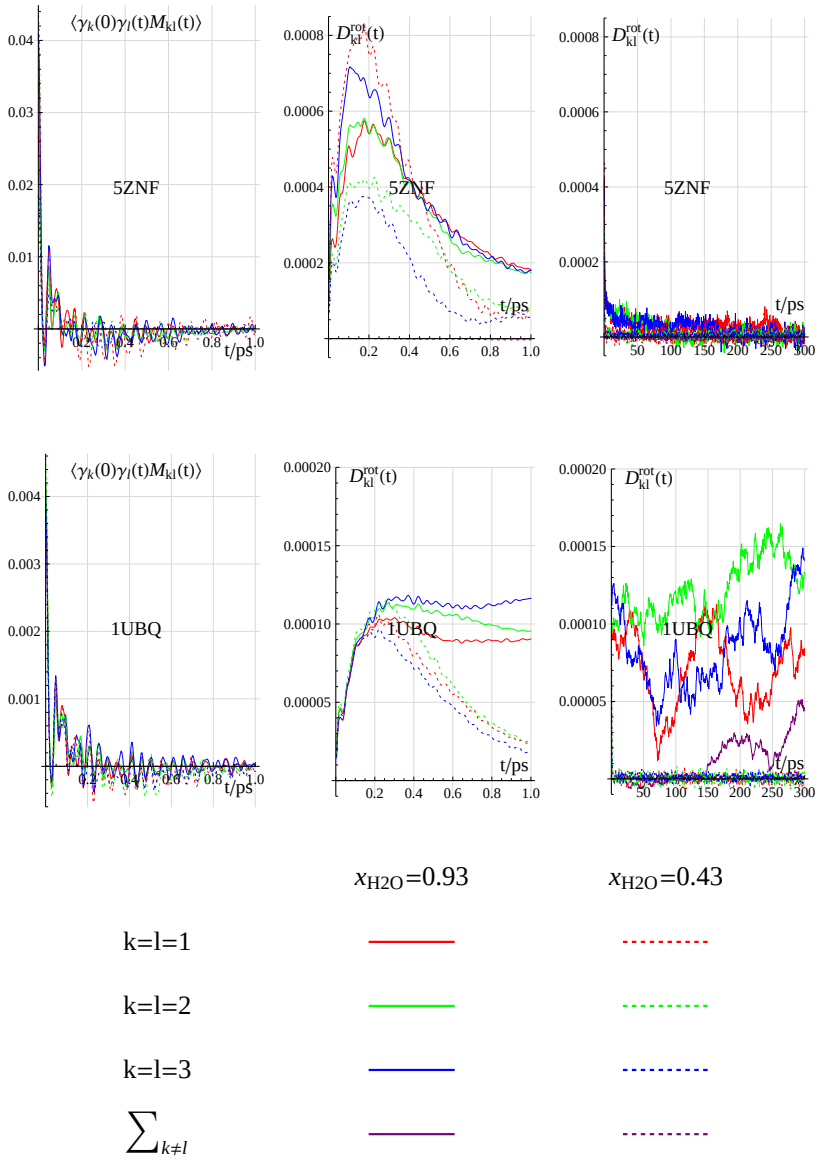


FIG. 12. Left column: angular velocity autocorrelation function, decomposed into diagonal terms ($k = l = 1, 2, 3$) and the sum of off-diagonal terms ($k \neq l$). Middle and right column: Integral of the rotational velocity autocorrelation function, short-term and long-term behaviour.

250THz for $k = 1, 3$. TRIF peaks at 120THz for $k = 1$, and 50THz for $k = 2, 3$. The intra molecular geometry of TIP3 was kept fixed; consequently, the high frequency peak is absent. In general, increase of ionic strength may change the amplitude of the sharp high frequency peak but does not alter its position, while the broad low frequency peak is shifted to higher frequencies. Therefore, we conclude that the high frequency peak represents remnant features of intra molecular vibrations handed over by the least-squares fit of the

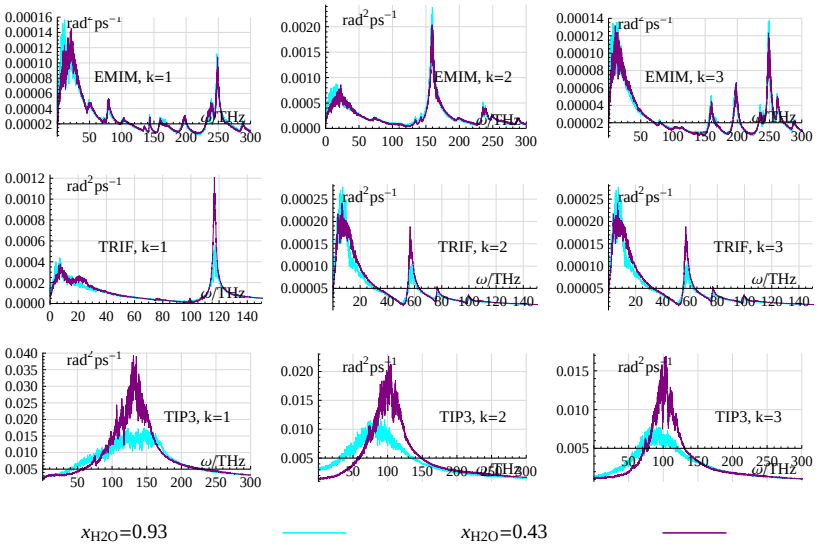


FIG. 13. Power spectra of the rotational angular velocity autocorrelation functions.

Horn algorithm. They are averaged out upon integration when computing the diffusion coefficients. The low frequency oscillations are definitely of intermolecular character as they shift with the variation of ionic strength.

Figure 14 compares the various approximations to the elements of the diffusion tensor analogous to the translation above. Again we find the principal relation $\tilde{D}^{\mathrm{rot}}_{kk} < \bar{D}^{\mathrm{rot}}_{kk} \approx D^{\mathrm{rot}}_{kk}$. Furthermore, this figure is augmented by the "diffusion coefficients" derived from the linear slope of the angular displacement $\langle \Delta\phi^2_k(t) \rangle$ (cf. Equ. 30, Equ. 31, and Equ. 32). The complete angular displacement curves as a function of ionic strength are given in Figure 15.

From Figure 14 we learn that the diffusion coefficients obtained from the integral of the velocity coordinates, $\tilde{D}^{\mathrm{rot}}_{kk}$ (Equ. 25) is of lower statistical quality than the values derived from the linear slope of $\langle \Delta\phi^2_k(t) \rangle$. This is somewhat puzzling because the third column in Figure 11 shows nicely converging diffusion coefficients. One has to bear in mind, however, that these results have been obtained including the metric whose relaxation damps the integrand in the asymptotic time regime. Once this metric is neglected, as in $\tilde{D}^{\mathrm{rot}}_{kk}$, the integral becomes noisy and looses statistical quality.

Although the size and shape of EMIM, TRIF and TIP3 are quite different their rotational behaviour, which can be deduced from all figures discussed so far, is the following: TRIF is an exact symmetric top with two equivalent axes $k = 2, 3$. EMIM is fairly close to a symmetric top with two similar axes $k = 1, 3$. TIP3, however, behaves as a general top with

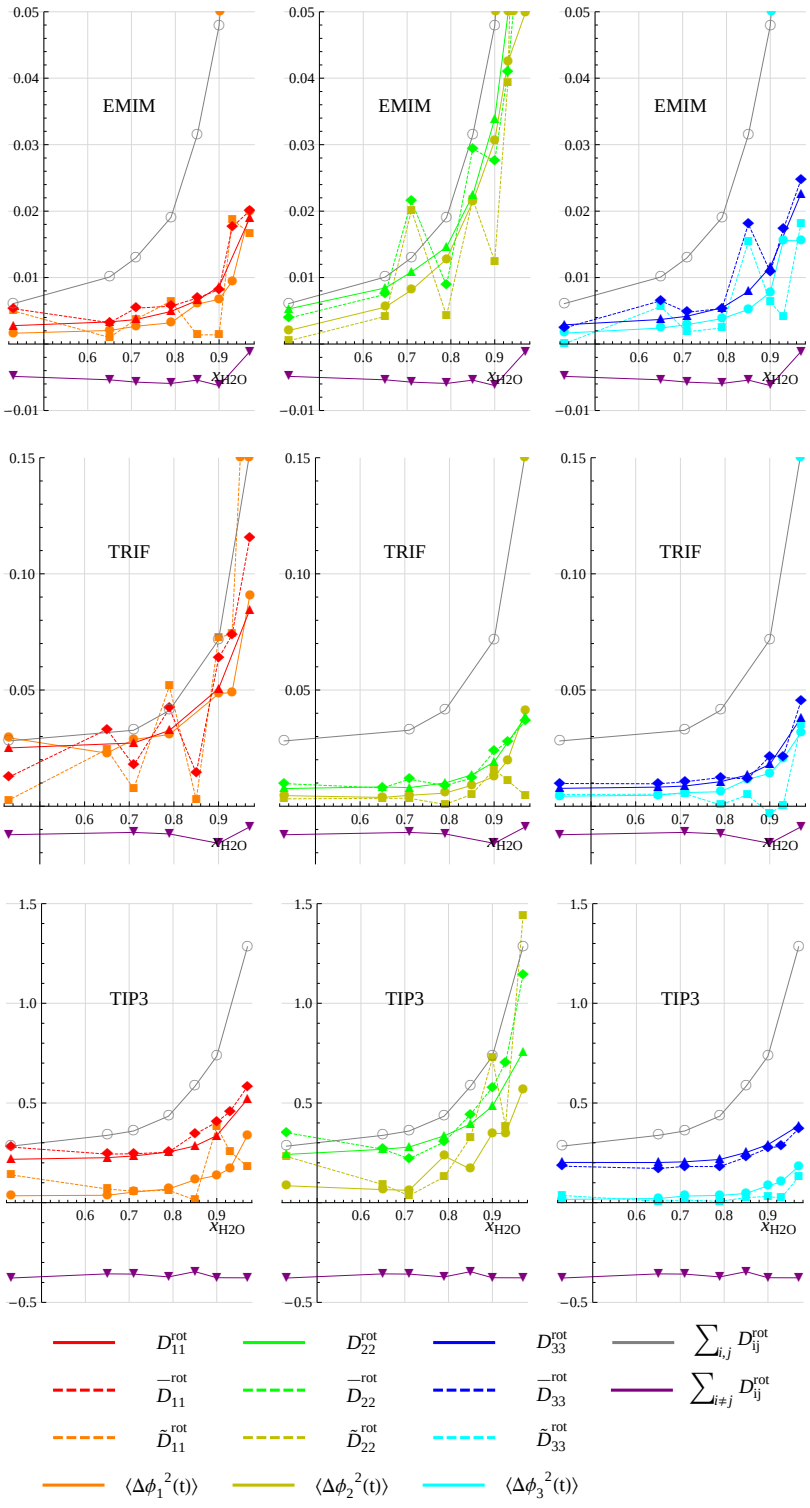


FIG. 14. Partial rotational diffusion coefficients (diffusion tensor elements), D_{kk}^{rot}, and approximations, $\bar{D}_{kk}^{\mathrm{rot}}$, and $\tilde{D}_{kk}^{\mathrm{rot}}$, as well as the sum of the off-diagonal elements and the total sum, the rotational diffusion coefficient (times three). Additionally, the angular MSD, $\langle\Delta\phi_k^2(t)\rangle$, is given.

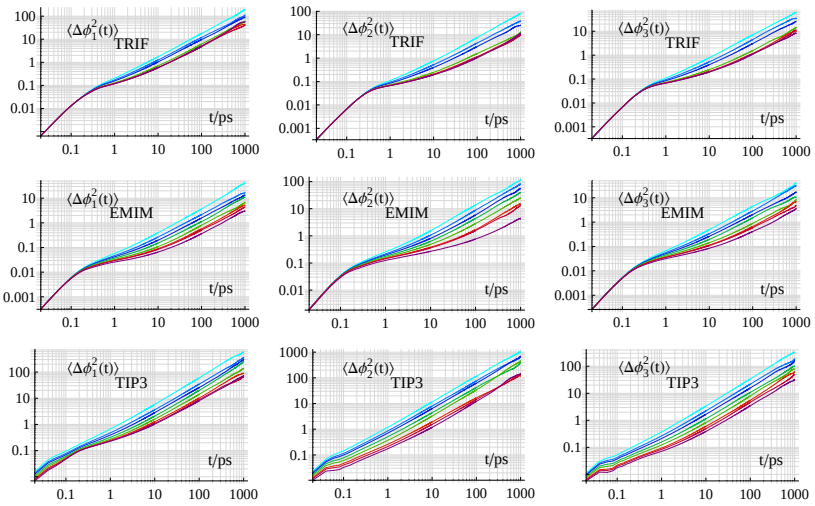


FIG. 15. Double logarithmic plot of the angular MSD. The same legend as in Figure 5 applies.

three different axes. Within the accuracy achievable by a 200ns simulation the two proteins 1UBQ and 5ZNF come close to spherical tops with three nearly equivalent axes.

We close this section with some remarks concerning the relations between rotations of an axis and about an axis. The cumulant expansion (Equ. 29) postulates that the product of the relaxation time of an axis times the sum of the diffusion coefficients for rotation about the other two axes should be unity. Figure 16 checks this postulate: It is found to be valid for high water mole fractions but breaks down with higher ionic strength. Apart from the question of validity of the cumulant expansion for ionic liquids whose complexity implies multi exponential behaviour, there are two possible explanations: On the one hand, the relaxation times $\langle \tau_k \rangle$ for the axis coinciding with the dipole moment vector obey viscosity scaling. The complementary diffusion coefficients, however, don't. Therefore, the product cannot remain constant, i.e. be independent of the ionic strength. On the other hand, the increasing importance of the off-diagonal elements of the diffusion tensor cannot be ignored.

D. Collective Structure and Dynamics

So far we have analysed the structure in terms of pair correlations and the dynamics in terms of single particle motion. The strong electrostatic forces in ionic liquids, however, inevitably require a collective point of view. Here, the term "collective" stands for the many-body correlations exceeding a description by single molecules or pairs. We start with

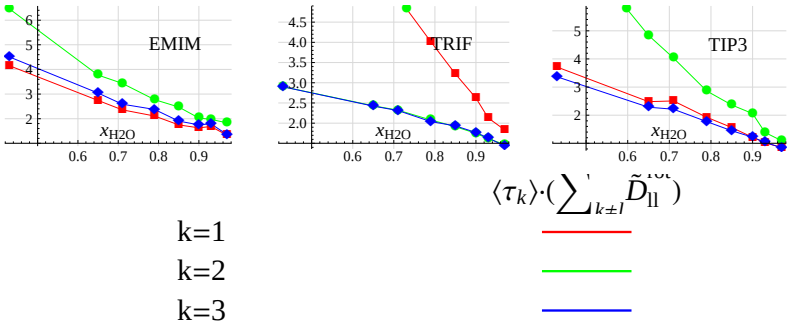


FIG. 16. Product of the relaxation time of an axis times the sum of the diffusion coefficients for rotation about the other two axes.

a specific example. As a measure of structural collectivity in liquid systems Kirkwood introduced the following quantity[12]:

$$G_k = \frac{1}{N\mu^2}\sum_{i=1}^{N}\langle\vec{\mu}_i\cdot\vec{M}_D\rangle = \frac{1}{N\mu^2}\langle M_D^2\rangle \tag{44}$$

. Here, $\vec{\mu}_i$ is the dipole moment of molecule i with length μ. The collective of all dipoles is given by $\vec{M}_D = \sum_i \vec{\mu}_i$. The normalisation by the number of dipoles N makes G_k an intensive quantity, while μ^2 ensures correlations between dipole moments of unit length ($\hat{\mu} = \vec{\mu}/\mu$). G_k, as a measure of collective order, may be resolved into contributions from concentric spherical shells[24]

$$G_k(R) = \frac{1}{N\mu^2}\sum_{i=1}^{N}\langle\vec{\mu}_i\cdot\vec{M}_D(R)\rangle = \frac{1}{N}\sum_{i}\sum_{\substack{j\\ r_{ij}<R}}\langle\hat{\mu}_i\hat{\mu}_j\rangle \tag{45}$$

where

$$\vec{M}_D(R) = \sum_{\substack{j\\ r_{ij}<R}}\vec{\mu}_j \tag{46}$$

is the total dipole moment of the sphere of radius R. A reformulation as an integral gives

$$G_k(R) = \int_0^R \frac{1}{N}\sum_i\sum_j\langle\hat{\mu}_i\hat{\mu}_j\rangle\,\delta(r-r_{ij})\,\mathrm{d}r \tag{47}$$

$$= 1 + \rho\int_0^R g^{110}(r)\,4\pi r^2\mathrm{d}r \tag{48}$$

with the dipolar radial distribution function $g^{110}(r)$

$$g^{110}(r) = \frac{1}{4\pi\rho r^2 N}\sum_i\sum_{j\neq i}\langle\hat{\mu}_i\hat{\mu}_j\cdot\delta(\vec{r}-\vec{r}_{ij})\rangle \tag{49}$$

which is a generalisation of the standard radial distribution function Equ. 38. For a cubic box of length L, spherical shells are complete up to the inscribed sphere of radius $R = L/2$. Beyond this threshold shells are incomplete but integration can be still continued until the final value $R_{\max} = \sqrt{3}/2L$. Thus, from Equ. 44 the asymptotic value of the R-dependent G_k-factor is

$$\lim_{R \to R_{\max}} G_k(R) = \frac{\langle \vec{M}_D^2 \rangle}{N\mu^2} \tag{50}$$

. Figure 17 shows the R-dependent G_k-factor as a running integral of the dipolar RDF for three selected mole fractions, with and without solute.

As the relative variation of the respective volumina is small compared to the water mole fractions the decrease of the latter implies a decrease in water density. Simultaneously, the height of the first peak in the dipolar RDF grows, i.e. the ordering of water dipoles is enhanced. However, this growth is slightly less than the decrease in density. In combination this lowers the G_k-factor. In addition, negative correlations in the dipolar RDF at larger distances increase and thus also reduce the magnitude of the G_k-factor.

In those cases, where the mole fractions match exactly, there is clear evidence that an additional protein solute reduces the G_k-factor by reducing the height of the dipolar RDF. In any case, marginal differences in the dipolar RDF are clearly visible in the respective G_k-factor. This comes from the parabolic amplifier r^2 (Equ. 48) which blows up even modest contributions at large distances. For this reason it is hardly possible to obtain an asymptotic value of G_k by the integration of the dipolar RDF for EMIM. Here, G_k is so close to unity because the dipolar RDF is rather modest and consequently noisy. The parabolic amplifier r^2 inflates this noise thus prohibiting a reliable computation of the deviation from unity.

In case of very small G_k-factors we have found out that the value derived from Equ. 50 is a better route. In this way we find values of typically one third for TRIF.

1. Transition Single Particle Dynamics – Collective Dynamics

The G_k-factor as a measure of collective orientational structure enables a transition from single particle dynamics to collective dynamics via the Kievelson-Madden relation[20]

$$\langle \tau_M \rangle = G_k \cdot \langle \tau_\mu \rangle \tag{51}$$

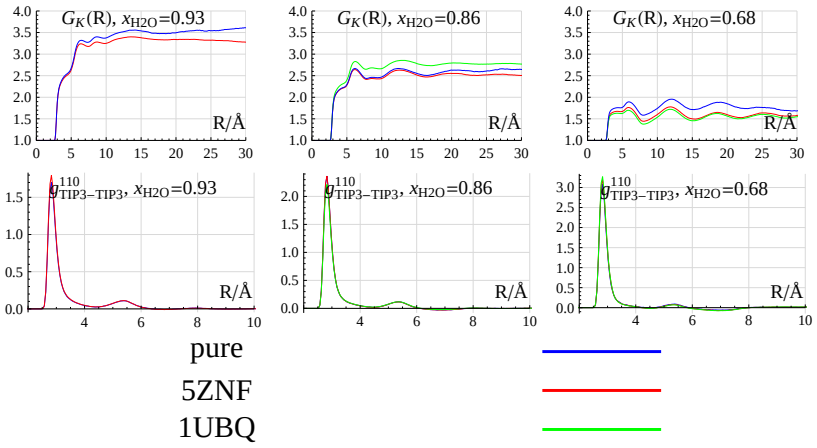


FIG. 17. R-dependent G_k-factor (upper row) and the dipole-dipole spatial correlation function (lower row), $g^{110}(r)$, for TIP3 in the three systems, pure IL-water, 5ZNF and 1UBQ in IL-water. Note that the average water mole fraction of the two or three systems is given.

which connects the relaxation time of the collective dipole moment $\langle\tau_M\rangle$ with that of the single molecule dipole moment $\langle\tau_\mu\rangle$. The three molecular species, EMIM, TRIF, and TIP3, represent all three possibilities: As G_k is close to one for EMIM $\langle\tau_M\rangle$ and $\langle\tau_\mu\rangle$ should be equal. As can be seen from the first row of Figure 18 this is fairly valid for the pure system as well as for the two protein solutes. The only two exceptions are at mole fractions $x_{\mathrm{H_2O}} = 0.89$ and 0.68 in the 5ZNF system.

For TRIF, collective relaxation should be faster (!) than single dipole relaxation. This is indeed the case as we learn from the pure system in the second row of Figure 18. However, the ratio $\langle\tau_M\rangle/\langle\tau_\mu\rangle$ is not constant but increases with increasing ionic strength. For two protein systems, 5ZNF and 1UBQ, the underlying multi exponential model of relaxation produces some peculiarities. The problem is with those TRIF molecules which undergo a tight binding to the protein which results in a high relaxation time[6]. Therefore, we have tried to remove the longest relaxation time from the multi exponential fit. The resulting $\langle\tau_M^{\mathrm{mod}}\rangle$ are represented by dashed lines in Figure 18. Of course, this reduces the average collective relaxation time but the remnant value now scales with the single particle relaxation time and the respective ratio as a function of mole fraction seems to be constant. Referring back to Section IV B viscosity scaling seems to apply to collective dynamics as well.

For TIP3, collective relaxation should be slower than single particle reorientation. The third row of Figure 18 confirms this. For the pure system, the ratio $\langle\tau_M\rangle/\langle\tau_\mu\rangle$ resembles

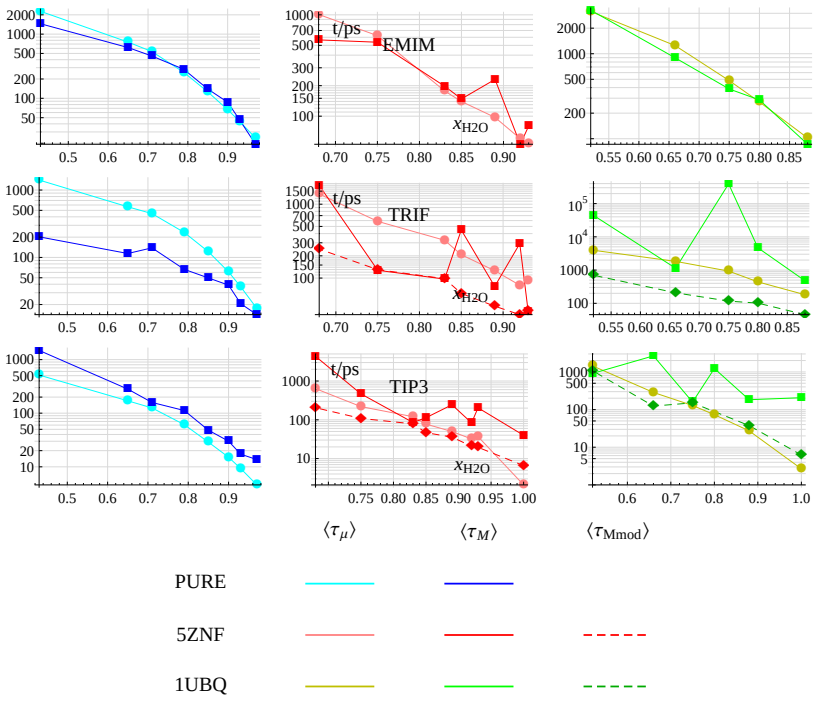


FIG. 18. Average relaxation times of the single molecule dipole moment $\langle\tau_\mu\rangle$ and of the collective dipole moment $\langle\tau_M\rangle$, as well as the modified collective dipole moment relaxation times $\langle\tau_M^{\mathrm{mod}}\rangle$ without the long time constants.

the behaviour of the G_k-factor although reduced by a factor of two thirds. TIP3 molecules solvating protein solutes again produce peculiarities which make it difficult to identify a clear-cut ratio. Elimination of the long time constants in the multi exponential model, as already done for TRIF, also results in a fairly constant ratio and $\langle\tau_M^{\mathrm{mod}}\rangle$ again obeys viscosity scaling. The respective modified G_k-factor, however, is around unity.

E. Rotational Dielectric Relaxation $\epsilon(\omega)$

In the previous section we have already discussed the average relaxation time $\langle\tau_M\rangle$ of the total rotational dipole moment

$$\vec{M}_D = \sum_i \vec{\mu}_i = \sum_i \sum_\alpha q_{i,\alpha}(\vec{r}_{i,\alpha} - \vec{r}_{i,\mathrm{COM}}) \tag{52}$$

and its single particle analogue $\langle\tau_\mu\rangle$. For charged species, one is always concerned with the origin dependence of the dipole moment. Therefore, we use $\vec{\mu}_i$ which is defined with respect to the centre of mass (COM) for interpretations only. When summing over all molecular

contributions $\vec{\mu}_i$, i.e. computing the total moment, $\vec{M}_D$, the origin dependence cancels out as shown in numerous studies before[29]. The average relaxation time $\langle\tau_M\rangle$ only permits a global characterisation of the time correlation function

$$\phi_D(t) = \langle\vec{M}_D(0) \cdot \vec{M}_D(t)\rangle \tag{53}$$

. In this section we will give $\phi_D(t)$ in full detail but in the frequency range instead of the time domain. With a slight modification of the Fourier transform to the Fourier-Laplace transform and taking the derivative of $\phi_D(t)$ instead of the original function we get a direct link to the rotational part of the dielectric spectrum[29]:

$$\varepsilon(\omega) - 1 = -\frac{4\pi}{3Vk_BT}\int_0^\infty \mathrm{d}t\, e^{i\omega t}\,\frac{\mathrm{d}\phi_D(t)}{\mathrm{d}t} \tag{54}$$

For a multi exponential fit of $\phi_D(t) = \sum_j A_j e^{-t/\tau_j}$ the imaginary part of the dielectric constant $\varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega)$ is given by

$$\varepsilon''(\omega) = \frac{4\pi}{3Vk_BT}\,\omega\,\sum_j \frac{A_j\tau_j}{1+\omega^2\tau_j^2} \tag{55}$$

. For three selected mole fractions of the pure (dashed lines) and the 5ZNF system (solid lines) $\varepsilon''(\omega)$ is shown in Figure 19.

From the two main peaks in the spectrum the high frequency peak can be attributed to the solvent because it remains qualitatively the same in the systems with and without protein solute. The amplitude remains almost constant while a slight shift to lower frequencies is observed as a consequence of the retardation by the solute.

With increasing ionic strength the solvent peak is shifted to lower frequencies and reduced in height. The shift comes from the enhanced viscosity retarding molecular motion and, therefore, collective dipolar reorientation. The reduction can be tracked back to the replacement of the water component with a high dipole density by the less polar ions. This phenomenon can be illustrated in more detail by Figure 20 which shows the effects of the decomposition of $\vec{M}_D$

$$\vec{M}_D(t) = \sum_i \vec{M}_{D,i}(t) \tag{56}$$

which implies

$$\phi_D(t) = \sum_i\sum_j \langle\vec{M}_{D,i}(0) \cdot \vec{M}_{D,j}(t)\rangle \tag{57}$$

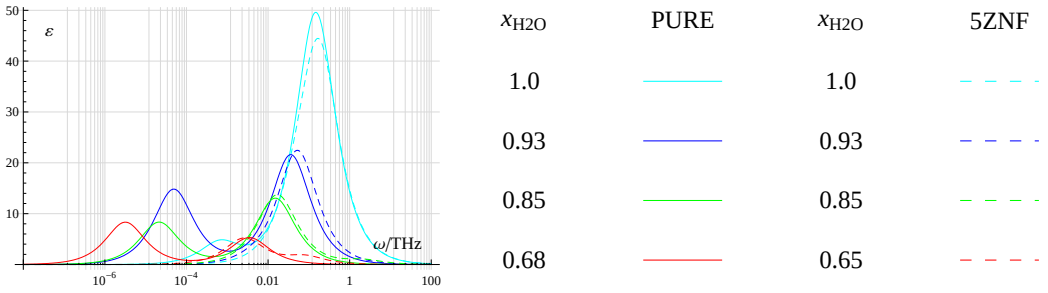


FIG. 19. Imaginary part of the rotational contribution to the dielectric spectrum $\varepsilon''(\omega)$ of hydrated ILs with and without solute (5ZNF and pure).

. Here, the subscripts stand for the three molecular species, EMIM, TRIF, and TIP3. When summing the three self terms, $\langle \vec{M}_{D,i}(0) \cdot \vec{M}_{D,i}(t) \rangle$ (dashed line in Figure 20) we get the qualitative shape of the total curve which is modified quantitatively by the cross terms, $\langle \vec{M}_{D,i}(0) \cdot \vec{M}_{D,j}(t) \rangle$ $(i \neq j)$. The contribution from TRIF – characterised by a upcoming double peak structure – remains more or less constant on an absolute scale. We have already observed this in a previous study[6] for the static dielectric constant: The increasing number of TRIF molecules is almost compensated by the strong anti-parallel alignment of neighbouring TRIF pairs (Figure 5).

The apparently slower motion of the protein – as compared to the solvent species – results in a solute peak at much lower frequencies. Its shift as a function of ionic strength goes synchronously with the solvent peak in such a way that the relative distance between the solvent and the solute peak remains roughly constant. In other words the spectrum as a whole is shifted towards lower frequencies. The amplitudes of the solvent peak do not change monotonically. We have discussed these peculiarities in previous study in detail[5]. We close this section noting that a precise determination of the position of the solute peak would require extremely long simulations up to the microsecond regime. This is the reason why we do not give a spectrum for the even slower dynamics of the ubiquitin system.

V. CONCLUSION

From the analysis of simulation raw data for the hydrated ionic liquid EMIM-TRIF mixed with TIP3 at various mole fractions and influenced by different solutes, none, 5ZNF, and 1UBQ, the following conclusions can be drawn:

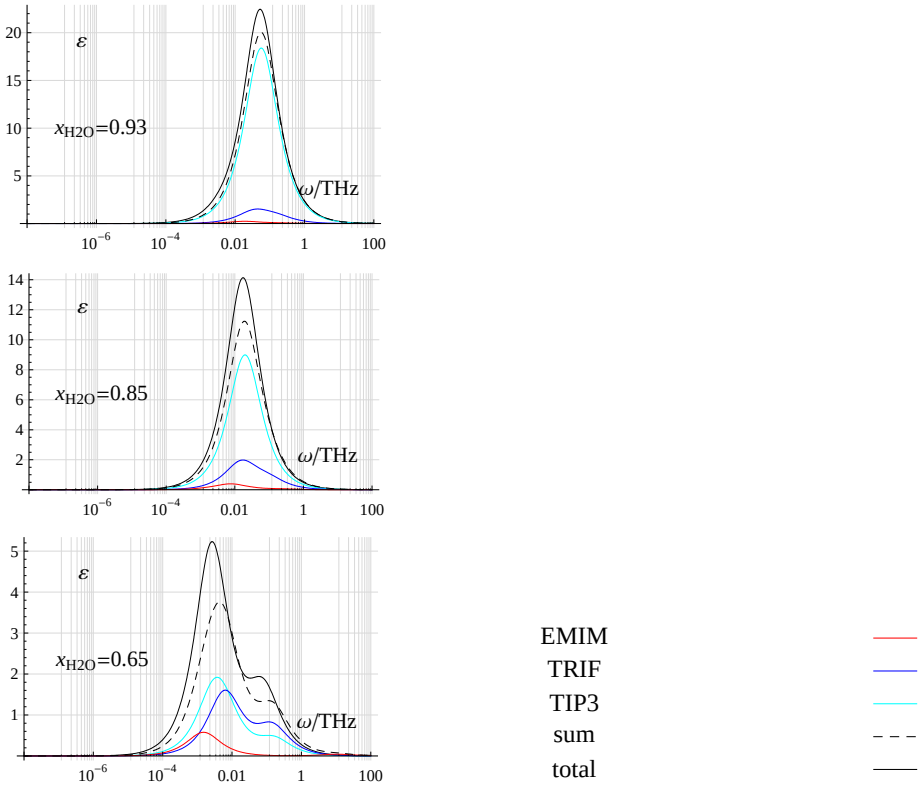


FIG. 20. Imaginary part of the rotational contribution to the dielectric spectrum $\varepsilon''(\omega)$ of pure hydrated ILs. The total rotational contribution is decomposed into contributions from the solvent species, EMIM, TRIF, and TIP3. Additionally, their sum is also given which indicates the extent of correlation between the collective dipole moment of different solvent species.

- The charge ordering function shows the usual behaviour caused by the screening effect of water which can also digitised by an appropriate fit.
- For a three component solvent there are six RDFs which sometimes show specific peculiarities:
  - With increasing ionic strength the EMIM-EMIM RDF shifts towards shorter distances. This rather counterintuitive effect is the result of the strong anion-water network which expels cations thus bringing them closer together.
  - Decomposition of RDFs into contributions from different Voronoi shells reveals further interesting features. The second peak in the TRIF-TRIF RDF which develops with increasing ionic strength is essentially constituted by the second Voronoi shell. Upon angular resolution a water bridged anti-parallel structure

can be identified.

- Restricting single particle dynamics to the motion of the centre of mass and the reorientation of the dipolar axis seems to follow hydrodynamic rules, i.e. respective motional parameters obey viscosity scaling. Axes perpendicular to the dipolar axes do not follow the hydrodynamic regime.

- Introducing a body-fixed frame as alternative to the usual laboratory frame single particle translation and reorientation can be analysed in higher detail: For example, one can compute angular velocities for the rotation about specific axes from which, by integration, an angular mean square displacement (MSD), completely analogous to the translational MSD, can be derived.

- Even more important, the overall diffusion coefficient, may be translational or rotational, can be decomposed into contributions along, or from, different axes. To our best knowledge this is the first IL study of such a systematic decomposition in the body-fixed frame. The main findings are:

  - The relaxation of the metric is essential.
  - Correlation between different axes, the so-called off-diagonal terms increase with increasing ionic strength. They reduce the overall diffusion coefficient because of their negative sign. Thus the small translational and rotational diffusion coefficients of pure ionic liquids are caused by the strongly correlated motion along different axes.
  - The so-called cumulant expansion links the rotation about axes with the rotation of axes. However, it only works at very low ionic strength.

- The G_K-factor can be used as a valuable measure of collective orientational structure. Its resolution into spherical shells reveals long range correlations, especially at high ionic strength.

- The different character of the three solvent species (EMIM,TRIF, and TIP3) show up in a G_K-factor close to unity, below unity, and above unity.

- This offers three possibilities to test the Kievelson-Madden relation. It is fulfilled semi-quantitatively for the pure systems, but spoiled by solvent molecules bound to the

solute in the 5ZNF and 1UBQ systems. Computation of a modified G_K-factor, freed from the impact of bound molecules, enables a modified Kievelson-Madden relation.

- The full time dependence of the rotational total dipole moment correlation function is analysed via the dielectric spectrum: From the two essential peaks, the solvent peak is shifted and reduced with increasing ionic strength. The solute also causes a shift of the solvent peak but to a moderate degree. The solute peak is also shifted with increasing ionic strength but no systematic trend of the amplitude is found. However, in this regime we are facing the limits of a 200ns simulation.

VI. ACKNOWLEDGEMENTS

The computational work was performed on the "Vienna Scientific Cluster" (www.zid.tuwien.ac.at/vsc) of the University of Vienna, the Vienna University of Technology, and the University of Natural Resources and Applied Life Science Vienna. We thank for generous allocation of computer time. This work was supported by Project No. P19807 of the FWF Austrian Science Fund.

---

[1] Berni J. Alder and W. Edward Alley. Generalized hydrodynamics. *Physics Today*, 37(1):56–63, 1984.

[2] Bruce J. Berne and Robert Pecora. *Dynamic light scattering.* Dover Publications, Mineola, New York, 2000.

[3] O. Borodin and G. D. Smith. *J. Phys. Chem. B*, 110:11481–11490, 2006.

[4] Mario G. Del Pópolo and Gregory A. Voth. On the structure and dynamics of ionic liquids. *The Journal of Physical Chemistry B*, 108(5):1744–1752, 2004.

[5] M. Haberler, C. Schröder, and O. Steinhauser. Solvation studies of a zinc finger protein in hydrated ionic liquids. *Phys. Chem. Chem. Phys.*, 13:6955–6969, 2011.

[6] Michael Haberler and Othmar Steinhauser. On the influence of hydrated ionic liquids on the dynamical structure of model proteins: a computational study. *Phys. Chem. Chem. Phys.*, 13:17994–18004, 2011.

[7] C. G. Hanke, N. A. Atamas, and R. M. Lynden-Bell. *Green Chem.*, 4:107–111, 2002.

[8] C. G. Hanke, S. L. Price, and R. M. Lynden-Bell. *Mol. Phys.*, 99:801–809, 2001.

[9] E. Helfand. Transport coefficients from dissipation in a canonical ensemble. *Phys. Rev.*, 119:1, 1960.

[10] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, Apr 1987.

[11] Wei Jiang, Yanting Wang, and Gregory A. Voth. Molecular dynamics simulation of nanostructural organization in ionic liquid/water mixtures? *The Journal of Physical Chemistry B*, 111(18):4812–4818, 2007. PMID: 17388424.

[12] John G. Kirkwood. The dielectric polarization of polar liquids. *J. Chem. Phys.*, 7(10):911–919, October 1939.

[13] Marco Klähn, Geraldine S. Lim, Abirami Seduraman, and Ping Wu. On the different roles of anions and cations in the solvation of enzymes in ionic liquids. *Phys. Chem. Chem. Phys.*, 13:1649–1662, 2011.

[14] R. Kubo. Statistical-mechanical theorie of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.*, 12:570, 1957.

[15] Ryogo Kubo. Generalized cumulant expansion method. *Journal of the Physical Society of Japan*, 17(7):1100–1120, 1962.

[16] Hanbin Liu, Kenneth L. Sale, Blake A. Simmons, and Seema Singh. Molecular dynamics study of polysaccharides in binary solvent mixtures of an ionic liquid and water. *The Journal of Physical Chemistry B*, 115(34):10251–10258, 2011.

[17] J. N. C. Lopes, J. Deschamps, and A. A. H. Pádua. Modeling ionic liquids using a systematic all-atom force field. *J. Phys. Chem. B*, 108(6):2038–2047, 2004.

[18] José N. A. Canongia Lopes and Agílio A. H. Pádua. Nanostructural organization in ionic liquids. *J. Phys. Chem. B.*, 110(7):3330–3335, Feb 2006.

[19] R. M. Lynden-Bell and William A. Steele. A model for strongly hindered molecular reorientation in liquids. *The Journal of Physical Chemistry*, 88(26):6514–6518, 1984.

[20] Paul Madden and Daniel Kivelson. *Adv. Chem. Phys.*, 56:467, 1984.

[21] C. J. Margulis, H. A. Stern, and B. J. Berne. Computer simulation of a "green chemistry" room-temperature ionic solvent. *The Journal of Physical Chemistry B*, 106(46):12017–12021, November 2002.

[22] N. M. Micaêlo and C. M. Soares. Protein structure and dynamics in ionic liquids. insights from molecular dynamics simulation studies. *J Phys Chem B*, 112(9):2566–2572, Mar 2008.

[23] T. I. Morrow and E. J. Maginn. Molecular dynamics study of the ionic liquid 1-n-butyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Chem. B*, 106:12807–12813, 2002.

[24] M. Neumann and O. Steinhauser. *Chem. Phys. Lett.*, 102:508, 1983.

[25] G. Neumayr, C. Schröder, and O. Steinhauser. Relaxation of voronoi shells in hydrated molecular ionic liquids. *J. Chem. Phys.*, 131(17):174509, Nov 2009.

[26] Durba Roy, Nikhil Patel, Sean Conte, and Mark Maroncelli. Dynamics in an idealized ionic liquid model. *J Phys Chem B*, 114(25):8410–8424, Jul 2010.

[27] C. Schröder, M. Haberler, and O. Steinhauser. On the computation and contribution of conductivity in molecular ionic liquids. *J. Chem. Phys.*, 128(13):134501–134501, Apr 2008.

[28] C. Schröder, G. Neumayr, and O. Steinhauser. On the collective network of ionic liquid/water mixtures. iii. structural analysis of ionic liquids on the basis of voronoi decomposition. *J Chem Phys*, 130(19):194503, May 2009.

[29] C. Schröder, T. Rudas, G. Neumayr, S. Benkner, and O. Steinhauser. On the collective network of ionic liquid/water mixtures: I. orientational structure. *J Chem Phys*, 127(23):234503, Dec 2007.

[30] C. Schröder, T. Rudas, G. Neumayr, W. Gansterer, and O. Steinhauser. Impact of anisotropy on the structure and dynamics of ionic liquids: a computational study of 1-butyl-3-methyl-imidazolium trifluoroacetate. *J Chem Phys*, 127(4):044505, Jul 2007.

[31] C. Schröder and O. Steinhauser. The influence of electrostatic forces on the structure and dynamics of molecular ionic liquids. *J Chem Phys*, 128(22):224503, Jun 2008.

[32] C. Schröder and Othmar Steinhauser. Using fit functions in computational dielectric spectroscopy. *J Chem Phys*, 132(24):244109, Jun 2010.

[33] C. Schröder, C. Wakai, H. Weingärtner, and O. Steinhauser. Collective rotational dynamics in ionic liquids: a computational and experimental study of 1-butyl-3-methyl-imidazolium tetrafluoroborate. *J. Chem. Phys.*, 126(8):084511, Feb 2007.

[34] Sergio M Urahata and Mauro C C Ribeiro. Structure of ionic liquids of 1-alkyl-3-methylimidazolium cations: a systematic computer simulation study. *J Chem Phys*, 120(4):1855–1863, Jan 2004.

[35] J. Vila, P. Ginés, E. Rilo, O. Cabeza, and L. M. Varela. Great increase of the electrical conductivity of ionic liquids in aqueous solutions. *Fluid Phase Equilibria*, 247(1-2):32–39, September 2006.

[36] Tianying Yan, Yanting Wang, and Craig Knox. On the dynamics of ionic liquids: Comparisons between electronically polarizable and nonpolarizable models I. *J. Phys. Chem. B*, 114:6905, 2010.

6 Final Discussion

The effects of hydrated ionic liquids (IL) on biomolecular solvation have been investigated by means of molecular dynamics simulations. This allowed to explain the phenomena associated with solvation, such as protein stability, diffusion, and even the dielectric constant, by means of the atomic structure and dynamics. Three series of hydrated ILs with the water mole fraction varying between 1.0 and 0.5 were simulated: one series containing only the solvent (1-ethyl-3-methyl-imidazolium as cation, trifluoromethanesulfonate as anion, and water), one the solvent and the zinc finger protein as solute (PDB-code 5ZNF), and one the solvent and ubiquitin as solute (PDB-code 1UBQ).

Solute Structure

The secondary structure of the zinc finger protein and ubiquitin has been examined by traditional (DSSP, Stride, and RMSD) and novel methods (Voronoi contactmatrices). At a relatively high water content, i.e. high water mole fraction, the proteins displayed the highest instability. Amino acids that are flanking secondary structures and which are exposed to solvent molecules were spotted as the main source of instability. The dipole moment of the zinc finger protein and the intramolecular energy of both proteins was elevated at this point. These findings identified a point on the scale of mole fractions as the transition from dipolar screening of electrostatic interactions, as found in aqueous solvents, to charge screening, typical for ionic liquids.

Structure of Solute Neighbourhood

Voronoi decomposition of space was used to define molecular neighbourhood unambiguously and without need for parameters. Furthermore, the solvent distribution in immediate neighbourhood to the solute, i.e. the first solvation shell, could be mapped to individual amino acids on the solute surface. Thus, anions were identified to be resident at specific amino acids of opposite charge whereas the cations populated the entire solute surface without preference for negatively charged amino acids.

Dynamics of Solute Neighbourhood

The dynamics of the solute's neighbourhood, the first solvation shell, was described by the residence function. From this, the mean residence time was calculated for various mole fractions. This mean residence time was found to increase exponentially – for all three solvent molecule species in consideration – with increasing ionic liquid content, i.e. decreasing water mole fraction. This exponential scaling of solvation shell dynamics goes in hand with an exponential rise of the viscosity.

Structure of the Solvent

The three dimensional arrangement of solvent molecules, the solvent packing, has been presented in terms of radial distribution functions and charge ordering functions. This demonstrated, first, the dielectric screening at higher water mole fractions, and second,

the strong and highly ordered anion-water network which expels the bulky cations which in turn get drawn closer to each other. This is interpreted as nanostructures reported already in other studies.

Dynamics of the Solvent in the Laboratory-Fixed Frame

The translational and rotational diffusion coefficients were calculated for the series of hydrated ILs at varying mole fractions and applied to hydrodynamic relations. The translational diffusion scales with viscosity as expected from the Stokes-Einstein relation. From the rotational diffusion, only the dipole moment axis relaxes at times that scale with the viscosity. Therefore it can be concluded that the dipole moment dynamics has an unexpectedly direct influence on the hydrodynamic properties.

Dynamics of the Solvent in the Body-Fixed Frame

The translational and rotational motion of each molecule was also transformed into the molecule's body-fixed frame. This allowed to decompose the motion of each molecule along and about body-fixed axes. The solvent molecules as well as the solutes displayed a diffusive translational motion at higher water mole fractions which changes to an oscillatory pattern with increasing IL content. Also this aspect is a consequence of the strong interactions in ILs. Moreover, the interference between motions along different axes is found to be mostly responsible for the reduced diffusion coefficients in hydrated ILs with low water content. A theory of rotational diffusion was developed linking the rotational velocity autocorrelation functions to the angular displacement, analogous to the Einstein relation of translational diffusion.

Collective Structure and Dynamics of the Solvent

The strong interactions among the solvent molecules entail collective behaviour which was demonstrated by means of the Kirkwood G_K-factor. This value can be calculated from the radial dipole orientation function representing the structure as well as from the mean relaxation times of the single particle and collective dipole moments. A good correspondence between the two approaches was observed for the hydrated ILs without solute. The biomolecules, however, drastically hindered the collective dipole moment reorientation. To compensate for this effect a modified relation between single particle and collective dynamics has been established.

Mesoscopic View: The Dielectric Spectrum

The rotational contribution to the generalised dielectric constant and spectrum was investigated for the three series of hydrated ILs. The solutes, 5ZNF and 1UBQ, had a drastic impact in that they increased the dielectric constant to around $\varepsilon = 120$. Furthermore, the scaling of collective dynamics with the viscosity encouraged to estimate dielectric spectra in a regime exceeding the time window covered by the simulation runs.

7 Gepetto

7.1 Introduction

Gepetto (Grid Enabled Parallel Enhanced Trajectory TOol) is an efficient tool for the temporal and spatial analysis of molecular dynamics trajectories.

Key analysis methods of Gepetto include:

- Timeseries of single particle or collective observables
- Auto- and cross-correlation of observable timeseries
- Mean square displacement of observable timeseries
- Simple radial distribution functions of whole atomgroups as well as generalised to spatial orientation correlation functions
- 3D-Spatial distributions of single particle observables
- Voronoi decomposition of space by Delaunay tessellation

The program is specialised to deal with molecular dynamics trajectories from the simulation software Charmm which is most used within our group. Typically the data of trajectories is split into two different file formats: The static properties of every atom are stored to the structure file (*.psf*-file) such as mass, charge, bonds and angles with other atoms. The dynamic properties, the coordinates and the velocities at consecutive timesteps, are stored to trajectory files (*.dcd*- or *.trj*-files). Based on the instructions given in an input script Gepetto will read the trajectory data and calculate observables for selected atoms.

Key features of Gepetto include:

- *Gep*-script language for analysis instructions.
- Atom selection for choosing and grouping subsets of atoms to atomgroups.
- Optimised calculation of redundant intermediate or end results.

- Support for calculation of single particle (atomgroup) observables in arbitrary reference frame.
- Support for periodic boundary conditions.
- Prepared and optimised for parallel code.

The development of this program started with the work of Gregor Neumayr at the *Institute of Computational Biological Chemistry* in early 2007. In cooperation with Thomas Taylor the design of the program was overhauled in late 2007. The author started contributing to maintenance and development by late 2008.

7.2 General Program Flow

Figure 4 provides a bird's eye view on the main design and workflow of GEPETTO:

First the program **parses** and translates the user input script that contains instructions for the analysis of a trajectory in the form of one or more tasks. A task is specified by multiple parameters: Locations of the structure file, the trajectory file(s), and filename(s) for the report file(s), as well as atom selections to use and observables to calculate.

Next, GEPETTO builds atomgroups based on the **atom selection** which can be specified explicitly or in terms of selection patterns that match atoms of the structure file. An atomgroup aggregates the properties of one or more atoms. Usually an atomgroup is a molecule and the aggregated property is the total mass of all the atoms, the centre-of-mass, or the dipole moment. Sometimes it also makes sense to form an atomgroup of an amino acid of a protein, of a functional group, or just of a particular atom. For this step the information of the structure file is required.

The **optimisation** step organises subroutine calls and memory allocation to ensure that repeatedly required intermediate and end results are calculated and stored efficiently. Moreover, it enables the implementation of parallel code in future versions. For example, the trajectory can be split into multiple parts that can be processed separately – an "embarrassingly parallel workload" in terms of parallel computing. On one side the optimisation enables high performance computation on large molecular dynamics systems. On the other side, this optimisation part increases the complexity of the program which makes maintenance and further development challenging.

During the **calculation** part, the previously selected subroutines are applied sequentially on the trajectory data to calculate the specified observables. The results are then stored to the optimised memory. For example, the centre-of-mass of two atomgroups, say the sidechain of amino acid 32 and 33 of ubiquitin, is calculated for every selected timestep of the trajectory and stored to memory in form of a timeseries.

The **postprocessing** part deals with calculations that require results from the trajectory. Histograms are normalised and timeseries can be further processed to calculate the mean square displacement or a correlation function. Furthermore, mean square displacements or correlation functions for alike atomgroups can be averaged, e.g. the dipole moment correlation functions of all selected water molecules are averaged to yield one average dipole moment correlation function.

Finally, the desired observables in form of histograms, timeseries, correlation functions, and the like are **written to report files**.

A more detailed documentation on the individual parts outlined here can be found in the GEPETTO user and implementation guide written by Gregor Neumayr [27] and in the developer guide in Section 7.4.

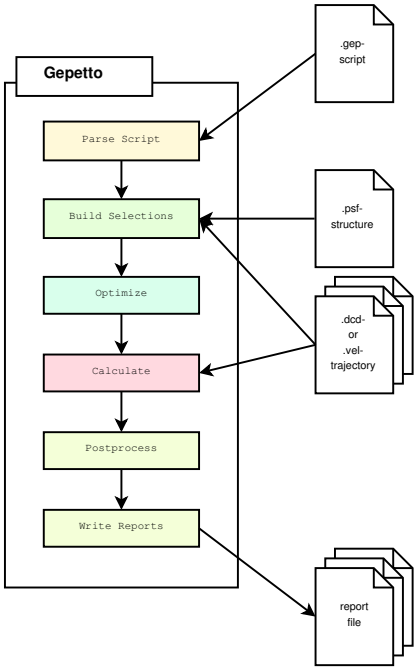


Figure 4: Basic outline of the program flow: The classes and subroutines of the source code for Gepetto have been grouped into six categories that are passed in sequence.

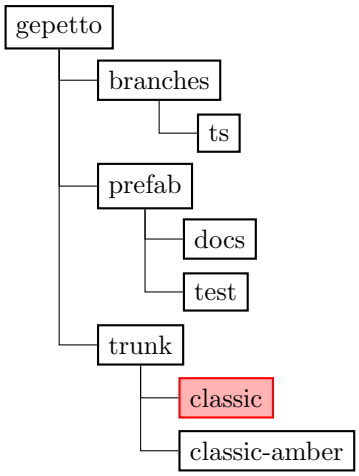


Figure 5: Folder structure of GEPETTO in the SUBVERSION repository. The directory containing the most stable version is marked in red, `classic`.

7.3 Installation Guide

In general, there is no executable GEPETTO version available - GEPETTO has to be compiled from the source files. Also, there are no "release" versions so far. Instead, there is a continuous development of the program that is documented in the repository of the version control system SUBVERSION. This repository is currently installed at the computer with the name `raid3.mdy.univie.ac.at` and can be accessed with the program SSH.

File Hierarchy

The file hierarchy of GEPETTO in the repository can be described by three major groups (see Figure 5):

1. The directory `trunk` contains the source code of the most stable version of GEPETTO, `classic` (see Figure 7), and a slightly modified version, `classic-amber`, that can deal with trajectories produced by the MD simulation software AMBER.

2. The directory `prefab` contains the documentation in HTML-format and test cases, both *.gep*-scripts as well as trajectories, for GEPETTO (see Figure 6). The user documentation can be accessed with a HTML-browser from
`gepetto/prefab/docs/html/quickbook/index.html`
and the developer documentation from the file
`prefab/docs/html/doxygen/index.html`.

3. The directory `branches` is used to experiment with the source code of GEPETTO without consequences on the most stable version in the `trunk`.

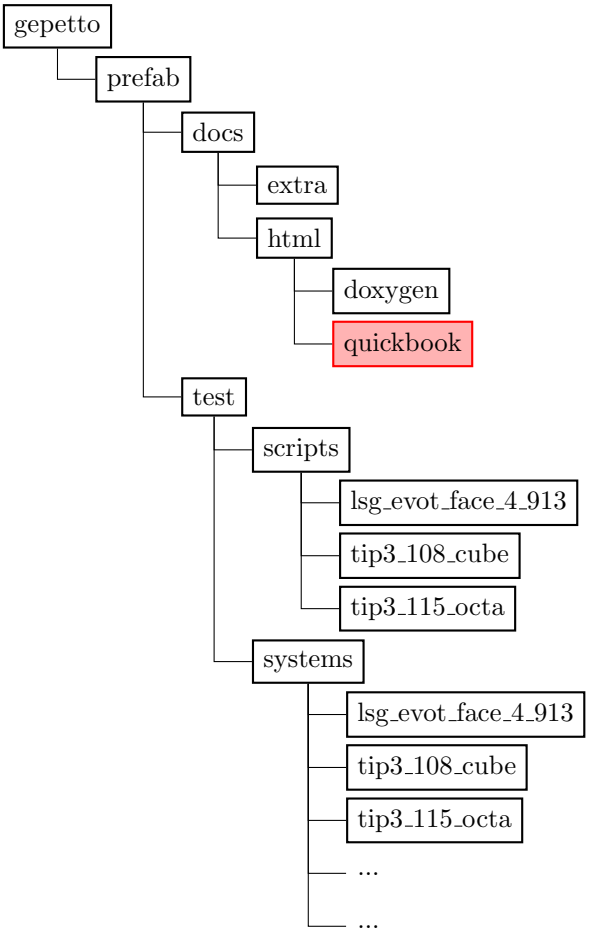


Figure 6: Folder structure for `prefab` containing "prefabricated" HTML-documentation (marked in red), test MD-systems and test scripts.

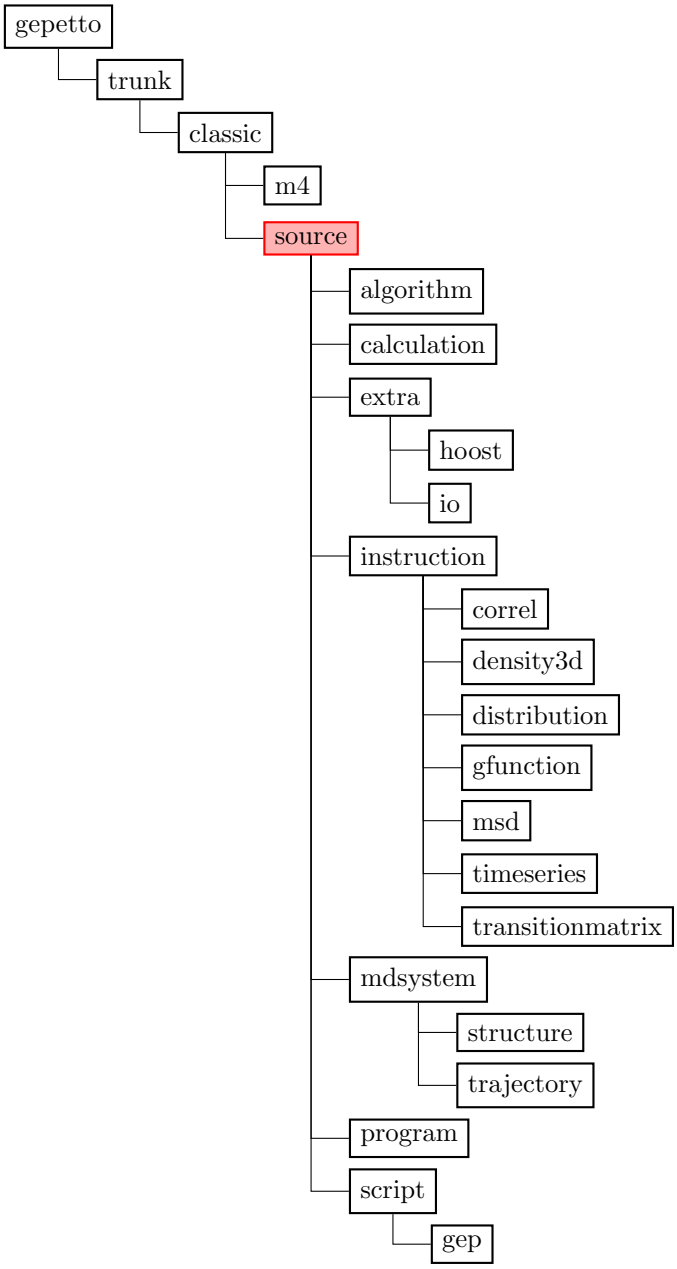


Figure 7: Folder structure of the source code of GEPETTO. The folders roughly reflect the categories defined in Figure 4: `script` → "Parse Script"; `instruction`, `mdsystem` → "Build Selections"; `instruction`, `calculation` → "Optimize"; `algorithm`, `calculation` → "Calculate"; `instruction` → "Postprocess"; `instruction` → "Write Reports".

Step-By-Step Installation

Note that the subsequent commands apply to the Unix like operating system "Debian".

1. Get BOOST and FFTW
   GEPETTO depends on two external libraries that have to be installed beforehand:

   - BOOST C++ libraries (www.boost.org, version 1.49.0)
     e.g. as a precompiled package with header files (therefore "-dev")
     `sudo apt-get install libboost-dev`
   - FFTW library (www.fftw.org, version 3.3)
     e.g.
     `sudo apt-get install libfftw3-dev`

2. Get GEPETTO
   The source files for GEPETTO are stored to a version control system repository that is located on the machine `raid3.mdy.univie.ac.at` at the *Institute of Computational Biological Chemistry*.

   - To access the repository the program SUBVERSION is required.
     `sudo apt-get install subversion`
   - Copy the source files from the repository to a new local directory called `gepetto`:
     `svn co svn+ssh://raid3/raid/svn-repo/gepetto/trunk/ gepetto`

3. Compile and install GEPETTO
   If the GEPETTO source files are obtained as described before the next steps are:

   (a) `cd` to the new directory `gepetto`.
   `cd gepetto`

   (b) Create a build directory, `cd` to it, execute the configure script, compile (`make`) and install (`make install`).

   ```
   mkdir build && cd build
   ../configure
   make && make install
   ```

   (c) If the Boost libraries and/or the FFTW library are installed in unusual directories (other than `/usr/lib` or similar) the following options have to be set when executing the configuration script. Note that the specified path (`/path/to/library-directory`) must contain the two directories `lib` and `include` which contain the library binaries and header files.

   ```
   --with-boost=/path/to/boost-directory
   --with-fftw3=/path/to/fftw3-directory
   ```

   e.g.

   ```
   ../configure --with-fftw3=\
   /scratch/michael/fftw-3.3alpha1/build_barrel
   ```

(d) If the compiled GEPETTO binary shall be executed on other machines it is necessary to link all libraries statically which requires the following options when executing the configuration script.
`--enable-static`
e.g.
`../configure --enable-static`

4. The following options for the configuration script ensures that debug code and symbols are included into the compiled binary.

   ```
   --enable-debug
   --enable-io_debug
   ```

   e.g.
   `../configure --enable-debug`

5. The following "make targets" are also implemented and can be invoked after execution of the configuration script:

   - Copy binaries to build directory
     `make install`
   - Create the DOXYGEN and QUICKBOOK HTML-documentation
     `make html`
   - Create the QUICKBOOK HTML-documentation
     `make html_qbk`
   - Create the QUICKBOOK HTML-documentation on one page (useful for printing)
     `make html_qbk_1p`
   - Create the ETAGS file for quick code navigation in EMACS and VI
     `make etags`
   - Remove the object files that were created during the compilation process
     `make clean`

7.4 Developer Guide: How-To

This section describes how to implement a new analysis method based on the existing program design. Its target readers are computer scientists with (little) knowledge in programming in C++. This "How-To" covers two objectives:

1. It enables the reader to develop a new analysis method for GEPETTO using existing data structures and algorithms.

2. It introduces her to the concepts and design of the program in a hands-on like approach.

A thorough and systematic description of the source code of GEPETTO can be found in the developer guide written by Gregor Neumayr [27] which is updated and included in the source code.

Preparation

For experimental development the folder `branches/ts` in the repository should be used (which by May 2012 is an exact copy of `trunk/classic`). Changes to the source code that are committed to the repository but turn out to corrupt the program will not affect the most stable version in the folder `trunk/classic`. As soon as the experimental code is tested and doesn't interfere with other parts of the program it can be added to the source code in the trunk. This can be done with SUBVERSION and will not be explained here.

Furthermore, it is convenient to compile GEPETTO with debug code and symbols included. The debug code mainly causes the program to run verbosely, i.e. to print a lot of information to the terminal from which GEPETTO is run. Therefore, the options `--enable-debug` and `--enable-io_debug` should be included when running the configure scripts during the installation procedure. Of course, the BOOST and FFTW libraries are still required to be installed beforehand (see Section 7.3).

```
svn co svn+ssh://raid3/raid/svn-repo/gepetto gepetto
cd gepetto/branches/ts
mkdir debug-build
cd debug-build
../configure --enable-debug --enable-io_debug
make
```

The debug symbols are stored to the executable `gepetto` and links machine code instructions to the original program source code. This permits to run the program line by line as in the source code using debug programs, like the GNU DEBUGGER, GDB. From experience it can be said that this will prove extremely useful and a brief description is given in Section 7.4.

First Debug Run

Now the executable, `gepetto`, is located in the build directory which is also the current working directory and can be executed

```
./gepetto
```

which prints out a brief description for the usage of the program. Among others the verbosity options are described which permit to fine-tune the amount of debug information printed by GEPETTO.

A script file has to be passed as an argument for a more meaningful use of GEPETTO. In this example a test-script and -system, `tip3_108_cube`, from the repository directory `gepetto/prefab/test` will be used:

```
cd ../../../prefab/test/scripts/tip3_108_cube
../../../../branches/ts/debug-build/gepetto test.gep
```

Before commenting on the debug information that gets printed to the terminal the script `test.gep` (see below) will be described briefly: The trajectory that shall be analysed is specified by the *.dcd*-file and the *.psf*-file, as well as by the coordinate frame selection (`firstframe`, `maxframes`, and `graining`) that shall be considered. The actual analysis is described by the task which includes a reference to the trajectory (`traj1`), the type and the observable, the core and surround selection, and the report filename.

```
trajectory traj1 {
  filelist {
    filepath "../../prefab/test/cases/tip3_108_cube/tip3_108_cube_*.dcd"
    filetype dcd_charmm
  }
  filelist {
    filepath "../../prefab/test/cases/tip3_108_cube/tip3_108_cube.psf"
    filetype psf
  }
  firstframe 1
  maxframes 0
  graining 100
}

task {
  trajectory traj1
  type timeseries
  observable com
  selection {
    type core
    ag { residuetype TIP3 residuenumber 70, 95 }
  }
  selection {
    type surround
    ag { residuetype TIP3 residuenumber 70, 95 }
  }
  report { filepath "test" }
}
```

This analysis script instructs GEPETTO to compute the centre-of-mass timeseries of the molecules with the residue name `TIP3` and residue numbers 70 and 95 and store them to the file `test.COM` (see below) in multiple columns: (1) time in ps, (2)-(4) x-, y-, and z-coordinate of residue 70 in $\mathring{A}$, (5)-(7) x-, y-, and z-coordinate of residue 95 in $\mathring{A}$.

```
0 3.91699 -5.48856 -7.49094 -1.17518 -2.85262 -4.83261
0.1 3.79069 -5.22563 -7.085 -1.53184 -2.66836 -4.7799
0.2 4.03266 -5.17584 -6.19942 -1.54506 -2.73242 -5.14329
0.3 4.16416 -5.31385 -5.97743 -0.815894 -2.54577 -5.16673
0.4 4.44379 -5.34672 -5.98038 -0.515953 -2.81362 -4.80185
0.5 4.07017 -5.21226 -6.44836 -0.984694 -2.81774 -4.94154
0.6 3.90552 -5.29705 -6.88662 -1.73315 -2.72483 -4.84345
0.7 3.95057 -5.64767 -7.28095 -1.85408 -2.29513 -4.66067
0.8 3.78561 -5.82157 7.21528 -1.68109 -2.33963 -4.80644
0.9 3.60411 -5.49928 6.87258 -1.68957 -2.54236 -5.11584
1 3.34843 -5.68801 7.02739 -2.20061 -2.86569 -4.93722
```

In fact, a single selection statement in `test.gep` would be sufficient to specify the molecules of interest. Nevertheless GEPETTO requires two selections, one of type `core` and the other `surround`, since it was initially conceived to calculate distances between two selections of test points, i.e. radial distribution functions.

Having described the input *.gep*-script and the report file it is time to turn to the debug information that is printed to the terminal. Even for small tasks the amount of printed debug information becomes huge and therefore only selected bits will be presented. The debug information is always preceded by a line stating `DEBUG` or `DEBUG: PARSER`, just as warnings, status-, logging- and other information-messages are preceded by lines stating `WARNINGS`, `STATUS`, `LOGS`, and `INFORMATION`, respectively.

As the computer executes the code of GEPETTO it follows the program flow that is described in Section 7.2 and depicted in Figure 4. This is documented in the debug information by clearly marked headings which will be discussed briefly one after the other.

`SCRIPT` and `TRANSLATION OF SCRIPT`
: information is generated during parsing where the text in the *.gep*-script is checked and translated into a parse tree, a data structure that reflects the grammar of the *.gep*-script language. This parse tree is printed nicely formatted with a root, branches, and leaves.

`SELECTIONS`
: information is printed while the atoms of the MD system are grouped as specified in the input script: `create()`. Furthermore, an integer valued flag (`FlagsN`) is assigned to each atomgroup based on the requirements of the task: calculation of the centre-of-mass. Usually it is not necessary to alter any code in this part of the program and hence the debug information is kept short by default.

`OPTIMIZATION`
: Optimisation constitutes an important part in GEPETTO and therefore relevant

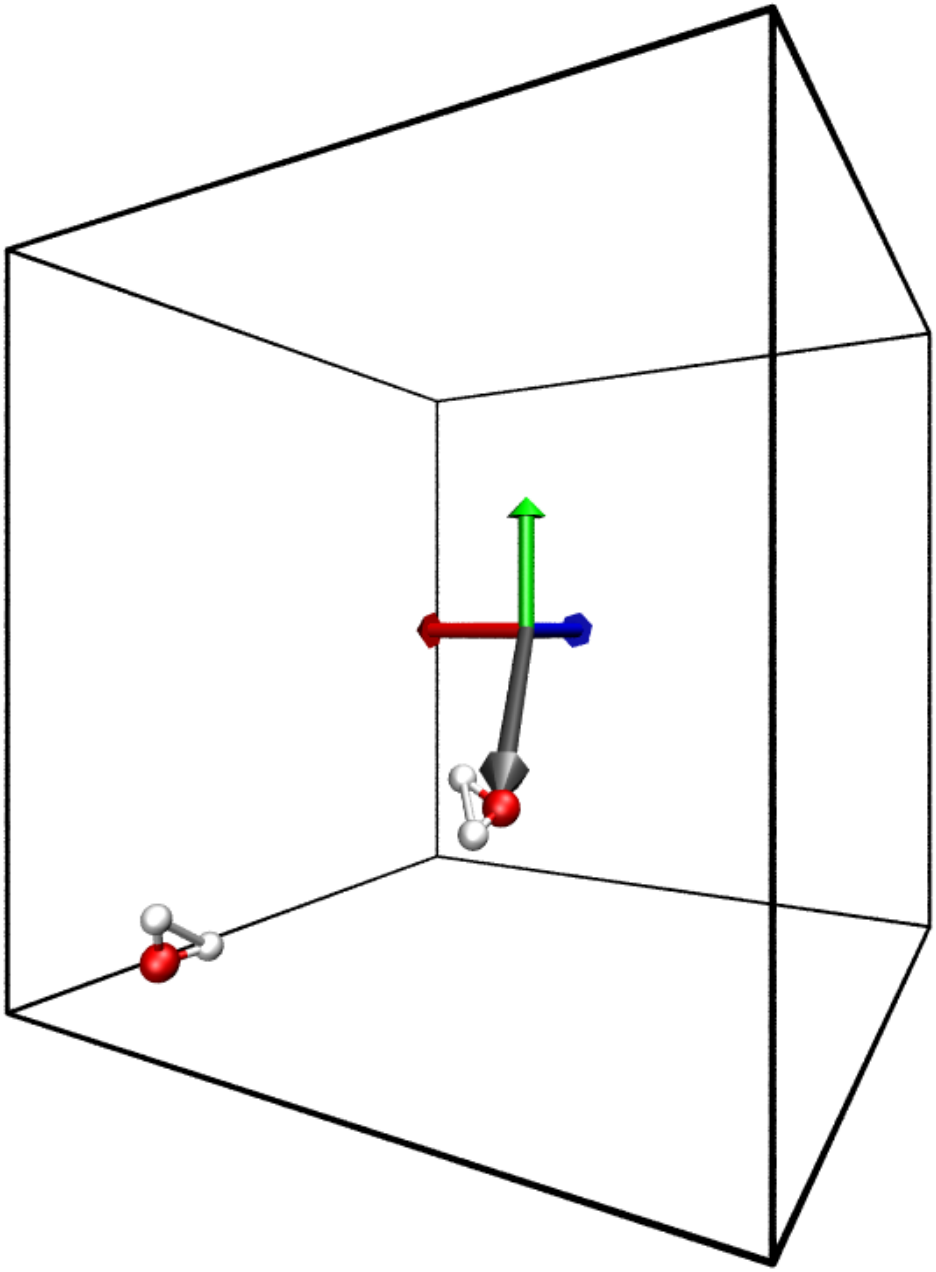


Figure 8: Snapshot of the trajectory of the system `tip3_108_cube`: Only the molecules with the residue name `TIP3` and residue numbers 70 and 95 are illustrated. Additionally, (i) the cubic simulation cell is outlined, (ii) the basis vectors at the origin are indicated by the green, the red, and the blue arrow, as well as (iii) the position vector of the centre-of-mass of a molecule is indicated by a gray arrow.

information is lengthy and split into multiple parts. From these, parts affecting the timeseries, the optimisation for algorithms of order $O(N)$ and $O(M)$, and tasks are of interest for this example:

```
OPTIMIZATION FOR TIMESERIES
OPTIMIZATION FOR O(N)
CALCULATIONFLAGGROUPS
TASKS
OPTIMIZATION FOR O(M)
```

`CALCULATION`
: The main function of GEPETTO is to calculate observables from the coordinates and/or velocities of selected atoms. This is done in the `CALCULATION` part: in this example the centre-of-mass of two particular atomgroups, the atoms having the residue type `TIP3` and the residue number 70, and those with residue number 95, for every 100th coordinate frame in the trajectory. In the debug information one can identify the names of the most important functions that are called during the calculation: `doCalculations()`, `calcN()`, `calcCOM()`, `algoM::calcM()`, `recCOM()`, as well as, in this case, the values that are stored to the data structure for the timeseries.

`POSTPROCESSING` and `WRITING RESULTS TO FILES`
: information is scarce since in this example there is actually no further treatment of the data required.

Implementation of New Analysis Method

Here, the steps required to implement a new analysis method are documented with the help of an example: The timeseries of the centre-of-mass (COM) and the dipole moment (DIP) of a molecule shall be calculated in the body-fixed (BF) frame of reference. Thus it is possible to see how the movements of the atoms of a single molecule influence the position of its COM as well as its DIP. To make things clear this task will be mathematically described and illustrated by graphics.

As in the previous section, the test-scripts and -cases from the repository directory `gepetto/prefab/test` will be used to check the progress of the development. For this task the MD system `LSG_EVOT_FACE_4_913` is appropriate since it features non-rigid molecules, such as EVOT (1-ethyl-3-methylimidazolium), FACE (trifluoroacetate), and the peptide LSG (LYS-SER-GLU). A snapshot of the trajectory of the test case MD system is displayed in Figure 9.

As for the mathematical definition, the COM of a molecule i is

$$\vec{r}_{\mathrm{COM},i} = \frac{1}{\sum_\alpha m_{i,\alpha}} \sum_\alpha \vec{r}_{i,\alpha} \cdot m_{i,\alpha} \tag{7}$$

where $\vec{r}_{i,\alpha}$ is the position vector and $m_{i,\alpha}$ the mass of atom α of molecule i. The DIP is defined with respect to the COM

$$\vec{\mu}_{\mathrm{i}} = \sum_\alpha (\vec{r}_{i,\alpha} - \vec{r}_{\mathrm{COM},i}) \cdot q_{i,\alpha} \tag{8}$$

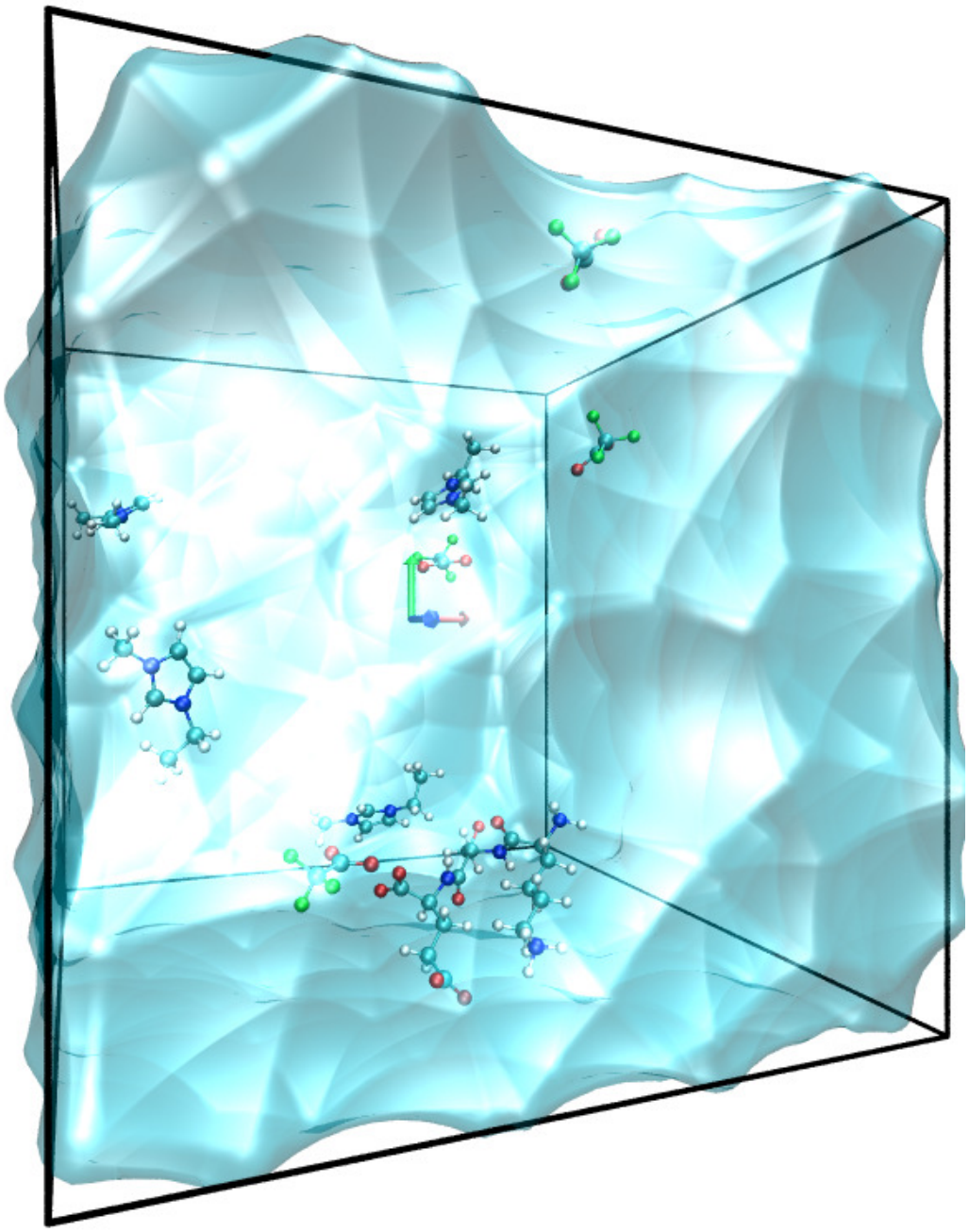


Figure 9: Snapshot of the trajectory of the test case MD-system: Non-rigid molecules (EVOT, FACE, and LSG) are displayed as CPK-models. The rigid water molecules (TIP3P model) are represented by their surface in light blue. The unit vectors along the x-, y-, and z-axis of the LF frame are displayed as a red, green, and blue arrow, respectively (the unit vectors are scaled by three for convenience).

where $q_{i,\alpha}$ is the partial charge of atom α of molecule i. Figure 10 illustrates the vector from the origin to the COM, $\vec{r}_{\mathrm{COM},i}$, as well as the DIP vector, $\vec{\mu}_{\mathrm{i}}$, with its initial point at the COM.

In short, the BF frame of a molecule i is defined by the translation vector and the rotation matrix that is calculated from a least-squares superposition from the molecule i to a reference molecule of the same type.

A rigorous definition requires a more detailed description: A set of points $S_{\mathrm{A}} = \{\vec{r}_{\mathrm{A},1}, ..., \vec{r}_{\mathrm{A},N}\}$ can be superimposed on a reference set with corresponding points $S_{\mathrm{B}} = \{\vec{r}_{\mathrm{B},1}, ..., \vec{r}_{\mathrm{B},N}\}$ such that every pair of points $\{\vec{r}_{\mathrm{A},\alpha}, \vec{r}_{\mathrm{B},\alpha}\}$ lies as close as possible. First the centre-of-geometry (COG) of both sets of points has to move to the origin, i.e. $\vec{r}_{\mathrm{A},\alpha} - \vec{r}_{\mathrm{COG,A}}$, and $\vec{r}_{\mathrm{B},\alpha} - \vec{r}_{\mathrm{COG,B}}$ for each α. Then, the set S_{A} has to be rotated such that the sum of the squared distances between corresponding points in the two sets S_{A} and S_{B} is minimised. For this optimisation problem the algorithm of Horn [28] is implemented in Gepetto. Altogether, the mathematical expression for the rotation matrix defining the BF frame is

$$R_{\mathrm{A}} = \arg\min_{R'_{\mathrm{A}}} \sum_{\alpha} \left((\vec{r}_{\mathrm{B},\alpha} - \vec{r}_{\mathrm{COG,B}}) - R'_{\mathrm{A}} \cdot (\vec{r}_{\mathrm{A},\alpha} - \vec{r}_{\mathrm{COG,A}}) \right)^2 \tag{9}$$

Now, the points $\vec{r}_{\mathrm{A},\alpha}$ can be identified as atomic coordinates of a molecule $i = \mathrm{A}$. For large molecules, such as proteins, it makes sense to select only a subset of its atoms for a superposition, e.g. only the C_α-atoms within secondary structures. As already mentioned, the rotation matrix R_{A} and the translation vector $\vec{r}_{\mathrm{COG,A}}$ define the BF frame for the molecule $i = \mathrm{A}$. The coordinates in the LF frame ($\vec{r}_{\mathrm{A},\alpha}$) can thus be transformed into coordinates in the BF frame ($\vec{r}^{\,\mathrm{BF}}_{\mathrm{A},\alpha}$) by

$$\vec{r}^{\,\mathrm{BF}}_{\mathrm{A},\alpha} = R_{\mathrm{A}}(\vec{r}_{\mathrm{A},\alpha} - \vec{r}_{\mathrm{COG,A}}) \tag{10}$$

The COM and the DIP vectors can be transformed analogously

$$\vec{r}^{\,\mathrm{BF}}_{\mathrm{COM,A}} = R_{\mathrm{A}}(\vec{r}_{\mathrm{COM,A}} - \vec{r}_{\mathrm{COG,A}}) \tag{11}$$

$$\vec{\mu}^{\,\mathrm{BF}}_{\mathrm{A}} = R_{\mathrm{A}}\vec{\mu}_{\mathrm{A}} \tag{12}$$

Figure 11 illustrates the BF frame which is defined by the least squares superposition of a subset of the atoms of EVOT, the imidazolium-ring atoms and the hydrogens directly bonded to the ring, to corresponding atoms of the reference. Moreover, the difference between the COM of the entire molecule and the COG of atom subset becomes evident when comparing Figure 10 and Figure 11: The gray arrow for the COM and the COG, respectively, points to different locations of the EVOT molecule.

Now that the description of the new analysis method is covered the focus can turn to the implementation in the source code of Gepetto. There is already boilerplate code provided, marked by `NEWANALYSIS`, that will be adapted in this example. The relevant files can be easily extracted using the command line tool grep to search for occurrences of the word `NEWANALYSIS` within the source code (see Section 7.4). Furthermore, Section 7.4 describes how to relate the source code to the program execution with the help of gdb.

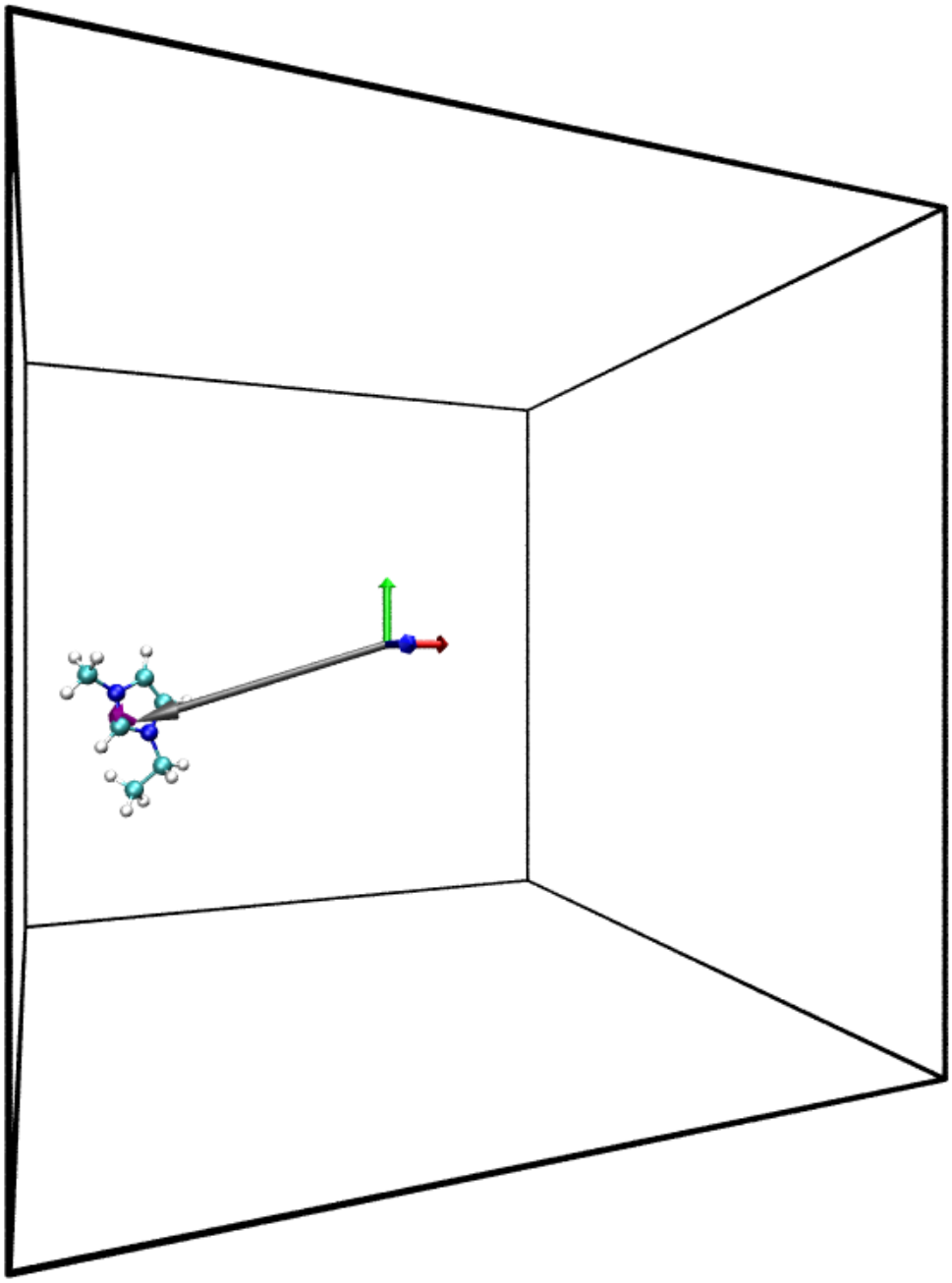


Figure 10: Same snapshot as in Figure 9: Only the molecule EVOT with the residue number 4 (as in the *.psf*-file) is displayed. Additionally, a gray arrow represents the vector pointing to the COM, $\vec{r}_{\mathrm{COM},4}$, and a purple arrow with its initial point at the COM represents the DIP, $\vec{\mu}_4$ (the DIP vector is scaled by three for convenience).

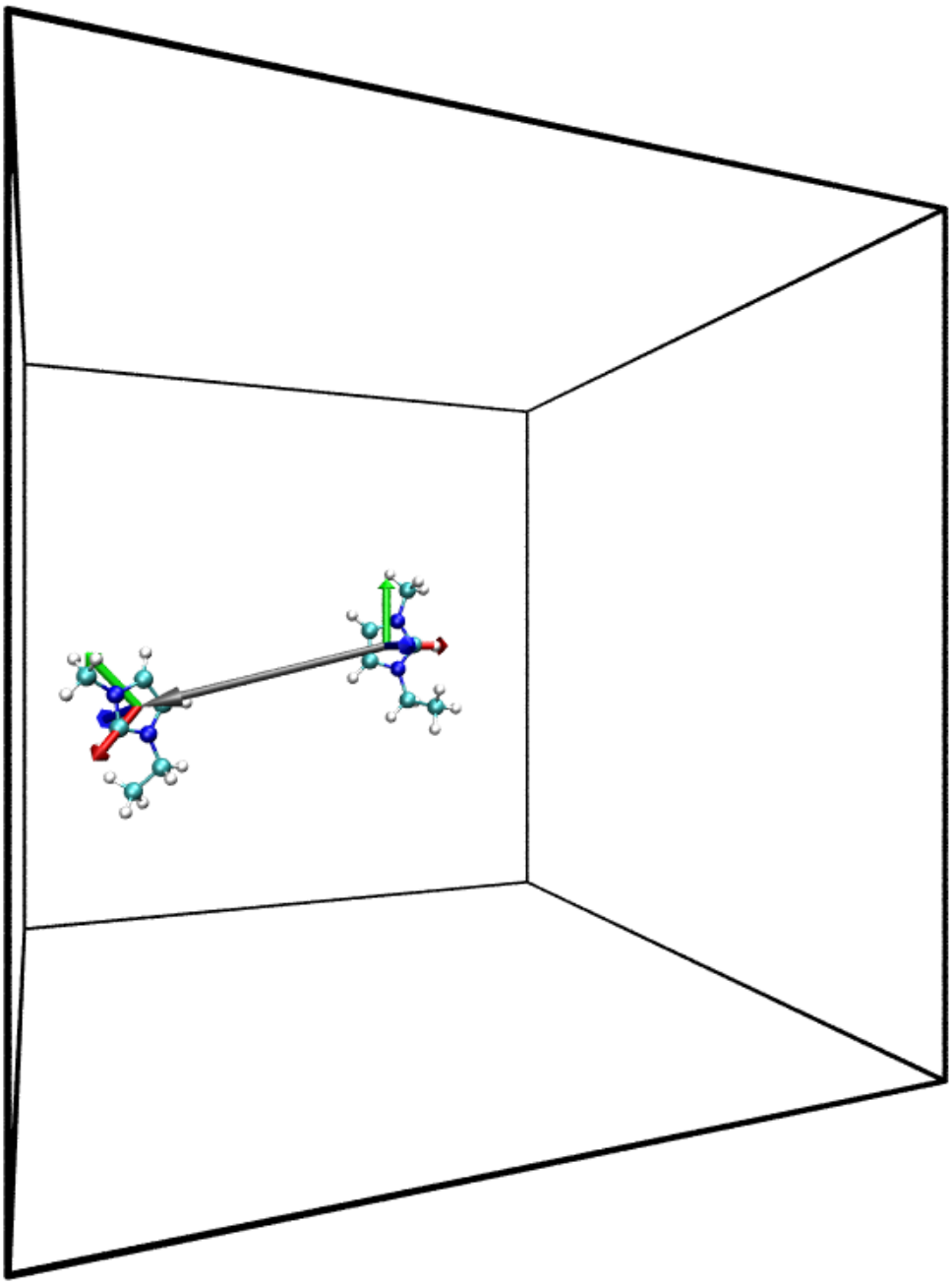


Figure 11: Same snapshot as in Figure 9: The molecule EVOT with the residue number 4 (as in the *.psf*-file) is displayed as well as the reference molecule used to define the BF frame (at the origin). A gray arrow is shown representing the vector pointing to the COG of the atoms used for the superposition, $\vec{r}_{\mathrm{COG,A}}$. The x-, y-, and z-axis of the LF frame are displayed in the middle of the simulation box, scaled by three, just as in the previous figures. Additionally, the unit vectors along the x-, y-, and z-axis of the BF frame are displayed, with coordinates in the LF frame.

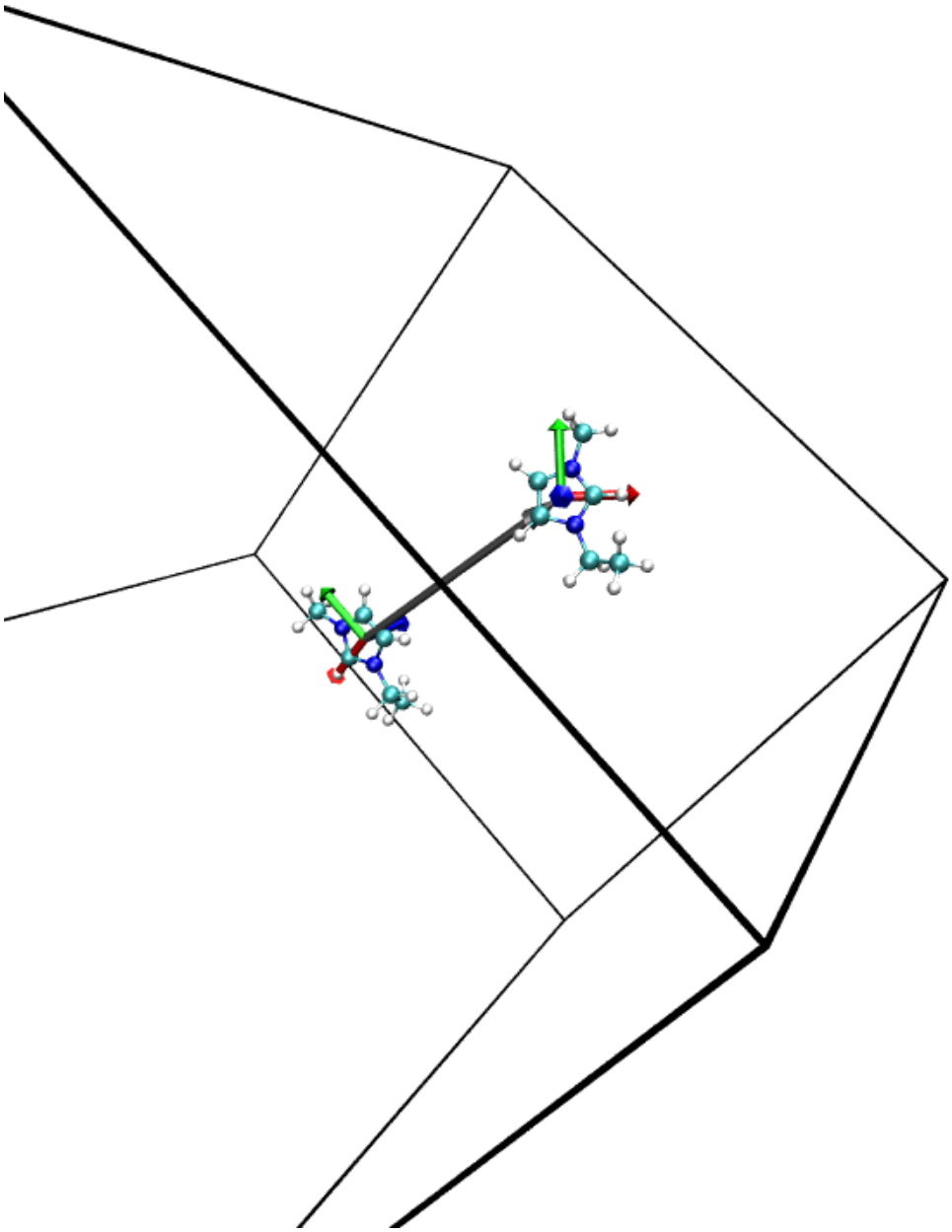


Figure 12: Same scene as Figure 11 but this time the molecule EVOT with the residue number 4 is moved to the centre and rotated such that the BF axes are oriented as the LF axes in Figure 11.

Two files from the source code have decisive impact:
`source/instruction/observable.cpp` and
`source/algorithm/algorithm_M.cpp`

`source/instruction/observable.cpp`
contains the source code for the class `Observable` which collects information of the observable that will be calculated, in this case the COM and DIP in the BF frame for a given atom selection. During the `TRANSLATION OF SCRIPT` part of the program flow an object `Task` will be created for each `Task` in the *.gep*-script and `Observable` is a member variable of `Task` corresponding to the value of the keyword `Observable` in the *.gep*-script. The following functions are invoked in the constructors of these classes.

First, the dimensions of the observable per trajectory frame have to be specified in order to set the memory size for the timeseries accordingly. The function `Observable::setDimensionType` sets these dimensions within a `switch` control structure based on the name of the observable.

```
void Observable::setDimensionType()
{
  switch( ObservableName )
  {
  case NEWANALYSIS:
    N1Type=Nr1Dim;
    N2Type=Nr1Dim;
    N3=3;
    break;
  // case statements for other observables
  case COM:
    N1Type=Nr1Dim;
    N2Type=NrCoreAtomGroups;
    N3=3;
    break;
  }
}
```

In terms of memory size, the new observable is going to be quite similar to `COM`; it has a value for each selected atomgroup (`N2Type=NrCoreAtomGroups`) and the value itself has multiple entries, 3 for the COM and 3 for the DIP, 6 in sum (`N3=6`). Therefore, the code will be adapted analogously to the observable `COM`.

```
  case NEWANALYSIS:
    N1Type=Nr1Dim;
    N2Type=NrCoreAtomGroups;
    N3=6;
    break;
```

Later in the program flow, during the `OPTIMIZATION` part the memory for the data structure to store the timeseries is allocated (`caTimeseries::resize`). At that point

the temporal and spatial SELECTION part has already finished and the number of coordinate frames and atomgroups is known so that it is possible to calculate the actually required memory size.

Secondly, flags can be set that make GEPETTO efficiently calculate the DIP and COM in the LF frame for each selected atomgroup. These basic calculations are frequently required and are executed at the very beginning of each iteration in the CALCULATION part. The function Observable::getFlagsN sets the flags again based on the name of the observable. The flags correspond to the bits of an unsigned int, that is if a bit is set to 0 or 1 (a design frequently used in C-programs).

```
unsigned int Observable::getFlagsN() const
{
  unsigned int f(0);
  switch( ObservableName )
  {
  case NEWANALYSIS:
  {
    f |= mode::COM;
    break;
  }
  // case statements for other observables
  }
  return f;
}
```

Here, mode::COM sets the first bit to 1 and causes GEPETTO to calculate the COM for each selected atomgroup. Analogously, mode::DIP has to be added to get the DIP as well. The values on the right-hand side of the |= operator, mode::COM and mode::DIP, are declared by the enumeration mode::CalculationFlag in the file source/globals.h. The case statement for NEWANALYSIS now looks this way

```
  case NEWANALYSIS:
  {
    f |= mode::COM;
    f |= mode::DIP;
    break;
  }
```

Eventually, during the OPTIMIZATION part the flags are used to select the functions to calculate the COM and DIP, algoN::GetFunctions, as well as the function that implements the new analysis method, algoM::GetFunction.

Before continuing, it makes sense to compile (make) and run GEPETTO in the debug mode and to pay attention to the altered debug information. For this purpose, the *.gep*-script newanalysis-3.gep in the scripts directory of the MD system LSG_EVOT_FACE_4_913 (see Section 7.4) instructs GEPETTO to calculate a timeseries of the observable newanalysis (which still has to be implemented).

```
gepetto newanalysis-3.gep
```

The altered debug information can be found it the parts

```
OPTIMIZATION FOR TIMESERIES
OPTIMIZATION FOR O(N)
CALCULATIONFLAGGROUPS
TASKS
CALCULATION
```

`source/algorithm/algorithm_M.cpp`
contains the source code for the functions that are called during the `CALCULATION` part of GEPETTO. The function `algoM::calcNEWANALYSIS` gets selected in `algoM::GetFunction` when the task in the *.gep*-script has the observable `NEWANALYSIS`.

Now the boilerplate code iterates over the atomgroups of the core selection and defines the index of the current atomgroup for the arrays containing the COM and DIP as the integer variable `j`. Furthermore, a statement to print debug information can be identified.

```
int algoM::calcNEWANALYSIS(Srp<calculationFlagGroupM>& cGFlag,
                           const Srp<caAtomGroup>& cAG,
                           const caStructure& caAGS)
{
#ifdef DEBUG
  out::debug <<std::level(VERBOSE_STANDARD) <<"algoM::calcNEWANALYSIS() "
             <<"calculationFlagGroupM="<<*cGFlag <<"\n";
#endif
  for(uint i(0);i<cGFlag->coreCollIdSize;i++)
  {
    uint j(cGFlag->coreCollationIds[i]);
  }
  return 0;
}
```

Possibly the easiest way is to copy code from similar functions, e.g. `algoM::recCOM` and insert it appropriately. The code below also includes some debug output that can explain some of the used variables.

```
uint N3=cGFlag->seriesHeader[8];
for(uint i(0);i<cGFlag->coreCollIdSize;i++)
{
    uint j(cGFlag->coreCollationIds[i]);
    cGFlag->seriesData[N3*i]  +=cAG->AgComX[j];
    cGFlag->seriesData[N3*i+1]+=cAG->AgComY[j];
    cGFlag->seriesData[N3*i+2]+=cAG->AgComZ[j];
#ifdef DEBUG
    out::debug <<i " " <<j <<" " <<cAG->AgComX[j]
               <<" "  <<cAG->AgComY[j]
               <<" "  <<cAG->AgComZ[j] <<"\n";
#endif
```

Again, at this point it makes sense to compile and run GEPETTO with the *.gep*-script `newanalysis-3.gep` and to pay attention to the altered debug information. Additionally, the *.gep*-script may be altered to investigate the effects on the program execution, for example the selections (`core` and `surround` have to be the same, as discussed in Section 7.4).

```
selection { type core
  ag { residuetype EVOT, FACE residuenumber * }
}
selection { type surround
  ag { residuetype EVOT, FACE residuenumber * }
}
```

Inspecting the report file, `newanalysis-3.NEWANALYSIS`, it becomes clear that only three of the six possible entries into the timeseries are filled namely those for the COM. The function `calcNEWANALYSIS` only added the COM, calculated in `algoN::calcCOM` and accessed by the variable `cAG->AgComX[j]`, etc. , to the timeseries, `cGFlag->seriesData[N3*i]`, etc. The COM is calculated with coordinates of the LF frame so it still has to be transformed to the BF frame and the same has to be implemented for the DIP.

So instead, the position vector for the COG of the atoms used for superposition has to be subtracted from the COM, as described in Equ. 11. This vector is stored to `cAG->AgTranslationX[j]`, etc. if a suited reference is given in the *.gep*-script.

```
    double tx = cAG->AgComX[j] - cAG->AgTranslationX[j];
    double ty = cAG->AgComY[j] - cAG->AgTranslationY[j];
    double tz = cAG->AgComZ[j] - cAG->AgTranslationZ[j];
```

Then, the resulting vector is rotated and stored to the entries of the timeseries.

```
cGFlag->seriesData[N3*i+0] +=
      tx * cAG->AgRotationXX[j] +
      ty * cAG->AgRotationXY[j] +
      tz * cAG->AgRotationXZ[j];
cGFlag->seriesData[N3*i+1] +=
      tx * cAG->AgRotationYX[j] +
      ty * cAG->AgRotationYY[j] +
      tz * cAG->AgRotationYZ[j];
cGFlag->seriesData[N3*i+2] +=
      tx * cAG->AgRotationZX[j] +
      ty * cAG->AgRotationZY[j] +
      tz * cAG->AgRotationZZ[j];
```

The DIP is position independent and therefore only has to be rotated.

```
cGFlag->seriesData[N3*i+3] +=
      cAG->AgDipX[j] * cAG->AgRotationXX[j] +
      cAG->AgDipY[j] * cAG->AgRotationXY[j] +
      cAG->AgDipZ[j] * cAG->AgRotationXZ[j];
```

```
cGFlag->seriesData[N3*i+4] +=
      cAG->AgDipX[j] * cAG->AgRotationYX[j] +
      cAG->AgDipY[j] * cAG->AgRotationYY[j] +
      cAG->AgDipZ[j] * cAG->AgRotationYZ[j];

cGFlag->seriesData[N3*i+5] +=
      cAG->AgDipX[j] * cAG->AgRotationZX[j] +
      cAG->AgDipY[j] * cAG->AgRotationZY[j] +
      cAG->AgDipZ[j] * cAG->AgRotationZZ[j];
```

Compiling and running GEPETTO with the *.gep*-script `newanalysis-3.gep` will produce a segmentation fault, i.e. the program tries to access memory that hasn't been reserved for that execution of GEPETTO. As indicated, a reference for superposition is missing in the *.gep*-script which has to be added in form of a *.pdb*-file.

```
  selection { type core
    ag { residuetype EVOT residuenumber 4 }
    orient { filepath "../../system/lsg_evot_face_4_913/evot_ring.-1.pdb" }
  }
  selection { type surround
    ag { residuetype EVOT residuenumber 4 }
  }
```

For EMIM, for instance, it may be reasonable to only use the atoms of the imidazolium-ring as reference, therefore the other atom entries in the *.pdb*-file can be deleted. Furthermore, the residue ID entry (residue sequence number, defined by www.wwpdb.org) has to be set to `-1` to use the reference for every atomgroup of the core selection.

The segmentation fault can also be investigated by means of the debugger GDB.

```
gdb ./gepetto
...
(gdb) run newanalysis-3.gep
...
Program received signal SIGSEGV, Segmentation fault.
0x0000000000415f45 in algoM::calcNEWANALYSIS (cGFlag=..., cAG=...,
    caAGS=...) at ../../../source/algorithm/algorithm_M.cpp:446
446     double tx = cAG->AgComX[j] - cAG->AgTranslationX[j];
```

The commands `bt`, `l`, and `p` (see Section 7.4) provide further clues on the origin of the segmentation fault. With `bt` the program stack in memory is printed and with `p` the values of the variables can be examined. Note that the operator `->` cannot be used in connection with `Srp` in GDB.

```
(gdb) bt
#0  0x0000000000415f45 in algoM::calcNEWANALYSIS (cGFlag=..., cAG=...,
    caAGS=...) at ../../../source/algorithm/algorithm_M.cpp:446
#1  0x0000000000413033 in algoM::calcM (cFGT=..., cHB=..., postN_AGV=...,
    preN2_AGV=..., caAS=..., caAGS=...)
    at ../../../source/algorithm/algorithm_M.cpp:49
```

```
#2  0x000000000046c32b in FrameType::doCalculations (this=0x96e140,
    curFrames=..., calculationH=...)
    at ../../../source/instruction/frametype.cpp:80
#3  0x00000000004c3745 in TaskList::calcStuff (this=0x87a340)
    at ../../../source/instruction/tasklist.cpp:634
#4  0x0000000000409976 in main (argc=2, argv=0x7fffffffe1b8)
    at ../source/program/main.cpp:117
(gdb) l
441 {
442 uint j(cGFlag->coreCollationIds[i]);
443 // cGFlag->seriesData[N3*i]  +=cAG->AgComX[j];
444 // cGFlag->seriesData[N3*i+1]+=cAG->AgComY[j];
445 // cGFlag->seriesData[N3*i+2]+=cAG->AgComZ[j];
446     double tx = cAG->AgComX[j] - cAG->AgTranslationX[j];
447     double ty = cAG->AgComY[j] - cAG->AgTranslationY[j];
448     double tz = cAG->AgComZ[j] - cAG->AgTranslationZ[j];
449 cGFlag->seriesData[N3*i+0] +=
450       tx * cAG->AgRotationXX[j] +
(gdb) p tx
$1 = 2.1016366816536609e-317
(gdb) p cAG->AgTranslationX[j]
There is no member or method named AgTranslationX.
(gdb) p cAG
$2 = (const Srp<caAtomGroup> &) @0x96e210: {Helper = 0x96e880,
    Coupled = true}
(gdb) p (*cAG).AgTranslationX[j]
Cannot access memory at address 0x0
(gdb) p (*cAG).AgTranslationX
$3 = (data_t *) 0x0
```

The last lines identify the variable `cAG->AgTranslationX` as set to `NULL`. This variable corresponds to $\vec{r}_{\mathrm{COG,A}}$ and GEPETTO needs a reference coordinate set to correctly calculate these.

As soon as the *.gep*-script is fixed the execution proceeds normal with additional output from the superposition function `algoN::calcORI`: first, the translation vector `TRLxyz` and the rotation matrix `ROTxx..zz` that define the BF frame and second, the atoms of each atomgroup transformed to the BF frame and in *.pdb*-format. Figure 13 is created using these coordinates as well as the regular output, COM and DIP in the BF frame, $\vec{r}^{\,\mathrm{BF}}_{\mathrm{COM,A}}$ and $\vec{\mu}^{\mathrm{BF}}_{\mathrm{A}}$. Figure 14 displays the timeseries of the COM and DIP in the BF frame of all four EVOT atomgroups as found in the report file `newanalysis-3.NEWANALYSIS`.

Finding a Word with grep

Assuming that the current directory is the same as in Section 7.4 the command

```
grep -r NEWANALYSIS ../source/ | grep -v svn
```

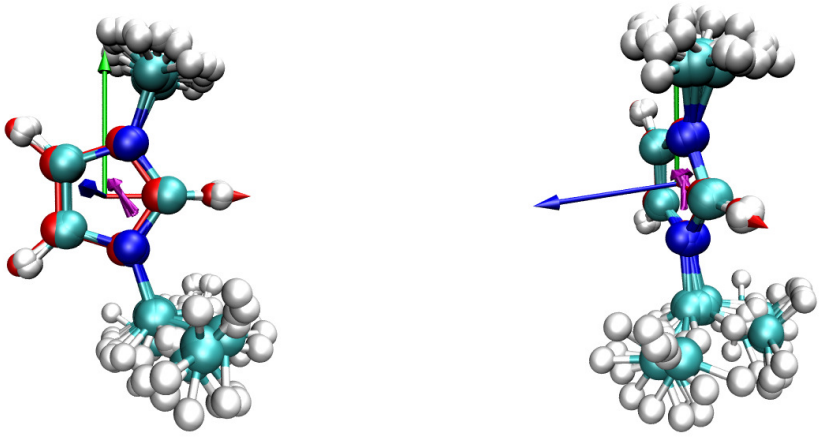


Figure 13: Superposition of the 4 EVOT molecules at $t \in \{0\text{ps}, 5\text{ps}, 10\text{ps}\}$, as well as the reference, the imidazolium ring atoms, in red. Additionally, the respective dipole moment vectors are drawn in purple with their initial points at the COM and the unit vectors along the BF frame axes (scaled by three) are included. On the right side the view point is shifted to display the conformations of the ethyl-group.

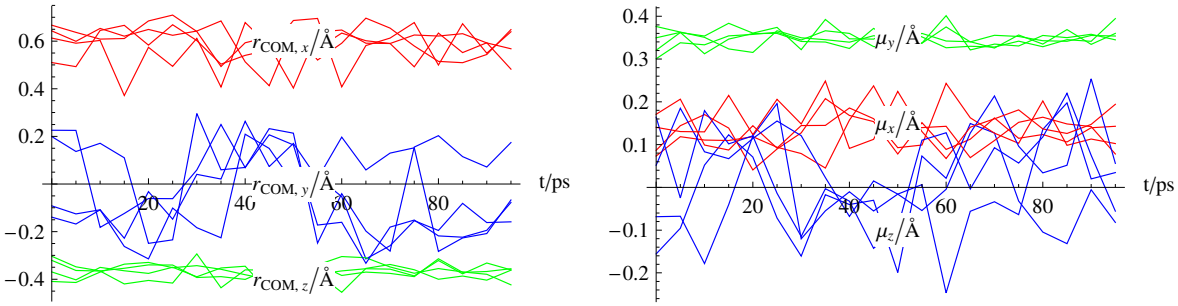


Figure 14: Timeseries of the BF frame COM and DIP coordinates of the 4 EVOT molecules. Note that the colours correspond to the unit vectors along the BF frame axes in Figure 13.

lists all lines of the files in the source directory that contain the word `NEWANALYSIS`.

Relating the Source Code to the Program Execution using gdb

If GEPETTO is installed with the debug option as described in Section 7.4 it is possible to investigate the program execution with the debugging program GDB (The GNU Project Debugger). One of many features of GDB is that it allows to execute the program line by line just as written in the source code. The command

```
gdb ./gepetto
```

starts the debugger and loads the debug information from GEPETTO. A breakpoint can be set at a particular code fragment in order to investigate it closely. In Section 7.4 the function `Observable::setDimensionType` is of interest and the next command tells GDB to stop the execution of GEPETTO when this function is invoked

```
(gdb) break Observable::setDimensionType
```

GEPETTO requires a *.gep*-script file with a task to follow the program flow as described in Section 7.2 and Figure 4. For this example the *.gep*-script in file `newanalysis-3.gep` will be used. With the next command GDB executes GEPETTO until it reaches a breakpoint

```
(gdb) run ../test/scripts/newanalysis-3.gep
```

Now, the functions and variables that are in the current scope are can be investigated interactively and the program execution can be continued line by line:

`list` or `l` prints the source code corresponding to the machine code in execution.

`print ObservableName` or `p ObservableName` prints the value of a member variable.

`backtrace` or `bt` prints the functions that have been called to reach the current breakpoint starting in `main()` (a stack frame display) and the file and line in the source code that correspond (see below).

`frame x` or `fr x` selects the stack frame `x`. The stack frame listing is given by `bt`.

`next` or `n` executes the machine code corresponding to the next line in the source code in the current scope.

`step` or `s` executes the machine code corresponding to the next line in the source code which may be in another function and a new scope.

`continue` or `c` continues to program execution.

The result of the command `bt` allows to relate the source code to the program flow. The frame number is in the left most column, followed by the function name, the file name, and the line number.

```
#0  Observable::setDimensionType (this=0x7fffffffd980)
    at ../../../source/instruction/observable.cpp:53
#1  0x0000000000477ffe in Observable (this=0x7fffffffd980,
    A=Observable::NEWANALYSIS)
    at ../../../source/instruction/observable.h:95
#2  0x00000000005049f0 in Script_Gep::toObservable (this=0x8795b0,
    token=@0x954b10) at ../../../source/script/gep/gep.cpp:84
#3  0x000000000050fa74 in Script_Gep::translate (this=0x8795b0,
    aResource=...) at ../../../source/script/gep/gep.cpp:1057
#4  0x00000000004bec42 in TaskList::parse (this=0x879340)
    at ../../../source/instruction/tasklist.cpp:27
#5  0x00000000004091ee in main (argc=2, argv=0x7fffffffe1b8)
    at ../source/program/main.cpp:93
```

newanalysis-3.gep

```
trajectory traj1 {
  filelist {
    filepath "../../system/lsg_evot_face_4_913/lsg_evot_face_4_913_3.dcd"
    filetype dcd_charmm
  }
  filelist {
    filepath "../../system/lsg_evot_face_4_913/lsg_evot_face_4_913.psf"
    filetype psf
  }
  firstframe 1
  maxframes 0
  graining 50
}

task {
  trajectory traj1
  type timeseries observable newanalysis
  selection { type core
    ag { residuetype EVOT residuenumber * }
#    orient { filepath "../../system/lsg_evot_face_4_913/evot_ring.-1.pdb" }
  }
  selection { type surround
    ag { residuetype EVOT residuenumber * }
  }
  report { filepath "newanalysis-3" }
}
```

References

[1] KR Seddon. Room-temperature ionic liquids: Neoteric solvents for clean catalysis. *Kinetics and Catalysis*, 37(5):693–697, SEP-OCT 1996. Conference on New Trends in Chemical Kinetics and Catalysis, Novosibirsk, Russia, Nov 21-24, 1995.

[2] Christian Schröder. *Computational studies of molecular ionic liquids*. 2011.

[3] Nicola J. Roberts, Amanda Seago, John S. Carey, Richard Freer, Christopher Preston, and Gary J. Lye. Lipase catalysed resolution of the lotrafiban intermediate 2,3,4,5-tetrahydro-4-methyl-3-oxo-1h-1,4-benzodiazepine-2-acetic acid methyl ester in ionic liquids: comparison to the industrial t-butanol process. *Green Chem.*, 6:475–482, 2004.

[4] Michael Zavrel, Daniela Bross, Matthias Funke, Jochen Büchs, and Antje C. Spiess. High-throughput screening for ionic liquids dissolving (ligno-)cellulose. *Bioresource Technology*, 100(9):2580 – 2587, 2009.

[5] Natalia V Plechkova and Kenneth R Seddon. Applications of ionic liquids in the chemical industry. *Chem Soc Rev*, 37(1):123–150, Jan 2008.

[6] J. Gorke, F. Srienc, and R. Kazlauskas. Toward advanced ionic liquids. polar, enzyme-friendly solvents for biocatalysis. *Biotechnology and Bioprocess Engineering*, 15(1):40–53, 2010.

[7] M. Moniruzzaman, K. Nakashima, N. Kamiya, and M. Goto. Recent advances of enzymatic reactions in ionic liquids. *Biochemical Engineering Journal*, 48(3):295–314, February 2010.

[8] Raffaele Saladino, Claudia Crestini, Samanta Pino, Giovanna Costanzo, and Ernesto Di Mauro. Formamide and the origin of life. *Physics of Life Reviews*, 9(1):84 – 104, 2012.

[9] M. Feig. *Modeling Solvent Environments: Applications to Simulations of Biomolecules*. Wiley-VCH, 2010.

[10] A. D. McNaught and A. Wilkinson. Iupac. compendium of chemical terminology, 2nd ed. (the "gold book"). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata.

[11] Johnny Deschamps, Margarida F. Costa Gomes, and A. A. H. Padua. Molecular simulation study of interactions of carbon dioxide and water with ionic liquids. *ChemPhysChem*, 5(7):1049–1052, 2004.

[12] Gerhard König. *Methodological Studies concerning Free Energy Simulations*. PhD thesis, University of Vienna, 2010.

[13] C. Hardacre, J. D. Holbrey, M. Nieuwenhuyzen, and T. G. A. Youngs. Structure and solvation in ionic liquids. *Acc Chem Res*, 40(11):1146–1155, Nov 2007.

[14] Alexander Wulf, Ralf Ludwig, Padmanabhan Sasisanker, and Hermann Weingärtner. Molecular reorientation in ionic liquids: A comparative dielectric and magnetic relaxation study. *Chemical Physics Letters*, 439(4-6):323 – 326, 2007.

[15] Hiroyuki Tokuda, Kikuko Hayamizu, Kunikazu Ishii, Abu Bin Hasan Susan, and Masayoshi Watanabe. Physicochemical properties and structures of room temperature ionic liquids. 1. variation of anionic species. *J. Phys. Chem. B*, 108:16593–16600, 2004.

[16] A. Samanta. Solvation dynamics in ionic liquids: What we have learned from the dynamic fluorescence stokes shift studies. *The Journal of Physical Chemistry Letters*, 1(10):1557–1562, 2010.

[17] C. Schröder, C. Wakai, H. Weingärtner, and O. Steinhauser. Collective rotational dynamics in ionic liquids: a computational and experimental study of 1-butyl-3-methyl-imidazolium tetrafluoroborate. *J. Chem. Phys.*, 126(8):084511, Feb 2007.

[18] M. P. Allen and D. J. Tildesley. *Computer simulations of liquids*. Oxford Press, New York, 1989.

[19] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education, 2001.

[20] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: the biomolecular simulation program. *J Comput Chem*, 30(10):1545–1614, Jul 2009.

[21] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[22] Michael Haberler. Computing the static conductivity of ionic liquids. Master's thesis, University of Vienna, 2008.

[23] C. Schröder, G. Neumayr, and O. Steinhauser. On the collective network of ionic liquid/water mixtures. iii. structural analysis of ionic liquids on the basis of voronoi decomposition. *J Chem Phys*, 130(19):194503, May 2009.

[24] M. Haberler, C. Schröder, and O. Steinhauser. Solvation studies of a zinc finger protein in hydrated ionic liquids. *Phys. Chem. Chem. Phys.*, 13:6955–6969, 2011.

[25] Michael Haberler and Othmar Steinhauser. On the influence of hydrated ionic liquids on the dynamical structure of model proteins: a computational study. *Phys. Chem. Chem. Phys.*, 13:17994–18004, 2011.

[26] Michael Haberler, Christian Schröder, and Othmar Steinhauser. Hydrated ionic liquids with and without solute: The influence of water content and protein solutes. 2012. in revision.

[27] Gregor Neumayr. *Structural Decomposition and Structural Relaxation of Solvation Shells of Hydrated Molecular Ionic Liquids and Protein Solutions.* PhD thesis, University of Vienna, 2010.

[28] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, Apr 1987.

8 CV

Michael Haberler

Kleingasse 22/21, 1030 Vienna – Tel.: +43 676 9404774 – Email: michael.haberler@gmx.at

Research Experience

05/2008 – 07/2012 **Department of Computational Biological Chemistry, University of Vienna**
Thesis: *Solvation of Biomolecules in Ionic Liquid - Water Mixtures*
Adviser: Prof. O. Steinhauser
Synopsis: Conducted and analysed molecular dynamics computer simulations of zinc finger protein and ubiquitin in ionic liquid - water mixtures.

03/2007 – 03/2008 **Department of Computational Biological Chemistry, University of Vienna**
Diploma Thesis: *Computing the Static Conductivity of Ionic Liquids*
Adviser: Prof. O. Steinhauser
Synopsis: Analysed molecular dynamics trajectories of ionic liquids to compute their electrical conductivity.

02/2006 – 05/2006 **Department of Physical Chemistry, University of Leicester**
Adviser: Prof. J. Woodward
Synopsis: Investigated radical reaction mechanism of acylphosphine oxide by laser flash photolysis and time resolved EPR-spectroscopy.

10/2005 – 01/2006 **Department of Biochemistry, University of Leicester**
Adviser: Prof. M. Pfuhl
Synopsis: Created and purified myosin mutants to investigate interaction with myosin binding protein by NMR-spectroscopy.

07/2005 – 08/2005 **Department of Genetics, University of Vienna**
Adviser: Prof. M. Jantsch
Synopsis: Conducted crossing experiments and RNAi screens on C. Elegans mutants displaying defects during mitosis to map responsible gene.

Teaching Experience

09/2008 – 01/2009, 09/2009 – 01/2010 **University of Vienna**
Teaching Assistant "Mathematics for Molecular Biologists"

Education

05/2008 – 07/2012 **Doctor of Science, Molecular Biology**, University of Vienna
Expected date of graduation: July 4th, 2012

09/2006 – **BSc, Physics**, University of Vienna
61/180 finished

09/2002 – 03/2008 **Magister rer. nat., Molecular Biology**, University of Vienna
Graduated with Distinction