Protein Hydration Shells Studied by Molecular Dynamics Simulations and Small-Angle Scattering: Structural, Thermal, and Surface Effects

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Faktultät
der Universität des Saarlandes

von
Johanna-Barbara Linse

Saarbrücken
2026

Tag des Kolloquiums:	28.05.2026
Dekan:	Prof. Dr.-Ing. Dirk Bähre
Berichterstatter:	Prof. Dr. Jochen S. Hub
	Prof. Dr. Martin H. Müser
Vorsitz:	Prof. Dr. Gregor Jung
Akad. Mitglied:	Dr. Martin Brinkmann

Acknowledgments

Over five years ago, I was given the incredible opportunity to begin my PhD journey. During this time, much has changed – both scientifically and personally. There were challenges, uncertainties, and moments of doubt, but I was never alone. Many remarkable people supported, accompanied, and inspired me along the way. This thesis, like any scientific work, is therefore not the achievement of a single person but the result of collaboration, encouragement, and shared effort. I would therefore like to take this opportunity to thank all those who contributed to this work and to my development during these years.

First and foremost, I would like to thank Jochen, my supervisor. Thank you for trusting me over five years ago to bring this exciting project to completion and for giving me the chance to embark on this wonderful PhD journey. Your unwavering belief in my abilities and your insightful guidance have been the cornerstones of this work. Your patience, encouragement, and critical perspective not only enriched my research but also helped me grow into a more independent and passionate scientist. Your door was always open, and I am deeply grateful for the freedom you gave me, as well as for the opportunities to present my work at international conferences.

Next, I would also like to thank Martin Müser for kindly agreeing to serve as the second reviewer of this thesis. I am very grateful for your time, your interest in my work, and your willingness to take on this role.

Furthermore, I would like to sincerely thank all collaborators of my publications, in particular Philip A. Anfinrud, Hyun Sun Cho, and Friedrich Schotte, for the wonderful collaboration and the many insightful discussions from which I learned a lot.

I am incredibly grateful to have been part of such a wonderful research group. Thank you to everyone who accompanied me for all or part of my PhD journey. Therefore, I would like to thank all current and former members of the Computational Biology group with whom I had the pleasure to work: Miloš, Leonie, Robert, Katharina, Jeremy, Chetan, Gari, Maciej, Noora, Joel, Tobias, Alejandro, Leonhard, Lukas, Kristian and many others. Thank you for the collaborative spirit, the mutual support, and for simply making everyday research life more enjoyable. I will also fondly remember our car rides, during which some of the most creative – and occasionally quite crazy – ideas emerged, from unconventional conference hats to future group

events (archery on horseback still comes to mind). And I will always cherish the memories of evenings in pubs, group retreats, excursions, company runs, Christmas parties, conferences, and countless shared moments beyond science. I consider myself very fortunate to have walked this path with you.
Miloš, thank you for helping me with all the tools during my first months as a PhD student, and thank you for all the conversations we had at conferences where we met again. Maciej, thank you for all your help with the software and computational tools. Without your support, parts of the project would have taken considerably longer. Leonie, thank you for always taking the time for discussions about SAXS and SANS and for the wonderful time we spent together in Greece. Katharina, thank you for all the conversations we had in the office, for your help in organizing group events and for always having an open ear, especially when things did not go as planned. Gari, thank you for keeping the office well supplied with homemade baked goods and for the hikes. Chetan, thank you for the many scientific and non-scientific conversations, often just before one of us was about to leave the office, and for listening and giving advice whenever I felt stuck or out of ideas.
I would also like to thank Bettina for her invaluable help with countless forms, contract extensions, applications, and administrative matters, as well as for the many pleasant non-scientific conversations in her office or in the hallway.
Many thanks to Maciej and Christian for taking care of the computing resources – clusters and nodes that are essential for our work – and for always helping when something stopped working.
I am grateful to Jochen, Katharina, Leonie, and Hendrik for proofreading (parts of) my thesis and for their helpful comments and suggestions.
Outside the institute, I would like to thank my friends and roommates, including Julius, Alex, Lisa-Marie, Luka, Nathalie, and many others, for their support, patience, and for reminding me that there is a life beyond the PhD.
Zuletzt möchte ich meiner Familie von ganzem Herzen danken. Danke, dass ihr mir zugehört habt, wenn ich wieder einmal über Dinge gesprochen habe, die ihr oft gar nicht (vollständig) verstanden habt. Allein durch euer Zuhören haben sich manchmal neue Ideen entwickelt oder Lösungen meiner Probleme ergeben. Eure bedingungslose Liebe und Unterstützung haben mich durch diese Zeit getragen. Vielen Dank Petra, Roland, Jonathan, Marie-Therese und Sophia-Selina, dass ihr mich immer unterstützt und mir Rückhalt gegeben habt. Ohne euch wäre diese Arbeit nicht möglich gewesen. Dieser Erfolg ist daher ebenso euer Verdienst wie meiner.

Abstract

Hydration shells surrounding proteins are crucial for biological functions because they influence the structure, dynamics and interactions of proteins. To understand these hydration shells, we combined molecular dynamics (MD) simulations with models of light and heavy water and small angle X-ray and neutron scattering (SAXS/SANS) experiments. Modified heavy water pair potentials – SPC/E-HW, TIP3P-HW, TIP4P/2005-HW – reproduce important physicochemical differences to light water, enabling accurate prediction of biophysical experiments with heavy water. Contrast analysis from experiments and simulations across different force fields reveals that the properties of hydration shells depend on the protein shape, charge and (surface) amino acid composition. Overall several tested force fields in MD simulations reproduce the hydration shell contrasts in remarkable agreement with experiments. Simulations also show that acidic and polar residues enhance the hydration contrast, while apolar residues cause water depletion. Intrinsically disordered proteins (IDPs) exhibit weaker, conformation-dependent hydration compared to folded domains. Temperature-ramp SAXS experiments show that the hydration shells weaken with increasing temperature due to increased disorder and displacement of surface-bound water. These findings emphasize the structural specificity of hydration shells and their detectability through SAXS/SANS and MD simulations.

Zusammenfassung

Hydrationshüllen um Proteine beeinflussen deren Struktur, Dynamik und Wechselwirkungen maßgeblich. Um Hydrationsschalen zu verstehen, kombinieren wir Molekulardynamik (MD)-Simulationen mit leichten und schweren Wassermodellen sowie Kleinwinkel-Röntgen- und -Neutronenstreuexperimenten (SAXS/SANS). Modifizierte Paarpotenziale für schweres Wasser reproduzieren wichtige physikochemische Unterschiede zu leichtem Wasser und ermöglichen so eine genaue Vorhersage von biophysikalischen Experimenten mit schwerem Wasser. Kontrastanalysen von Experimenten und Simulationen mit verschiedenen Kraftfeldern zeigen, dass die Eigenschaften der Hydrationsschalen von der Form, Ladung und (Oberflächen-)Aminosäurezusammensetzung der Proteine abhängen. Die meisten getesteten Kraftfelder in Simulationen reproduzieren die Hydrationsschalenkontraste in bemerkenswerter Übereinstimmung mit Experimenten. Simulationen zeigen zudem, dass saure und polare Aminosäurereste den Hydrationskontrast verstärken, während apolare Reste Wasserverarmung verursachen. Intrinsisch ungeordnete Proteine weisen eine schwächere, konformationsabhängige Hydration auf als gefaltete Proteine. Temperaturrampen-SAXS-Experimente zeigen, dass Hydrationsschalen mit steigender Temperatur aufgrund erhöhter Unordnung und Verdrängung von oberflächengebundenem Wasser schwächer werden. Diese Ergebnisse unterstreichen die strukturelle Besonderheit der Hydrationsschalen und ihre Nachweisbarkeit durch SAXS/SANS und MD-Simulationen.

Contents

List of Figures

List of Tables

Acronyms

MD molecular dynamics
SAS small-angle scattering
SAXS small-angle X-ray scattering
SANS small-angle neutron scattering
WAXS wide-angle X-ray scattering
SWAXS small- and wide-angle X-ray scattering
SEC-SAXS size-exclusion-chromatography SAXS
IDP intrinsically disordered proteins
IDR intrinsically disordered regions
DNA deoxyribonucleic acid
RNA ribonucleic acid
NMR nucleic magnetic resonance
cryo-EM cryo-electron microscopy
vdW van der Waals forces
LJ Lennard-Jones
QM quantum mechanical
PBC periodic boundary condition
PDB protein data bank
PDDF pair distance distribution function
LINCS LINear Constraint Solver
CHARMM Chemistry at Harvard Molecular mechanics
AMBER Assisted Model Building with Energy Refinement
PME Particle-mesh Ewald summation
Å Ångström
$\boldsymbol{R_g}$ radius of gyration
$\boldsymbol{I(0)}$ forward scattering (or zero-angle) intensity

Publications

Paper I
Johanna-Barbara Linse, Jochen S. Hub
Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW
Journal of Chemical Physics, Volume 154, Issue 19, 194501 (2021)

Paper II
Johanna-Barbara Linse, Jochen S. Hub
Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data
Communications Chemistry, Volume 6, Article number: 272 (2023)

Paper III
Johanna-Barbara Linse, Tobias M. Fischbach, Jochen S. Hub
How protein hydration depends on amino acid composition, peptide conformation, and force fields
Biophysical Journal, Volume 125, Issue 1, Pages 255-269 (2026)

Paper IV
Johanna-Barbara Linse, Hyun Sun Cho, Friedrich Schotte, Philip A. Anfinrud, Jochen S. Hub
Depletion of the protein hydration shell with increasing temperature observed by small-angle X-ray scattering and molecular simulations
Journal of the American Chemical Society, Volume 147, Issue 51, 47117-47125 (2025)

Additional Publications

Publications that I have co-authored during my PhD time, but are not part of this thesis

Paper III
Jakub Macošek, Bernd Simon, Johanna-Barbara Linse, Pravin Kumar Ankush Jagtap, Sophie L. Winter, Jaelle Foot, Karine Lapouge, Kathryn Perez, Mandy Rettel, Miloš T. Ivanović, Pawel Masiewicz, Brice Murciano, Mikhail M. Savitski, Inga Loedige, Jochen S. Hub, Frank Gabel, Janosch Hennig
Structure and dynamics of the quaternary hunchback mRNA translation repression complex
Nucleic Acids Research, Volume 49, Issue 15, Pages 8866–8885 (2021)

Paper IV
Nele Merret Hollmann, Pravin Kumar Ankush Jagtap, Johanna-Barbara Linse, Philip Ullmann, Marco Payr, Brice Murciano, Bernd Simon, Jochen S. Hub, Janosch Hennig
Upstream of N-Ras C-terminal cold shock domains mediate poly(A) specificity in a novel RNA recognition mode and bind poly(A) binding protein
Nucleic Acids Research, Volume 51, Issue 4, Pages 1895–1913 (2023)

Contribution

Paper I
J.-B.L. performed and analyzed MD simulations.
J.-B.L. and J.S.H. designed the project and wrote the article.

Paper II
J.-B.L. performed and analyzed MD simulations.
J.-B.L. and J.S.H. designed the project and wrote the article.

Manuscript I
J.S.H. designed the research and the contributed software.
T.M.F. and J.-B.L. performed and analyzed data.
J.-B.L. and J.S.H. and wrote the article.

Manuscript II
J.S.H. motivated the project.
J.-B.L. performed and analyzed MD simulations.
H.S.C., F.S. and P.A.A. performed the SAXS experiments and analysis.
J.B.L., H.S.C., F.S., P.A.A. and J.S.H. designed the project.
J.-B.L., H.S.C., P.A.A. and J.S.H. wrote the article.

Chapter 1

Introduction

> *Water is life's matter and matrix, mother and medium. There is no life without water.*
>
> – Albert Szent-Gyorgyi

Water is one of the most important molecules that make life possible. Moreover, water is the only substance that is more present in all three phases – liquid, ice and gas – in nature than any other substance, which further emphasizes its unique properties and importance.

The importance of water extends to the functioning of biological macromolecules, like proteins and nucleic acids, mainly due to the fact that most of these biological macromolecules need to be in an aqueous environment in order to fold correctly and perform their tasks efficiently. Already James D. Watson and Francis H. C. Crick have stated in the second half of the 20th century that water molecules are important for the structure of nucleic acids [1]. Even dry enzymes need a few water molecules to survive and keep functioning [2]. In an aqueous environment, water molecules are surrounding the surface of the biological macromolecules, filling the gaps between amino acids or nucleic acids within the structure of the biomolecule, and interacting with the biomolecule itself [3, 4]. Consequently, biomolecules affect water molecules, and vice versa, the water influences the behavior of the biomolecules [5].

Biological macromolecules, including proteins, perform important tasks in our cells and thereby make life possible. Proteins comprising of amino acids are the main components of living cells. The various specialized functions that proteins perform in cells range from contributing to cellular stability and structure to regulating crucial cellular processes such as the biochemical metabolism. Consequently, proteins play a key role in most processes of life.

Water molecules in close proximity to the biological macromolecules interact with the amino or nucleic acids on its surface, mainly with polar and charged groups. The water molecules in this area are interacting with the biological molecule through hydrogen bonds, electrostatic interactions and van der Waals forces. As a result, these water molecules in close proximity to the biological macromolecule form a

shell, the so-called hydration shell, around the biological macromolecule. The water molecules in this hydration shell have significantly different properties compared to the water molecules in the bulk water, which are not directly influenced by the biological macromolecule.

The interactions between water and biological macromolecules are significant for many chemical and biochemical reactions, as they can play an essential role the stability, dynamics and functions of biological macromolecules [6]. Water molecules in close proximity to the protein can (i) assure correct protein-protein, protein-RNA and protein-DNA recognition, (ii) act as a reagent in biochemical processes, (iii) assist in proton and electron transfer, (iv) increase the flexibility of proteins by breaking internal H-bonds, (v) ensure proper substrate binding at active sites, and (vi) prevent the protein from becoming trapped in misfolded states [3, 7–10], to name just a few examples. Consequently, water molecules plays a crucial role in the function of biological macromolecules and the hydration shell can be considered as an integral part of the biological macromolecules.

Because proteins are involved in so many life-essential functions, it is important to understand proteins and protein-water interactions at the molecular level in order to understand the complex processes that take place in living organisms. Understanding these complex, life-essential processes on a molecular level provides important information about the mechanisms of diseases, which is are crucial to fight against diseases and to develop new or more effective drugs and treatments [11–14]. Since almost all proteins in the human body reside in an aqueous environment and thus water plays a non-negligible role in protein function, the explicit consideration of water in studies and applications of protein function holds the potential to improve processes like drug discovery and docking [8].

Experimental structural biology techniques can be used to obtain information on protein structures that is necessary to study proteins and protein-water interactions on a molecular level [15–17]. Four widely used experimental techniques in structural biology are nucleic magnetic resonance (NMR) spectroscopy, cryo-electron microscopy (Cryo-EM), X-ray crystallography and small angle scattering (SAS) scattering. Each of these methods can be used on its own or in combination with one or several others. These four techniques – NMR, Cryo-EM, X-ray crystallography and SAS – differ in several points, like the sample size that can be studied, the amount of sample that is needed to perform the experiment, and the spatial resolution limit that can be reached with the technique (see also Table 1.1). In the following, four experimental techniques are described in more detail:

1. **X-ray crystallography** is a technique to determine the three dimensional atomic structure of a crystalline substance with atomic resolution [15]. The crystal, which could consists of proteins, is irradiated with X-rays that scatter on all atoms and thereby produce a diffraction pattern. By analyzing the diffraction pattern it is possible to determine the arrangement of atoms in the crystal under investigation.
2. **Cryo-EM** is an imaging technique that enables the investigation of the atomic structure of ordered regions in biological macromolecules in high resolution, but cannot detect disordered regions [16]. The sample is initially rapidly frozen (shock-frozen) to preserve the (flexible) biomolecules in their near-native states without damaging them, e.g. through the formation of ice crystals, in order to then, in a subsequent step, examine the shock-frozen sample with electron beams. The resulting images can be used to built three-dimensional models of a biomolecule.
3. **NMR** is a spectroscopy technique to obtain information about the structure and dynamics of biomolecules in solution. The samples can be examined in NMR under almost native conditions, but isotopically labeled biological macromolecules are often required, as this technique utilizes the magnetic properties of certain atomic nuclei. All nuclei that contain an even number of protons and neutrons have a nuclear spin of zero and are therefore not magnetic and not NMR-active. Consequently, all nuclei with a non-zero nuclear spin are magnetic and therefore NMR-active. Furthermore, this technique can currently only be used for smaller molecules, as otherwise the superposition of signals can lead to problems in interpretation. By analyzing the couplings between the nuclear spins, it is thus possible to obtain detailed information about the structure and dynamics of biomolecules. Thus, an NMR experiment provides information about distances, angles and orientations averaged over all biomolecules in the sample. The totality of biomolecules in the sample is also referred to as the entire molecular ensemble. Moreover, disordered regions can also be observed in NMR, but often these disordered regions cannot be assigned [18, 19]. Overall, NMR is an indirect method for determining the high-resolution atomic structure of biological macromolecules, as it derives the molecular structure and dynamics from the magnetic properties of the nuclei rather than providing direct images of the molecular arrangements.
4. **Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS)** are two different diffraction techniques used to study the

shape and size of biomolecules in solution. In SAS (Small angle scattering) methods, such as SAXS and SANS experiments, samples that have sizes from 10 Å (small molecules) to 10^3 Å (viruses) can be analyzed [20, 21]. In both methods, the sample is exposed to a monochromatic beam, X-rays in the case of SAXS and neutrons in the case of SANS experiments. The information obtained with the two experimental techniques SAXS and SANS is different, because X-rays scatter at the electrons and neutrons scatter at the atomic nuclei. Therefore, these two techniques complement each other very well and are often used in tandem. Both techniques provide a one-dimensional scattering curve as a result of measuring the intensity of the scattered radiation at small scattering angles. Because each molecule in the ensemble of biomolecules in solution has a different orientation, a one-dimensional scattering curve from a SAS experiment only provides information with low-resolution about the overall shape and dynamics of biological macromolecules and its hydration shell in solution [22–24]. For the interpretation of experimental SAS curves, mathematical and physical models are commonly used, such as rigid-body models, dummy-atom models, and all-atom MD simulations. These models often require fitting parameters – for example, the maximum diameter of the biomolecule, the radius of gyration, or the density and thickness of the hydration shell – in order to account for solution-specific effects and to reproduce the experimental scattering data accurately [25–28]. Because SAS curves contain relatively limited structural information, it is particularly important to avoid overfitting when applying models and fitting parameters. This can be achieved by incorporating additional physico-chemical data to improve the robustness and accuracy of the interpretation. Furthermore, systematic errors resulting from solute-solute interference, buffer subtraction uncertainties, minor radiation damage, and contributions from excluded solvent and hydration shell must also be considered. Therefore, the interpretation of SAS data is still a key challenge [29]. In the last years the method has evolved significantly. For example, size-exclusion-chromatography SAXS (SEC-SAXS) can be used to examine inhomogeneous solutions and ensure sample integrity by separating molecules in the solution according to their size [30, 31]. In case of SANS the concentration of deuterium oxide (D_2O) in the buffer can be adjusted, which enables the investigation of structures in heterogeneous samples. Due to the continuous development and improvements, SAS methods have become increasingly popular in recent years. However, if information about the ar-

Technique (PDB deposits)		Size (Sample state)	Resolution Limits	Amounts	Notes
NMR (14020)		<100 kDa (solution)	3-4 Å	µmoles/milligrams	Requires isotopically labeled recombinant protein, disordered regions can be observed but may not be assigned
X-ray Crystallography (176296)		Limited by crystal quality	<1-3 Å	µmoles/milligrams	Mutant constructs necessary for many membrane proteins, disordered regions invisible. Gold standard for structural water
Cryo-EM: Single particle (16009)		>100 kDa (vitrified ice)	Mostly >3 Å	nanomoles/µgrams	Resolution and size limits improving, best samples have symmetry, disordered regions invisible
SAXS		>10 kDa (solution)	>20 Å	nanomoles/µgrams	Native material can usually be used

Table 1.1: Comparison of experimental methods that are used in structural biology to study the structure and dynamics of biological macromolecules (Tab. adopted from Ref. 32). The numbers of PDB deposits were updated based on current data from RCSB PDB data bank website (rcsb.org) [33].

rangement of the atoms are of interest, it is possible to study the wide angles using the wide-angle X-ray scattering (WAXS) method.

The methods cryo-EM, NMR and X-ray crystallography provide detailed structural information at high resolution. In contrast, SAS methods are providing structural information only with low resolution. Regarding the elucidation of molecular structures, cryo-EM and X-ray crystallography are direct methods, whereas NMR and SAS are indirect methods (Fig. 1.1). X-ray crystallography, cryo-EM and NMR spectroscopy need special sample preparation, such as crystallized (X-ray crystallography), shock-frozen (cryo-EM) or labeled (NMR). Therefore, the sample is seldom observed in its native environment.

To address this limitation, these high-resolution methods are often combined with other experimental such as small-angle scattering techniques (SAXS/SANS) or computational techniques such as molecular dynamics (MD) simulations. Unlike X-ray crystallography, cryo-EM or NMR, these methods do not require freezing, crystallization or labeling of samples and the samples can be studied under near-native conditions.

Molecular dynamics simulations as a computational method play a crucial role in gaining atomistic insights into biomolecular structures and ensembles in solution. MD simulations therefore can complement other experimental techniques in struc-

tural biology, e.g. to refine structures of proteins in solution from NMR data or several structures of a biological ensemble, as it is often only possible to obtain with high-resolution experimental techniques a single structure of the biomolecule in one of its local energy minima. MD simulations, initially developed in the 1950s, are a computational method that uses Newton's equation of motion to calculate the motions of water, ions, and biological macromolecules or more complex biological systems [34]. In order to describe the potential energy surface and thus determine the forces acting on individual particles, so-called force fields are used in MD simulations.

One of the very fist steps was done by Bernie J. Alder and Thomas E. Wainwright back in 1957, who pioneered the simulation of liquid behavior using MD, focusing on the interaction of hard spheres [35]. A few years later, in 1974, Frank H. Stillinger and Aneesur Rahman performed the first MD simulation of water [36]. Two pivotal milestones contributed significantly to the progress of MD in the realm of protein research. Firstly, the publication of the first three-dimensional protein structure obtained through X-ray crystallography in 1958 by John C. Kendrew et al. [37]. This breakthrough paved the way for further exploration of protein dynamics. Secondly, in 1977, J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus

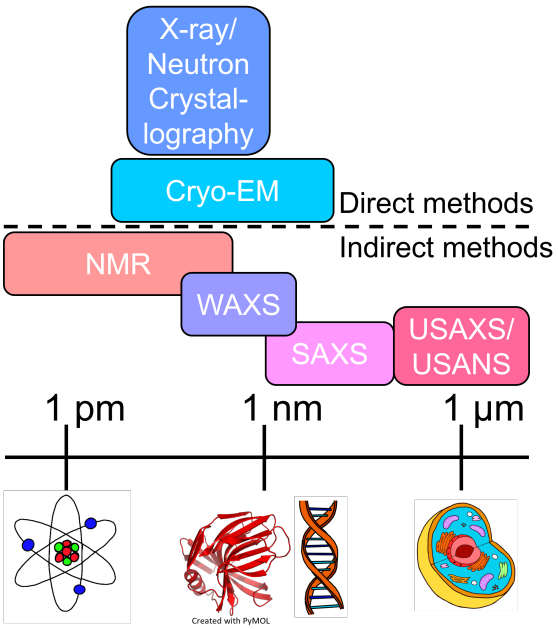


Figure 1.1: Techniques used in structural biology to study the structures of biological macromolecules and their dynamics, compared according to the spatial resolutions and subdivided into high and low resolution techniques.

accomplished the pioneering simulation of folded protein dynamics [38]. Their work marked a crucial advancement in understanding protein behavior through computational means.

Since then, the scope of MD has expanded beyond liquids and nowadays encompasses the study of larger molecules like biomolecules and proteins and it is nowadays an integral tool in structural biology. Overall, MD is particularly valuable for calculating structural and dynamic properties such as the position and motion of each atom in a at any given time that can be challenging to measure experimentally. This allows to study interactions on a molecular level. Since MD simulations can track the the motion of individual atoms over time, MD simulations are often referred to as a computational microscope [39, 40]. By simulating the motion of atoms and explicitly modeling the interactions with water molecules, researchers can observe the dynamic behavior of proteins and gain insights into the underlying mechanisms driving biomolecular processes such as protein folding, conformational changes or ligand binding [5, 6, 8, 41–43]. MD simulations thus make it possible to investigate phenomena that are difficult to study experimentally, such as folding, unfolding and conformational transitions of proteins. Hence, MD simulations provides intricate details about the three-dimensional structures and dynamics of biomolecules. Such insights have proven instrumental in the development of therapeutic agents for diseases resulting from protein misfolding, including Alzheimer's [44] and Parkinson's disease [45].

However, it is important to note that MD simulations have limitations. The accuracy of the force fields, which describe the interactions between atoms, and the choice of simulation parameters can influence the reliability of the results. Additionally, the timescales accessible to MD simulations are often limited by the availability of computational resources, making it challenging to capture rare events or processes that occur on longer timescales. However, many of these limitations can be partly overcome using enhanced sampling techniques such as umbrella sampling [46], metadynamics [47] or transition path sampling [48–51].

The combination of multiple techniques, can help researchers to overcome the limitations of individual methods, avoid over-fitting of the data [52], and obtain a more complete understanding of biomolecular structures and dynamics [53, 54]. In many cases, one single method alone cannot reveal all aspects of a biomolecule's structure, dynamics, and interactions, making it necessary to combine multiple techniques for a complete understanding [55–58]. Therefore, the integration of experimental data, bio-informatics, biomolecular modeling, and simulations is necessary

to obtain comprehensive information about the structure, dynamics, and other properties of biological macromolecules. The combination of multiple techniques thus enables researchers to study biomolecular systems under near-native conditions, providing valuable insights into their behavior in physiological environments.
For example, MD is often used to refine crystal structures and the combination of NMR with MD is described as a powerful tool to gain atomic insights into the structure and dynamics of biological macromolecules [59].
Especially the low-resolution SAS method needs additional physio-chemical data to avoid over-fitting and to interpret the data. In order to obtain atomistic interpretations of experimental SAS curves, MD simulations are a common approach. In this thesis for the calculation of SAS curves from MD trajectories a modified, freely available version of the simulation software GROMACS [28, 60] is used to avoid additional solvent related fitting parameters that can lead to inaccuracies. Thus it is possible to make accurate predictions of SAS patterns from MD simulations with explicit solvent. Furthermore, with this code, it is also possible to couple the MD simulation to experimental data to guide the simulation into regions of experimentally observed structures and thereby to overcome force field inaccuracies [61].
One of the key advantages of MD simulations is also the ability to observe protein-water interactions at a level of detail that is difficult to obtain experimentally. Simulations can provide atomistic insights into the hydration patterns around proteins, the dynamics of water molecules within the protein's vicinity, and the energetics of protein-water interactions. Additionally, MD simulations can capture the behavior of water molecules in confined environments such as protein binding sites or active sites, shedding light on the role of water in ligand recognition and enzymatic catalysis.
In conclusion, MD simulations provide a powerful approach to study the interactions between water and proteins. They offer atomistic insights into the dynamic behavior and structural changes of proteins in solution, as well as the role of water in protein stability, folding, and function. By combining computational simulations with experimental techniques, researchers can gain a comprehensive understanding of the complex interplay between water and proteins, ultimately advancing our knowledge of biological systems.
Hydration shells biomolecules have been probed by a complementary set of experimental techniques that access structure and dynamics on different length- and time-scales. Ultrafast vibrational spectroscopy (fs-ps IR and 2D-IR) re-

solves hydrogen-bond rearrangements and spectral diffusion of the OH/OD stretch on sub-ps to ps timescales and has been used extensively to report local reorientation and heterogeneity of hydration water near side chains and backbone groups [62–71]. Surface-sensitive variants probe interfacial hydration at membranes and air–water interfaces [72–76]. Quasi-elastic neutron scattering and incoherent neutron scattering quantify translational and rotational mobility and residence times of hydration waters on ps–ns scales and Å length scales [77]. Terahertz (THz) spectroscopy and broadband dielectric relaxation (GHz–THz) probe collective, low-frequency intermolecular modes and have been used to infer effective hydration-layer thicknesses and collective slowdowns [78–81]. Complementary structural constraints come from X-ray and neutron scattering, often combined with modeling to extract radial distribution functions and coordination numbers [82, 83]. NMR and site-selective IR probes add local residence-time and orientation information [84–86]. Collectively these experiments demonstrate slowed, heterogeneous hydration in the first shell while differing in how far and how strongly perturbations extend beyond it [42, 87].

Molecular simulations have evolved alongside experiment to provide atomistic insight into hydration structure and dynamics [42, 87, 88]. Classical MD using fixed-charge force fields remains the foundation for modeling aqueous systems, successfully reproducing key qualitative features such as enhanced local density, hydrogen-bond ordering, and slowed dynamics in the first hydration shell [89–91]. However, such models can misrepresent dielectric properties or ion-specific effects, for example, due to limitations in the representation of intermolecular interactions and polarization [92–96]. The continuous development of classical force fields and water models – such as SPC/E, TIP4P/2005, and their recent versions – as well as more advanced sampling techniques have progressively improved the predictive power of simulations and their agreement with experiments [89, 97, 98]. Simulations thus play a central role in connecting macroscopic observables to microscopic structure and dynamics, with accuracy governed by the balance between model complexity, parameterization, and computational tractability.

From experiments and simulations we now have robust, convergent findings: (i) water in the first hydration shell of proteins and polar biomolecular groups shows clear structural perturbations (altered O-O/O-H distributions and orientation) and slower rotational and translational dynamics. It was commonly reported to be about $\sim 2-4\times$ slower reorientation relative to bulk for many protein surfaces, with much larger slowdowns near strongly charged or tightly coordinated sites [62, 67, 68,

74–76, 99, 100]; (ii) the majority of structural perturbation decays rapidly, typically within $\sim$0.5$-$1.0 nm ($\approx$1$-$3 water layers) for small solutes and many protein patches, although the magnitude and range depend on surface chemistry, curvature, and local charge density [101–103]; (iii) collective dielectric and THz responses sometimes indicate a larger "effective" hydration footprint because they integrate weaker, longer-range dynamical correlations [100, 104]. Reconciling these collective measures with localized structural probes is an ongoing task.

Nevertheless several open questions persist. The precise spatial extent of altered dynamics – i.e., whether proteins have a thin ($\sim$1 layer) or a thicker ($\sim$1 nm or larger) dynamical hydration shell – remains debated because different probes weight distinct modes and because data analysis models influence inferred thicknesses [87, 101]. Other remaining gaps include quantifying the spatial and temporal heterogeneity of hydration dynamics, linking specific hydration patterns to biomolecular function, and reconciling discrepancies between structural and dynamical probes [100, 105, 106]. Methodological challenges also still arise from limited spatial resolution in scattering, spectral overlap and assignment ambiguities in IR/THz data, and model-dependence in simulations, particularly in capturing polarization, proton transfer, and rare water-exchange events [87, 96, 106]. Continued integrated experimental-computational studies that target the same observables and adopt common metrics for residence times, shell definitions, and dynamical heterogeneity will be critical to closing the remaining gaps.

The Aim of this thesis is to contribute to a better understanding of the water-protein interactions on a molecular level and the spacial structure of protein hydration shells. Some experimental techniques that provide insights into protein-water interactions are carried out with both light and heavy water (D_2O). To interpret such experiments involving heavy water with the help of MD simulations, we parameterized accurate pair potentials for heavy water by reparameterizing existing three- and four-site water models (see Chapter 5). By reparameterizing the pair potentials, we attempted to reproduce the most important structural, thermodynamic, and dynamic properties of D_2O, so that these models provide a reliable basis for the simulation and interpretation of experiments conducted with heavy water. In a subsequent step, in Chapter 6, we tested to what extent the overall structure of the hydration shell is reproduced in MD simulations. We performed MD simulations of five globular proteins to calculate scattering profiles that explicitly account for the contributions of the hydration shell. A comparison of these profiles with the consensus SAS curves from high-precision SAS experiments revealed an excellent agreement between the simulations and the experiments with respect to the hydration shell contrast, as indicated by the SAS-derived radius of gyration (R_g) values. This allowed us to demonstrate that MD simulations not only reproduce the scattering properties of proteins in solution but also provide an accurate representation of the overall structure of their hydration shells. After demonstrating that the hydration shell is fairly accurately represented in MD simulations, we next investigated how different surface amino acids influence the structure of the hydration shell. To this end, we introduced targeted mutations of selected surface residues in both a small globular protein and an intrinsically disordered protein (IDP) and analyzed how the different amino acid types influence the properties of the hydration shell. The results of these comparisons are presented in Chapter 7. Our simulations reveal that around acidic residues the most pronounced hydration shells are formed, followed by basic, polar, and apolar residues, whereby apolar residues tend to cause local water depletion. Because proteins do not all exist and function only at room temperature, but also at extreme temperatures, we looked in Chapter 8 at the influence of the temperature on the protein hydration shell. The results of this study show that an increase in temperature leads to a depletion of the hydration shell, which is a consequence of an increased dynamic disorder within the hydration shell and a partial displacement of surface-coordinated water molecules.

Three- and four-site models for heavy water.
Heavy water, also called deuterium oxide (D_2O), is water with heavy hydrogen isotopes, deuterium (D), in place of the light hydrogen atoms (protium, H) that are present in light water (H_2O). The distinct physical and chemical properties of heavy water [107] make it a powerful tool for biophysical experiments, enabling more detailed studies of biomolecular dynamics. For example, D_2O is often used as a solvent in SANS experiments because it offers the possibility of contrast variation, which can be useful for studying (heterogeneous) biomolecular structures [108–110]. Accurate modeling of D_2O is helpful not only for SANS experiments, but also for other techniques such as NMR spectroscopy and Fourier transform infrared (FTIR) spectroscopy, where D_2O is also frequently used to measure and determine various properties of biomolecules [111, 112]. To be able to interpret biophysical experiments with D_2O precisely, reliable models for heavy water are therefore essential. Because we were not able to reproduce the properties of an already existing model for heavy water [113], we have parameterized new force fields for heavy water based on three widely applied three- and four-sided water models for light water, namely SPC/E [89], TIP3P [90], and TIP4P/2005 [97]. Both our three-sided models SPC/E-HW and TIP3P-HW reproduce the differences of heavy water relative to light water at room temperature. Our four-sided model TIP4P2005-HW captures the differences over a wide temperature range. These models ensure in MD simulations that the unique properties of D_2O are correctly represented and thus experiments with D_2O as a solvent can be accurately modeled. The pair potentials for heavy water developed in this study enable a precise interpretation of experiments performed with heavy water with the help of MD simulations.

Scrutinizing the protein hydration shell.
It has been known for quite some time that the protein and the water molecules in its vicinity influence each other, affecting protein functions such as folding, molecular recognition and enzymatic activity [7, 114, 115]. These interactions also lead to the formation of water layers around the protein with higher and lower densities compared to the bulk water, the so-called hydration shell. The dynamics of the water molecules in the hydration shell of proteins have already been extensively studied with the help of spectroscopic experiments. However, their overall structure and contrast with respect to the bulk water remained unclear due to a lack of precise experimental data. SAS signals are sensitive to the protein hydration shell, as the altered density in the protein hydration shell contributes to the contrast measured by

SAS [42, 101, 116–123]. Therefore, by using SAS experiments with X-rays (SAXS) or neutrons (SANS), it was shown that the hydration shell of many globular proteins has an increased density compared to the bulk water. One studied structural parameter that can be calculated from SAS curves and captures the changes in the hydration shell is the radius of gyration (R_g). The R_g is a measurement of the compactness of a protein and provides information about the mass distribution within a protein. The increased hydration shell density compared to bulk water leads to an increased or decreased radius of gyration in SAXS or SANS in D_2O experiments, respectively. We were able to reproduce these altered R_g values with calculated SAXS and SANS curves from MD simulations with explicit solvent. By combining explicit solvent SAS calculations from MD simulations and high-quality SAS experiments [124], we have validated 18 different protein force field/water model combinations with respect to the protein hydration shell and investigated the overall structure and contrasts of protein hydration shells of five globular proteins, namely lysozyme, RNaseA, xylanase, urate oxidase and glucose isomerase. Overall, we were able to show that the hydration shell contrasts from MD simulations are in remarkable agreement with those from high-precision SAS experiments for most, but not all tested force fields. Thus, MD simulations reproduce the overall structure of the protein hydration shell relatively accurately.

Hydration shells of globular and disordered proteins: Effects of surface amino acids.

Despite the importance of the hydration shell on the dynamics and structures of proteins, a detailed molecular understanding of how the hydration shell is shaped by the surface chemistry of proteins is still limited, mainly because direct experimental studies of the hydration shell structure are limited. To assess how the 20 proteinogenic amino acids affect the hydration shell structure, we use atomistic MD simulations combined with explicit-solvent SAXS predictions to quantify their impact on the radius of gyration and overall solute-solvent contrast. To isolate the effects of surface chemistry, we introduced systematic surface mutations in two model systems: the GB3 domain, a small globular protein, and XAO, a short intrinsically disordered peptide (IDP). From MD simulations, we derived amino-acid-specific contrast scores for both R_g values and the hydration shell contrasts. Our results show that acidic amino acids produce the strongest hydration shell contrast, followed by basic, polar, and finally apolar residues, which generated local water depletion layers. These trends are consistent across different water models

and conformation-dependent – particularly in XAO, where hydration was generally weaker due to poor water-peptide packing at the IDP surface. These findings are in qualitative agreement with experimental hydration effects [125–138] and reveal a hierarchy of hydration behavior that reflects underlying chemical interactions. Together, these findings are highlighting the complex interplay between sequence, structure, and hydration, providing a detailed, quantitative framework for how the amino acid surface composition shapes the hydration shell.

Depletion of the protein hydration shell with increasing temperature.
The dynamic properties of the hydration shells of proteins have been extensively studied using experimental techniques, but how the hydration shell structure responds to temperature changes is less well understood. In particular, the extent to which the hydration shell is disrupted or depleted by an increase in temperature is not fully understood, although this is important for fundamental processes such as protein folding, cold denaturation and biomolecular phase behavior [139–143]. To address this problem, we have combined temperature-ramp small angle X-ray scattering (T-ramp SAXS) with atomistic MD simulations to investigate how temperature modulates hydration shell structure in two folded proteins: the GB3 domain and the villin headpiece (HP35). Using a customized setup, our collaborators, Hyun Sun Cho, Friedrich Schotte and Philip A. Anfinrud, collected SAXS data at the BioCARS (Advanced Photon Source) beamline over a broad temperature range (260 K-345 K) that included both supercooled conditions and partial de-convolution. Complementary all-atom MD simulations with SAXS calculations using explicit solvents then allowed us to interpret the experimental data at atomic resolution. Our results show that increasing the temperature leads to a consistent decrease in both the protein contrast and the radius of gyration, reflecting the depletion of the hydration shell. This depletion results not only from an increased dynamic disorder within the hydration shell, but also from the partial displacement of surface-coordinated water molecules. These results demonstrate the strong temperature sensitivity of the hydration shells in folded proteins and provide a structural basis for temperature-dependent phenomena such as cold denaturation, thermophoresis, and biomolecular phase separation.

Chapter 2

Biochemical basics

Proteins play key roles in several processes of live and are one of the main constituents of cells. The proteins are comprised of building blocks called amino acids. The chemical characteristics of amino acids determine the properties and functions of the proteins they form. The bonds between amino acids can be both covalent and non-covalent and are critical for the structure and function of proteins. In the subsequent sections of this chapter covalent and non-covalent bonds, amino acids and proteins are described in detail.

2.1 Covalent and non-covalent bonds

Covalent and non-covalent bonds together determine the structure, stability, and function of biological macromolecules.

Covalent bonds are strong, directional chemical bonds formed by the sharing of electrons between atoms [144]. In proteins, covalent peptide bonds link amino acids into stable chains (Fig. 2.1), preserving permanent backbone of the molecule – the primary structure. Outside of chemical reactions, these bonds remain stable under physiological conditions and define the permanent framework of the molecule.

Non-covalent interactions are weaker but essential for the three-dimensional architecture and dynamic behavior of biomolecules [146]. They act within and between biomolecules and include ion pairs (electrostatic attractions between oppositely charged groups), hydrogen bonds (specific interactions between a hydrogen atom and an electron pair donor), and van der Waals forces (vdW) (attractions between permanent, induced, and/or instantaneous dipoles, including London dispersion, Keesom, and Debye forces). Non-covalent forces stabilize secondary and tertiary structures, mediate conformational flexibility, and enable transient molecular recognition events such as ligand binding or protein-protein interactions. While covalent bonds define the fundamental framework of a biomolecule, non-covalent interactions provide flexibility and enable reversible processes such as folding, structural rearrangements, ligand binding, and protein–protein interactions.

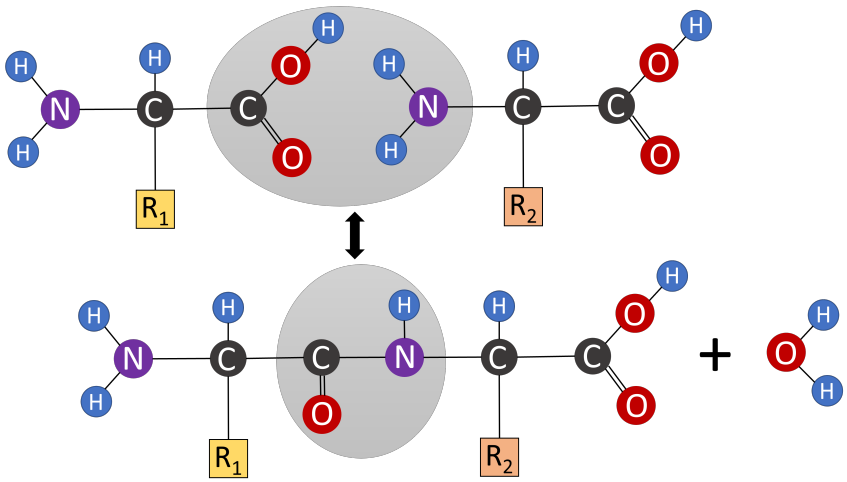


Figure 2.1: Formation of a peptide bond between two amino acids, by condensation of an amino and a carboxyl group with release of a water molecule. (Fig. adapted from Ref. 145.)

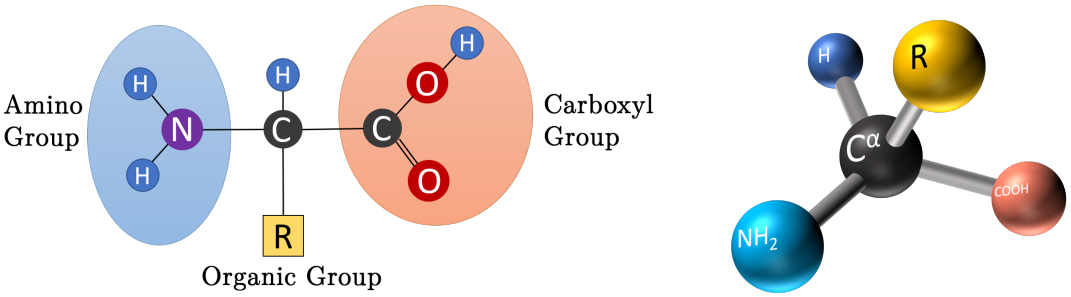
The interplay between covalent and non-covalent bonding enables biomolecules to adopt complex shapes and perform diverse biological functions.

2.2 Amino acids

Amino acids occur in living cells, are the building blocks for proteins and play essential roles in many biological processes. Over 400 different amino acids exist [147], but in the universal genetic code of living organisms are only 20 amino acids encoded, the proteinogenic amino acids. These 20 proteinogenic amino acids are the main building blocks of proteins, consisting of a central carbon atom, an amino group, a carboxyl group and an organic group, unique to each amino acid (Fig. 2.2). The (unique) organic group is bound to the central α-carbon atom. Therefore, the 20 proteinogenic amino acids belong to the class of the α-amino acids.

All proteinogenic amino acids in nature are also chiral, except glycine. Chirality in case of amino acids means that depending on the arrangement of the molecular groups around the α-carbon atom, amino acids can occur in two forms that cannot be converted into each other by rotation. If the Fischer projection is used to describe the chirality, the amino group is located to the right of the α-carbon atom in a D-amino acid (lat. **D**extro), but to the left in an L-amino acid (lat. **L**evo). In proteins the amino group is almost always on the left hand side of the α-carbon, L-α-amino acids, while in rare cases it can also be on the right hand side, D-α-amino acids [146]. Hence the 20 proteinogenic amino acids are all L-α-amino acids except

glycine. Glycine is the smallest and simplest proteinogenic amino acid and, due to its chemical structure, it is also the only proteinogenic amino acid that does not have a chiral center, resulting in the absence of L- and D-forms.
Often the proteinogenic amino acids are classified into four different groups based on the organic groups: polar, non-polar, basic and acidic amino acids (Fig. 2.3).



(a) (b)

Figure 2.2: (a) 2D and (b) 3D representation of the basic L-α-amino acid structure. The amino group (light blue), carboxyl group (red), organic group (yellow) and the additional hydrogen atom (dark blue), are bound to the α-carboxyl (black). (Fig. adopted from Ref. 145.)

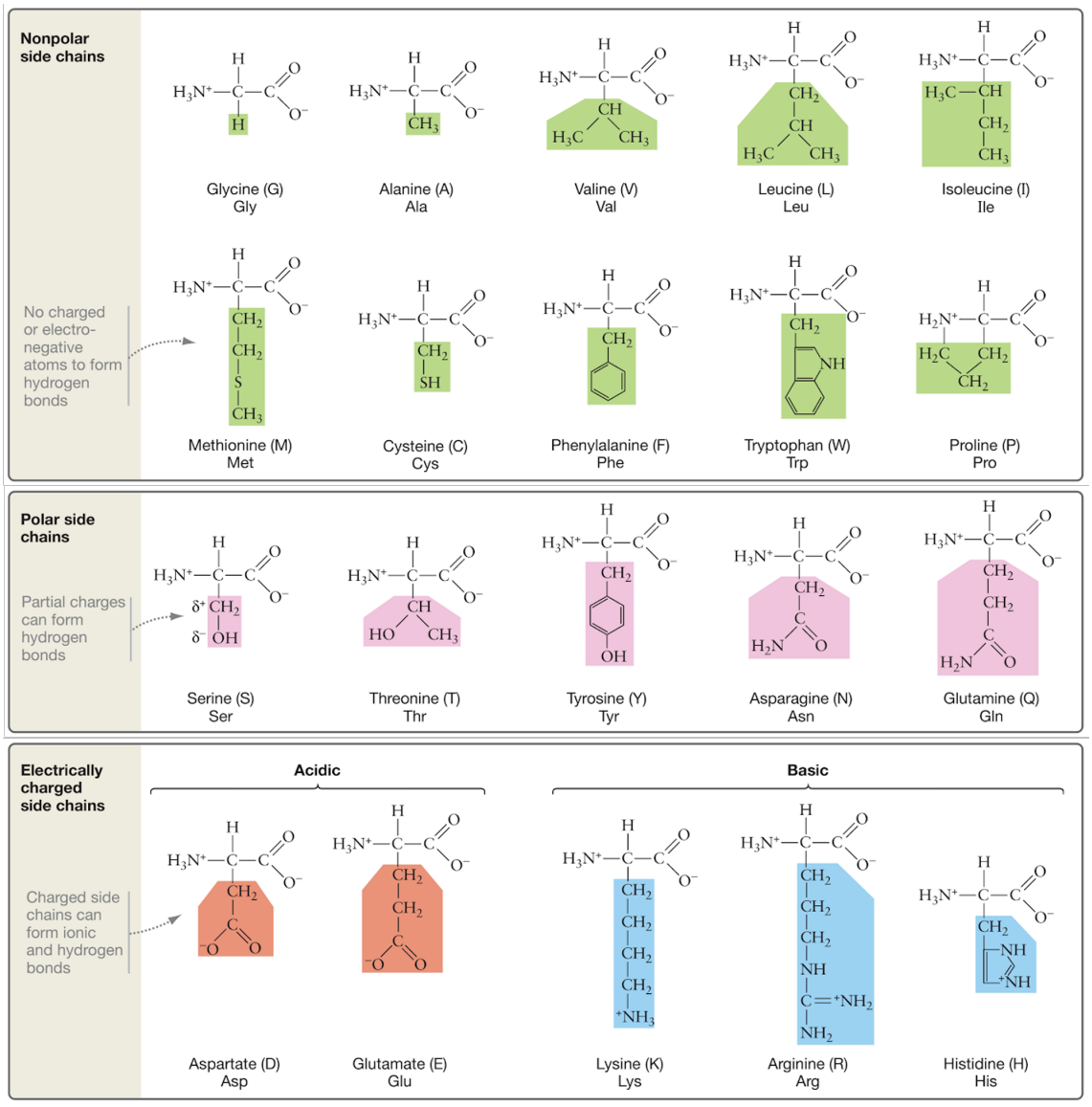


Figure 2.3: The 20 proteinogenic amino acids classified into the four different groups: non-polar (green), polar (pink), acidic (red) and basic (blue) amino acids (Fig. adopted from Ref. 148).

2.3 Proteins

The term protein (gr. proteion = of the most) was first used in 1938 by the Swedish chemist Jöns Jacob Berzelius in a letter to the Dutch chemist Gerardus Mulder, who shortly thereafter introduced the term protein for the very first time in a publication [146, 149, 150]. Proteins are three-dimensional structures and one of the main and most diverse function carriers in cells that are biologically active only in (one of) their native conformation(s). Many proteins spontaneously fold into their native conformation, driven by non-covalent (intramolecular) interactions that guide the polypeptide chain toward a low free energy state. However, some proteins require molecular chaperones to assist in proper folding and prevent misfolding. The native structure of a protein usually corresponds to a local free energy minimum that is both stable and functional under physiological conditions, although it does not always represent the absolute global minimum. Although many proteins are folding into stable and specific three-dimensional structures, there are proteins or parts of proteins that exhibit high flexibility and do not adopt a single defined structures under physiological conditions. Therefore, these intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) exist as dynamic ensembles of conformations. This variety of conformations enables them to interact with diverse binding partners, such as other proteins, nucleic acids and small molecules, and thereby the IDPs and IDRs can adopt a stable, three-dimensional structure.

The main building blocks of proteins are the 20 proteinogenic amino acids as described in the previous section 2.2. With these 20 proteinogenic amino acids it is possible to synthesize $20^3 = 800$ tripeptides. For a small protein with 100 amino acid residues there exist $2^{100} = 1.27 \cdot 10^{130}$ possibilities for its amino acid sequence; compared to the estimated number of atoms (10^{79}) in the universe this is a huge amount [146]. In a polypeptide chain two amino acids are connected through peptide bonds, illustrated in Fig. 2.1. The sequence of the amino acids along a polypeptide chain is determined by the nucleotide sequence in the deoxyribonucleic acid (DNA). Many proteins that are occurring in living cells are not encoded directly in the DNA and formed by chemical, post-translational modifications of the proteinogenic amino acids.

Proteins themselves are colorless in the visible spectrum, but some proteins, have a non-protein component, e.g. metal ions or low-molecular-weight organic compounds, for their biological activity that is tightly bound to the protein by a covalent or non-covalent bond and which could give the protein complex a color. Examples of colored

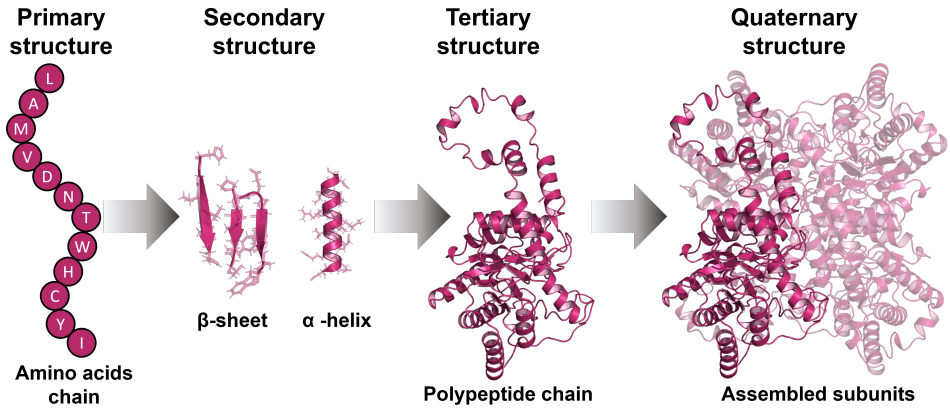


Figure 2.4: The four levels of protein structure: primary, secondary, tertiary and quaternary structure (Fig. adapted from Ref. 160).

proteins with a non-protein component are (i) hemoglobin, which has bound low-molecular organic compounds which in turn has bound metal ions, namely iron ions, and appears red when interacting with oxygen [151–153], (ii) bacteriorhodopsin, which has bound a low-molecular organic compound, a retinal, and thus appears purple when activated by visible light [154–157], and (iii) the green fluorescent protein, which has also bound a low-molecular organic compound, the photoprotein aequorin, and appears green when optically activated [158, 159].

The structure of proteins, with the exception of IDPs, can be divided into four levels: (i) primary, (ii) secondary, (iii) tertiary and (iv) quaternary structure, illustrated in Fig. 2.4. (i) The amino acid sequence of a polypeptide chain, the primary structure of a protein, determines the folding of a protein and therefore, its spatial arrangement. This results in, on the one hand, (ii) the secondary structure of a protein, comprised of periodically folded α-helices and β-sheets stabilized through hydrogen bonds, and, on the other hand, (iii) the tertiary structure, comprised of aperiodic spatial arrangements of the entire polypeptide chain stabilized through disulfide (especially in extracellular proteins), ionic and hydrogen bonds, as well as vdW interactions. (iv) Since the molecular mass of many proteins is between 10 and 1000 kDa, and the molecular mass of a single polypeptide chain consisting of several hundred amino acid residues is between 10 and 100 kDa, it can be concluded, that many proteins consist of multiple polypeptide chains (subunits) stabilized by non-covalent bonds (e.g. hydrophobic effects of apolar groups, hydrogen bonds, and van der Waals forces), electrostatic interactions between charged groups, and intramolecular bonds. The three dimensional structure of the protein with all its subunits is called the quaternary structure.

Chapter 3

Molecular dynamics simulations

Molecular dynamics (MD) simulations are used in many areas of science and technology, including chemistry, physics, materials science, and biology. MD simulations can provide insights into the structure, behavior and properties of complex systems, such as proteins, DNA, ribonucleic acid (RNA), and materials. Furthermore, MD simulations can also be used to study phase transitions and the behavior of materials under extreme conditions. By means of MD simulations, for example (i) thermodynamic properties (e.g. temperature and pressure), (ii) structural properties (e.g. bond lengths and angles) and (iii) dynamical properties (e.g. diffusion coefficients and reaction rates) can be studied. Overall, MD simulation is a computational method for studying the behavior of complex systems by simulating interactions and motions of the system on a molecular/atomistic level over time (Fig. 3.1).

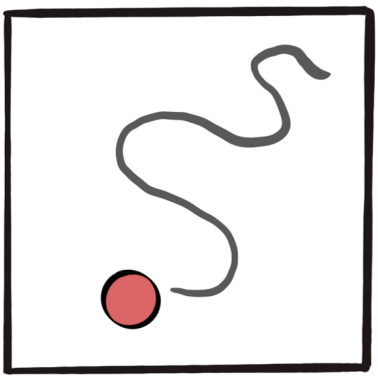


Figure 3.1: Schematic illustration of the motion of a particle over time. The particle is shown in red at its current position, and the gray, swirly line is the trajectory up to the current position.

3.1 Approximations in MD simulations

To simulate the motion of atoms over time according to their interactions with each other at speeds much lower than light and under the influence of external fields, one has to solve the time-dependent Schrödinger equation for the Hamilton operator $\hat{H}$ for the whole system [161–163], given by

$$i\hbar\frac{\partial}{\partial t}\psi = \hat{H}\psi \ , \tag{3.1}$$

where ψ is the wave function and $\hbar$ the reduced Planck constant. Due to the computing power available nowadays, quantum mechanical calculations based on the Schrödinger equation are only feasible for (very) small systems and short time scales.

Therefore, approximations must be used to make simulations computationally feasible, while still providing useful insights into complex molecular systems.
The Born-Oppenheimer approximation allows the motion of atomic nuclei to be treated independently of the motion of electrons. This approximation is based on the fact that electrons have a much lower mass than the nuclei and can therefore be treated as if they react instantaneously to changes in the positions of the nuclei [164]. Therefore, based on the Born–Oppenheimer approximation, electrons are not treated explicitly in the simulations; instead, partial charges representing the electronic distribution are assigned to the atomic nuclei.
Additionally, the motion of the nuclei can be treated classically according to the Ehrenfest theorem [165], because the extent of their wave packets is small compared to the scale over which the potential changes, and their wave packets remain localized due to very short decoherence times [166]. Therefore, the classical Newtonian equations of motion can be used to calculate the nuclei's trajectories, while atoms are modeled as spheres to simplify their interactions. The Newton's equation of motion for the j-th atom is given by

$$m_j \cdot \frac{\partial^2 \vec{X}_j}{\partial t^2} = \vec{F}_j(\vec{X}_1, ..., \vec{X}_N) \ , \tag{3.2}$$

where m_j is the mass, X_j the position and $\vec{F}_j$ the force acting on the j-th atom. The forces acting on the j-th atom are given by the negative derivative of the potential energy surface V with respect to it position:

$$-\frac{\partial V(\vec{X}_1, ..., \vec{X}_N)}{\partial \vec{X}_j} = \vec{F}_j(\vec{X}_1, ..., \vec{X}_N) \ . \tag{3.3}$$

The potential energy surface, which determines the forces acting on each individual particle, is used to simulate the motions of the atoms in a system. To define the potential energy surface a further approximation, so called force fields are used. A force field, described in more detail in Sec. 3.2, consists of a set of empirically validated (simple) mathematical functions that together with their (corresponding) parameters describe the potential energy of a system as a function of the positions of the atoms.
Since most of the systems to be studied by MD simulations have hundred to hundred thousands of atoms, an analytical solution of Newton's equation of motion is not possible, hence numerical methods - usually algorithms from the Verlet family, described in more detail in Sec. 3.4 - are used when performing simulations.

In MD simulations, the systems are small compared to macroscopic systems, so the computational effort and simulation time become feasible. For this reason, the periodic boundary conditions (PBC) described in section 3.7 are used in the MD to minimize or avoid boundary and surface effects. With PBCs, the small system under consideration is effectively surrounded by infinite periodic copies of itself, approximating an infinitely large system and the particles thus experience forces as in a large system, thereby avoiding surface effects.
All the approximations described here are necessary to make simulations computationally feasible and at the same time to gain useful insights into complex molecular systems.

3.2 Force Fields

The description of the potential energy V of the system under investigation by means of simple mathematical terms is referred to as a force field. Thus, a force field provides an easy-to-calculate description of the potential energy V. Usually, empirical force fields, based on experimental data, exact quantum mechanical (QM) calculations or ab initio methods, are used. They are built on the approximation, that atoms are rigid bodies and take into account effects like bonding, repulsion and electrostatic interactions. Nowadays, several different molecular dynamics force fields are available. Many all-atom force fields share a broadly similar functional form, describing the potential energy surface of the system as the sum of bonded terms

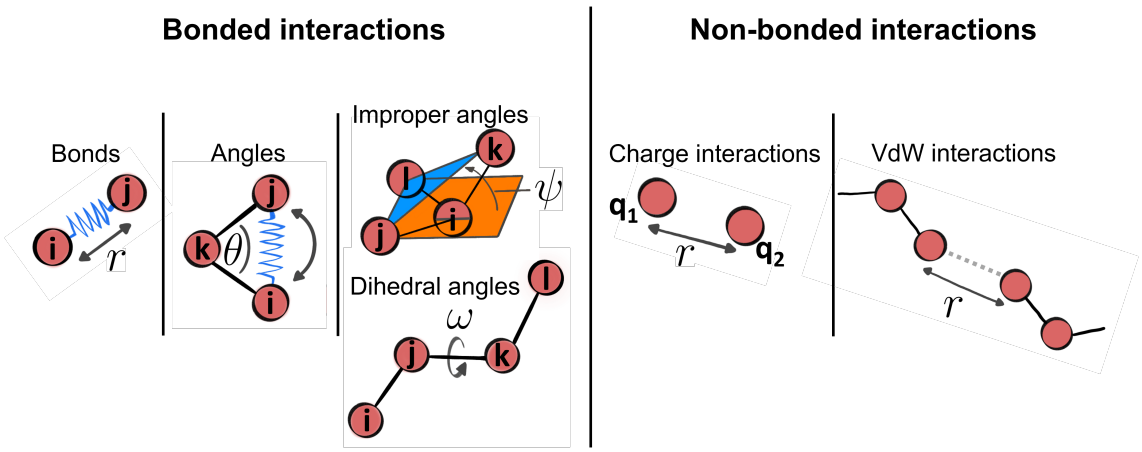


Figure 3.2: Illustration of the bonded and non-bonded terms used for calculating the potential energy V and the resulting force F in MD simulations. The equations of the single contributions is described in Eq. (3.4).

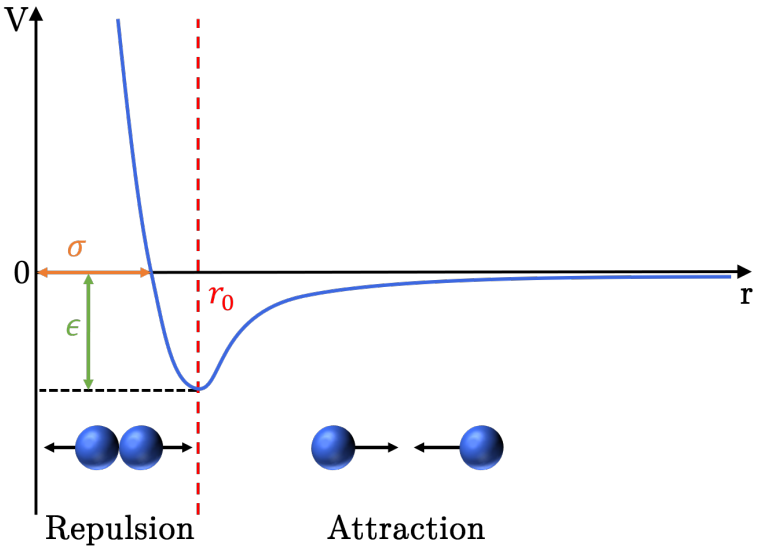


Figure 3.3: The Lennard-Jones Potential V (blue) plotted against the distance r between two particles. For short distances a strong repulsive force, the Pauli repulsion, and for longer distances attractive forces, the London dispersion (also called attractive van der Waals forces), are acting on the particles. The potential has its root at $r = \sigma$ (orange) and its minimum at $r_0 = \sqrt[6]{2}\sigma$ (red) with $V(r_0) = -\epsilon$ (green). (Fig. adopted from Ref. 145.)

(bonds, angles, dihedrals) and non-bonded terms (vdW and electrostatic interactions), with differences primarily in parameter values and certain functional details. Other force fields, however, may include additional interaction terms, polarizability, coarse-grained representations, or implicit solvent models, which alter the functional form.

The potential energy V and the resulting force F are usually composed of (i) bonded and (ii) non-bonded terms, as illustrated in Fig. 3.2. (i) The bonded terms are composed of all intramolecular interactions that depend on the bond conformation of the molecules, including bond stretches, bond angles, and torsion angles. (ii) The non-bonded terms are composed of the Coulomb potential, the Van der Waals attraction and the Pauli repulsion. The latter two are often combined into the Lennard-Jones potential, shown in Fig. 3.3. The Coulomb and LJ-potentials are explicitly calculated up to a cut-off radius r_c for all atom pairs ij.

Since the Coulomb potential decays only with $\frac{1}{r}$ and is very long-ranged, it requires most of the computational time. These long-range interactions are therefore supplemented with a sum in reciprocal space, the Particle-mesh Ewald summation (PME) [167–169], for longer distances and the periodic images. The PME method can in

GROMACS also be used for the LJ-potentials [170], but often a simple and efficient approximation – the plain cut-off – is used, since most of the force fields were parameterized with a plain cut-off.
The potential energy of the system can thus be formulated as the sum of the energies of the individual contributions, which in turn are determined by force fields based on the ball-spring model

$$V = V_{\text{bonded interactions}} + V_{\text{non-bonded interactions}} \tag{3.4}$$
$$= V_{\text{bonded}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{improper}} + V_{\text{LJ}} + V_{\text{Coulomb}} \tag{3.5}$$
$$= \sum_{\text{bonds}} \frac{1}{2} k_r (r - r_0)^2 \tag{3.6}$$
$$+ \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{3.7}$$
$$+ \sum_{\text{dihedral}} \frac{1}{2} k_{\text{tor}} \left(1 + \cos(n\omega - \gamma)\right) \tag{3.8}$$
$$+ \sum_{\text{improper}} \frac{1}{2} k_\psi (\psi - \psi_0)^2 \tag{3.9}$$
$$+ \sum_i \sum_{i<j} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.10}$$
$$+ \sum_i \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \ . \tag{3.11}$$

Here k_r, k_θ, k_{tor} and k_ψ denote the binding, bending, torsional and improper stiffness, respectively. r is the current distance between to atoms – the bond length – and θ, ω, ψ are the current atomic, dihedral and improper angles between the atoms, respectively. r_0 is the equilibrium length, and θ_0 and ψ_0 are the equilibrium angles. In addition to the force constant k_{tor}, a multiplicity n and a phase shift angle γ are required to describe the periodic term that is needed for the bonded dihedral interactions. ϵ refers to the depth of the Lennard-Jones well and σ is the distance at which the interaction energy between two particles, as described by the Lennard–Jones potential, is zero. q_i and q_j are the charges of the respective atoms. And ϵ_0 denotes the electric constant (vacuum permittivity), and ϵ_r the dielectric constant (relative permittivity).
The set of parameters of a force field is selected to best describe experimental data or ab initio methods like exact QM calculations, that are used for the parameterization. The parameterization of empirical force fields is often costly and typically tailored

to a limited class of systems, such as proteins and biomolecules. Moreover, because these force fields are fitted to reproduce quantities at specific conditions – like a certain temperature – their accuracy tends to decrease outside that range, meaning they usually provide reliable results only within a (narrow) temperature range due to their empirical nature. Therefore, different or modified force fields are often needed in practice and in fact that is why, the testing and further development of force fields is a continuous ongoing challenge [171–174].
The two most common force fields that are nowadays used in MD simulations of biological systems are:

1. AMBER (Assisted Model Building with Energy Refinement): The original force field, which was designed to simulate peptides and nucleic acids, was published in 1984 by Peter Kollmann and his group [175]. The Amber force field has been updated and refined, that is why nowadays multiple versions are available, including Amber99SB [176], Amber03 [177, 178] and Amber14SB [179] and others [171, 172, 174, 180]. Thus, it is nowadays used for a wide range of simulations, including biomolecules, such as proteins, nucleic acids, and carbohydrates.
2. CHARMM (Chemistry at Harvard Molecular mechanics): The force field was developed by Professor Martin Karplus and his co-workers in the Department of Chemistry at the Harvard University and was published in 1983 [181]. In MD simulations CHARMM force fields are extensively used for nucleic acids, lipids, carbohydrates, proteins and other biomolecules. Over the years, several versions of the CHARMM force field, like c22* [182, 183] and c36 [184], have been developed, each containing (slightly) different parameter sets optimized for specific types of molecules or systems.

All atom versus united atoms

In MD simulations one commonly chooses the level of resolution for the modeled system: all-atom, united-atom, or coarse-grained.

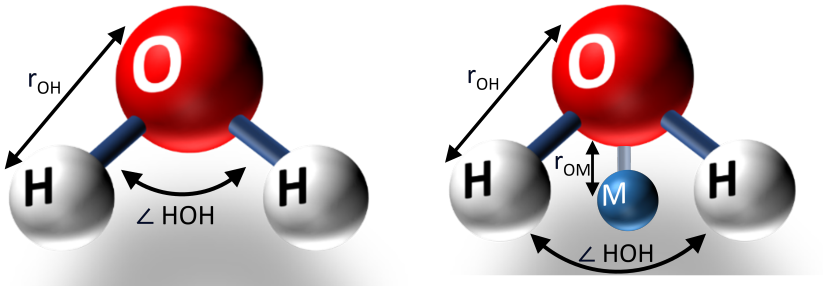
- In all-atom simulations every atom is represented explicitly, including hydrogens, allowing for highly detailed modeling of molecular interactions and structural dynamics. The force center of each atom is localized at or near the position of its nucleus. This approach provides the most accurate depiction of physical behavior but comes at a high computational cost, limiting system size and simulation timescales. Typical all-atom force fields include CHARMM [181, 184], AMBER [173, 175, 177] and OPLS-AA [185].

- In united-atom simulations, non-polar hydrogen atoms are combined with their bonded heavy atoms (e.g. CH, CH_2 and CH_3) and treated as a single interaction site. This reduce the number of particles, simplify the force calculations and speed up simulations while retaining an essentially atomistic description of the molecule. This approach improves computational efficiency while retaining much of the chemical accuracy, making it useful for systems like lipids and hydrocarbons.
- In coarse-grained simulations several atoms are combined into a single bead. Often several heavy atoms and their hydrogens are mapped into one interaction site. The resulting model has far fewer degrees of freedom and uses effective interaction potentials tuned to reproduce larger-scale behavior. Coarse-grained models permit much longer time- and length-scale simulations but sacrifice atomic detail and chemical specificity. A popular example of a coarse grained force field is MARTINI [186, 187].

3.3 Water models

In MD simulation, water models are used to simulate the behavior of water molecules in different phases and environments, such as solutions, clusters, or in gas. In simulation systems, proteins are often present in water or aqueous solutions, that most closely resembles the environment in which they occur in biological systems. Simulation systems therefore usually contain a large number of water molecules, and it is thus of paramount importance to have reasonably good models that reproduce the experimental behavior of water as closely as possible. Important properties of water to reproduce include its density, diffusion coefficient, isothermal compressibility, average number of hydrogen bonds per molecule, potential energy, and heat of vaporization. Nowadays there exist rigid, flexible and polarizable water models that try to reproduce the properties of water as good as possible. Many of these water models are rigid water models, meaning that the positions of the hydrogen and oxygen atoms in relation to each other are fixed to match the known geometry of water. Thus, the bonded interactions are treated with constraints and the rigid water models are therefore (mainly) based on the internal geometry and on the intermolecular, non-bonded interactions with the atoms of other molecules calculated using a combination of Lennard-Jones (LJ) and Coulomb potentials (see also Sec. 3.2). Since the time-step size in MD simulations is determined by the fastest bond oscillations, particularly those of hydrogen bonds, rigid water models are often used in MD simulations in order to be able to use larger time steps (described in more detail in Sec. 3.2).

The most commonly used water models are currently three-site and four-site rigid water models. In the three-site water models, such as SPC/E [89] and TIP3P [90],



(a) (b)

Figure 3.4: Schematic representation of three- and four-site classical rigid water models.

each atom has a point charge and the oxygen atom has additional LJ parameters (Fig. 3.4(a)). In the four-site water models, like TIP4P [90], TIP4P/2005 [97] and OPC [188], the negative charge of oxygen is located at a virtual atom instead of at the position of the oxygen atom (Fig. 3.4(b)). This improves the electrostatic properties of the model, such as reproducing the dielectric constant of water more faithfully. These classical water models never reproduce all experimental properties correctly, but most water models reproduce many of the desired properties with fairly decent accuracy (even though usually only in a limited range of temperatures). Therefore it is very important to choose the right water model depending on the simulation (system).

The choice of the right water model depends on several factors, such as the type of system to be simulated, the desired properties of the water and the force field that is used, as the force fields have been optimized against one or multiple selected water models. The validation required to ensure the reliability of simulation results and the development of new water models is a continuous ongoing challenge.

3.4 Integration schemes

In MD simulations, a N-body problem must be solved for which there is in general no analytical solution for the equation of motion. Therefore, it must be solved numerically in small time increments Δt:

$$\vec{x}_k(t_i) \text{ with } t_i = i \cdot \Delta t \text{ for } i = 0, 1, ..., n_{\text{steps}} \ . \tag{3.12}$$

For the calculation of the atomic positions at any time-step, the knowledge of the potential energy surface of a system, which is defined by the interactions between the particles, is a prerequisite, since the forces acting on the particles can be derived from this potential energy surface by partial differentiation $\vec{F} = -\nabla_{\vec{r}} V(\vec{r})$. The forces $\vec{F}$ acting on the particles are related to the acceleration $\vec{a}$ of the particles and their mass m by Newton's law ($\vec{F} = m\vec{a}$). If the

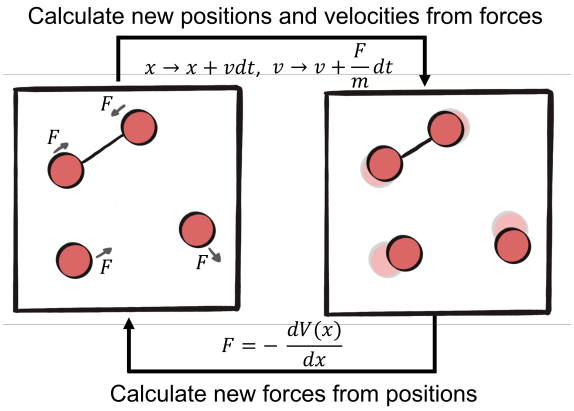


Figure 3.5: In each simulation step the new positions of the particles are calculated. Adapted from Ref. 189.

initial positions and velocities are known and no forces/accelerations act on the particles, the trajectory of each particle could be calculated analytically. However, it must be taken into account that a change in the position of the particles also affects the potential energy and thus the forces acting on the particles. Because the many-body nature of typical MD systems prevents analytical integration, the trajectory is propagated numerically in small time increments Δt. Thus, the 'new' position and velocity of a particle are calculated from its 'old' position, velocity and the corresponding forces acting on the particle (Fig. 3.5). However, this approach introduces (small) errors due to the finite size of Δt and the finite precision of the calculations.

The algorithms most commonly used in MD are the Verlet, Velocity-Verlet, and Leapfrog algorithms. All three algorithms belong to the Verlet family of algorithms and can be derived by a third-order Taylor expansion of the time-dependent position of the particle. They all meet the general requirements for a numerical algorithm, such as efficiency of the computation, accuracy of the computed parameters and low memory usage. For the three algorithms mentioned here, the total truncation error is $\mathcal{O}(\Delta t^2)$ [1] [191, 192] for both, the positions and velocities. Additionally the memory requirements are approximately the same for all three algorithms. More importantly, the Verlet-family algorithms are symplectic integrators, which means that the volume in phase-space is preserved. They are also time-reversible like Newton's equations, meaning that a simulation that is run with a negative time-step from the end of any trajectory will retrace that trajectory until it reaches its initial conditions. Thus, the most important criteria of an algorithm for MD simulations are met, namely time reversibility (proven by Tuckerman, Berne and Martyna in 1992 using the the Liouville formalism [193]), calculation of velocity, and approximate conservation of energy, angular momentum, and total momentum. The order of operations to calculate positions, velocities and forces from each other varies for each variant of the Verlet algorithms and is depicted in Fig. 3.6.

For the speed of the simulations the size of the time-step is a limiting factor. It has to be chosen much smaller than the fastest motion in the system to describe it meaningfully

$$\Delta t \ll t_{fastest} \ . \tag{3.13}$$

[1] In 1892 the $\mathcal{O}$ notation was established by Paul Bachmann to express the growth rate or the magnitude of complex functions by a simple function [190].

In atomistic simulations the fastest atomic motions in a biomolecule are the hydrogen bond stretching vibrations, which are in the order of $10^{-14}\,\mathrm{s} = 10\,\mathrm{fs}$ [194]. To meaningfully describe these fastest motions of a system, the size of the time-step must be set an order of magnitude smaller. Thus, for atomistic simulations, the time-step should be 1 fs. With such a small time-step of 1 fs many time increments are needed to reach the desired simulation timescales, resulting in enormous computational costs. Therefore, in practice to avoid the fastest bond oscillations, those of the hydrogen bonds, hydrogen bond lengths are often constrained using LINCS (**LIN**ear **C**onstraint **S**olver) [195, 196]. As a consequence, the freezing of the fast degrees of freedom with LINCS allows the use of larger time-steps, making the simulation more stable and much faster. The use of constrained hydrogen bond lengths thus makes it possible to use a time-step of 2 fs instead of 1 fs, making the simulation twice as fast.
One can increase the time-step even further by removing the hydrogens completely by converting the hydrogens into virtual sites. This means that the hydrogens are not treated as independent particles. Instead, their positions are recalculated at each time-step using fixed geometric rules based on the positions of nearby heavy atoms. Since these hydrogens are not longer directly involved in the force integration, their high-frequency vibrations are effectively removed, allowing the use of a larger time-step of 4 fs without compromising the stability of the simulation. This makes it feasible to achieve longer simulation times with less computational costs.

Verlet-algorithm

The Verlet algorithm, which can be derived from a Taylor expansion of the position of the particle $x(t)$, is an iterative two-step procedure that calculates the position of the next time-step $x(t + \Delta t)$ using the value of the current time-step $x(t)$ and the value of the previous time-step $x(t - \Delta t)$

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{\vec{F}(x(t))}{m}\Delta t^2 + \mathcal{O}(\Delta t^4) \ , \tag{3.14}$$

where the force F can be calculated from the (known) potential energy as

$$F(x(t)) = -\frac{\partial V(x(t))}{\partial x} \ . \tag{3.15}$$

Since the parameter $x(t-\Delta t)$ in Eq. (3.14) is usually not given or not known for the

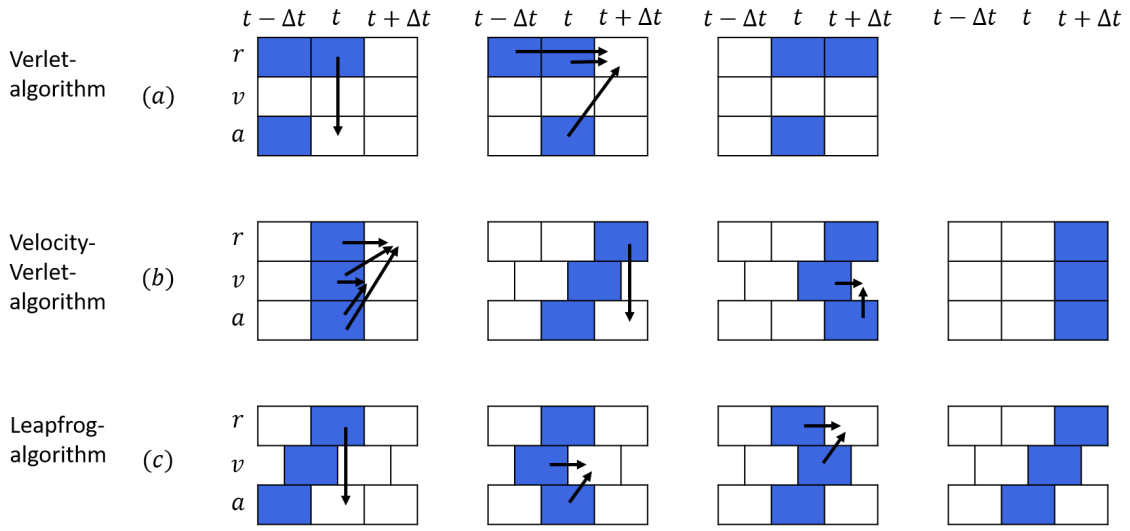


Figure 3.6: Schematic representation of the algorithms from the Verlet-family used in MD simulations: (a) Verlet-algorithm, (b) Velocity-verlet-algorithm, (c) Leapfrog-algorithm. The blue fields are indicating the stored parameters at each time-step. Arrows indicate that a variable is required to calculate the field the arrow points to (Fig. adapted from Ref. 197).

initial time-step at $t = 0$, this first time-step has to be calculated separately, e.g. it is possible to calculate this step by performing one step using Euler's method

$$x(t + \Delta t) = x(t) + v(t)\Delta t \ . \tag{3.16}$$

The calculation of the velocity v, which is not needed to perform the integration but necessary to determine the kinetic and total energies, has to be done separately and over two time-steps:

$$v(t) = \frac{x(t + \Delta t) - x(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \ . \tag{3.17}$$

The numerical properties of the Verlet-algorithm are not quite optimal, since on the one hand a subtraction is used and on the other hand the Δt^2 term becomes very small for small Δt values, which can lead to inaccuracy's. Furthermore, note that the velocities of the particles are not calculated directly, which are often needed, to monitor the total or kinetic energy of the system or to control the temperature.

Velocity-Verlet-algorithm

A modified version of the Verlet-algorithm is the Velocity-Verlet-algorithm, in which the velocity is explicitly calculated and thus the total energy can easily be calculated for every time-step. This modified variant is using the velocities instead of $x(t-\Delta t)$. Thus, the value of $x(t+\Delta t)$ is calculated from the current value and the current instantaneous velocity and acceleration:

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \ . \tag{3.18}$$

The acceleration is calculated from the force:

$$a(t) = \frac{F(x(t))}{m} = -\frac{1}{m}\frac{\partial V(x(t))}{\partial x} \ . \tag{3.19}$$

Half of the velocity-update for the next time-step is calculated from the current acceleration and half from the acceleration for the next time-step:

$$v(t+\Delta t) = \underbrace{v(t) + \frac{1}{2}a(t)\Delta t}_{v\left(t+\frac{\Delta t}{2}\right)} + \frac{1}{2}a(t+\Delta t)\Delta t \ . \tag{3.20}$$

Compared to the Verlet-algorithm, the Velocity-Verlet-algorithm is self-starting, since no value from a previous step is used. The numerical properties are slightly better than those of the Verlet-algorithm since the subtraction no longer occurs, making it less prone to rounding errors, but still potentially very small Δt^2 terms are added to the positions.

Leapfrog-algorithm

In 1970 the leapfrog-algorithm was developed by R. W. Hockney 1970 [198] and is another equivalent formulation of the Verlet-algorithm [199]. The positions and accelerations are calculated at full time-steps, whereas the velocities on half time-steps. The algorithm takes its name from the children's game "Leapfrog", in which one has to jump over the other (the connection becomes clear if the procedure of the algorithm is recorded as in Fig. 3.7).

$$x(t + \Delta t) = x(t) + v\left(t + \frac{\Delta t}{2}\right) \cdot \Delta t \tag{3.21}$$

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + a(t) \cdot \Delta t \ . \tag{3.22}$$

For the calculation of the total energy of a system, the kinetic energy and the potential energy must be available at the same time-step $x(t)$ and thus, the velocity has to be available at full time-steps $x(t)$. The full time-step velocity can be approximated as

$$v(t) = v\left(t - \frac{\Delta t}{2}\right) + \frac{1}{2}a(t) \cdot \Delta t \ . \tag{3.23}$$

Note that the next half time-step velocity is related to the full time-step velocity as

$$v\left(t + \frac{\Delta t}{2}\right) = v(t) + \frac{1}{2}a(t) \cdot \Delta t \ . \tag{3.24}$$

The Leapfrog algorithm is self-starting, which means no special treatment of the first time-step is required and it reduces the effect of rounding errors, since no subtractions or Δt^2 terms are used.

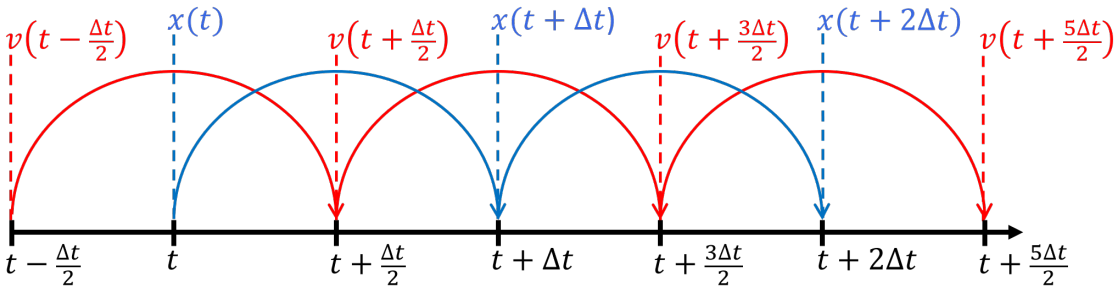


Figure 3.7: Schematic representation of the leapfrog-algorithm. The positions are calculated at full time-steps (blue) and the velocities at half time-steps (red). (Fig. adopted from Ref. 145.)

3.5 Statistical ensembles

Experiments in the laboratory are usually performed at known thermodynamic states defined by particle number N, pressure P, volume V, temperature T and energy E. Therefore, it is important to represent these thermodynamic states as accurately as possible in the simulations. A statistical description of a system in which certain thermodynamic variables – such as temperature, pressure, or volume – are kept constant is referred to as a thermodynamic ensemble. Fig. 3.8 visualize the three most used ensembles in MD:

- NVE - microcanonical ensemble: Describes an isolated system enclosed in a fully isolated container. The system does not exchange heat or pressure with its surroundings.
- NVT - canonical ensemble: Describes a system that is in contact with a heat bath. It allows energy exchange in the form of heat.
- NPT - isothermal-isobaric ensemble: Describes a system that is in contact with a heat bath and a pressure bath. It allows energy exchange in the form of heat and by adapting the volume.

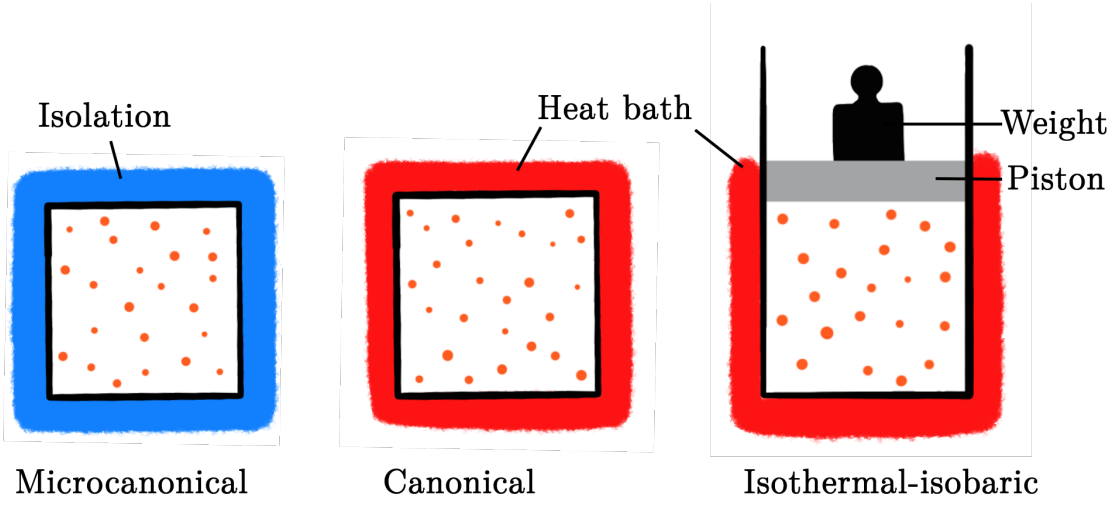


Figure 3.8: Visualization of the three most common statistical ensembles used MD: (i) Micro-canonical ensemble, with constant number of particles (red dots), volume and energy (isolation in blue); (ii) Canonical ensemble, with constant number of particles (red dots), volume and temperature (isothermal bath in red); (iii) Isobaric-isothermal ensemble, with constant number of particles (red dots), pressure (piston with a weight) and temperature (isothermal bath in red). (Fig. adopted from Ref. 145.)

3.6 Temperature and pressure coupling

The conservation of energy in MD simulations is only approximately given, due to the numerical algorithms used, in which drifts in the kinetic energy and thus in the internal energy can occur due to numerical approximations and errors in the integration of the equations of motion. Since the simulations are to be carried out at a specific thermodynamic ensemble, as mentioned in Sec. 3.5, it is necessary to counteract this energy drift. This is achieved by using thermostats, which act as external heat baths to control temperature and maintain energy stability, while barostats serve as pressure baths to regulate pressure and ensure the correct ensemble conditions.

For the temperature, the velocity-rescaling thermostat [200] is often used nowadays, which is basically the Berendsen thermostat [201], but with an additional stochastic term that allows for temperature fluctuations and thus reproduces a correct thermodynamic ensemble.

To reach the target pressure, the Berendsen barostat [201] is often used during the equilibration of the simulation system, because it rapidly adjusts the pressure of the system toward the desired target value through smooth, exponential scaling. This rapid response helps to stabilize the system efficiently at the beginning of the simulation. However, the Berendsen method does not reproduce the correct physical pressure fluctuations, meaning it only approximates the target ensemble which can lead to inaccurate thermodynamic sampling.

For this reason, the Parrinello-Rahman barostat [202] is often used during the subsequent MD production runs. Unlike Berendsen, it allows both volume and shape fluctuations of the simulation box, enabling the system to accurately sample the NPT ensemble. This results in physically realistic pressure fluctuations and more reliable thermodynamic properties. Although Parrinello-Rahman takes longer to stabilize the system, its accurate statistical sampling makes it essential for production runs where precise ensemble behavior is required.

3.7 Periodic boundary conditions

Because the computational effort of performing an MD simulation scales with the number of atoms, typical simulation systems are much smaller than typical experiments. In most cases, the calculation is performed with a single copy of the biomolecule that is surrounded by just enough water to prevent artifacts from intermolecular interactions. To avoid surface effects, which would have a large effect on

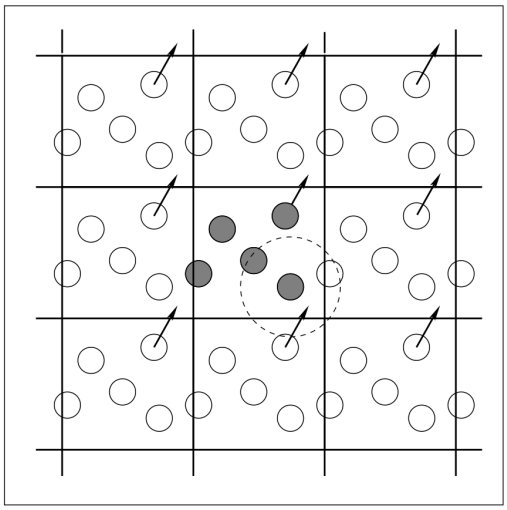
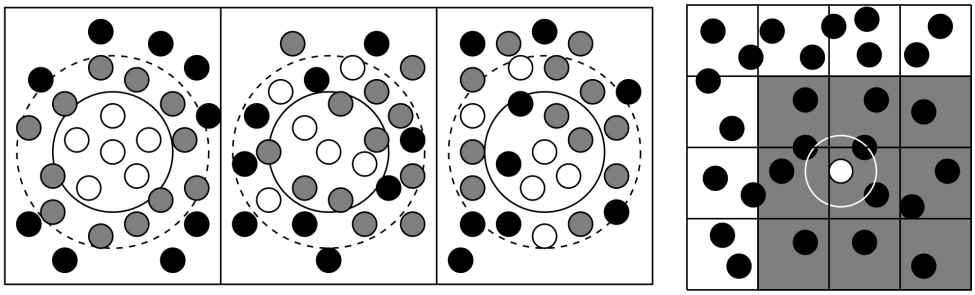


Figure 3.9: Particles in a two-dimensional square-shaped simulation box (gray) are surrounded by eight neighbor boxes, that are copies of the central box translated in all spatial directions. The dashed circle marks the cut-off, which is used to protect particles from interacting with copies of themselves or with other particles and simultaneously with their copies. For a cubic box the cut-off must be smaller than half the box lengths to avoid self-interactions. When a particle leaves the box to one side, the particle enters the box on the opposite side with the same velocity (circles with arrows) (Fig. adopted from Ref. 203).

the biomolecule because the surface would be very close, one typically employs periodic boundary conditions. Under periodic boundary conditions (PBCs), the small system under consideration is located in an infinitely large system, which consists of copies of the small system shifted in all spatial directions. The use of PBC thus allows to perform simulations with a relatively small number of particles, while the particles experience forces as in an infinite system. If an atom moves out of the simulation box, it moves to the neighboring box and on the opposite side a periodic image of the particle enters the box. This treatment also has the side-effect of keeping the particle number in the system constant. Using the minimum image convention, the forces on a particle are calculated by using the closest image of each particle. Therefore, box size must be larger than twice the cut-off, because otherwise the particle will interact with itself. In Fig. 3.9, the concept of PBCs is illustrated using square boxes in two dimensions.

3.8 Neighbor lists

Neighbor lists are used in MD to efficiently compute pairwise interactions between particles in a system. The use of neighbor lists reduces the number of calculations and speeds up the runtime of MD simulations. A neighbor list is a list containing all neighboring particles for each simulation particle. When calculating the interactions between particles, the neighbor list is used to check which particles are within the boundary distance. Only the interactions between particles that are listed as neighbors need to be calculated, because the interactions between distant particles are set to zero if they are outside of the cut-off distance and can be neglected. Thus, the number of pairwise interactions to be calculated can be considerably reduced, which in turn reduces the computational costs and makes it possible to simulate larger and more complex systems. The neighbor list can be constructed using various algorithms, such as the Verlet list or the cell list method (see Fig. 3.10). When using the Verlet list method, first a shell is constructed around each particle in the simulation system with a radius r_ν that is slightly larger than the largest non-bonded cutoff r_C of the particle of interest. Subsequently, neighborhood lists has to be created and saved for each particle in the system, storing all particles within the shell of the particle under consideration. The cell list method is often used for larger simulation systems. In this method, the simulation space is divided into boxes of equal size with an edge length of r_C, which corresponds to the largest non-bound cutoffs of the particles in the system. Using this technique, a particle can only interact with particles in the same box or in adjacent boxes. The neighbor lists are then updated at regular intervals to account for changes in the positions of the particles. In case of the Verlet list only the positions of particles that have moved beyond a certain distance since the last update need to be checked. A schematic representation of the Verlet list and the cell list methods are shown in Fig. 3.10.



(a) (b)

Figure 3.10: (a) Visualization of the Verlet-list in two dimensions. At the starting point (leftmost panel) the Verlet list is set. The atoms are shown as spheres, where the atoms inside the cutoff-range r_C (solid line) are shown in white; the atoms in the outer skin cut-off range r_ν (dashed line represents) are shown in gray; and all atoms further away from the central atom than r_ν are shown in black. The positions of the moving particles after a certain time are shown in the middle box and the right box is depicting the positions even later. The Verlet lists are to be updated when a particle travels a distance greater than $r_C - r_\nu$; this means that the lists must be updated before the black atoms move in the solid circle and thus before the situation in the left image occurs (Fig. adopted from Ref. 203).
(b) Visualization of the cell-list technique in two dimensions. This technique is usually used for larger simulation systems. In it the simulation space is divided into boxes of equal size with edge length r_C (cut-off range of the particles). This list technique exploits the fact that a particle (white) can interact only with particles in the same box or with particles in the immediate neighboring boxes (boxes with which the white particle interacts are colored gray) (Fig. adopted from Ref. 203).

3.9 Steps to run MD simulations

To perform a MD simulation five main steps that are shown in Fig. 3.11 are required and these steps are described in more detail below.

1. Set up the system: Create the molecular system of interest, including the types and initial positions of atoms or molecules, and any required boundary conditions or constraints. In most cases, the initial structure of the biomolecule under investigation has been obtained by an experimental method such as crystallography, cryo-EM or NMR. If the structure of the biomolecule has already been published, it can be found in the protein data bank (PDB) [33, 204, 205], otherwise it is also possible to receive it directly from the experimentalists. Since several years, 3D structures of proteins can also be predicted directly from the amino acid sequence using AI systems such as alpha fold [206, 207] or RoseTTA fold [208]. However, the predictions of the AI systems should be checked critically because these systems still have problems predicting e.g. membrane proteins or asymmetric protein complexes [209–211]. Regardless of which method was used to determine the protein structure, usually only a single structure is obtained, in most cases the native structure, which is located at a (deep) minimum of the potential surface. In this step an appropriate force field has to be selected that best describes the interactions between the atoms in the system. The force field consists of mathematical functions that determine the potential energy and forces acting on the particles as described in section 3.2.
2. Energy minimization: Perform a number of steps of gradient descent along the negative gradient of the energy to optimize the initial configuration of the system and reduce unrealistic atomic overlaps or deformed geometries.
3. Equilibration: Allow the different degrees of freedom of the system to equilibrate by applying a series of simulation steps, such as a short molecular dynamics simulation at constant temperature and/or pressure.
4. Production run: Run a longer MD simulation at the desired conditions (e.g., specific temperature, pressure, or external fields) to generate meaningful data and observe the dynamics of the system over time. During the simulation, Newton's equation of motion is solved numerically over small time-steps, describing how the positions and velocities change over time. This implies that at each time-step, positions and velocities are updated based on the forces acting on them, determined from the interactions between the particles.

5. Data Analysis and Interpretation: Analyze the output trajectory of the MD simulation, this includes extracting thermodynamic properties, calculating atomic or molecular motions, and studying structural changes or interactions in the system. Interpret the simulation results in the context of the research questions or objectives. Compare the simulated data with experimental observations or other theoretical models to validate the accuracy and reliability of the simulation.

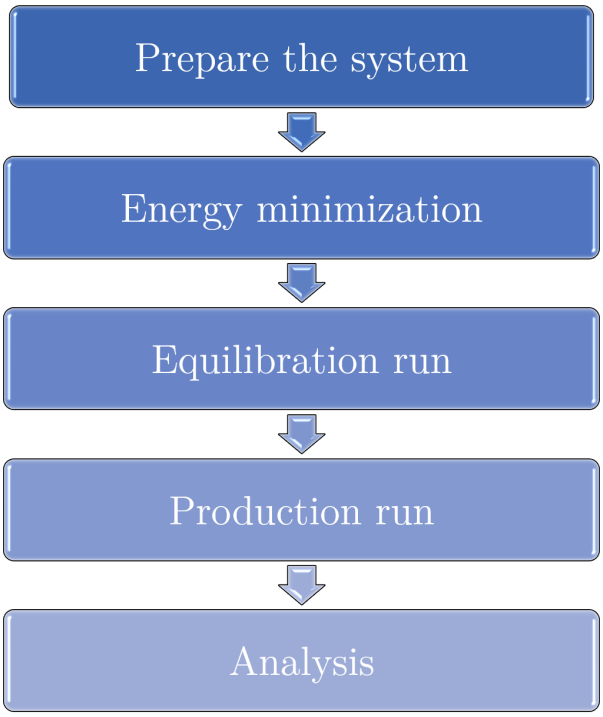


Figure 3.11: Schematic visualization of the steps required for a MD simulation. (Fig. adopted from Ref. 145.)

Chapter 4

Combining experiments with MD simulations

The combination of experimental techniques such as SAXS, SANS, and NMR with MD simulations can provide a powerful and insightful tool for the study of biomolecular structures and dynamics, since no single method alone can provide all the necessary information in this multidisciplinary field. By combining experimental techniques with MD simulations, researchers can validate and refine experimental structural models with MD simulations and similarly, experimental results can guide and improve the accuracy of MD simulations. The synergy between experimental data and computational models can help to overcome the limitations of each technique leading to a more accurate and comprehensive understanding of biomolecular systems.

4.1 Small-angle scattering

Small-angle scattering (SAS) is an experimental diffraction technique that can be used to study biological macromolecules in solution. The main part of this section is based on Ref. 212 & 213. The sample, typically a protein or a macromolecular complex in solution, is exposed to a highly collimated monochromatic beam – often a single wavelength or a narrow band of wavelengths – of X-rays (SAXS) or neutrons (SANS). As the waves penetrate matter they will interact with individual atoms (Fig. 4.1), producing secondary waves that can interfere constructively or destructively. X-rays scatter at electrons while neutrons scatter at nuclei. The coherent, elastically scattered radiation is then detected by a two-dimensional detector, while the direct beam

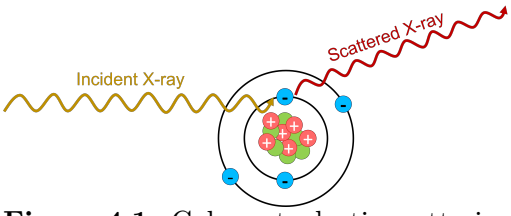


Figure 4.1: Coherent, elastic scattering of a X-ray at the electrons. The incoming X-ray beam (dark orange) is scattered at the electron (blue) without losing energy, which means that the scattered X-ray has the same energy as the incident X-ray beam.

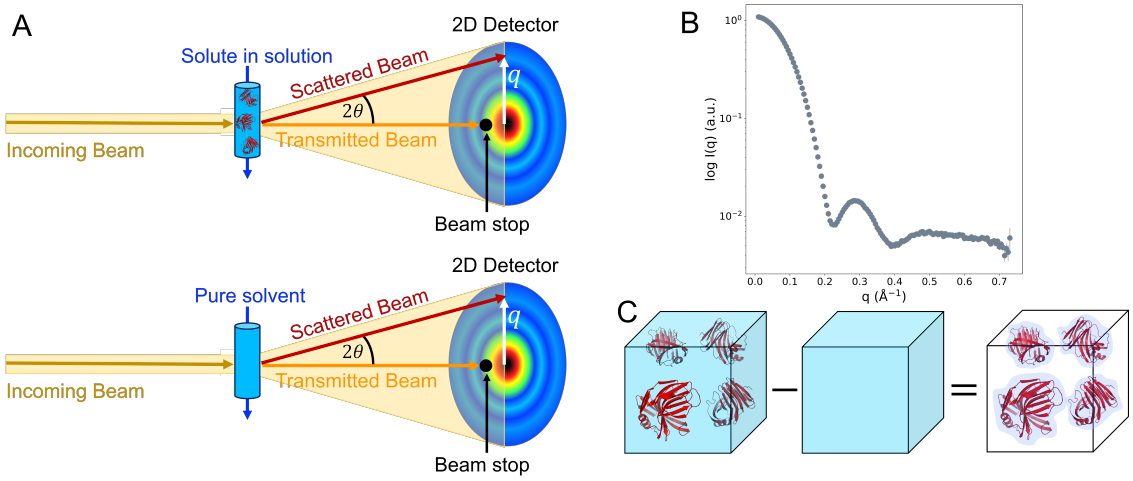


Figure 4.2: (A) Schematic representation of the experimental setup of SAS measurements. The sample is exposed to a highly collimated beam (dark yellow) of X-rays or neutrons. The scattered radiation (dark red) is recorded on a detector. The radiation passing through the sample unscattered (orange), the direct beam, is absorbed by a beam stop (black point). (B) One dimensional scattering curve calculated by circular averaging from the detector data. The scattering intensity is plotted against wave vector q. Data for xylanase (PDB ID: 2DFC, SASBDB [214] entry: SASDP35) are taken form Ref. 212. (C) Schematic representation of the meaning of subtracting the intensity of the solvent from that of the biomolecule in the solvent. As a result, the scattering intensity of the biomolecule and the hydration shell is obtained, which provides information about both.

is absorbed by a beam stop. In small-angle scattering experiments, small angles are often between 0.1° and 10° measured, leading to a resolution in the nanometer range. A schematic representation of the experimental setup of SAS measurements is shown in Fig. 4.2A.

The collected scattering patterns on the detectors are circular, because the scattering objects, e.g. biomolecules, should be randomly oriented in the solution. Therefore, by using circular averaging, the 2D scattering pattern detected by the detector can be reduced to a 1D scattering curve by plotting the measured scattering intensity I against the momentum transfer q as shown in Fig. 4.2B. The momentum transfer q is given by

$$q = \frac{4\pi}{\lambda} \sin(\theta) \ , \tag{4.1}$$

where λ is the wavelength of the incoming radiation and θ is half the angle between the incoming and scattered radiation [212]. Since λ is fixed during an experiment and θ is small, the 1D scattering plot $I(q)$ – I versus q – is basically the scattering

intensity I as a function of the scattering angle 2θ [213]. Thus, due to the fact that θ is independent of the wavelength of the incident radiation, unlike q, it is also possible to compare scattering curves from SAS experiments performed with different wavelengths λ.
In a scattering experiment, all atoms in the sample are contributing to the observed scattering intensity. Thus, the solvent has also a significant contribution to the total scattering intensity. To obtain information about the biomolecule without the bulk solvent the scattering intensity of the biomolecule diluted in the solvent and that of the pure solvent are measured separately. Subsequently, the intensity of the solvent is subtracted from that of the biomolecule in the solvent to obtain the scattering intensity of the biomolecule and the hydration shell as shown in Fig. 4.2C, because both the density of the protein and the density of the hydration shell differ from those of the pure solvent. As a result, scattering information are obtained from the biomolecule under investigation as well as from its hydration shell.
In SAS experiments, the information content is limited by the rotational averaging of the molecules rather than by the experimental resolution [213], which is why the method is often described as low resolution method. Nevertheless, SAS experiments can provide highly precise measurements of parameters such as the radius of gyration (R_{g}) and the overall shape of the scattering particle(s).
The sample under investigation has to contain monodisperse, identical biomolecules in order to obtain high-quality scattering data, from which precise structural parameters – such as an accurate radius of gyration (R_{g}) – can be calculated.
From the scattering geometry, the momentum transfer is given by

$$q = \frac{4\pi}{\lambda}\sin\,(\theta)\ ,$$

where λ is the wavelength and 2θ the scattering angle. Small real-space distances between scattering centers correspond to large values of q (high scattering angles), while large distances – such as the overall size and shape of a biomolecule – are probed at small q (low scattering angles) (Fig. 4.3) [213].

Pair distance distribution function (PDDF)

The structure of a biomolecule in solution determines the shape of the scattering profile $I(q)$, which, however, cannot be interpreted intuitively due to its complex form. In order to analyze the structure of the biomolecule anyway, it is recommended to Fourier-transform $I(q)$ to obtain the pair distance distribution function (PDDF) –

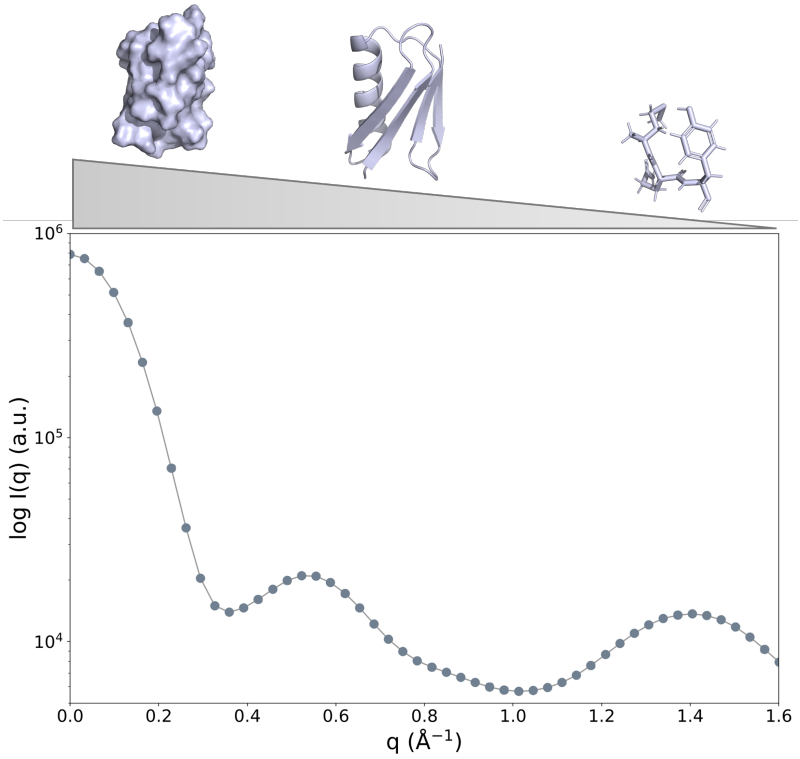


Figure 4.3: A scattering plot of a GB3 domain (PDB code: 2OED [215]) is shown, where the scattering intensity I is plotted against q. In the low q range, between 0 and $0.2\,\text{Å}^{-1}$, information about the shape, density and size of the protein can be obtained; information about the spatial arrangement and the overall fold of the protein can be obtained in the range between $0.2\,\text{Å}^{-1}$ and $1\,\text{Å}^{-1}$; and in the range above $1\,\text{Å}^{-1}$ information about the atomistic structure can be determined.

also called the radial Patterson function $p(r)$ – describing the probability distribution of the interatomic vector lengths $\vec{r}$ within the biomolecule [212]. In SAS experiments, the q values can not be measured up to infinity – as required for a Fourier-transform on $I(q)$ – and therefore, an indirect Fourier-transform has to be applied instead as initially published by O. Glatter in 1977 [216]. Since the indirect Fourier-transform on $I(q)$ requires an estimated value of the maximum particle dimension $D_{\max}$ and the assumption that PDDF is zero at $r = 0$, the thus PDDF is depending on the accuracy of these assumptions. The PDDF profiles calculated in this way are sensitive to symmetry and domain structure within proteins and thus can provide valuable information about the shape and volume of the biomolecule. Since the PDDF function is the probability distribution of the distance between any two points within the biomolecule, it provides information about the size, shape and internal structure of the biomolecule.

Radius of gyration and zero-scattering intensity

The intensity of the beam scattered at zero angle ($2\theta = 0\,^\circ$), the forward scattering (or zero-angle) intensity $I(0)$, is the difference between the number of scattering particles in the solute including the protein hydration shell $N_{\text{solute,PHS}}$, and pure solvent N_{solvent} squared inside a volume of the same size:

$$I(0) = [N_{\text{solute,PHS}} - N_{\text{solvent}}]^2 \ . \tag{4.2}$$

By describing the number of scattering particles in terms of volumes V and densities ρ, $I(0)$ can be described by

$$I(0) = [V_{\text{solute}} \cdot (\rho_{\text{solute}} - \rho_{\text{solvent}}) + V_{\text{PHS}} \cdot (\rho_{\text{PHS}} - \rho_{\text{solvent}})]^2 \ , \tag{4.3}$$

where $N_{\text{solute}} = V_{\text{solute}} \cdot (\rho_{\text{solute}} - \rho_{\text{solvent}})$ and $N_{\text{PHS}} = V_{\text{PHS}} \cdot (\rho_{\text{PHS}} - \rho_{\text{solvent}})$ are the number of scattering particles relative to the pure buffer in the biomolecule and in the hydration shell, respectively. Overall, $I(0)$ is equal to the square of the total number of scattering particles in the scattered volume and can therefore be used to estimate the molecular mass of the solute.

In SAS experiments $I(0)$ cannot be measured (directly) because the scattered radiation cannot be distinguished from the direct beam – the radiation that passed through the sample unscattered. Therefore, $I(0)$ can only be determined by extrapolation from experimental SAS data.

R_{g}^2 is defined as the average scattering density weighted squared distance of all scat-

terers to the center of mass. It measures the compactness of a protein structure and provides information about the mass distribution within a biomolecule. Therefore, R_{g} is sensitive to the shape and compactness of biomolecule and thus, two different biomolecules with the same volume, but different (three-dimensional) shapes may have different R_{g} values.

There are two methods to determine the R_{g} and $I(0)$, namely via the Guinier approximation or the PDDF. In the Guinier approximation, only the q and $I(q)$ values in the very small q-region – the Guinier region – are considered, since this small q-region of the scattering curve is characteristic for the overall dimension of the biomolecule. Therefore, the the $I(q)$ values in the small q-region must be of high accuracy in order to obtain high quality R_{g} values with the Guinier method. The Guinier approximation was derived in 1939 by André Guinier and is given by [217]:

$$I(q) = I(0)\exp\left(\frac{-(q^2 R_{\mathrm{g}}^2)}{3}\right) . \tag{4.4}$$

Thus, R_{g} and $I(0)$ can be determined by a linear fit of $\ln[I(q)]$ versus q^2, the slope of the straight line can be used to compute R_{g} and the y-axis intercept represents the zero-scattering intensity $I(0)$. The Guinier approximation is particularly useful for globular particles where the relationship between the scattering intensity and the scattering vector holds in the Guinier region, the very small angular range. This relationship is valid for globular proteins only as long as $qR_{\mathrm{g}} < 1.3$ [212, 218, 219]. For elongated proteins and peptides, this upper limit is lower. Since the Guinier fit must be linear, especially for globular proteins, it can also be used to quickly and efficiently determine sample quality [212]. Deviations from linearity usually indicate strong intermolecular effects, such as aggregation or repulsion, polydispersity of the samples, or improper background subtraction.

Unlike the Guinier method, the PDDF method uses the entire scattering profile, which makes the PDDF method less sensitive to intermolecular interactions. The PDDF method can be used for particles with small Guinier regions and for noisier data. To obtain accurate results with the PDDF method, the data in the high q range must be of high quality, and there must still be reasonably good low q data. Then, especially for experimental data, this method provides more accurate values of R_g and $I(0)$ compared to the Guinier fit. By using the PDDF, the R_{g} can be

calculated by

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 p(r) dr}{2 \int_0^{D_{max}} p(r) dr} \ , \tag{4.5}$$

where R_g is the second moment of the $p(r)$ function [212].
$I(0)$ is the zero moment of the PDDF function, which corresponds to the area under the curve [212]:

$$I(0) = 4\pi \int_0^{D_{max}} p(r) dr \ . \tag{4.6}$$

The relatively uncomplicated determination of R_g and $I(0)$ by means of scattering measurements allows structural information to be obtained rapidly. The possibility to calculate R_g and $I(0)$ with different mathematical models, such as the Guinier approximation or the PDDF, allows a validation of the parameters and an assessment of the internal consistency and robustness of the experimental small angle scattering data.

SAXS and SANS - Two complementary methods

Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) are two techniques for investigating the structural properties of biomolecules in solution at the nanoscale that can complement each other very well. Due to the use of X-rays in SAXS and neutrons in SANS, the two techniques SAXS and SANS show some distinct differences that result in both advantages and disadvantages.
In SAXS experiments X-rays are used to examine the sample. The X-rays are scattering mostly at the electrons in the sample, resulting in electron density contrast. The scattering factors for X-rays are positive and increase with the atomic number, which means that no difference can be detected between hydrogens and deuteriums.
In SANS experiments neutrons are used for probing that interact mostly with the atomic nuclei. Neutrons are thus not dependent on electron density and therefore provide a different type of contrast, namely the neutron scattering length-density contrast. The coherent scattering lengths of atoms can be positive and negative and are irregular. For example, the coherent scattering length of hydrogen is slightly negative ($-0.3741\ 10^{-12}\text{cm}^2$ [220]) and the coherent scattering length of deuterium is highly positive ($0.6671\ 10^{-12}\text{cm}^2$ [220]), resulting in a large difference between the coherent scattering lengths of hydrogen and deuterium. Therefore, it is possible

to utilize deuterium substitution to influence contrast in so-called contrast variation studies, described in the next section, making it a powerful tool for studying complex biological macromolecules.
Nowadays, SAXS experiments can be performed at synchrotrons or free-electron lasers or at in-house instruments, whereas SANS experiments can only be performed at neutron facilities, as a neutron source is required. Therefore, the access to the experimental SAXS instruments is much easier than to SANS instruments. Another advantage of SAXS over SANS experiments is that generally smaller sample quantities are required in SAXS experiments compared to SANS experiments. Furthermore, in contrast to neutrons, X-rays have a lower incoherent scattering background. Therefore, data analysis in SANS experiments can be more complicated than in SAXS experiments, since the background subtraction in SANS experiments has to be done very carefully.
Both experimental techniques, SAXS and SANS, can be carried out in batch mode or with size exclusion chromatography (SEC) [221, 222]. By using the batch mode, the sample is added directly to the sample cell, whereas by using the SEC mode, the sample first pass through a separation column, in which the molecules in solution can be separated according to their size or molecular weight. One advantage of using the SEC mode is that the buffer scattering is measured in-line during the separation, reducing errors from preparing and measuring a separate buffer sample and improving the accuracy of buffer subtraction, while also allowing separation of aggregates, dimers, and monomers. However, the SEC mode requires larger sample quantities than the batch mode.
Overall, both experimental techniques, SAXS and SANS, have their advantages and disadvantages, but can complement each other very well.

Contrast variation in SANS

In SANS experiments, it is possible to vary the scattering length density of the buffer as well as of parts of the biomolecules [213]. This is an advantage over other scattering techniques, like SAXS. In contrast variation experiments, hydrogens are replaced by heavy hydrogens, so-called deuteriums. This allows certain parts of the protein to be made ´invisible' and thus the other parts can be observed individually. In SAS experiments with contrast variation, solvents with concentrations of 0%, 40%, and 100% D_2O are often used. At 0% D_2O, hydrogenated biomolecules generally exhibit strong positive contrasts and are fully visible. At approximately 40% D_2O, the solvent scattering length density can match that of hydrogenated pro-

tein, making those parts effectively invisible while deuterated regions remain visible, as shown in Fig. 4.4. At 100% D_2O, the solvent scattering length density is close to that of a fully deuterated protein, so that the deuterated regions may become invisible, while hydrogenated regions usually remain visible with negative contrast. However, the exact match points depend on the type of biomolecule (e.g. protein, RNA or DNA) and its deuteration level. Therefore, scattering data obtained with different D_2O concentrations in the solvent can provide valuable insights into the shapes and orientations of deuterated and non-deuterated parts of biomolecules or biomolecular complexes.

4.1.1 Calculating SAS curves from explicit solvent MD simulations

The method used in this work to calculate SAS data from explicit solvent MD simulations was proposed by Po-chia Chen and Jochen S. Hub in 2014 [28] and is based on the methodology of Park et al. from 2009 [224]. The detailed procedure of the method is described in Ref. 28. In this method, $I(q)$ is calculated in the same way as in SAS experiments, namely derived from the difference between the solute and the solvent:

$$I(q) = I_{\text{solute}}(q) - I_{\text{solvent}}(q) \ . \tag{4.7}$$

Therefore, similar to experiments, two systems are needed in simulations: (i) one that contains the solute in solvent and (ii) another that contains only the solvent. The simulation box containing the protein must be large enough to avoid interactions with itself over the periodic boundary. Thus the simulation box corresponds to a monodisperse solution as also used and needed in SAS experiments.

To calculate the SAS curves from explicit solvent MD simulation trajectories, an envelope is first constructed around the protein, which includes both the solute and all solvent molecules that belong to the hydration shell. All solvent molecules outside of the envelope are thus bulk solvent-like. By subtracting the intensity of the pure solvent envelope from the intensity of the envelope containing the system with the protein, a SAS curve is obtained which can be compared with an experimental SAS curve. That is sufficient to use only the envelope instead of the whole simulation system to calculate the SAS curve of the solute and the hydration shell was shown in Ref. 28. Thus, by using the envelopes for the calculation of the SAS curves, all atoms inside the envelope contribute to the calculation of the SAS curves and

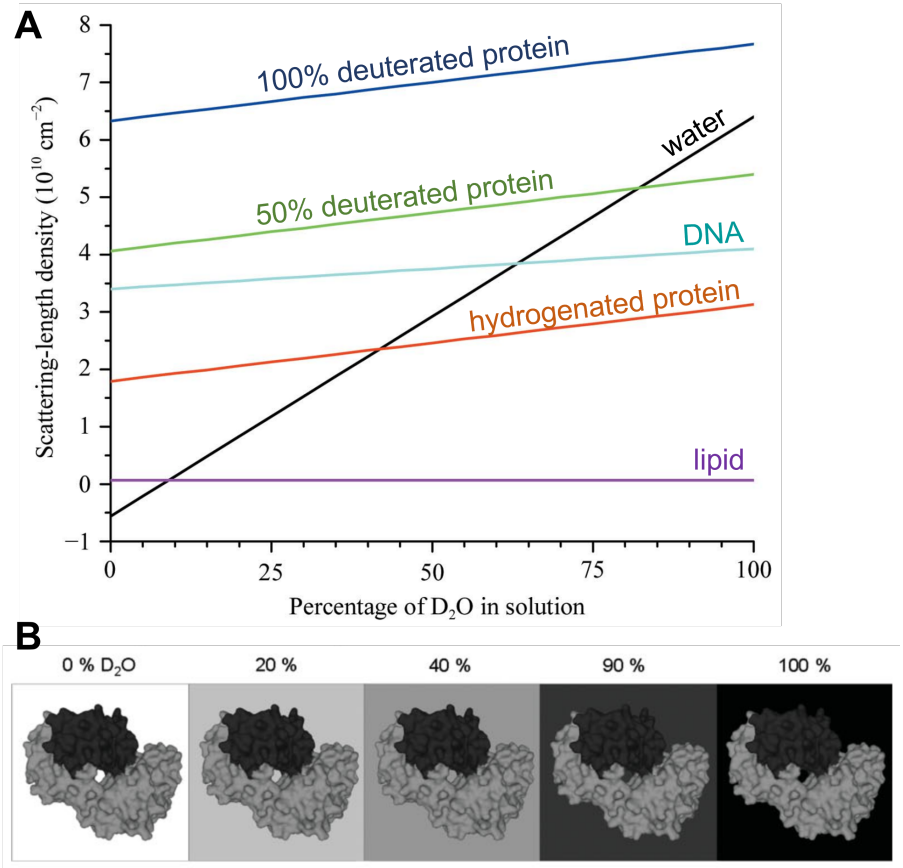


Figure 4.4: (A) The scattering length densities for different biological substances are shown as a function of D_2O, in different colors: water (black), hydrogenated protein (dark orange), protein with 50% of the non-exchangeable protons replaced by deuterium (green), protein with 100% of the non-exchangeable protons replaced by deuterium (blue), DNA with an equal distribution of base pairs (cyan) and a lipid dimyristoyl phosphatidylcholine (violet). (Fig. adapted from Ref. 223.) At the points where the biomolecules and the water have the same scattering density – the intersections between the colored lines and the black line – the biomolecules are invisible in SAS experiments. (B) For a protein which has a hydrogenated part (light grey) and a deuterated part (almost black), this means that the deuterated part and the hydrogenated part are differently visible in different D_2O concentrations in the solution; the hydrogenated protein part is invisible in a solution containing around 40% D_2O and the 100% deuterated protein part is invisible in a solution containing 100% D_2O. (Fig. adopted from Ref. 213.)

all atoms outside the envelope that are bulk-solvent like do not contribute to the calculation of the SAS curve. Overall, the use of envelopes not only saves a lot of computational time, but also reduces statistical noise.
Since there are no free parameters when calculating the SAS curve with the explicit-solvent method, the experimental curve is fitted to the simulated curve when comparing with experiments, using an absolute scale f that accounts for the usually unknown absolute intensity of the experimental SAS curve and an absolute offset c that accounts for both buffer uncertainties as well as uncertainties due to incoherent scattering in SANS experiments:

$$I_{\text{fit}}(q) = f \cdot I_{\text{exp}}(q) + c \ . \tag{4.8}$$

In this method, explicit solvent molecules are used in the simulations, which is the reason that no free parameters for the hydration shell or the excluded solvent are needed during the calculation of the SAS curves. Therefore, the calculated values for I_0 and R_{g} are not adjusted by using fitting parameters.
Other software programs such as Crysol[225, 226], Cryson[227], PepsiSAXS[228], SoftWAXS[229], FoXS[230, 231] or AquaSAXS[232] that are used to evaluate experimental SAS data are applying methods such as a shell with a certain thickness, grid elements, the use of water dummy beads, density maps or voxelization to adjust the hydration shell and excluded solvent using one or several parameters. Since the structure and shape of the hydration shells is very complicated and can vary between different proteins significantly, the use of fitting parameters for the hydration shell and the excluded solvent can lead to inaccurate results. Therefore, the non-use of fitting parameters for the hydration shell and the excluded solvent in the explicit-solvent method is a significant advantage in contrast to many other software programs used to evaluate experimental SAS data.

Restraining MD simulations with experimental SAS data

If the experimental and simulated small- and wide-angle X-ray scattering (SWAXS) curves from unconstrained simulations do not match, possible reasons include experimental inaccuracies, limitations of the force fields or water models, insufficient sampling, or structural differences between the simulated and experimental sturcture, such as variations between solution and crystal structures or the presence of different conformational states (e.g. ground vs. excited state). To search the structural ensemble that represents the experimental data best, if the assumption is that there

are inaccuracies in the simulation, is to couple the MD simulation to a target curve obtained from an experimental SAS curve. In the SWAXS-driven MD method developed by Po-chia Chen and Jochen S. Hub in 2015 [61] the coupling of an experimental target curve to one MD simulation is done by adding a biasing potential energy term $V_{\mathrm{SAS-bias}}$ to the force field potential energy term V_{ff} derived for the conformation of the biomolecule, if the simulated SAS curve does not match with the experimental one:

$$V_{\mathrm{exp-biased\ ff}} = V_{\mathrm{ff}} + V_{\mathrm{SAS-bias}}. \tag{4.9}$$

Only in the case if the simulated SWAXS curve and the one from the experiment matches, the biasing potential $V_{\mathrm{SAS-bias}}$ is 0 resulting in an unchanged energy landscape. In all other cases, the energy landscape in the simulation is changed by adding this biasing potential; in other words, the simulation is carried out with an experiment-biased force field. With this method it is thus possible to overcome sampling issues and force field inaccuracies as shown in Ref. 61. And in the case where the biomolecule under investigation has mostly one conformation, the SWAXS-driven MD method with a single replica is often sufficient.
However, if the biomolecule is intrinsically disordered or has intrinsically disordered regions, the experimental data represent an ensemble from a rather heterogeneous ensemble. By coupling the experimental data of such (partly) flexible biomolecules to only one single replica, it might still be possible that the simulated SAXS curve will match the experimental one. If the experimental SAXS curve matches with the simulated one by performing a single-replica SWAXS-driven MD simulation for a (partly) flexible biomolecule, the agreement may still be misleading. This is because the experimental data often represent a heterogeneous ensemble of conformations, whereas a single replica can converge to a narrow set of structures that fit the data but fail to capture the true structural diversity. To correctly interpret the data, the target SWAXS curve is coupled to multiple replicas in parallel, allowing each replica to explore different regions of conformational space while collectively matching the experimental ensemble. In this multi-replica approach developed by Markus R. Hermann and Jochen S. Hub in 2019 [233], the potential energies in equation (4.9) are now dependent not just on one, but on several (n) replicas $X_1, .., X_n$:

$$V_{\mathrm{exp-biased\ ff}}(X_1, ..., X_n) = V_{ff}(X_1, ..., X_n) + V_{\mathrm{SAS-bias}}(X_1, ..., X_n). \tag{4.10}$$

Therefore, also the biasing potential energy term $V_{\mathrm{SAS-bias}}(X_1, ..., X_n)$ is now cal-

culated from the ensemble of the multi-replicas, and is thus an replica-averaged restraint. In multi-replica SAXS-driven MD simulations, the experimental curve is matched to the ensemble-averaged scattering curve over all replicas, rather than to only one single trajectory. If the number of replicas is sufficiently large and the simulations are long enough, the system requires only the smallest possible biasing potential to reproduce the experimental data [233]. This minimal-bias approach follows Jaynes' maximum entropy principle [233, 234], yielding the most unbiased probability distribution that still satisfies the experimental constraints.

Chapter 5

Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW

Reprinted from Journal of Chemical Physics, Volume 154, Issue 19, 194501 (2021) (Ref. 235), with the permission of AIP Publishing. We thank the journal for permission to reprint it.

The authors of this paper are Johanna-Barbara Linse and Jochen S. Hub, with the following contributions:

J.-B.L. performed and analyzed MD simulations.
J.-B.L. and J.S.H. designed the project and wrote the article.

The Journal of Chemical Physics

AIP Publishing

RESEARCH ARTICLE | MAY 17 2021

Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW

Johanna-Barbara Linse ; Jochen S. Hub

J. Chem. Phys. 154, 194501 (2021)
https://doi.org/10.1063/5.0050841

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW

Cite as: J. Chem. Phys. 154, 194501 (2021); doi: 10.1063/5.0050841
Submitted: 19 March 2021 • Accepted: 23 April 2021 •
Published Online: 17 May 2021

Johanna-Barbara Linse and Jochen S. Hub[a)]

AFFILIATIONS

Theoretical Physics and Center for Biophysics, Saarland University, Campus E2 6, 66123 Saarbrücken, Germany

[a)]Author to whom correspondence should be addressed: jochen.hub@uni-saarland.de

ABSTRACT

Heavy water or deuterium oxide, D_2O, is used as a solvent in various biophysical and chemical experiments. To model such experiments with molecular dynamics simulations, effective pair potentials for heavy water are required, which reproduce the well-known physicochemical differences relative to light water. We present three effective pair potentials for heavy water, denoted SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW. The models were parameterized by modifying the widely used three- and four-site models for light water, with the aim of maintaining the specific characteristics of the light water models. At room temperature, SPC/E-HW and TIP3P-HW capture the modulations relative to light water of the mass and electron densities, heat of vaporization, diffusion coefficient, and water structure. TIP4P/2005-HW captures, in addition, the density of heavy water over a wide temperature range.

I. INTRODUCTION

Water with the chemical formula D_2O is called heavy water or deuterium oxide. In heavy water, the common light hydrogen atoms (protium, ^1H) are replaced with heavy hydrogen atoms (deuterium, D or ^2H, a hydrogen isotope with a nucleus composed of one proton and one neutron). Despite nearly identical equilibrium structures of H_2O and D_2O molecules, heavy water and light water exhibit different physical properties.[1] Hydrogen bonds in heavy water are stronger than those in light water,[2] which causes enhanced cluster building and thereby an increased structural order at low temperatures in D_2O.[3] At higher temperatures, the viscosity and the heat capacity of D_2O are increased relative to H_2O. D_2O has an increased melting temperature compared to H_2O. The maximum density of D_2O is reached at an increased temperature of 11.2 °C instead of 4 °C.[4]

Heavy water is used in various biophysical experiments. For instance, D_2O is used as a solvent for biomolecules in small-angle neutron scattering (SANS) experiments.[5–7] Because deuterium exhibits a largely increased neutron scattering length as compared to protium, changing the relative H_2O/D_2O concentration modulates the contrast between the biomolecule and the buffer in SANS experiments. This property may render certain biomolecular subunits invisible during the so-called contrast variation experiments. Likewise, D_2O has been used for nuclear magnetic resonance (NMR) spectroscopy[8] and Fourier transform infrared spectroscopy (FTIR).[9] To allow accurate modeling of such experiments by molecular dynamics (MD) simulations, reliable effective pair potentials of heavy water are required.

Models for light water have been developed since a few decades. The simple point charge (SPC) model proposed by Berendsen *et al.* in 1981 was among the first models but remains widely used.[10] The popular TIP3P and TIP4P models by Jorgensen *et al.* followed in 1983.[11] The TIP3P model is a three-site water model, like SPC, but it uses the experimentally observed HOH angle of 104.52° instead of the ideal tetrahedral angle of 109.47° adopted by SPC. TIP4P is a four-site model. In 1987, Berendsen *et al.* proposed the extended simple point charge model (SPC/E), a reparameterization of the SPC model with polarization corrections, thereby taking the different water dipoles in solution and in vacuum into account when comparing simulations with the experimental heat of vaporization.[12] A modified version of TIP3P for the Chemistry at Harvard Macromolecular Mechanics (CHARMM) force field has further been implemented, which has Lennard-Jones (LJ) interactions also assigned to the hydrogen atoms, in contrast to the original TIP3P model.[13] All these models were optimized to reproduce properties such as the density, potential energy, and heat of vaporization purely for liquid water at 25 °C and atmospheric pressure. In 2005, Abascal and Vega proposed an optimized version of TIP4P, denoted TIP4P/2005, by taking the polarization corrections into account and

by fitting against experimental data from a wide temperature and pressure range.[14] The excellent agreement with experimental data over a wide temperature range came with the price of slightly worse agreement with the experimental heat of vaporization at room temperature as compared to SPC/E. Until today, the development of models for light water remains an active field, for instance, with the aim of including electronic polarization[15] or modeling surface effects.[16]

Fewer studies focused on models for heavy water. As the simplest approach, merely the mass of the hydrogen atoms has been doubled to investigate the vibrational spectra of heavy water[17] or to simulate heavy water permeation across aquaporins.[18] However, this approach cannot account for the modified properties of D_2O as compared to H_2O. Therefore, a model for heavy water on the basis of the SPC/E model has been suggested, denoted SPC/HW.[19] To model the larger dipole of heavy water, SPC/HW uses an increased negative partial charge of $q_O = -0.87e$ for the oxygen atom as compared to $-0.8476e$ used by SPC/E while leaving all other parameters except for the hydrogen mass unchanged. The SPC/HW model has been employed to study the effect of heavy water on lipid membrane properties,[20] the membrane–water interface,[21] the structure and dynamics of anions,[22] and the structure and stability of nanocrystals.[23] However, with several versions of the Gromacs simulation software[24] (3.05 and newer), we could not reproduce the diffusion coefficient, heat of vaporization, or density reported in Ref. 19, irrespective of the chosen cutoffs for Lennard-Jones (LJ) interactions, the Coulomb method [particle-mesh Ewald (PME) or plain cutoff], scheme for neighbor lists, corrections for missing dispersion interactions behind the LJ cutoff, or the temperature coupling scheme. We did not test the Gromacs version prior to version 3.05, which has been published in 2001. The properties of SPC/HW obtained with a current Gromacs version are presented below. This discrepancy prompted us to develop new models for heavy water.

We present three effective pair potentials for liquid heavy water developed on the basis of highly used three- and four-site water models for light water: SPC/E, TIP3P, and TIP4P/2005.[11,12,14] Our strategy was not to parameterize new models from scratch but instead to modify these H_2O models as little as possible, only to the extend needed to reproduce experimental properties of D_2O. This way, we aimed to maintain the specific characteristics of these popular H_2O models, and we aimed to improve the transferability of the new D_2O models to applications beyond pure-water systems, in particular toward biomolecular systems. We denote the new models SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW.

II. METHODS

A. Simulation systems and parameters

MD simulations were carried out with the Gromacs software,[24] versions 2019.6 and 2020.3. Cubic boxes with a box length of 3 nm were created and subsequently filled with water molecules. For SPC/E,[12] SPC/HW,[19] SPC/E-HW, TIP3P,[11] and TIP3P-HW, the box contained 884 water molecules. For TIP4P/2005[14] and TIP4P/2005-HW, the box contained 909 molecules. The energy of each simulation system was minimized within 500 steps with the steepest decent algorithm. Subsequently, the systems were equilibrated for 100 ps. The simulations were carried out at temperatures 276.95, 283.15, 288.15, 293.15, 298.15, 303.15, 308.15, 313.15, 318.15, 323.15, 333.15, 343.15, 353.15, 363.15, and 373.15 K. The temperature was controlled using velocity rescaling ($\tau = 0.1$ ps).[25] The pressure was controlled with the Berendsen barostat ($\tau = 1$ ps)[26] and with the Parrinello–Rahman barostat ($\tau = 5$ ps)[27] during equilibration and production simulations, respectively. Whereas the experimental data used here for validation were partly reported for 1 bar and partly for 1 atm, we simulated with 1 bar throughout this study for simplicity. The geometry of water molecules was constrained with the SETTLE algorithm.[28] Dispersive interactions and short-range repulsion were described by a Lennard-Jones potential with parameters ε and σ as follows:

$$V_{\mathrm{LJ}}(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^{6}\right], \tag{1}$$

which was cut off at a distance of $r = 1$ nm. The pressure and energy were corrected for missing dispersion interactions beyond the cutoff. We found that, owing to the applied dispersion corrections, using a longer LJ cutoff had only a small effect on the computed water properties. Neighbor lists were updated with the Verlet scheme. Coulomb interactions between point charges q_1 and q_2,

$$V_{\mathrm{Coul}}(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r}, \tag{2}$$

were computed with the smooth particle-mesh Ewald (PME) method.[29,30] Here, ε_0 denotes the vacuum permittivity. We used a Fourier spacing of ~0.12 nm, which was optimized by the Gromacs mdrun module at the beginning of each simulation. Systems at 298.15 K, for which the compressibility was computed, were simulated for 150 ns with a 0.5 fs integration time step. All other systems were simulated for 20 ns with a 1 fs time step. Simulations were carried out in single precision on Intel Xeon E-2136 processors, while all nonbonded interactions including PME were offloaded to an Nvidia GTX 1070Ti graphics card.

Statistical errors were computed for simulations at 298.15 K by binning the trajectory into 15 ns time blocks. The physical properties were computed for each block. The values reported below represent the average and standard error over the blocks.

TIP4P/2005-HW was optimized with the ForceBalance software.[31,32] Simulations submitted within the ForceBalance framework were carried out with Gromacs, version 2019.6. MD parameters were chosen as described above, except that simulations were carried out for 20 ns for all temperatures listed in the reference data in Table I. During the optimization steps, ForceBalance used a variant of the Newton–Raphson algorithm with a trust radius between 0.025 and 0.25. An additional penalty, which corresponds to ridge regression, was applied to prevent large steps in parameter space. The maximum number of iterations was set to 10 000.

B. Calculation of physical properties

Following previous work,[33] we approximated the heat of vaporization with

$$\Delta H_{\mathrm{vap}} \approx -E_{\mathrm{pot}} + RT, \tag{3}$$

where E_{pot} is the average potential energy per water molecule, R is the gas constant, and T is the temperature. This approximation is valid at atmospheric pressures. For SPC/E, SPC/E-HW, TIP4P/2005,

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

TABLE I. Experimental parameters of liquid heavy water used for the ForceBalance optimization of the TIP4P/2005-HW model.

Temperature (K)	Pressure (atm)	Density[4] (kg/m^3)	ΔH_{vap}[36] (kJ mol^{-1})	Thermal expansion coefficient[38] (10^{-4} K^{-1})	Isothermal compressibility[37] (10^{-6}atm^{-1})	Isobaric heat capacity[38] (cal/mol/K)
278.15	1.0	1105.5	45.942		51.49	
283.15	1.0	1105.7	45.746	−0.27	49.74	20.252 50
288.15	1.0	1105.6	45.546		48.38	
293.15	1.0	1105.0	45.343	1.21	47.37	20.305 20
298.15	1.0	1104.4	45.138		46.52	
303.15	1.0	1103.4	44.926	2.42	45.88	20.295 60
308.15	1.0	1101.9	44.712		45.37	
313.15	1.0	1100.1	44.495	3.43	45.1	20.257 30
318.15	1.0	1097.9	44.275		44.97	
323.15	1.0	1095.7	44.051	4.28	44.91	20.199 85
328.15	1.0	1093.1	43.823		44.98	
333.15	1.0	1090.5	43.591	5.02	45.16	20.137 60
338.15	1.0	1087.5	43.356		45.51	

and TIP4P/2005-HW, the potential energy averaged over the simulations was corrected by adding the polarization energy[12]

$$\Delta E_{pol} = (\mu - \mu_0)^2/2\alpha, \tag{4}$$

with the dipole moment of the model μ, the dipole moment of an isolated water molecule μ_0 = 1.85 D, and the isotropic scalar polarizability[12,34] α = 1.608 10^{-40} F m. For TIP3P or TIP3P-HW, no correction owing to the polarization energy was applied, following the original parameterization scheme.[11]

The self-diffusion coefficients were computed from the slope of the mean-square displacement of water molecules using a least-squares fit to the interval between 5 and 50 ps. The number of hydrogen bonds (H-bonds) was obtained with the Gromacs module gmx hbond. A cutoff of 30° was used for the hydrogen-donor–acceptor angle and of 0.35 nm for the donor–acceptor distance.

The isothermal compressibility was calculated via[14]

$$\kappa_T = \frac{\langle V^2\rangle - \langle V\rangle^2}{k_B T \langle V\rangle}, \tag{5}$$

where V is the simulation box volume, k_B is the Boltzmann constant, and $\langle\cdot\rangle$ denotes the average over the simulation frames. The compressibilities computed here for light water models agree with the literature.[16,35]

Electron densities were computed from the mass densities using the molar mass.

C. Parameterization strategy for SPC/E-HW and TIP3P-HW

The SPC/E and TIP3P models were adjusted with the aim of matching the physical properties of heavy water molecules and the properties of liquid heavy water as follows: The mass of the deuterium atoms was adjusted to set the D_2O mass to 20.0275 g/mol.[39] The O–H bond length and H–O–H angles were taken from the respective H_2O models. The partial charges of oxygen and deuterium were adjusted to match the experimental ratio between the dipole moments μ of light water and heavy water, which was reported as μ_{D_2O}/μ_{H_2O} = 1.01 both in a benzene solution and in the gas phase.[3,40]

Next, to refine the Lennard-Jones (LJ) parameters, we ran 20 ns MD simulations and systematically varied the σ and ε parameters of the oxygen atom close to the LJ parameters of the respective light water model. Finally, σ and ε were selected to (i) closely match the experimental density, (ii) closely match the change in the H_{vap} value of heavy water relative to light water, (iii) to reasonably match the diffusion coefficient, and (iv) to remain close to the parameters of the light water in order to preserve the characteristics of the light water model.

D. Parameterization strategy for TIP4P/2005-HW

The mass of the deuterium atoms was again chosen to match the D_2O mass of 20.0275 g/mol.[39] The H–O–H angle and O–H distance were taken from TIP4P/2005. All other parameters were optimized with the ForceBalance software,[31,32] with the aim of matching experimental data over a wide temperature range between 287.15 and 338.15 K. We adapted the charge of the dummy atom q_M, the distance between the oxygen and the dummy atom r_{OM}, and the Lennard-Jones parameters σ and ε of the oxygen atom. More restrictive optimization protocols, for instance, with fixed partial charges, did not yield acceptable water densities over a wide temperature range. The reference data used by ForceBalance are listed in Table I. We carried out several ForceBalance runs with slightly different weights for the target data, trust radius, and convergence criteria, which converged to different parameter sets. We selected a parameter set that well reproduced the experimental density and, simultaneously, reasonably well reproduced the heat of vaporization, diffusion coefficient, dipole moment, and radial oxygen–oxygen distance distribution.

III. RESULTS

As a reference, we first recomputed the physical properties of the widely used H_2O models SPC/E, TIP3P, and TIP4P/2005

at room temperature (Table II).[10,11,14] The computed results agree with the literature[12,14,32,44] to the extent expected when using slightly different simulation parameters. The results demonstrate the previously documented strengths and weaknesses of these models. For instance, the diffusion coefficient and compressibility of TIP3P are too large as compared to the experiment (cf. Table II, last column), whereas the density of TIP3P is too low. The properties of SPC/E and TIP4P/2005 reveal better agreement with the experiment, although deviations are still evident. Specifically, the heat of vaporization of TIP4P/2005 is 2 kJ/mol larger than expected from the experiment.

Table III presents the newly derived force field parameters for SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW, as well as, for reference, the parameters of the respective H_2O models and the previously proposed SPC/HW model.[19] Simulation topology files of the D_2O models in Gromacs format are provided in the supplementary material. The properties of D_2O obtained with the new models are listed in Table IV and discussed in the following.

A. Dipole moment

The experimental dipole of a D_2O molecule is ~1% larger as compared to an H_2O molecule.[3,40] For SPC/E-HW and TIP3P-HW, this relative increase was adopted by adjusting the partial charges. For TIP4P/2005-HW, in contrast, the dipole was not optimized but was an outcome of the ForceBalance protocol. Accordingly, the dipole of TIP4P/2005-HW is only 0.5% larger relative to TIP4P/2005, which is a smaller increase than expected from the experiment.[3,40]

B. Mass and electron density

The mass densities of liquid D_2O obtained with SPC/E-HW and TIP4P/2005-HW at 1 bar and 298.15 K agree with the experimental value within less than 0.2% (Table IV). In contrast, the density obtained with TIP3P-HW is 1.2% below the experimental value, in line with the too low density of TIP3P by 1.1% (see Table II).

Figure 1 (left) presents the mass densities of all water models considered in this study over a wide temperature range. Evidently,

TABLE II. Calculated and experimental properties of H_2O at 298.15 K. Calculated properties correspond to 1 bar and experimental properties to 1 atm: mass density, average potential energy during MD simulations, polarization correction, polarization-corrected potential energy, heat of vaporization, diffusion coefficient, isothermal compressibility, and average number of hydrogen bonds per molecule.

		SPC/E	TIP4P/2005	TIP3P	Expt.
Density	kg/m^3	998.810(5)	997.090(5)	985.929(5)	997.0480[a,b]
$-E_{pot}^{MD}$	kJ/mol	46.819(2)	47.829(2)	40.100(2)	
ΔE_{pol}	kJ/mol	5.22125	4.32095		
$-E_{pot}$	kJ/mol	41.597(2)	43.508(2)	40.100(2)	41.5[b,c]
ΔH_{vap}	kJ/mol	44.076(2)	45.987(2)	42.579(2)	43.990[a,d]
D	10^{-5} cm^2/s	2.522(2)	2.104(2)	5.478(4)	2.2999[d,e]
κ_T	10^{-6} bar^{-1}	46.3(2)	46.5(3)	58.3(2)	45.225[f,g]
⟨# H-bonds⟩		3.6025	3.66229	3.35226	3.62 ± 0.1[d,h]

[a]Reference 4.
[b]Reported for 1 atm.
[c]Reference 11.
[d]Pressure not reported.
[e]Reference 41.
[f]Reference 42.
[g]Reported for 1 bar.
[h]Reference 43.

TABLE III. Parameters of SPC/E, SPC/E-HW, SPC/HW,[19] TIP3P, TIP3P-HW, TIP4P/2005, and TIP4P/2005-HW.

	Mass O (g/mol)	Mass H Mass D (g/mol)	q_M (e)	q_O (e)	q_H q_D (e)	σ (nm)	ε (kJ mol^{-1})	∠HOH ∠DOD (°)	r_{OM} (Å)
SPC/E	15.9994	1.008 000		−0.8476	0.4238	0.316 557	0.650 194	109.47	
SPC/E-HW	15.9994	2.014 054		−0.8564	0.4282	0.318 776	0.573 885	109.47	
SPC/HW[19]	15.9994	2.014 054		−0.8700	0.4350	0.316 557	0.650 194	109.47	
TIP3P	15.9994	1.008 000		−0.8340	0.4170	0.315 057	0.636 390	104.52	
TIP3P-HW	15.9994	2.014 054		−0.8424	0.4212	0.317 156	0.565 396	104.52	
TIP4P/2005	16.0000	1.008 000	−1.1128	0.0000	0.5564	0.315 890	0.774 898	104.52	0.154 648 5
TIP4P/2005-HW	16.0000	2.013 754	−1.1220	0.0000	0.5610	0.316 590	0.749 730	104.52	0.156 349 7

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

TABLE IV. Experimental and calculated parameters of liquid D_2O at 298.15 K and 1 bar: mass density, average potential energy during MD simulations, polarization correction, polarization-corrected potential energy, heat of vaporization, diffusion coefficient, compressibility, and average number of hydrogen bonds per molecule.

		SPC/HW[19]	SPC/E-HW	TIP4P/2005-HW	TIP3P-HW	Expt.
Density	kg/m^3	1125.307(5)	1106.169(5)	1103.998(5)	1092.168(6)	1104.4[a,b]
$-E_{pot}^{MD}$	kJ/mol	51.283(2)	48.486(2)	48.660(2)	41.215(2)	
E_{pol}	kJ/mol	6.59701	5.742 64	4.510 86		
$-E_{pot}$	kJ/mol	44.686(2)	42.744(2)	44.149(2)	41.215(2)	
ΔH_{vap}	kJ/mol	47.165(2)	45.223(2)	46.629(2)	43.694(2)	45.138[c,d]
Diffusion coefficient	10^{-5} cm^2/s	1.370(2)	1.691(2)	1.613(2)	4.246(4)	1.87–1.9[b,e,f]
Compressibility	10^{-6} bar^{-1}	41.6(1)	44.2(2)	47.0(2)	57.7(3)	46.5[f,g]
⟨# H-bonds⟩		3.7013	3.656 17	3.686 84	3.403 227	3.76 ± 0.1[d,h]

[a]Reference 4.
[b]Reported for 1 bar.
[c]Reference 36.
[d]Pressure not reported.
[e]References 45–48.
[f]Reported for 1 atm.
[g]Reference 37.
[h]Reference 43.

both TIP3P and TIP3P-HW underestimate the densities at room temperature, and the computed densities decay too rapidly with increasing temperature. In contrast, SPC/E and SPC/E-HW favorably match the experimental densities at room temperature; the temperature dependence of the densities is more realistic as compared to the TIP3P variants yet still enhanced relative to the experiment. TIP4P/2005 and TIP4P/2005-HW favorably match the experimental densities over the whole temperature range between 276.95 and 373.15 K.

The difference of the mass densities of H_2O and D_2O is dominated by the increased mass of deuterium relative to protium. A more sensitive comparison between H_2O and D_2O is given by the electron density since H_2O and D_2O carry the same number of electrons. As shown in Fig. 1 (right), the experimental electron density

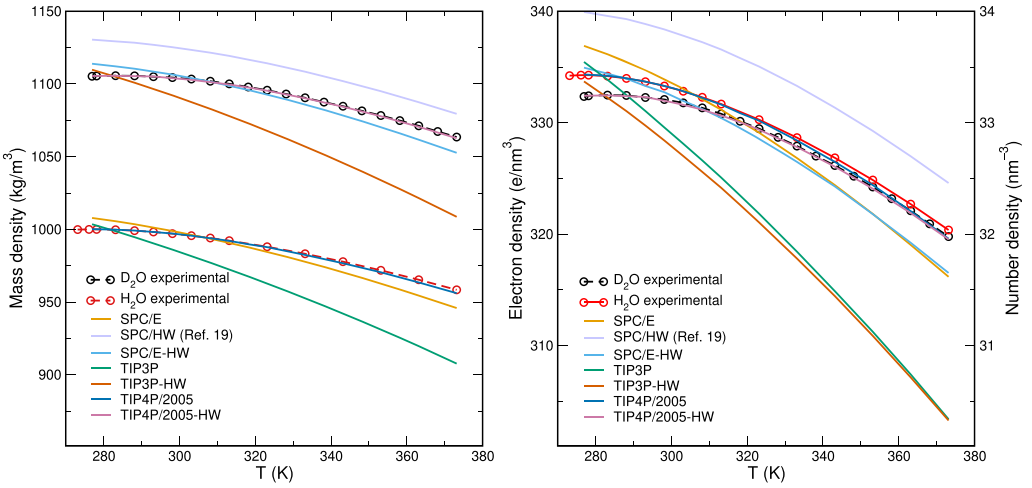


FIG. 1. Mass densities (left) and electron densities (right) of liquid H_2O and D_2O as a function of temperature at 1 bar. The right ordinate in the right panel shows the number density of water molecules. Experimental data[4] are shown as circles and dashed lines and calculated data as solid lines. Experimental data of H_2O and D_2O correspond to pressures of 1 atm and 1 bar, respectively. See the legend for the color code.

is slightly decreased for D_2O relative to H_2O, and this difference is closely reproduced by the TIP4P/2005-HW and TIP4P/2005 models. The SPC/E and TIP3P variants qualitatively reproduce the decreased electron density of heavy water. TIP4P/2005 and TIP4P/2005-HW capture the electron densities of light water and heavy water over the whole temperature range, respectively.

C. Heat of vaporization

The experimental ΔH_{vap} value of heavy water is increased by 2.6% relative to light water (Tables II and IV).[4,36] This increase is, by design of our parameterization strategy, well captured by SPC/E-HW relative to SPC/E (2.60%) and by TIP3P-HW relative to TIP3P (2.61%). For TIP4P/2005-HW relative to TIP4P/2005, ΔH_{vap} is increased by only 1.4% since (i) we focused on reproducing the density and ΔH_{vap} over a wide temperature range rather than the relative increase in the ΔH_{vap} value of heavy water and (ii) the ΔH_{vap} value of TIP4P/2005 is already too large at room temperature (Tables II and IV).[14] Hence, for future studies, using SPC/E-HW or TIP3P-HW may be more suitable than using TIP4P/2005-HW to study differences of thermodynamic properties between light water and heavy water.

Critically, ΔH_{vap} reported in Tables II and IV includes the polarization corrections for the SPC/E and TIP4P/2005 variants but not for the TIP3P variants, following the original force field derivations. Hence, the ΔH_{vap} value of both TIP3P and TIP3P-HW is in fact too low compared to experimental conditions.

D. Self-diffusion

The self-diffusion coefficient of liquid D_2O at 1 atm and 298.15 K was reported as 1.87×10^{-5} cm^2/s [47] or 1.90×10^{-5} cm^2/s.[46] In addition, Wilbur *et al.* reported the values of 1.22×10^{-5} cm^2/s at 283.15 K and 2.00×10^{-5} cm^2/s at 303.15 K, which is compatible with a value of $\sim 1.8 \times 10^{-5}$ cm^2/s at 298.15 K assuming an approximately linear temperature dependence over this range. The self-diffusion coefficients obtained with our D_2O models are listed in Table IV. All heavy water models qualitatively reproduce the experimentally observed reduced self-diffusion coefficient of heavy water relative to light water. However, because it was difficult to match both ΔH_{vap} and density on the one hand and the self-diffusion coefficient on the other hand, we accepted larger discrepancies for the self-diffusion coefficient. Specifically, SPC/E-HW and TIP4P/2005-HW yield slightly too low diffusion coefficients compared to the experiment. In contrast, TIP3P-HW reveals a greatly increased diffusion coefficient, in line with the increased diffusion by the TIP3P model.

E. Isothermal compressibility

The isothermal compressibility κ_T of liquid D_2O is 46.5×10^{-5} atm^{-1} at 298.15 K and 1 atm, slightly increased relative to H_2O.[37] The calculated κ_T values of SPC/E-HW and TIP4P/2005-HW reasonably agree with the experimental value, in line with the respective H_2O models. In contrast, the compressibilities of both TIP3P

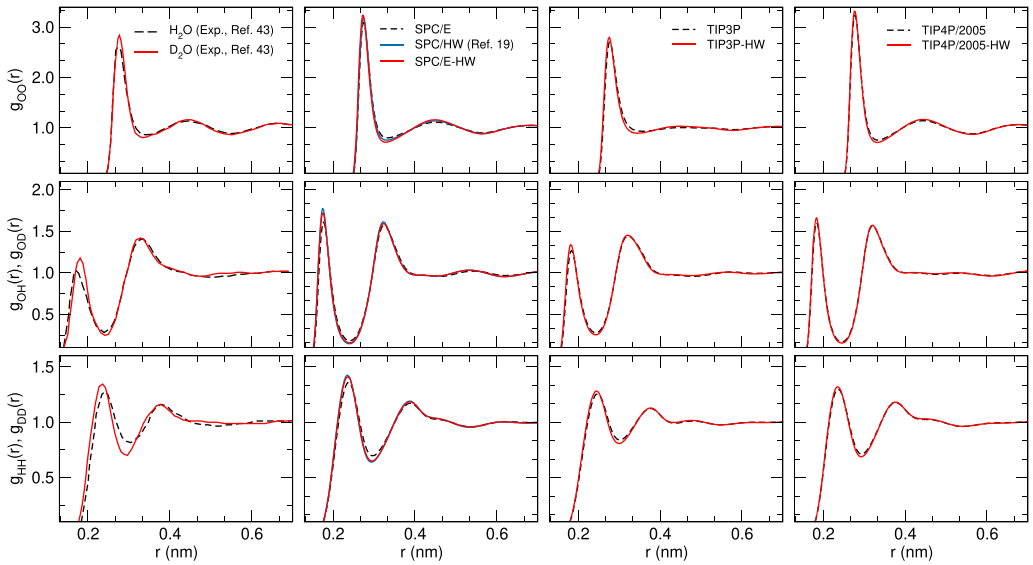


FIG. 2. O–O (first row), O–H (second row), and H–H (third row) radial distribution functions (RDFs) for SPC/E, SPC/HW,[19] SPC/E-HW, TIP3P, TIP3P-HW, TIP4P/2005, and TIP4P/2005-HW models at 1 bar and 25 °C. For reference, RDFs refined against experimental data (left column) were digitalized from Ref. 43.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

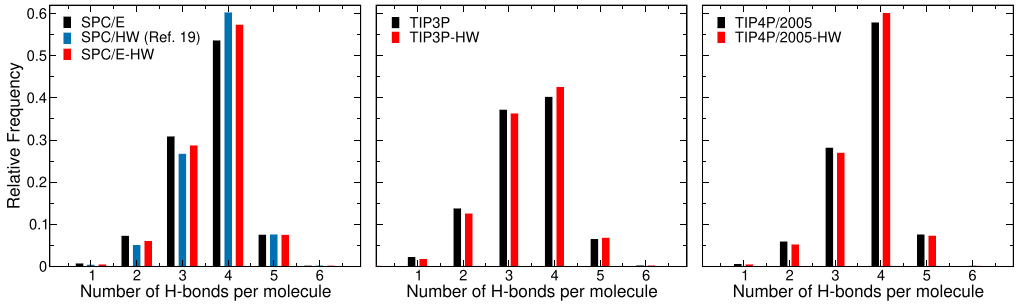


FIG. 3. Probability histograms for the number of hydrogen bonds per molecule in SPC/E, SPC/HW,[19] SPC/E-HW, TIP3P, TIP3P-HW, TIP4P/2005, and TIP4P/2005-HW at 1 bar and 25 °C.

and TIP3P-HW are strongly increased by ~25% relative to the experiment.

F. Water structure: Radial distribution function and number of hydrogen bonds per molecule

Using a combination of X-ray and neutron diffraction, Soper and Benmore showed that liquid D_2O is more structured than liquid H_2O, as quantified by more pronounced maxima and minima in atomic radial distribution functions (RDFs).[43] To probe the structure of our liquid D_2O models, we computed RDFs between oxygen atoms [$g_{OO}(r)$], oxygen and hydrogen atoms [$g_{OH}(r)$, $g_{OD}(r)$], and hydrogen/deuterium atoms [$g_{HH}(r)$, $g_{DD}(r)$], as presented in Fig. 2. Notably, the RDFs of D_2O models (Fig. 2, red lines) yield more pronounced maxima and minima relative to the respective H_2O model (Fig. 2, black dashed line). Hence, the liquid D_2O models exhibit the increased structure, in qualitative agreement with the experimental findings.

As a second measure for the degree of water structure, we computed the average number of hydrogen bonds (H-bonds) per water molecule. As expected from the increased water structure according to the RDFs, we find that all D_2O models yield an increased average number of H-bonds (Tables II and IV). These findings agree qualitatively with the diffraction data by Soper and Benmore (Fig. 2, left column).[43] Histograms over the average number of H-bonds per water molecule show that this shift is realized by an increased number of water molecules with four H-bonds at the cost of water molecules with only two or three H-bonds, consistently among the three D_2O models (Fig. 3).

IV. DISCUSSION

We presented three models for liquid heavy water, SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW. We parameterized SPC/E-HW and TIP3P-HW with the aim of (i) reproducing the relative differences between light water and heavy water, while (ii) changing the original water models as little as possible, thereby maintaining the characteristics of the original light water model. Consequently, SPC/E-HW and TIP3P-HW inherit the strengths and weaknesses of the SPC/E and TIP3P models. Specifically, both TIP3P and TIP3P-HW neglect the polarization contribution to the heat of vaporization, and they exhibit too low mass densities, too large diffusion coefficients, and too large isothermal compressibilities. In addition, liquid water modeled with TIP3P and TIP3P-HW lacks the internal water structure as revealed by the absence of a second solvation shell in the O–O RDFs and by the reduced number of H-bonds. Nevertheless, since the TIP3P model is widely used in biomolecular simulations, we expect TIP3P-HW to be useful for comparative simulation studies. SPC/E and SPC/E-HW exhibit an overall favorable agreement with experimental data.

TIP4P/2005-HW was parameterized with the aim of reproducing D_2O properties over a wide temperature range. Specifically, we aimed to reproduce the temperature-dependent density since the density plays a critical role in neutron scattering contrast variation experiments. The favorable agreement with experimental data at various temperatures limited the possibilities to reproduce the relative differences between D_2O and H_2O at room temperature. In consequence, the increase in the heat of vaporization of liquid D_2O relative to H_2O is not as precisely captured by TIP4P/2005-HW as compared to SPC/E-HW or TIP3P-HW.

This study was motivated by our inability to reproduce the reported properties of the SPC/HW model[19] with the Gromacs simulation software, irrespective of the Gromacs version and the choice of various simulation parameters. For instance, compared to Ref. 19, we obtained with the SPC/HW model an increased density (Fig. 1), more negative potential energies, and thereby a larger heat of vaporization, as well as smaller diffusion coefficients (Table IV). Therefore, we do not recommend SPC/HW[19] for simulations with Gromacs.

The most appropriate choice for a heavy water model in future studies will depend on the application. For studies that focus on room temperature and involve fine balances of thermodynamic properties such as solvation free energies, we anticipate that comparative simulations with SPC/E and SPC/E-HW may be most insightful because SPC/E and SPC/E-HW (i) favorably agree with a wide range of experimental data at room temperatures and (ii) accurately capture the increased ΔH_{vap} value of liquid D_2O relative to H_2O. For biomolecular simulations, simulations with TIP3P and TIP3P-HW

will provide a useful alternative because several biomolecular force fields were parameterized in conjunction with TIP3P. For studies involving D_2O and H_2O in wider temperature ranges, TIP4P/2005 and TIP4P/2005-HW are recommended. Taken together, the D_2O models presented here will be useful for gaining atomic and energetic insight into phenomena and experiments involving heavy water.

SUPPLEMENTARY MATERIAL

See the supplementary material for topologies, simulation systems, and simulation parameters in Gromacs format.

ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft (Grant No. HU 1971/3-1).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]G. Czakó, E. Mátyus, and A. G. Császár, "Bridging theory with experiment: A benchmark study of thermally averaged structural and effective spectroscopic parameters of the water molecule," J. Phys. Chem. A **113**, 11665–11678 (2009).
[2]S. Herrig, M. Thol, A. H. Harvey, and E. W. Lemmon, "A reference equation of state for heavy water," J. Phys. Chem. Ref. Data **47**, 043102 (2018).
[3]G. Némethy and H. A. Scheraga, "Structure of water and hydrophobic bonding in proteins. IV. The thermodynamic properties of liquid deuterium oxide," J. Chem. Phys. **41**, 680–689 (1964).
[4]*CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data. 2003–2004*, 84th ed., edited by D. R. Lide (CRC, Boca Raton, FL; London, 2003).
[5]M. M. Castellanos, A. McAuley, and J. E. Curtis, "Investigating structure and dynamics of proteins in amorphous phases using neutron scattering," Comput. Struct. Biotechnol. J. **15**, 117–130 (2017).
[6]O. Dunne, M. Weidenhaupt, P. Callow, A. Martel, M. Moulin, S. J. Perkins, M. Haertlein, and V. T. Forsyth, "Matchout deuterium labelling of proteins for small-angle neutron scattering studies using prokaryotic and eukaryotic expression systems and high cell-density cultures," Eur. Biophys. J. **46**, 425–432 (2017).
[7]F. Gabel, "Chapter thirteen—Small-angle neutron scattering for structural biology of protein–RNA complexes," Methods Enzymol. **558**, 391–415 (2015).
[8]K. E. Haslauer, D. Hemmler, P. Schmitt-Kopplin, and S. S. Heinzmann, "Guidelines for the use of deuterium oxide (D_2O) in ^{1}H NMR metabolomics," Anal. Chem. **91**, 11063–11069 (2019).
[9]G. Zuber, S. J. Prestrelski, and K. Benedek, "Application of Fourier transform infrared spectroscopy to studies of aqueous protein solutions," Anal. Biochem. **207**, 150–156 (1992).
[10]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Interaction Models for Water in Relation to Protein Hydration* (Springer, Dordrecht, 1981), pp. 331–342.
[11]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).
[12]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," J. Phys. Chem. **91**, 6269–6271 (1987).
[13]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. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," J. Phys. Chem. B **102**, 3586–3616 (1998).
[14]J. L. F. Abascal and C. Vega, "A general purpose model for the condensed phases of water: Tip4P/2005," J. Chem. Phys. **123**, 234505 (2005).
[15]G. Lamoureux, A. D. MacKerell, and B. Roux, "A simple polarizable model of water based on classical Drude oscillators," J. Chem. Phys. **119**, 5185–5197 (2003).
[16]Y. Qiu, P. S. Nerenberg, T. Head-Gordon, and L.-P. Wang, "Systematic optimization of water models using liquid/vapor surface tension data," J. Phys. Chem. B **123**, 7061–7073 (2019).
[17]C. P. Lawrence and J. L. Skinner, "Flexible TIP4P model for molecular dynamics simulation of liquid water," Chem. Phys. Lett. **372**, 842–847 (2003).
[18]A. B. Mamonov, R. D. Coalson, M. L. Zeidel, and J. C. Mathai, "Water and deuterium oxide permeability through aquaporin 1: MD predictions and experimental verification," J. Gen. Physiol. **130**, 111–116 (2007).
[19]J. R. Grigera, "An effective pair potential for heavy water," J. Chem. Phys. **114**, 8064–8067 (2001).
[20]T. Róg, K. Murzyn, J. Milhaud, M. Karttunen, and M. Pasenkiewicz-Gierula, "Water isotope effect on the phosphatidylcholine bilayer properties: A molecular dynamics simulation study," J. Phys. Chem. B **113**, 2378–2387 (2009).
[21]M. Pasenkiewicz-Gierula, K. Baczynski, M. Markiewicz, and K. Murzyn, "Computer modelling studies of the bilayer/water interface," Biochim. Biophys. Acta, Biomembr. **1858**, 2305–2321 (2016).
[22]G. Prampolini, P. Yu, S. Pizzanelli, I. Cacelli, F. Yang, J. Zhao, and J. Wang, "Structure and dynamics of ferrocyanide and ferricyanide anions in water and heavy water: An insight by MD simulations and 2D IR spectroscopy," J. Phys. Chem. B **118**, 14899–14912 (2014).
[23]H.-W. Wang, D. J. Wesolowski, T. E. Proffen, L. Vlcek, W. Wang, L. F. Allard, A. I. Kolesnikov, M. Feygenson, L. M. Anovitz, and R. L. Paul, "Structure and stability of SnO_2 nanocrystals and surface-bound water species," J. Am. Chem. Soc. **135**, 6885–6895 (2013).
[24]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1-2**, 19–25 (2015).
[25]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**, 014101 (2007).
[26]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," J. Chem. Phys. **81**, 3684–3690 (1984).
[27]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182–7190 (1981).
[28]S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," J. Comput. Chem. **13**, 952–962 (1992).
[29]T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems," J. Chem. Phys. **98**, 10089–10092 (1993).
[30]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," J. Chem. Phys. **103**, 8577–8593 (1995).
[31]L.-P. Wang, T. J. Martinez, and V. S. Pande, "Building force fields: An automatic, systematic, and reproducible approach," J. Phys. Chem. Lett. **5**, 1885–1891 (2014).
[32]L.-P. Wang, J. Chen, and T. Van Voorhis, "Systematic parametrization of polarizable force fields from quantum chemistry data," J. Chem. Theory Comput. **9**, 452–460 (2013).
[33]W. L. Jorgensen and J. D. Madura, "Temperature and size dependence for Monte Carlo simulations of TIP4P water," Mol. Phys. **56**, 1381–1392 (1985).
[34]V. P. Frontas'ev and L. S. Shraiber, "Study of the variation in the electron polarizability of molecules of ordinary and heavy water under the influence of temperature," J. Struct. Chem. **6**, 493–500 (1965).
[35]H. L. Pi, J. L. Aragones, C. Vega, E. G. Noya, J. L. F. Abascal, M. A. Gonzalez, and C. McBride, "Anomalies in water as obtained from computer simulations of the TIP4P/2005 model: Density maxima, and density, isothermal compressibility and heat capacity minima," Mol. Phys. **107**, 365–374 (2009).
[36]A. Crabtree and M. Siman-Tov, "Thermophysical properties of saturated light and heavy water for advanced neutron source applications," OSTI.GOV Technical Report ORNL/TM-12322 (1993), https://doi.org/10.2172/6306919.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

[37]F. J. Millero and F. K. Lepple, "Isothermal compressibility of deuterium oxide at various temperatures," J. Chem. Phys. **54**, 946–949 (1971).
[38]G. Ulrych, *Properties of Liquid Heavy Water* (Begell House, Inc., New York, 2016).
[39]International Association for the Properties of Water and Steam, IAPWS R16-17, Revised Release on the IAPWS Formulation 2017 for the Thermodynamic Properties of Heavy Water, 2018.
[40]A. Maryott and F. Buckley, *Table of Dielectric Constants and Electric Dipole Moments of Substances in the Gaseous State*, Circular of the National Bureau of Standards (U. S. Department of Commerce, National Bureau of Standards, 1953).
[41]M. Holz, S. R. Heil, and A. Sacco, "Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements," Phys. Chem. Chem. Phys. **2**, 4740–4742 (2000).
[42]G. S. Kell, E. Whalley, and W. G. Schneider, "The PVT properties of water," Philos. Trans. R. Soc., A **258**, 565–614 (1965).
[43]A. K. Soper and C. J. Benmore, "Quantum differences between heavy and light water," Phys. Rev. Lett. **101**, 065502 (2008).
[44]B. Hess and N. F. A. van der Vegt, "Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force field and water models," J. Phys. Chem. B **110**, 17616–17626 (2006).
[45]D. J. Wilbur, T. DeFries, and J. Jonas, "Self-diffusion in compressed liquid heavy water," J. Chem. Phys. **65**, 1783–1786 (1976).
[46]L. G. Longsworth, "The mutual diffusion of light and heavy water," J. Phys. Chem. **64**, 1914–1917 (1960).
[47]R. Mills, "Self-diffusion in normal and heavy water in the range 1–45.deg," J. Phys. Chem. **77**, 685–688 (1973).
[48]J. S. Murday and R. M. Cotts, "Self-diffusion in liquids: H_2O, D_2O, and Na," J. Chem. Phys. **53**, 4724–4725 (1970).

Chapter 6

Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data

This chapter was originally published in Communications Chemistry, Volume 6, Article number: 272 (2023) (Ref. 236). We thank the journal for permission to reprint it.

The authors of this paper are Johanna-Barbara Linse and Jochen S. Hub, with the following contributions:

J.-B.L. performed and analyzed MD simulations.
J.-B.L. and J.S.H. designed the project and wrote the article.

communications chemistry

ARTICLE

https://doi.org/10.1038/s42004-023-01067-1 OPEN

Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data

Johanna-Barbara Linse [1] & Jochen S. Hub [1]✉

Biological macromolecules in solution are surrounded by a hydration shell, whose structure differs from the structure of bulk solvent. While the importance of the hydration shell for numerous biological functions is widely acknowledged, it remains unknown how the hydration shell is regulated by macromolecular shape and surface composition, mainly because a quantitative probe of the hydration shell structure has been missing. We show that small-angle scattering in solution using X-rays (SAXS) or neutrons (SANS) provide a protein-specific probe of the protein hydration shell that enables quantitative comparison with molecular simulations. Using explicit-solvent SAXS/SANS predictions, we derived the effect of the hydration shell on the radii of gyration R_g of five proteins using 18 combinations of protein force field and water model. By comparing computed R_g values from SAXS relative to SANS in D_2O with consensus SAXS/SANS data from a recent worldwide community effort, we found that several but not all force fields yield a hydration shell contrast in remarkable agreement with experiments. The hydration shell contrast captured by R_g values depends strongly on protein charge and geometric shape, thus providing a protein-specific footprint of protein–water interactions and a novel observable for scrutinizing atomistic hydration shell models against experimental data.

[1] Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken 66123, Germany. ✉email: jochen.hub@uni-saarland.de

ARTICLE COMMUNICATIONS CHEMISTRY | https://doi.org/10.1038/s42004-023-01067-1

Water molecules play key roles in protein functions such as folding, molecular recognition, enzymatic activity, and proton transfer[1–3]. During such functions, water interacts with the geometrically rough and chemically heterogeneous protein surface by the formation of hydrogen bonds with polar and ionic groups as well as by long-ranged Coulomb and Van-der-Waals forces. Protein–water interactions together with water-internal interactions lead to the formation of a water layer with different structural and dynamic properties as compared to bulk water, termed protein hydration shell. The modified water dynamics in the hydration shell have been studied by NMR and Terahertz spectroscopy, time-dependent fluorescence Stokes shift, inelastic neutron scattering, molecular dynamics (MD) simulations, and several other techniques[4–13]. These data revealed that geometric constraints and the hydrogen bond network lead to a mild slowdown of water dynamics by factors of 3–5 and to an increased water ordering as compared to bulk water[14]. The importance of protein–water interactions is further augmented in crowded cellular environments, where macromolecules typically adopt 25–40% of the volume[15]. In such environments, up to 70% of the water is part of a biomolecular hydration shell[3], demonstrating that biology largely involves non-bulk-like water[16].

Whereas the dynamics of the protein hydration shell has been investigated in great quantitative detail by spectroscopy, the overall structure and contrast of the hydration shell is far less understood. Small-angle scattering (SAS) with X-rays (SAXS) or neutrons (SANS) revealed that the hydration shell of many proteins exhibit an increased density compared to the bulk[17], which imposes an increased or a decreased radius of gyration R_g as detected by SAXS or by SANS in D_2O, respectively. MD simulations with explicit solvent reproduced the modified R_g values and attributed this effect to an excess density of ~6%[18,19]. A recent combined SAXS/SANS study on ultra-charged proteins suggested that anionic surface amino acids (Asp/Glu) enhance the hydration shell density more as compared to cationic amino acids (Arg/Lys)[20]. Because data from SAS or spectroscopy provide information on the hydration shell with only low spatial resolution or low information content, atomic insight into protein hydration mostly relies on MD simulations[16,18,21–23]. However, whether current protein force fields and water models accurately capture the structure of the protein hydration shell is not known.

Apart from the pioneering study by Svergun et al.[17], SAS data has been hardly used as a probe for protein hydration, mostly because undesired scattering contributions from protein aggregation, radiation damage, or poor buffer matching are hard to quantify with sufficient accuracy from a single SAS experiment, despite progress in quality standards for SAS experiments and analysis[24,25]. Thus, to obtain high-precision SAS data for benchmarking and to test the reproducibility of SAS data, a recent worldwide round-robin study collected SAS data for five globular proteins at 12 SAXS and four SANS instruments, leading to a total of 247 SAS curves[26]. The five proteins considered in the round-robin study were selected to be relatively rigid to avoid complications owing to pronounced flexibility. Whereas SAS data for lysozyme and urate oxidase were subject to increased uncertainties (Supplementary Discussion), SAS data for ribonuclease A (RNaseA), xylanase, and glucose isomerase were reproducible at many beamlines, thus providing consensus SAS data with unprecedented accuracy.

Here, we tested whether all-atom MD simulations reproduce the increased density of the protein hydration shell. To this end, we carried out simulations using 18 different combinations of protein force field and water model (Table S1), computed SAS curves taking explicit solvent into account[18,27–30], and compared the derived R_g values to consensus SAS data from the round-robin study (Table S2)[26]. We included widely used variants of the CHARMM[31] and AMBER force fields families[32–35] and ten different water models, including several recently proposed force field combinations with increased protein–water dispersion interactions[35–38]. Overall, we find nearly quantitative agreement between simulation and experiments for many force field combinations; however, the calculations furthermore reveal deviations relative to experiments for certain protein force fields or water models. Thus, the quantitative comparison involving high-precision SAS data and explicit-solvent SAS calculations provides a novel route for scrutinizing the structure of the protein hydration shell.

Results

Explicit-solvent SAS calculations reveal the hydration shell effect on R_g. The three-dimensional solvent density around xylanase is illustrated in Fig. 1b and in Supplementary Movie 1, computed from a simulation carried out with the ff99SBws protein force field and the TIP4P/2005s water model (Fig. S1)[36,39]. The density reveals the first hydration layer, which is structured by the formation of favorable interactions between the solvent molecules and the protein surface (red/orange mesh), as well as the second hydration layer (blue mesh), which is more dispersed. By averaging the solvent density over the protein surface, the solvent density is obtained as a function of the distance from the Van-der-Waals surface of xylanase, revealing, in addition to the pronounced first and second hydration layer a shallow third layer at a distance of ~7 Å (Fig. 1c, solid magenta line), as reported by many previous MD studies (Ref. [19] and references therein). To test whether solvent density modulations owing to water–protein interactions differ from density modulation owing to the internal structure of bulk water, we carried out a simulation of bulk TIP4P/2005s water, in which water molecules within a volume similar to the volume of xylanase were restrained (Supplementary Methods). Density modulations around such restrained bulk water are by far smaller as compared to the density modulations at the protein surface (Fig. 1c, compare dashed dark green with solid magenta line), in line with previous reports[18]. Thus, explicit-solvent MD simulations yield the structure of the hydration shell that differs from the structure of bulk solvent and, thereby, manifests as a modified radius of gyration R_g detected by SAS experiments[17].

Using explicit-solvent SAS calculations[29,40], we computed from MD simulations SAXS curves, SANS curves in H_2O, and SANS curves in D_2O as function of momentum transfer q, where $q = 4\pi \sin(\theta)/\lambda$ with the scattering angle 2θ and the wavelength λ of the X-ray beam (Fig. 1d). Two approaches may be used to extract R_g from the SAS intensity curves $I(q)$: (i) via the the Guinier fit $\ln[I(q)/I_0] \approx -(qR_g)^2/3$ to the small-q region, where I_0 is the forward scattering intensity (Fig. 1d, inset); or (ii) via the the pair distance distribution function (PDDF), also referred to as $P(r)$ function, which is obtained from the SAS curve via a regularized inverse Fourier transform[41,42], providing the radius of gyration via $R_g^2 = \int r^2 P(r)\,dr / \left[2\int P(r)\,dr\right]$. From the simulations, we report R_g obtained with the Guinier fit, yet we validated the agreement with the R_g obtained from the PDDF. Both, R_g and I_0 are influenced by the contrast of the hydration shell relative to the bulk solvent. However, because the experimental uncertainties of R_g are by far smaller as compared to uncertainties of I_0, we validated MD simulations against experimental R_g values in this study.

Because our SAXS and SANS calculations take explicit water molecules in the hydration shell into account, the R_g and also I_0 values are fully controlled by the water and protein force fields (together with MD parameters such as cutoffs). We quantified the effect of the hydration shell on the R_g by computing the difference

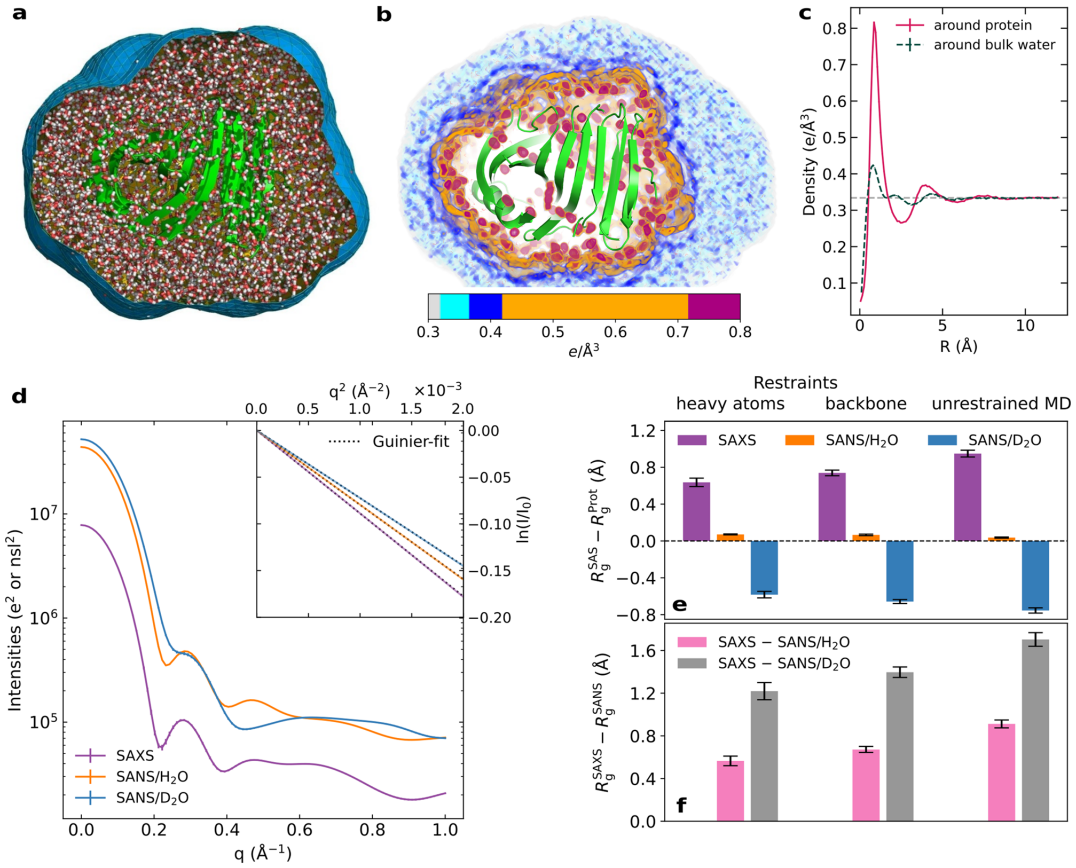


Fig. 1 Explicit-solvent MD reveals the hydration shell structure and modified R_g values from SAXS and SANS. a Simulation of xylanase obtained with ff99SBws and TIP4P/2005s water. Water molecules within the envelope (blue surface) contribute to SAS calculations. Water outside of the envelope is not shown for clarity. The protein is shown in green cartoon, water as red/white sticks. **b** Electron density of solvent inside the envelope in shades from light gray (bulk water) to blue to orange to red, revealing the first (orange and red) and the second (mostly blue) hydration layers. **c** Solvent density versus distance R from the Van-der-Waals surface of the protein, averaged over the protein surface (magenta solid line), revealing two pronounced and a third weak hydration shell. The solvent density around a volume of restrained bulk water (dark green dashed line) reveals by far smaller modulations, demonstrating that water–protein interactions lead to a more structured and more dense hydration shell compared to bulk water. The experimental bulk density of 0.334 e/Å^3 is shown by a gray dashed line. **d** Calculated intensity curves for SAXS (purple), SANS in H_2O (orange), and SANS in D_2O (blue) obtained from MD simulations. Curves are shown in absolute units of e^2 for SAXS and squared neutron scattering lengths (nsl^2) for SANS. Inset: Guinier plots of SAS curves (colored lines) and linear fits (dotted black lines) used to obtain the SAS-derived radii of gyration R_g. **e** Difference between SAS-derived R_g values and the R_g values of the pure protein (R_g^{Prot}) for SAXS, SANS/H_2O, and SANS/D_2O (color code as in panel d). R_g differences were computed from simulations with restrained heavy atoms (left), restrained backbone (middle), or from unrestrained MD (right). **f** Differences between R_g from SAXS and SANS/H_2O (pink), as well as from SAXS and SANS/D_2O (gray). All R_g differences are a footprint of the protein hydration shell. Statistical errors denote 1 SE.

$\Delta R_g = R_g^{SAS} - R_g^{Prot}$ between the R_g from the SAS curve, R_g^{SAS}, and R_g calculated from the atomic positions of protein atoms, R_g^{Prot}. The ΔR_g values calculated from simulations of xylanase with restraints on heavy atoms or on backbone atoms or from unrestrained MD simulations are shown in Fig. 1e, demonstrating that the hydration shell modulates R_g of xylanase by up to 0.9 Å.

Differences in R_g values obtained from SAXS relative to SANS/D_2O enables quantitative comparison between MD simulations and SAS experiments. The effect of the hydration shell on R_g is different in SAXS as compared to SANS experiments (Fig. S3). Because X-rays scatter at the electrons whereas neutrons scatter at the nuclei, SAXS curves report on the electron density contrast, whereas SANS curves report on the contrast of the neutron scattering length density. Many globular proteins exhibit a hydration shell with an increased electron density relative to the bulk solvent[17,18,29]. For such proteins, both, the protein and the hydration shell exhibit a positive electron density contrast relative to the bulk (Fig. S3a), leading in a SAXS experiment to an increased R_g ($\Delta R_g > 0$, Fig. 1e, purple bars). For SANS in D_2O, the protein exhibits a negative contrast of the neutron scattering length density whereas the hydration shell exhibits a positive contrast relative to bulk, resulting typically in a decreased R_g ($\Delta R_g < 0$, Fig. S3b, Fig. 1e, blue bars). For SANS in H_2O, the contrast of the protein is positive whereas the contrast of the

hydration shell is close to zero, leading to a small influence by the hydration shell on R_g ($\Delta R_g \approx 0$, Fig. S3c, Fig. 1e, orange bars).

Because the solution structure of proteins may differ from their crystal structure, R_g^{Prot} present in SAS experiments is not accurately known. Hence, we focused here on the difference $\Delta R_g^{SAS} = R_g^{SAXS} - R_g^{SANS}$ between the R_g values from SAXS (R_g^{SAXS}) and SANS (R_g^{SANS}). For the xylanase simulations discussed above, ΔR_g^{SAS} takes values up to 1.7 Å or 0.9 Å for SANS in D_2O or H_2O, respectively (Fig. 1f, gray or pink bars, respectively). In contrast to ΔR_g values, ΔR_g^{SAS} values do not depend on R_g^{Prot}, thus enabling direct comparison between simulation and experiment.

The hydration shell effect on R_g increases upon solvent relaxation on a flexible protein surface. We first evaluated the effect of atomic fluctuations on the hydration shell and, thereby, on ΔR_g and ΔR_g^{SAS}. To this end, a series of simulations of xylanase was carried out with increasing flexibility by using restraints on heavy atoms, on backbone atoms, or by running an unrestrained MD simulation. Figure 1e, f demonstrates that the fewer atoms of a protein were restrained, the larger is the influence of the hydration shell on ΔR_g, indicating a hydration shell with increasing contrast. To shed more light on the effect of protein flexibility on ΔR_g^{SAS}, we carried out additional simulations of restrained xylanase after unrestrained equilibration. Notably, after such unrestrained equilibration, applying restraints to the protein had on average no effect on ΔR_g^{SAS} (Fig. S4b, c). Thus, protein flexibility takes effect on ΔR_g and ΔR_g^{SAS} during the equilibration phase, enabling more favorable relaxation of water onto the protein surface as compared to simulations with restraints. With increasing conformational flexibility, the water molecules may favorably pack on the protein surface and penetrate cavities between side chains, which leads to an increasingly dense hydration shell. Below, we use results from unrestrained MD simulations for force field validations against experimental data.

Comparison of the hydration shell from 18 force field combinations with consensus SAS data. Next, we studied the effect of 18 different combinations of force fields for protein and water on the hydration shell, as quantified by ΔR_g and ΔR_g^{SAS} values. We considered widely used force field combinations such as CHARMM36m–TIP3P[31,43] as well as uncommon combinations such as CHARMM36m–SPC/E[31,44] (Table S1). Such uncommon combinations are generally not recommended because protein force fields have been parametrized with respect to specific water models; in this study, however, we considered such uncommon combinations with the aim to dissect effects of the protein force field and of the water model on the hydration shell. Figure 2a–c presents ΔR_g values from unrestrained MD simulations, i.e., the difference of R_g values from the Guinier analysis relative to R_g^{Prot} values. Absolute R_g values are shown in Fig. S5 and all computed R_g values are listed in Tables S3–S7. Among all force fields, ΔR_g values were positive, near-zero, and negative for SAXS, SANS/H_2O, and SANS/D_2O, respectively, in line with results in Fig. 1e. However, ΔR_g values vary considerably among different force fields by up to 0.5 Å, indicating different contrasts in the hydration shell. For instance, ff99SBws–TIP4P/2005s, which implements increased water–protein dispersion with the aim to obtain realistic ensembles of intrinsically disordered proteins[36,39], yields by far larger modulations of R_g (black bars) as compared to CHARMM36m with the CHARMM-modified TIP3P water model (cTIP3P, blue bars) or with the OPC water model[31,43,45]. These differences propagate into differences of ΔR_g^{SAS}, which we compare with consensus experimental SAS data in the following.

Figure 3b, d, f presents ΔR_g^{SAS} values obtained from SAXS relative to SANS/D_2O for the proteins RNaseA, xylanase, and glucose isomerase, computed with 18 different combinations of force fields for protein and water and using unrestrained simulations. For many force fields, excellent agreement is found between simulation and experiment (Fig. 3, horizontal lines), suggesting that many force fields yield a correct overall hydration shell contrast. Such agreement is remarkable considering that protein–water interaction potentials have not been refined against solution scattering data but rather against thermodynamic data such as hydration free energies[46]. However, Fig. 3b, d, f furthermore reveals considerable differences among protein force fields and water models. These ΔR_g^{SAS} differences reveal similar trends if derived from backbone-restrained instead of from unrestrained simulations (Fig. S6), demonstrating that ΔR_g^{SAS} variations among force fields are mostly not a consequence of different protein conformations but instead a consequence of different packing of water on the protein surface.

To test whether ΔR_g^{SAS} variations among force fields are explained by variations of the hydration shell densities, we computed solvent density profiles as function of distance R from the Van-der-Waals surface of xylanase using the force field combinations CHARMM36m–cTIP3P, ff15fb–TIP3P-FB, or ff99SBws–TIP4P/2005s (Fig. S7), which led to small, medium, or large ΔR_g^{SAS} values, respectively (Fig. 3d, blue, green, and black bar, respectively). In addition, density profiles around xylanase were compared with density profiles around volumes of restrained bulk water modeled with cTIP3P, TIP3P-FB, or TIP4P/2005s (Fig. S8), thereby comparing the hydration shell structures with the structure of bulk water[18]. We find that the height and width of the first solvent density peak at $R \approx 0.85$ Å correlate with ΔR_g^{SAS} values, both if taken from the total solvent density profiles (Fig. S7, S8) or if taken from the density profile relative to the bulk water structure (Fig. S8b, d). This analysis suggests that modulations of the hydration shell structure among different force fields indeed manifest in variations of ΔR_g^{SAS}.

Force field effects on the hydration shell are different for anionic compared to near-neutral proteins. Focusing first on the weakly charged proteins RNaseA and xylanase, ΔR_g^{SAS} values obtained with CHARMM36m are systematically lower relative to the experiment, irrespective of the applied water model SPC/E, TIP3P, or OPC3 (Fig. 3b, d). An exception is given by the RNaseA simulation with OPC, for which CHARMM36m led to a larger ΔR_g^{SAS} as compared to ff99SBws. However, this difference is inverted in simulations with backbone restraints (Fig. S6a, b), suggesting that the simulation with CHARMM36m–OPC adopted an unusual conformation, as confirmed by visual inspection of the trajectory (Fig. S9). Thus, in conformationally stable simulations, CHARMM36m imposed a less dense hydration shell as compared to the tested AMBER force field variants. Among all tested force fields, CHARMM36m–cTIP3P yields the lowest ΔR_g^{SAS} values, indicating a hydration shell with too low contrast. The AMBER force field variants ff14SB and ff99SB-ildn show good agreement with the experiment with any of the water models SPC/E, TIP3P, OPC3, TIP4P-D, or TIP4P/2005. Hence, surprisingly, among the tested force field combinations, the protein force field has a larger effect on the hydration shell contrast as compared to the water model.

Several older force fields are not suitable for simulating intrinsically disordered proteins (IDPs) as they impose overly

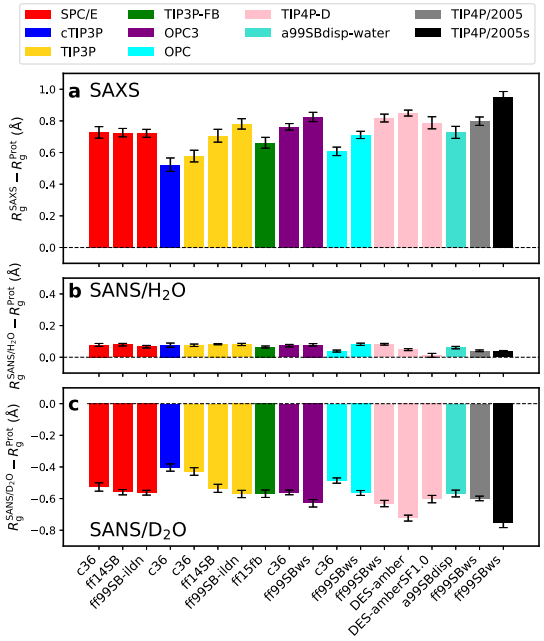


Fig. 2 SAS-derived R_g values of xylanase relative to R_g of the bare protein. Modulations ΔR_g of the SAS-derived R_g values relative to R_g of the bare protein from unrestrained simulations of xylanase, obtained with 18 different combinations of protein force field (labels along the abscissa) and water model (color code, see legend). **a** ΔR_g from SAXS, **b** from SANS in H_2O, and **c** from SANS in D_2O. Statistical errors (1 SE) were obtained from block averaging. For force field abbreviations, see Table S1.

collapsed IDP ensembles, which has been attributed to a lack of protein–water dispersion interactions[36,47]. Thus, several modifications of the AMBER force field family have been developed with refined protein–water dispersion interactions, leading to more expanded IDP ensembles. We considered the protein force fields DES-amber[35,37], DES-amber without scaled charges DES-amberSF1.0, a99SBdisp[38], and ff99SBws[36], each simulated in conjunction with the recommended water model TIP4P-D, a99SBdisp, or TIP4P/2005s (Table S1). For RNaseA and xylanase, the ΔR_g^{SAS} values of DES-amber and a99SBdisp are in good agreement with the experimental values, although smaller variations are visible (Fig. 3b, d). The combination of ff99SBws with TIP4P/2005s leads to large ΔR_g^{SAS} values, exceeding the experimental values. This may indicate that ff99SBws–TIP4P/2005s overestimates protein–water interactions, leading to overly dense hydration shells of RNaseA and xylanase (Fig. 3b, d, black bar).

Glucose isomerase is a globular protein decorated with a high density of acidic amino acids, leading to a large negative charge of $-60\,e$. Glucose isomerase exhibits large ΔR_g^{SAS} values of SAXS relative to SANS/D_2O, indicative of an exceptionally pronounced hydration shell (Figs. 3f and S6f). These findings agree qualitatively with Kim et al. who reported a pronounced hydration shell around a highly anionic variant of green fluorescent protein (GFP) but not around the wild-type or around a highly cationic variant of GFP[20]. All 18 force field combinations captured the increased ΔR_g^{SAS} values of glucose isomerase relative to RNaseA or xylanase. Many force field combinations yield excellent agreement with the experimental value, however, several combinations even seem to overestimate ΔR_g^{SAS}, indicative of a slightly exaggerated hydration shell contrast. Interestingly, variations among different force fields do not follow the same trend for glucose isomerase as observed for the near-neutral proteins RNaseA or xylanase; for instance, CHARMM36m simulations yield similar (with TIP3P) or even larger (with SPC/E or OPC3) ΔR_g^{SAS} values as compared to the simulations with the AMBER force fields ff14SB or ff99SB-ildn. Thus, force field effects on the hydration shell depend on the physicochemical characteristics of the protein surface, suggesting that comparisons based on several proteins are mandatory to scrutinize force field effects on the protein hydration shell.

SANS collected in H_2O is subject to poorer signal-to-noise ratio as compared to SANS in D_2O owing to the greatly increased incoherent scattering cross section of hydrogen relative to deuterium, leading to a pronounced scattering background. These challenges may lead to an inaccurate buffer subtraction and are reflected by largely increased uncertainties of the ΔR_g^{SAS} values obtained from consensus SAXS and SANS/H_2O curves (Fig. 3a, c, e, horizontal orange lines, Table S2)[26]. Thus, current SANS/H_2O data are not suitable for quantitative validation of the hydration shell from MD simulations. Instead, we suggest that MD simulations, after having validated the hydration shell against SAXS and SANS/D_2O data as done here (Fig. 3b, d, f), may be used in future studies to improve the accuracy of the buffer subtraction of SANS/H_2O experiments, thereby enabling a more quantitative analysis of SANS/H_2O curves.

Protein size and shape furthermore influence ΔR_g^{SAS} values. Apart from SAS data of RNaseA, xylanase, and glucose isomerase, the round-robin SAS benchmark study collected SAS data for lysozyme and urate oxidase[26]. However, because the R_g values of lysozyme and urate oxidase were subject to increased spread owing to problems with radiation damage and aggregation (SI Discussion), these data have not been used for validating MD simulations in this study. Instead, SAS calculations for lysozyme and urate oxidate provide additional insight on the effect of protein size and shape on ΔR_g^{SAS} values. MD simulations with 18 different force field combinations show by far larger ΔR_g^{SAS} values for lysozyme as compared to urate oxidase (Fig. S10b, d). We explain the large effect of the hydration shell on R_g of lysozyme with the small size of protein (14.3 kDa), leading to a relatively large contribution of the hydration shell to the overall contrast of the solute. Urate oxidase (i) is larger (136.3 kDa) than lysozyme, leading to a smaller contribution by the hydration shell to the overall contrast and (ii) exhibits a solvent-filled cavity; because the hydration shell in the cavity adds contrast close to the center of mass of urate oxidase, part of the hydration shell may even decrease the R_g. Thus, ΔR_g^{SAS} values are not only controlled by the surface properties of the protein as emphasized by the case of glucose isomerase, but also strongly influenced by the size and geometric shape of the protein[48].

Figure 4 summarizes ΔR_g and ΔR_g^{SAS} values for five proteins, obtained as a consensus average over six combinations of protein force field and water model (Methods), which exhibited good agreement with experimental data according to Fig. 3b, d, f. As discussed above, Fig. 4 highlights that ΔR_g and ΔR_g^{SAS} values strongly depends on the protein and are rationalized by varying surface composition, protein size, and protein shape. Thus, ΔR_g^{SAS} values report a footprint of the hydration shell that reflects protein-specific protein–water interactions and enable quantitative comparison with the experiment.

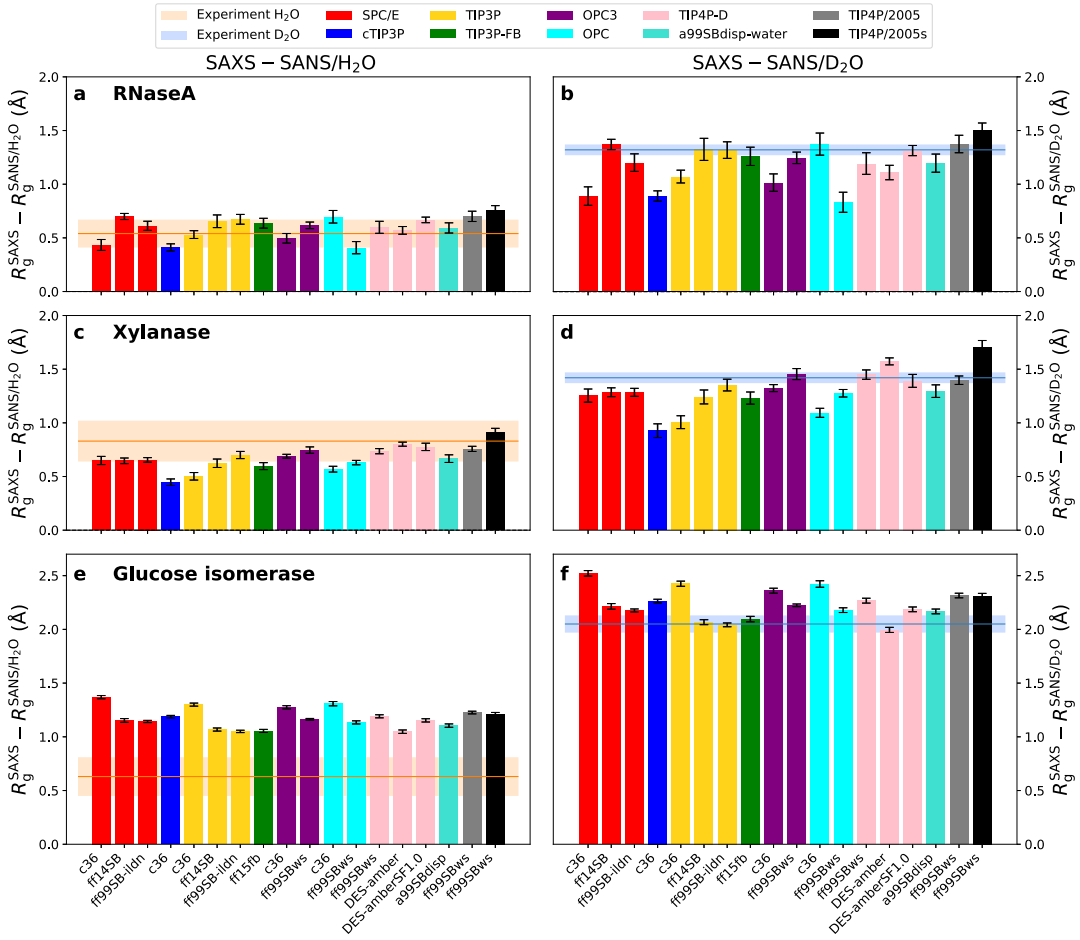


Fig. 3 Difference in R_g values from SAXS relative to SANS in H_2O or D_2O. Difference between R_g values from SAXS and SANS/H_2O (left column) or between SAXS and SANS/D_2O (right column) obtained from unrestrained simulations of **a**, **b** RNaseA, **c**, **d** xylanase, and **e**, **f** glucose isomerase. R_g values were obtained with 18 different combinations of protein force fields (labels along the abscissa) and water models (color code, see legend). Experimental consensus values and uncertainties from $P(r)$ analysis are shown as horizontal lines and shaded areas, respectively. Statistical errors denote 1 SE.

Discussion

MD simulations have been widely used to study the structure and dynamics of the protein hydration shell or to study proteins in crowded environments, where only a minor fraction of the solvent is bulk-like[19,49–51]. Whether the solvent structure in the hydration shell and in crowded environments matches with experimental conditions remains unclear. We showed that consensus SAS data combined with explicit-solvent SAS calculations provide a novel means to validate the hydration shell from MD simulations. To this end, we focused on the difference ΔR_g^{SAS} between the R_g values obtained with SAXS and SANS in D_2O, which (i) provides a measure for the contrast of the hydration shell relative to bulk solvent while it (ii) does not require accurate knowledge of the R_g of the bare protein.

Overall, we found remarkable agreement between ΔR_g^{SAS} from MD simulations and experiments for many combinations of protein force field and water model for the proteins RNaseA, xylanase, and glucose isomerase. These results demonstrate that the hydration shell contrast and, thereby, the packing of solvent on the protein surface is accurately captured by many modern force fields. However, we furthermore observed systematic differences among force field families. For the electrically nearly neutral proteins xylanase and RNaseA, simulations with CHARMM36m typically underestimated ΔR_g^{SAS} as compared to experimental values, in particular together with the widely used TIP3P or cTIP3P water models, indicative of underestimated hydration shell contrasts. Simulations with several AMBER variants revealed reasonable or even excellent agreement with experimental values. The ΔR_g^{SAS} values obtained with AMBER99SBws–TIP4P/2005s were larger compared to most other force fields and exceeded experimental values for several proteins, indicating that this force field yield a hydration shell with high contrast.

For the highly anionic glucose isomerase, and in contrast to simulations with near-neutral proteins, simulations with CHARMM36m revealed larger ΔR_g^{SAS} values as compared to most simulations with AMBER variants. These findings may suggest that CHARMM36m imposes tighter water packing on acidic

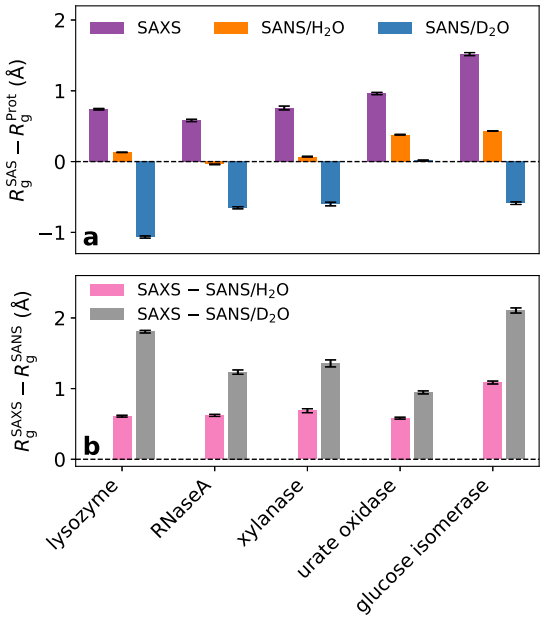


Fig. 4 Computational consensus ΔR_g values and ΔR_g^{SAS} values for five proteins. a Computational consensus ΔR_g values and **b** ΔR_g^{SAS} values for five proteins (see labels) obtained as average over six combinations of protein force field and water model that showed close agreement with experimental data according to Fig. 3. Color code is chosen following Fig. 1e, f.

residues as compared to AMBER force fields. However, we furthermore found that different coordination of ions[52] on acidic residues may contribute to variations of ΔR_g^{SAS} of glucose isomerase (Fig. S16).

Remarkably, the systematically different ΔR_g^{SAS} values with CHARMM36m relative to several AMBER variants were found irrespective of the applied water model, despite greatly different bulk properties of water models[53–55]. For instance, accurate bulk properties of a water model (such as OPC[45]) do not imply an accurate hydration shell contrast (if used together with CHARMM36m). Inversely, a water model with poor agreement with several bulk properties (such as TIP3P[45]) may yield an accurate hydration shell contrast (if used with ff14SB or ff99SB-ildn). Thus, ΔR_g^{SAS} yields structural information independent of the bulk properties that have been used to parametrize the water models. Protein–water interactions in MD simulations have frequently been validated using hydration free energies ΔG_{hyd} of amino acid analogues. However, ΔG_{hyd} values agree reasonably between CHARMM36m and recent AMBER variants[56], suggesting that ΔR_g^{SAS} provides a measure for protein–water interactions independent of ΔG_{hyd}. In addition, protein–water interactions have been discussed in the context of IDP simulations. Simulations with AMBER–TIP3P yield overly compact IDP ensembles, which has been attributed to a lack of protein–water dispersion interactions[36,37,47], whereas CHARMM36m–TIP3P[31] or AMBER variants with increased dispersion interactions[37] yields IDP ensembles in better agreement with experiments. Thus, there exists no simple correlation between ΔR_g^{SAS} and the spatial extent of IDP ensembles. Together, these observations suggest that ΔR_g^{SAS} values represent a footprint of protein–water interactions that is independent of previously considered observables, thus providing an additional observable for validating and further improving protein–water interactions in MD simulations.

Irrespective of the applied force field, ΔR_g^{SAS} values differed considerably between different proteins, in agreement with the data of the round-robin SAS study[26]. The highly anionic glucose isomerase exhibited the largest ΔR_g^{SAS} values among the five proteins considered in this study, indicative of a tightly packed hydration shell. These findings are in line with a SAS study of a highly anionic GFP variant[20] and demonstrate that the anionic aspartate and glutamate residues impose a densely packed hydration shell. Among the four proteins with zero or with a small net charge, lysozyme exhibited larger ΔR_g^{SAS} values as compared to urate oxidase, xylanase, and RNaseA (Fig. 4b). Urate oxidase exhibits the shape of a hollow cylinder with a large solvent-filled cavity, which may explain the low ΔR_g^{SAS} values (Fig. 4b) as well as a nearly vanishing ΔR_g for SANS/D_2O, in contrast to all other proteins (Fig. 4a). Thus, variations of ΔR_g^{SAS} are experimentally accessible footprints of protein-specific hydration shells reflecting specific geometric shapes or distributions of charged and polar moieties on the protein surface.

Because the ΔR_g^{SAS} values are in the range of only 1–2.5 Å, the comparisons presented here require highly accurate SAS data. Considering that SAS data may be subject to minor systematic errors, which may be difficult to detect, SAS data obtained at a single instrument may not yield the required accuracy, even if data collection and analysis follows established quality controls[24,25]. Instead, the use of consensus data collected at different SAS instruments, if possible, by independent researchers[26], is a rigorous means for obtaining data with unprecedented accuracy and, thereby, enables quantitative validation of the hydration layer as shown here. To validate the hydration shell of other biomolecules such as RNA or protein/RNA complexes, future benchmark studies similar to the round-robin study designed by Trewhella, Vachette, and coworkers would be of utmost value[26].

To enable quantitative comparison with the experiments, the MD simulations should match the experimental conditions and require control calculations. We carefully evaluated the effects of (i) protein flexibility (Figs. 1e, f and S4), (ii) use of salt as compared to use of only counter ions (Figs. S12–S15), (iii) refined sodium–carboxylate interaction parameters (Fig. S16) and (iv) Lennard-Jones cutoff settings (Fig. S17, see Supplementary Results). We found that these factors modulate ΔR_g^{SAS} only by a small fraction of an Ångström. Nevertheless, because such effects are clearly detectable in explicit-solvent SAS predictions of ΔR_g^{SAS}, they require consideration upon comparison with experiments.

Conclusions

We showed that the hydration shell contrast, as reported by SAS-derived R_g values, strongly depends on the geometric shape and surface composition of proteins, thus providing a probe of protein-specific protein–solvent interactions. As readout of the hydration shell structure, we focused on the difference in R_g values from SAXS relative to SANS experiments in D_2O (ΔR_g^{SAS}), which we computed from MD simulations with explicit-solvent SAS calculations to enable quantitative comparison with experimental SAS data. For many force fields, ΔR_g^{SAS} values from MD simulations revealed excellent agreement with consensus SAS data from a recent worldwide round-robin study[26], suggesting that simulations accurately capture the hydration shell contrast relative to the bulk. Because we furthermore observed differences among force fields, our calculations provide the basis for further improving the accuracy of protein–water interactions in molecular simulations. This study establishes the combination of high-precision SAS experiments with explicit-solvent calculations as a tool for scrutinizing atomistic models of the protein hydration shell.

Methods

Simulation setup and parameters. Initial structures of lysozyme, RNaseA, xylanase, glucose isomerase, and urate oxidase were taken from the protein data bank (PDB codes: 2VB1[57], 7RSA[58], 2DFC[59], 1MNZ[60] and 3L8W[61], respectively). Crystal waters were kept in the structures of lysozyme, RNaseA, and xylanase, whereas organic molecules of the crystallization buffer were removed. Missing amino acids of glucose isomerase and urate oxidase were added, such that the sequences were identical to the samples used in Ref. [26]. Specifically, one methionine was added to the N-terminus of glucose isomerase, and six amino acids (sequence (SLKSKL)) to the C-terminus of urate oxidase. Hydrogen atoms were added with the GROMACS module pdb2gmx. The starting structures were placed in a dodecahedral box, where the distance between the protein and the box edges was at least 2.0 nm, and the box was filled with explicit solvent. For the highly charged glucose isomerase (−60 e), sodium and chloride ions were added by replacing the appropriate number of water molecules with ions to obtain a salt concentration of 100 mM NaCl. For all other proteins, the system was neutralized by adding sodium or chloride counter ions, if not stated otherwise.

Each protein was simulated using 18 combinations of protein force field and water model (Table S1). Interactions of the proteins were described with one of the following force fields: AMBER14SB (ff14SB)[32], AMBER99SB-ildn (ff99SB-ildn)[33], AMBER99SBws (ff99SBws)[39], AMBER99sb (ff99SB)[62], AMBER15/force-balance (ff15fb)[34], DES-amber[35], DES-amberSF1.0[35], AMBER99SBdisp (a99SBdisp)[38], or CHARMM36m (c36)[31] (version July 2020). Water was described with one of the following models: SPC/E[44], CHARMM-modified TIP3P involving hydrogen atoms with Lennard-Jones interactions (cTIP3P)[43], TIP3P[63], TIP3P-FB[64], OPC3[65], OPC[45], TIP4P-D[37], a99SBdisp-water[38], TIP4P/2005[66], or TIP4P/2005s[36].

All MD simulations were carried out with the GROMACS software, version 2020.3[67]. After 400 steps of minimization with the steepest decent algorithm, the systems were equilibrated for 100 ps with harmonic position restraints applied to the heavy atoms of the proteins (force constant 1000 KJ mol^{-1}nm^{-2}). Subsequently, the production runs were started without restraints on the atoms or with restraints applied to the heavy atoms (force constant 2000 KJ mol^{-1}nm^{-2}) or applied to the backbone atoms (force constant 2000 KJ mol^{-1}nm^{-2}) of the protein. The equations of motion were integrated using a leap-frog algorithm[68]. The temperature was controlled at 298.15 K, using velocity rescaling ($\tau = 1$ ps)[69]. The pressure was controlled at 1 bar with the Berendsen thermostat ($\tau = 1$ ps)[70] and with the Parrinello-Rahman thermostat ($\tau = 5$ ps)[71] during equilibration and production simulation, respectively. The geometry of the water molecules was constrained with the SETTLE algorithm[72] and LINCS[73] was used to constrain all other bond length. A time step of 2 fs was used. Dispersive interactions and short-range repulsion were described by a Lennard-Jones potential. For simulations with AMBER variants, LJ interactions were cut off at 1 nm. For simulations with CHARMM36m, LJ forces were gradually switched off between 1 nm and 1.2 nm, if not stated otherwise. In simulations with AMBER variants, the pressure and energy were corrected of missing dispersion corrections beyond the cut-off. Neighbor lists were updated with the Verlet scheme. Coulomb interactions were computed with the smooth particle-mesh Ewald method[74,75]. We used a Fourier spacing of ~0.12 nm, which was optimized by the GROMACS mdrun module at the beginning of each simulation. Systems with restraints on heavy atoms or on the backbone were simulated for 50 ns, which is sufficient to sample fluctuations of water molecules or amino acid side chains (if present). Unrestrained simulations were carried out for 230 ns. Because the five proteins considered in the round-robin SAS study are relatively rigid, 230 ns were likely sufficient to sample the most relevant conformation space used to compute the increase of R_g owing to the hydration shell.

The 3D solvent density shown in Fig. 1b was computed with the rerun functionality of GROMACS-SWAXS using the environment variable GMX_WAXS_GRID_DENSITY=1 and GMX_WAXS_GRID_DENSITY_MODE=2[76]. The solvent density written in CUBE format was visualized with PyMol[77].

Explicit-solvent SAS calculations. The SAXS and SANS calculations were performed with GROMACS-SWAXS (version 2021.5), a modified version of the GROMACS simulation software that implements explicit-solvent SAXS[29] and SANS calculations[78]. GROMACS-SWAXS is furthermore used by the web server WAXSiS for automated explicit-solvent SAXS predictions[30] and is freely available at GitLab (https://gitlab.com/cbjh/gromacs-swaxs). For more details on the rationale behind explicit-solvent SAS calculations including differences relative to implicit-solvent SAS calculations, we refer to previous reviews[76,79]. A spatial envelope (Fig. 1a) was constructed at a distance of 9 Å from all protein atoms. Solvent atoms (water and ions) inside the envelope contributed to the calculated SAXS/SANS curves, thereby taking the hydration shell into account. The buffer subtraction was carried out using 2251 simulations frames of pure solvent simulation box, which was simulated for 50 ns and large enough to enclose the envelope. The orientational average was carried out using 200 **q**-vectors for each absolute value of q, and the solvent electron density was corrected to the experimental water density of 334 e/nm^3, as described previously[29]. In this study, a small number of only 200 **q**-vectors per absolute value of q was sufficient because we computed the SAS curves only up to small angles to carry out the Guinier analysis. The density correction is required to ensure accurate buffer matching of bulk solvent between the protein and the pure-water simulations, and to correct for the density of certain water models that differs from the experimental density. No fitting parameters owing to the hydration layer or excluded solvent were used, implying that the radius of gyration R_g was not adjusted by the fitting parameters but fully imposed by the force field (together with other MD parameters such as cutoffs, temperature, etc.).

SAXS and SANS curves were computed from 2251 simulation frames taken from the time interval between 5 ns and 50 ns or between 30 ns and 230 ns for restrained and unrestrained simulations, respectively. Thus, computed SAS curves and R_g values represent averages over protein and solvent fluctuations within the simulated time scales. Statistical errors of calculated SAS curves were obtained by binning the trajectories into 10 time blocks of 4.5 ns or 20 ns for simulations with or without restraints, respectively. Here, for restrained simulations, shorter simulations and time blocks used because restrained simulations exhibit shorter autocorrelation times as compared to unrestrained simulations. Likewise, the pure-solvent simulations were binned into independent blocks of 4.5 ns. Then the SAS curves were computed from independent pairs of solute and pure-solvent trajectories. Critically, the use of independent pure-solvent trajectory blocks is mandatory to exclude correlations between the SAS curves computed from time blocks. Reported error bars denote one standard error (1 SE).

SAXS data reported by the round-robin benchmark revealed only a marginal effect (if any) upon replacing H_2O with D_2O in SAXS experiments[26]. Thus, for the prediction of SANS/D_2O curves, we did not use force fields for heavy water[80] or force fields that would account for the deuteration of amino acids. Instead, we assigned the neutron scattering length of deuterium to water hydrogen atoms and to polar protein atoms. Hydrogen atoms of

the amine backbone groups were assumed to be deuterated with a probability of 90%.

To test whether computed R_g values from Guinier analysis agree with R_g values from the $P(r)$ function, we computed one SAXS curve of xylanase up to $q = 3\ nm^{-1}$, obtained the $P(r)$ with GNOM[81], and computed R_g from $P(r)$. The R_g values from Guinier and $P(r)$ analysis were identical and equaled 1.59 nm. Thus, we used computed R_g from Guinier analysis for the remainder of this study for simplicity.

Computational consensus R_g values were computed with the following combinations of protein and water force fields: ff14SB–TIP3P, ff99SB-ildn–TIP3P, ff15fb–TIP3P-FB, ff99SB ws–TIP4P-D, DES-amber–TIP4P-D, a99SBdisp–a99SBdisp-water.

Density profiles around proteins (Fig. 1c, solid line; Fig. S7) were computed from heavy atom-restrained simulations using gmx genenv module of GROMACS-SWAXS, by using an icosphere with 81920 triangular faces (option -nrec 6). The tool builds a series of envelopes with increasing distances from the Van-der-Waals surface of the protein and computes the electron density between pairs of adjacent envelopes[82]. Simulations used for computing density profiles around bulk water were set up as described in the Supplementary Methods (Fig. 1c, dashed line; Fig. S8).

Data availability

All data used in this manuscript are available from the authors. MD input files to reproduce this study are publicly available in a Zenodo repository at https://doi.org/10.5281/zenodo.10007526[83].

Code availability

MD simulation were carried out with GROMACS version 2020.3, which is freely available at https://www.gromacs.org. Explicit-solvent SAS calculations and density calculations were carried with GROMACS-SWAXS version 2021.5, which is freely available at https://gitlab.com/cbjh/gromacs-swaxs and documented at https://cbjh.gitlab.io/gromacs-swaxs-docs.

Received: 26 June 2023; Accepted: 20 November 2023;
Published online: 12 December 2023

References

1. Pal, S. K. & Zewail, A. H. Dynamics of water in biological recognition. *Chem. Rev.* **104**, 2099–2124 (2004).
2. Levy, Y. & Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389–415 (2006).
3. Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108 (2008).
4. Mondal, S., Mukherjee, S. & Bagchi, B. Protein hydration dynamics: much ado about nothing? *J. Phys. Chem. Lett.* **8**, 4878–4882 (2017).
5. Fogarty, A. C., Duboué-Dijon, E., Sterpone, F., Hynes, J. T. & Laage, D. Biomolecular hydration dynamics: a jump model perspective. *Chem. Soc. Rev.* **42**, 5672–5683 (2013).
6. Bagchi, B. Water dynamics in the hydration layer around proteins and micelles. *Chem. Rev.* **105**, 3197–3219 (2005).
7. Wüthrich, K. et al. NMR studies of the hydration of biological macromolecules. *Faraday Discuss.* **103**, 245–253 (1996).
8. Crilly, C. J., Eicher, J. E., Warmuth, O., Atkin, J. M. & Pielak, G. J. Water's variable role in protein stability uncovered by liquid-observed vapor exchange NMR. *Biochemistry* **60**, 3041–3045 (2021).
9. Laage, D., Elsaesser, T. & Hynes, J. T. Water dynamics in the hydration shells of biomolecules. *Chem. Rev.* **117**, 10694–10725 (2017).
10. Ebbinghaus, S. et al. An extended dynamical hydration shell around proteins. *Proc. Natl Acad. Sci. USA* **104**, 20749–20752 (2007).
11. Born, B., Kim, S. J., Ebbinghaus, S., Gruebele, M. & Havenith, M. The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss.* **141**, 161–173 (2009).
12. Sushko, O., Dubrovka, R. & Donnan, R. S. Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected. *J. Chem. Phys.* **142**, 055101 (2015).
13. Li, T., Hassanali, A. A., Kao, Y.-T., Zhong, D. & Singer, S. J. Hydration dynamics and time scales of coupled water-protein fluctuations. *J. Am. Chem. Soc.* **129**, 3376–3382 (2007).
14. Bellissent-Funel, M.-C. et al. Water determines the structure and dynamics of proteins. *Chem. Rev.* **116**, 7673–7697 (2016).
15. Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia Coli*. *J. Mol. Biol.* **222**, 599–620 (1991).
16. Harada, R., Sugita, Y. & Feig, M. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc.* **134**, 4842–4849 (2012).
17. Svergun, D. I. et al. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl Acad. Sci. USA* **95**, 2267–2272 (1998).
18. Merzel, F. & Smith, J. C. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl Acad. Sci. USA* **99**, 5378–5383 (2002).
19. Persson, F., Söderhjelm, P. & Halle, B. The geometry of protein hydration. *J. Chem. Phys.* **148**, 215101 (2018).
20. Kim, H. et al. SAXS/SANS on supercharged proteins reveals residue-specific modifications of the hydration shell. *Biophys. J.* **110**, 2185–2194 (2016).
21. Cheng, Y.-K. & Rossky, P. J. Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392**, 696–699 (1998).
22. Marchi, M., Sterpone, F. & Ceccarelli, M. Water rotational relaxation and diffusion in hydrated lysozyme. *J. Am. Chem. Soc.* **124**, 6787–6791 (2002).
23. Sterpone, F., Stirnemann, G. & Laage, D. Magnitude and molecular origin of water slowdown next to a protein. *J. Am. Chem. Soc.* **134**, 4116–4119 (2012).
24. Skou, S., Gillilan, R. E. & Ando, N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protoc.* **9**, 1727–1739 (2014).
25. Trewhella, J. et al. 2017 Publication Guidelines for Structural Modelling of Small-Angle Scattering Data from Biomolecules in Solution: An Update. *Acta Crystallogr D Struct. Biol.* **73**, 710–728 (2017).
26. Trewhella, J. et al. A Round robin approach provides a detailed assessment of biomolecular small-angle scattering data reproducibility and yields consensus curves for benchmarking. *Acta Cryst.* D78 (2022).
27. Park, S., Bardhan, J. P., Roux, B. & Makowski, L. Simulated X-ray scattering of protein solutions using explicit-solvent models. *J. Chem. Phys.* **130**, 134114 (2009).
28. Köfinger, J. & Hummer, G. Atomic-resolution structural information from scattering experiments on macromolecules in solution. *Phys. Rev. E* **87**, 052712 (2013).
29. Chen, P.-c & Hub, J. S. Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophys. J.* **107**, 435–447 (2014).
30. Knight, C. J. & Hub, J. S. WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res.* **43**, W225–W230 (2015).
31. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
32. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
33. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
34. Wang, L.-P. et al. Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. *J. Phys. Chem. B* **121**, 4023–4039 (2017).
35. Piana, S., Robustelli, P., Tan, D., Chen, S. & Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein-protein complexes. *J. Chem. Theory Comput.* **16**, 2494–2507 (2020).
36. Best, R. B., Zheng, W. & Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **10**, 5113–5124 (2014).
37. Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **119**, 5113–5123 (2015).
38. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl Acad. Sci. USA* **115**, E4758–E4766 (2018).
39. Best, R. B. & Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004–9015 (2009).
40. Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Curr. Opin. Struct. Biol.* **49**, 18–26 (2018).
41. Glatter, O. A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr.* **10**, 415–421 (1977).
42. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

43. MacKerell, A. D. J. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
44. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
45. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building water models: a different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
46. Hess, B. & van der Vegt, N. F. A. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J. Phys. Chem. B* **110**, 17616–17626 (2006).
47. Rauscher, S. et al. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *J. Chem. Theory Comput.* **11**, 5513–5524 (2015).
48. Kim, H. S. & Gabel, F. Uniqueness of models from small-angle scattering data: the impact of a hydration shell and complementary NMR restraints. *Acta Cryst. D* **71**, 57–66 (2015).
49. Petrov, D. & Zagrovic, B. Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLoS Comput. Biol.* **10**, e1003638 (2014).
50. Wang, P.-h, Yu, I., Feig, M. & Sugita, Y. Influence of protein crowder size on hydration structure and dynamics in macromolecular crowding. *Chem. Phys. Lett.* **671**, 63–70 (2017).
51. Heo, L., Sugita, Y. & Feig, M. Protein assembly and crowding simulations. *Curr. Opin. Struct. Biol.* **73**, 102340 (2022).
52. Yoo, J. & Aksimentiev, A. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* **20**, 8432–8449 (2018).
53. Sorenson, J. M., Hura, G., Glaeser, R. M. & Head-Gordon, T. What can X-ray scattering tell us about the radial distribution functions of water? *J. Chem. Phys.* **113**, 9149 (2000).
54. Onufriev, A. V. & Izadi, S. Water models for biomolecular simulations. *WIREs Comput. Mol. Sci.* **8**, e1347 (2018).
55. Kadaoluwa Pathirannahalage, S. P. et al. Systematic comparison of the structural and dynamic properties of commonly used water models for molecular dynamics simulations. *J. Chem. Inf. Model.* **61**, 4521–4536 (2021).
56. Qiu, Y., Shan, W. & Zhang, H. Force field benchmark of amino acids. 3. Hydration with scaled lennard-jones interactions. *J. Chem. Inf. Model.* **61**, 3571–3582 (2021).
57. Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. Triclinic lysozyme at 0.65 Å resolution. *Acta Cryst. D* **63**, 1254–1268 (2007).
58. Wlodawer, A., Svensson, L. A., Sjoelin, L. & Gilliland, G. L. Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochem.* **27**, 2705–2717 (1988).
59. Watanabe, N., Akiba, T., Kanai, R. & Harata, K. Structure of an orthorhombic form of xylanase II from Trichoderma reesei and analysis of thermal displacement. *Acta Cryst. D* **62**, 784–792 (2006).
60. Nowak, E., Panjikar, S. & Tucker, P. Atomic structure of Glucose isomerase. *To be published*.
61. Gabison, L. et al. Near-atomic resolution structures of urate oxidase complexed with its substrate and analogues: the protonation state of the ligand. *Acta Cryst. D* **66**, 714–724 (2010).
62. Tian, C. et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020).
63. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
64. Wang, L.-P., Martinez, T. J. & Pande, V. S. Building force fields: an automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **5**, 1885–1891 (2014).
65. Izadi, S. & Onufriev, A. V. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **145**, 074501 (2016).
66. Abascal, J. L. F. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123**, 234505 (2005).
67. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).
68. Hockney, R., Goel, S. & Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **14**, 148–158 (1974).
69. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
70. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
71. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
72. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
73. Hess, B. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
74. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
75. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
76. Chatzimagas, L. & Hub, J. S. in *Small Angle Scattering Part A: Methods for Structural Investigation. Methods in Enzymology*, Vol. 677 (ed. Tainer, J. A.) 433–456 (Elsevier, 2022).
77. Version 2.0 Schrödinger, LLC. *The PyMOL Molecular Graphics System*.
78. Chen, P.-c et al. Combined Small-Angle X-ray and Neutron Scattering Restraints in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **15**, 4687–4698 (2019).
79. Chatzimagas, L. & Hub, J. S. in *Small Angle Scattering Part B: Methods for Structural Interpretation. Methods in Enzymology*, Vol. 678 (ed. Tainer, J. A.) 23–54 (Elsevier, 2023).
80. Linse, J.-B. & Hub, J. S. Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW. *J. Chem. Phys.* **154**, 194501 (2021).
81. Franke, D. et al. *ATSAS 2.8* : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
82. Chen, P.-c & Hub, J. Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophys. J.* **107**, 435–447 (2014).
83. Linse, J.-B. & Hub, J. S. Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data (Simulation input files). https://doi.org/10.5281/zenodo.10007526 (2023).

Acknowledgements

We thank Jill Trewhella for insightful discussions. This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via grants HU 1971/3-1, HU 1971/3-2, and INST 256/539-1.

Author contributions

J.-B.L. performed and analyzed MD simulations. J.-B.L. and J.S.H. designed the project and wrote the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42004-023-01067-1.

Correspondence and requests for materials should be addressed to Jochen S. Hub.

Peer review information *Communications Chemistry* thanks Marcelo Poleto and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary material for: Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data

Johanna-Barbara Linse and Jochen S. Hub*

Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, 66123, Germany

E-mail: jochen.hub@uni-saarland.de

Supplementary Material Table of Contents

Supplementary Discussion

Choice of SAS data used for quantitative comparison with MD simulations

Trewhella, Vachette, et al. recently reported a worldwide round-robin benchmark on the reproducibility of SAXS and SANS experiments, involving experiments at 12 SAXS and four SANS instruments and leading to a total of 171 SAXS and 76 SANS curves on the following five proteins (Fig. S2): ribonuclease A (RNaseA, 13.690 kDa), hen egg white lysozyme (14.313 kDa), xylanase (20.844 kDa), urate oxidase (136.303 kDa), and glucose isomerase (172.910 kDa, also referred to as xylose isomerase).[1] Both, SAXS and SANS data were collected in batch mode or coupled to size-exclusion chromatography (SEC-SAXS, SEC-SANS). The R_g values were evaluated using Guinier analysis and based on the $P(r)$ function. Generally, good agreement was found between the experimental SAS curves and explicit-solvent SAXS predictions in the round-robin study.[1]

However, the quality and reproducibility of the SAS data differed among the five proteins. Specifically, R_g values for lysozyme varied substantially between different data sets, even if experiments were carried out in SEC mode. These variations have been attributed to an increased sensitivity of lysozyme to radiation damage[1] and are in line with variations of the lysozyme R_g in earlier reports.[2,3]

Urate oxidase exhibits increased propensity for aggregation and, thus, had to be shipped on ice and with a reduced concentration of only $5\,\mathrm{mg\,ml^{-1}}$.[1] In consequence, SAS data of urate oxidate was subject to poorer statistics as compared to the data of the other four proteins, in particular at wider scattering angles. Because the SAS curve over the whole q-range is used when computing R_g via the $P(r)$ function, such problems may lead to increased uncertainties of the R_g estimate. Furthermore, SANS in H_2O exhibited exceptionally high background, and yielded unrealistic R_g values from Guinier analysis. Owing to these problems with lysozyme and urate oxidase, we did not include these proteins in the quantitative comparison of ΔR_g^{SAS} values in Fig. 3. For the sake of completeness, the R_g^{SAS} values

for lysozyme and urate oxidase, together with the experimental R_g estimates from Guinier analysis and $P(r)$, are shown in Fig. S10.

Validation with R_g from $P(r)$ vs. Guinier analysis

In this study, we used the experimental R_g values obtained from the $P(r)$ function and not from the Guinier analysis for comparison with MD simulations. In contrast to the Guinier analysis that uses only the small-angle region of the SAS curve, the $P(r)$ function makes use of the entire SAS curve. Thus, R_g values from the $P(r)$ reported in the round-robin SAS benchmark study were less sensitive to remaining undetected protein–protein aggregation and subject to smaller statistical uncertainties as compared to R_g estimates from the Guinier analysis (Table S2).[1] Nevertheless, the ΔR_g^{SAS} values from simulations reveal reasonable agreement also with R_g from Guinier analysis (Fig. S11).

Supplementary Results: additional control SAS calculations

Counter ion cloud contributions outside the envelope are negligible

For the highly anionic glucose isomerase (GI), the counter ion cloud adds to the contrast of the hydration layer. According to Debye-Hückel theory, the counter ion cloud decays with the Debye length λ_D into the bulk solvent, which equals $\lambda_D = 9.7\,\text{Å}$ for a 100 mM NaCl solution. Hence, since only solvent atoms within the envelope at a distance of 9 Å from the protein (Fig. 1a, blue surface) contributed to our explicit-solvent SAS calculations, counter ion cloud effects were likewise only taken into account up to a distance of 9Å from the protein surface, whereas counter ion cloud contributions beyond 9Å were neglected.

To exclude that this approximation affects the computed SAS curves, we estimated the effect of the counter ion cloud at distances beyond 9 Å on R_g. Following Ref. 4, we modeled GI as a sphere with contrast, volume, and R_g taken from GI, and we used linearized Poisson–Boltzmann calculations to obtain the number densities of sodium and chloride as function of distance R from the sphere surface. By computing R_g including the counter ion

cloud up to $R = 9\,\text{Å}$ or up to a large distance, we found that R_g was modified by the cutoff at $R = 9\,\text{Å}$ by only $0.033\,\text{Å}$, $-0.062\,\text{Å}$ and $-0.03\,\text{Å}$ for SAXS, SANS in H_2O, and SANS in D_2O. Since these errors are within the statistical errors, these effects were neglected for further analysis.

Effect of salt relative to using only counter ions

Since GI carries a large negative charge ($-60\,e$), GI is surrounded by a counter ion cloud. The ion densities in the counter ion cloud decay into the bulk solvent with the Debye length λ_D, which is inversely proportional to the square root of the ion concentration in bulk solvent.[5] Thus, upon simulating with a smaller ion concentration relative to experimental conditions, the spatial extend of the counter ion cloud in simulation would exceed the experimental conditions; consequently, the counter ion cloud in simulations would, relative to the experiment, impose an electron density contrast at larger distances from the protein, which may lead to an overestimated R_g. We tested the effect of the ion concentration for GI by computing R_g, ΔR_g, and $\Delta R_\text{g}^\text{SAS}$ values from simulations with only counter ions, 100 mM NaCl salt, or 150 mM NaCl salt (Fig. S14). We find that, with increasing salt concentration, $\Delta R_\text{g}^\text{SAS}$ values decrease, as expected from a deceasing Debye length and, thereby, from a spatially more compact counter ion cloud. For instance, in simulations with CHARMM36m–TIP3P, upon replacing counter ions with 150 mM NaCl salt, the R_g from SAXS relative to SANS/D_2O decreases by ∼0.2Å. This trend is confirmed by GI simulations with all 18 force field combinations using purely counter ions (compare Fig. S15 with Fig. 3e/f). Thus, for quantitative comparison of $\Delta R_\text{g}^\text{SAS}$ values between simulation and experiment, an at least approximate match of the buffer conditions is mandatory, as used for our study.

In addition, we tested the effect of adding additional 150 mM NaCl to simulation of xylanase and RNaseA instead of using only counter ions (Figs. S12 and S13). For these near-neutral proteins, $\Delta R_\text{g}^\text{SAS}$ values agreed within error bars between simulations with salt and simulations with counter ions, as expected form the absence of a pronounced counter

ion cloud.

Effect of cutoff settings for Lennard-Jones interactions

Neglect of Lennard-Jones (LJ) interactions beyond the cutoff could in principle modify the density of the hydration shell. Thus, we tested the effects of LJ cutoffs for xylanase simulated with the CHARMM36m–TIP3P force field combination. The CHARMM36m force field has been parameterized with a gradual switch of LJ forces between 1 and 1.2 nm. We carried out xylanase simulations either (i) following the CHARMM recommendations (force-switch) or (ii) using a plain cutoff at 1 nm. For each cutoff setting, simulations were carried out with restraints on heavy atoms, restraints on the backbone, or using a unrestrained MD simulation (Fig. S17). For simulations with restraints on the backbone or for unrestrained MD simulations, we obtained larger ΔR_g^{SAS} values with the force-switch settings as compared to using a plain cutoff (Fig. S17, orange vs. yellow and lightblue vs. darkblue bars). Hence, upon validating force fields against ΔR_g^{SAS} values from experiments, it is critical to follow the cutoff settings that have been used during force field parametrization.

Notably, for simulations with restraints on heavy atoms, no such effect was visible (Fig. S17, pink vs. red bars). Thus, additional LJ interactions between 1 nm and 1.2 nm lead to larger ΔR_g^{SAS} only in the presence of amino acid side chain fluctuations. These findings demonstrate that the additional LJ interactions between 1 nm and 1.2 nm with force-switch settings lead to increased ΔR_g^{SAS} values not because of tighter packing of water on a given protein surface; instead, the additional interactions lead to more favorable protein–water conformations, which manifest in a larger density contrast of the hydration shell.

Effect of refined Lennard-Jones parameters to avoid Na^+–carboxylate overbinding in AMBER simulations (CUFIX)

Certain older models of monovalent cations such as Na^+ tend to overbind to carboxylate moieties, as present in aspartate and glutamate residues.[6,7] We found previously that such

overbinding of monovalent ions to anionic proteins may influence the R_g values obtained from explicit-solvent SAS calculations.[8] Whereas refined LJ parameters have been incorporated into the default CHARMM36m force field to avoid overbinding (denoted NBFIX or CUFIX), refined LJ parameters for AMBER force fields have become available only recently.[7]

As discussed in the Results, ΔR_g^{SAS} values for GI obtained with CHARMM36m were systematically larger as compared to values obtained with AMBER force fields, irrespective of the applied water model (Fig. 3e/f). We hypothesized that different binding properties of Na^+ ions in CHARMM36m as compared to AMBER simulations may partly explain such effect. To test this hypothesis, we carried out additional SAS calculations for GI with the ff14SB force field with CUFIX and compared the ΔR_g^{SAS} values with results from standard ff14SB or CHARMM36m (Fig. S16).[7] Upon modifying the LJ interaction with CUFIX in ff14SB simulations, (i) the density of Na^+ at the protein surface decreased relative to ff14SB and took lower densities as compared to ff99SB-ildn (Fig. S16g), (ii) ΔR_g^{SAS} values increased by $\sim$0.2Å or $\sim$0.1Å for SANS/D_2O or SANS/H_2O (Fig. S16b/c, pink vs. orange bars). However, the ΔR_g^{SAS} values with ff14SB/CUFIX remained below the results from CHARMM36m. Thus, refined Na^+–carboxylate Lennard-Jones interactions have only a small effect on ΔR_g^{SAS} values, and the increased ΔR_g^{SAS} value with CHARMM36m relative to several AMBER force fields is only partly explained by reduced binding of Na^+ ions to the protein surface.

Density correction for accurate buffer matching is critical for R_g predictions

Explicit-solvent SAS predictions implemented by GROMACS-SWAXS use a density correction for two reasons:[9,10] (i) Certain water models such as TIP3P underestimate the water density, which would lead to an overestimation of the contrast of the protein relative to bulk solvent and, thereby, to an overestimation of the forward scattering I_0. (ii) Since the bulk solvent density in the protein simulation may marginally differ from the density in the pure-solvent simulations, possibly owing to finite-size effects, spurious contrasts may emerge

at larger distances from the protein, which would lead to artifacts in the R_g estimate.

To avoid such artifacts, GROMACS-SWAXS corrects the solvent density both in the pure-solvent and in the protein simulation system. In the pure-solvent system, a uniform density is added to match a preselected bulk density ρ^0_{bulk}, which may be taken as the experimental density of pure water of $334\,\mathrm{e\,nm^{-3}}$. In the protein simulation, the bulk density $\rho^{\mathrm{MD}}_{\mathrm{bulk}}$ is obtained by averaging over simulation frames within the volume outside of the envelope. A correction factor is obtained as the difference relative to ρ^0_{bulk} as $f = \rho^0_{\mathrm{bulk}}/\rho^{\mathrm{MD}}_{\mathrm{bulk}}$ (see also Fig. S7a versus b). Then the solvent density inside of the envelope (which contributes to the calculated SAS curve) is corrected by the factor f.

Here, we tested the effect of the density correction on $\Delta R_g^{\mathrm{SAS}}$ values. To this end, we first replaced the preselected bulk density of $334\,\mathrm{e\,nm^{-3}}$ with the bulk density of the respective water model. Thereby, the protein–bulk contrast may not match the experimental value, however, accurate buffer matching between the protein and the pure-solvent simulations was still active. As expected, this led to a considerable change of the forward scattering I_0 for water models with inaccurate bulk densities such as TIP3P or cTIP3P; however, the $\Delta R_g^{\mathrm{SAS}}$ revealed only a marginal change that was not visible in the bar plots.

Second, we fully disabled the density correction, thereby not ensuring accurate buffer matching any more. In these calculations, we observed changes of $\Delta R_g^{\mathrm{SAS}}$ of SAXS relative to SANS/D_2O by up to 0.7 Å. In addition, the results revealed poor agreement with experiments. Thus, for accurate $\Delta R_g^{\mathrm{SAS}}$ and for comparison with experiments, corrections for accurate buffer matching are mandatory, as used in the present study.

Supplementary Methods

Solvent density profiles around restrained bulk water

As described by Merzel and Smith,[11] solvent density profiles around proteins reflect not only effects from water–protein interactions, but furthermore effects owing to the internal structure of water. In a similar spirit to the analysis by Merzel and Smith, we compared

solvent density profiles around xylanase with density profiles around bulk water atoms taking the approximate shape of xylanase. To this end, we superimposed an envelope built at the Van-der-Waals surface of xylanase with bulk water simulations, and selected all water molecules within this envelope (GROMACS-SWAXS output file: excludedvolume_0.pdb). These water molecules were restrained in the following by restraining the oxygen atoms with harmonic potentials (force constant 1000 kJ/mol/nm^2), henceforth referred to as restrained water (Fig. S8f, blue/white spheres). Next, simulations were set up as described in the Methods section in the main text, however, using the restrained water instead of the protein, and using the force field combinations c36–cTIP3P, ff15fb–TIP3P-FB, or ff99SBws–TIP4P/2005s. The geometries of restrained water molecules were constrained by using LINCS. Each system was simulated for 50 ns.

Density profiles (Fig. S8a) were computed from simulations of restrained water with the `gmx genenv` module of GROMACS-SWAXS, by using envelopes with 81920 triangular faces (option `-nrec 6`). The tool built a series of envelopes with increasing distances from the Van-der-Waals surface of the restrained water and computed the electron densities between pairs of adjacent envelopes.[12]

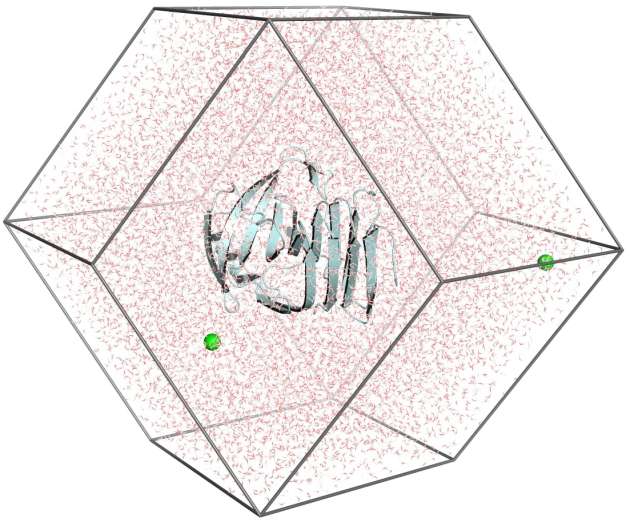
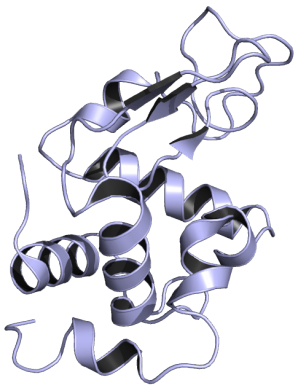
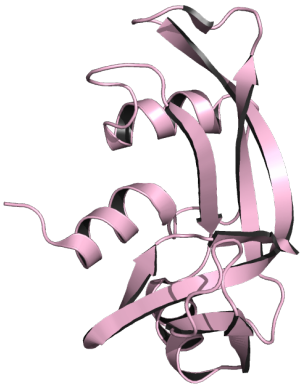


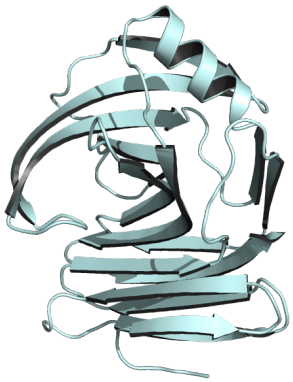
Figure S1: Simulation system of xylanase simulated with ff99SBws and TIP4P/2005s water in dodecahedral box. The protein is shown in cartoon representation, ions as spheres, and water as sticks. The box surfaces have a distance of at least 2 nm from the protein atoms.



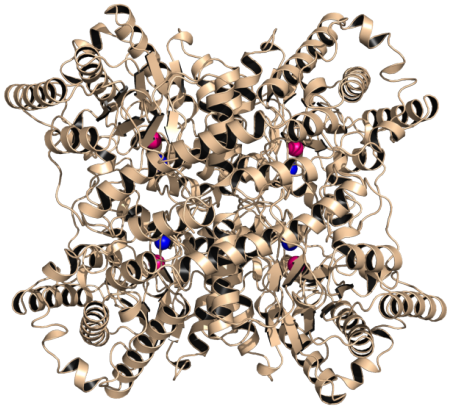
(a) Hen egg white lysozyme (PDB code 2VB1[13])



(b) Ribonuclease A (RNaseA; 7RSA[14])



(c) Xylanase (2DFC[15])



(d) Glucose isomerase. 1MNZ[16] after adding one methionine residue at the N- terminus. Ca^{2+} and Mg^{2+} ions inside the protein are shown as blue and pink spheres, respectively.

(e) Urate oxidase. 3L8W,[17] after adding six missing residues (sequence SLKSKL) at the C-terminus. Xanthine ligands are shown with blue sticks.

Figure S2: Cartoon representation of five proteins considered in this study (a) hen egg white lysozyme, (b) ribonuclease A (RNaseA), (c) xylanase , (d) glucose isomerase, and (e) urate oxidase.

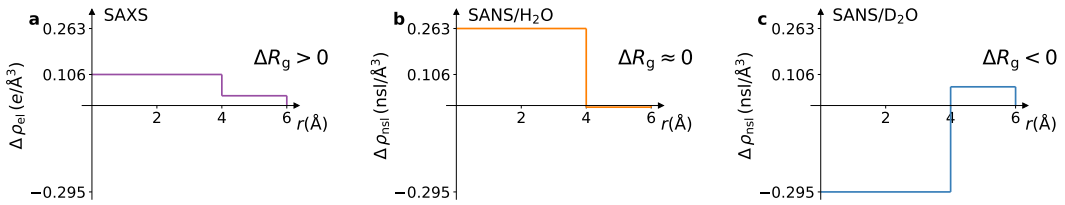


Figure S3: On the effect of the protein hydration shell on R_g as detected by SAXS, SANS in H_2O, or SANS in D_2O, illustrated with a hypothetical spherical protein with a (unitless) radius of 4 and a thickness of the hydration shell of 2. (a–c) Scattering contrast relative to bulk as function of distance r from the sphere center detected by (a) SAXS, (b) SANS/H_2O, and (c) SANS/D_2O experiments. (a) As detected by SAXS, both the protein and the hydration shell exhibit a positive electron density contrast relative to the bulk, resulting in an increased R_g owing to the hydration shell ($\Delta R_g > 0$). (b) During SANS/H_2O, the protein exhibits a positive contrast of the neutron scattering length density, while the contrast of the hydration shell is close to zero, leading to a small influence by the hydration shell on R_g ($\Delta R_g \approx 0$). (c) During SANS/D_2O, the protein exhibits a negative contrast whereas the hydration shell exhibits a positive contrast relative to bulk, resulting in a decreased R_g ($\Delta R_g < 0$).

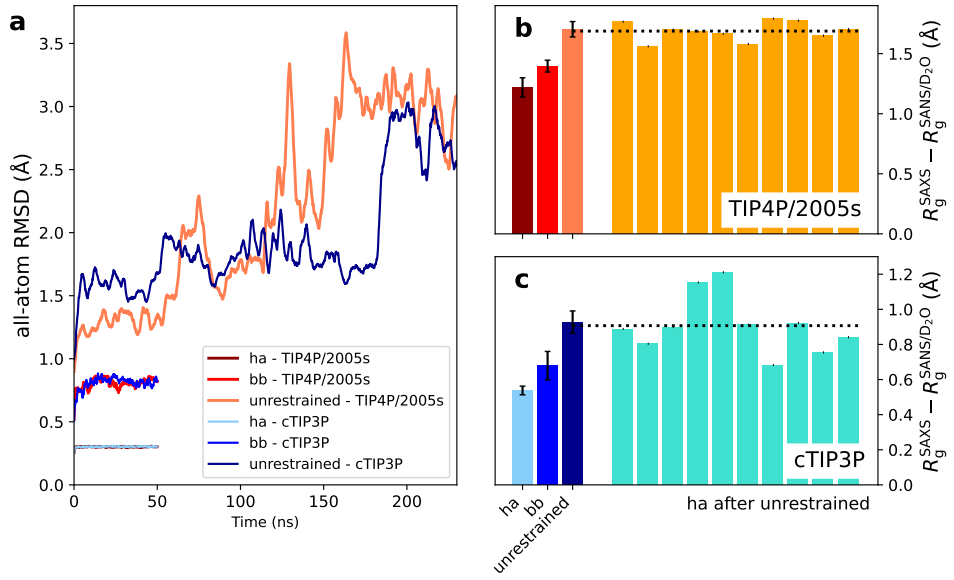


Figure S4: Effects of protein flexibility and hydration shell relaxation on R_g in simulations of xylanase. (a) All-atom RMSD with ff99SBws–TIP4P/2005s (brown shades) or CHARMM36m–cTIP3P (blue shades) with increasing flexibility. Simulations were carried out with restraints on heavy atoms (ha, see legend), with restraints on the backbone (bb), or without restraints (unrestrained). (b) Difference ΔR_g^{SAS} between R_g values from SAXS and SANS/D_2O for amber99SBws–TIP4P/2005s and (c) CHARMM36m–cTIP3P. Left three bars: ΔR_g^{SAS} from simulations with restraints on heavy atoms, restraints on the backbone, or from unrestrained simulations starting from the crystal structure (color code according to panel a), revealing an increased ΔR_g^{SAS} with increasing protein flexibility. Right ten bars: ΔR_g^{SAS} values from MD simulations with restraints on all heavy atoms starting from different times of the unrestrained simulation. Critically, the average ΔR_g^{SAS} values from these ten restrained simulations (dotted lines) agree with ΔR_g^{SAS} from the initial unrestrained simulation. Thus, not the protein flexibility per se, but instead the relaxation of protein–water packing during the initial unrestrained MD leads on average to increased ΔR_g^{SAS} values.

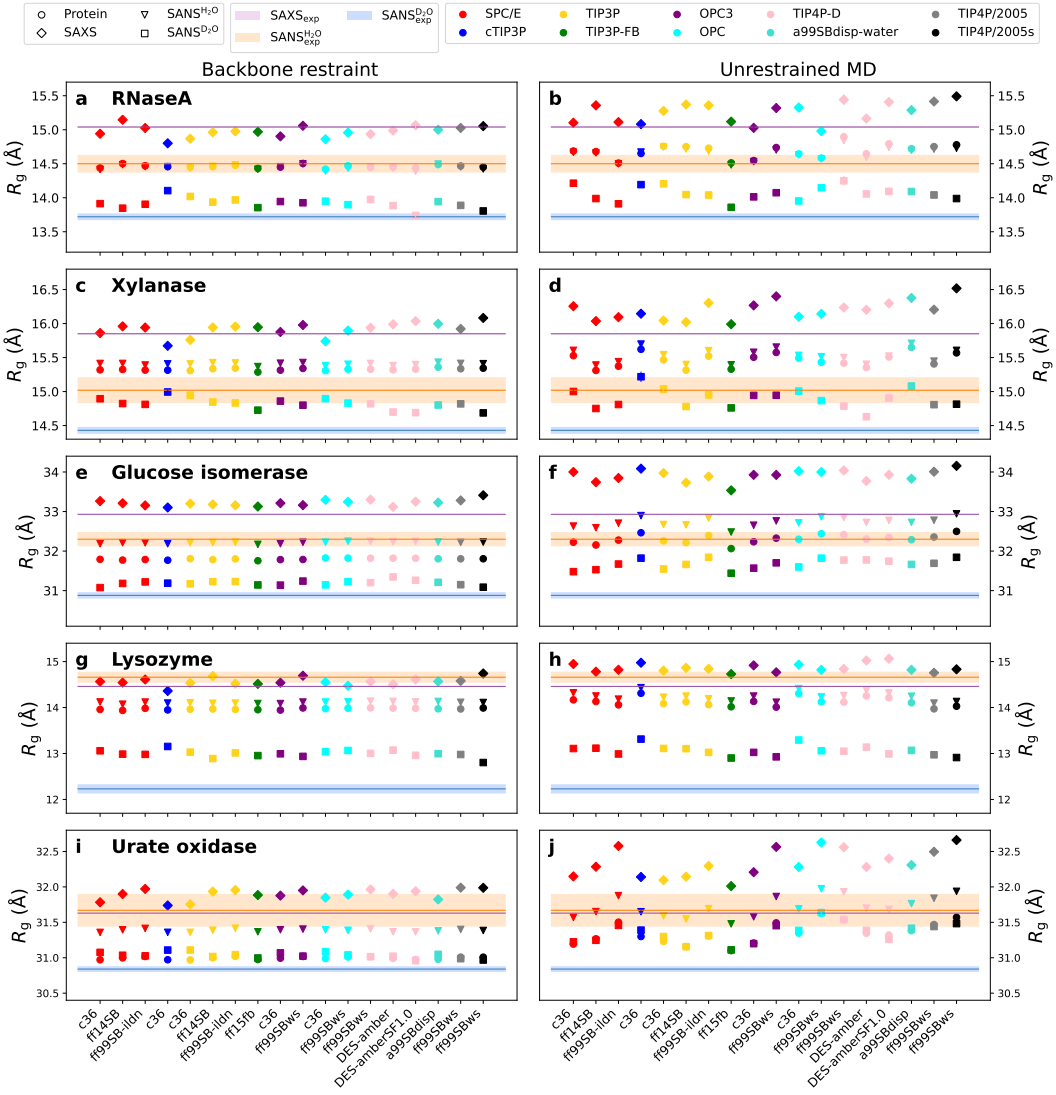


Figure S5: Absolute R_g values of (a/b) RNaseA, (c/d) xylanase, (e/f) glucose isomerase, (g/h) lysozyme, and (i/j) urate oxidase obtained from MD simulations with backbone restraints (left column) and without restraints (right column) with 18 different combinations of protein force fields (labels along the abscissa) and water models (color code, see legend). R_g^{SAS} values from unrestrained MD simulations for SAXS (diamonds), SANS/H_2O (triangles), and SANS/D_2O (squares). R_g of the pure protein (R_g^{Prot}) shown as circles. Experimental R_g^{SAS} values from $P(r)$ analysis are shown has horizontal lines for SAXS (purple), SANS/H_2O (orange), and SANS/D_2O blue. Shaded areas indicated experimental uncertainties of consensus SAS data.[1] Absolute R_g values shown here are not suitable for validating the hydration shell against experimental data because R_g of the bare protein (R_g^{Prot}) is not accurately known in the experiment.

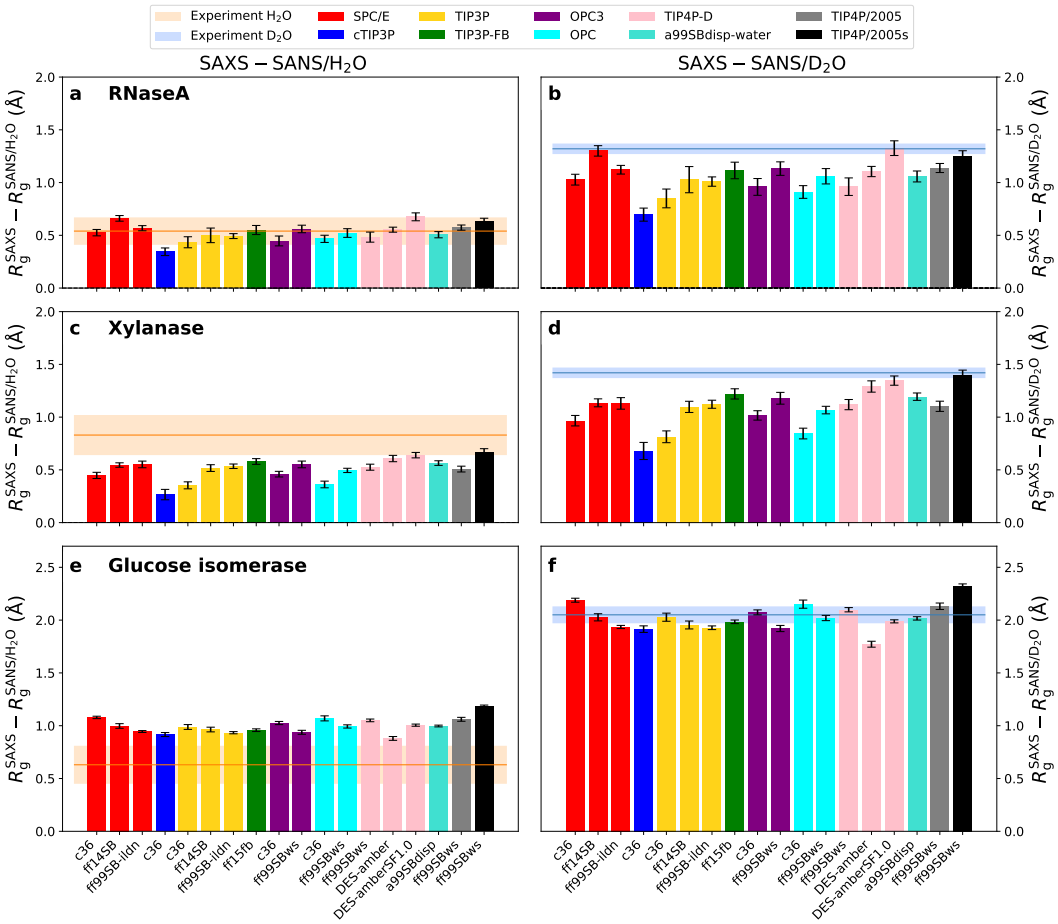


Figure S6: Same analysis as shown in Fig. 3, however based on simulations with backbone restraints instead of based on unrestrained simulations. Large variations of ΔR_g^{SAS} among different force field combinations are evident, even with identical backbone conformations. Thus, differences in protein–water interactions among different force fields and not different protein conformations in unrestrained simulations dominate the variations of ΔR_g^{SAS}. As a second finding, ΔR_g^{SAS} values from backbone-restrained simulations (this figure) are mostly smaller as compared to values from unrestrained simulations (Fig. 3), demonstrating that the equilibration of protein–water interactions during unrestrained simulations leads to a (slightly) increased packing of the hydration shell (see also Fig. S4).

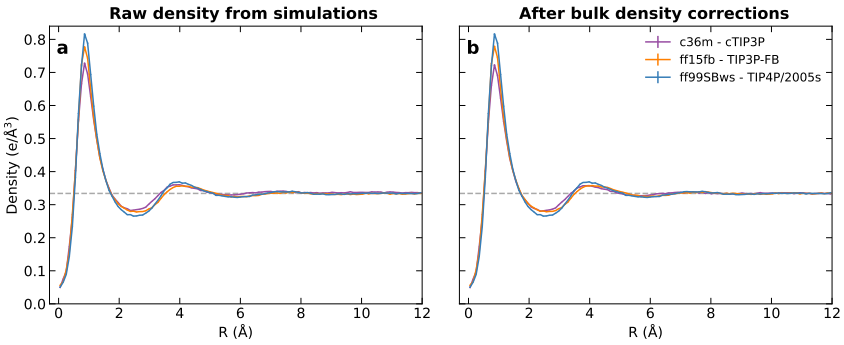


Figure S7: Typical solvent densities versus distance R from the Van-der-Waals surface and averaged over the protein surface, here shown for the protein xylanase obtained from simulations with c36–cTIP3P (purple), ff15fb–TIP3P-FB (orange), or ff99SBws–TIP4P/2005s (blue). (a) Densities taken from MD simulations or (b) densities scaled by a constant factor to correct the bulk density to 334 e/nm^{-3} (dashed horizontal line). The peaks representing the three hydration layers differ among the three combinations of protein force field and water model. The combination ff99SBws–TIP4P/2005s leads to the most pronounced hydration layer peaks, while c36–cTIP3P to the least pronounced peaks, which correlates with the effect of the hydration shells on R_g (compare with ΔR_g values in Fig. 2). Densities were computed with the gmx genenv module of GROMACS-SWAXS.

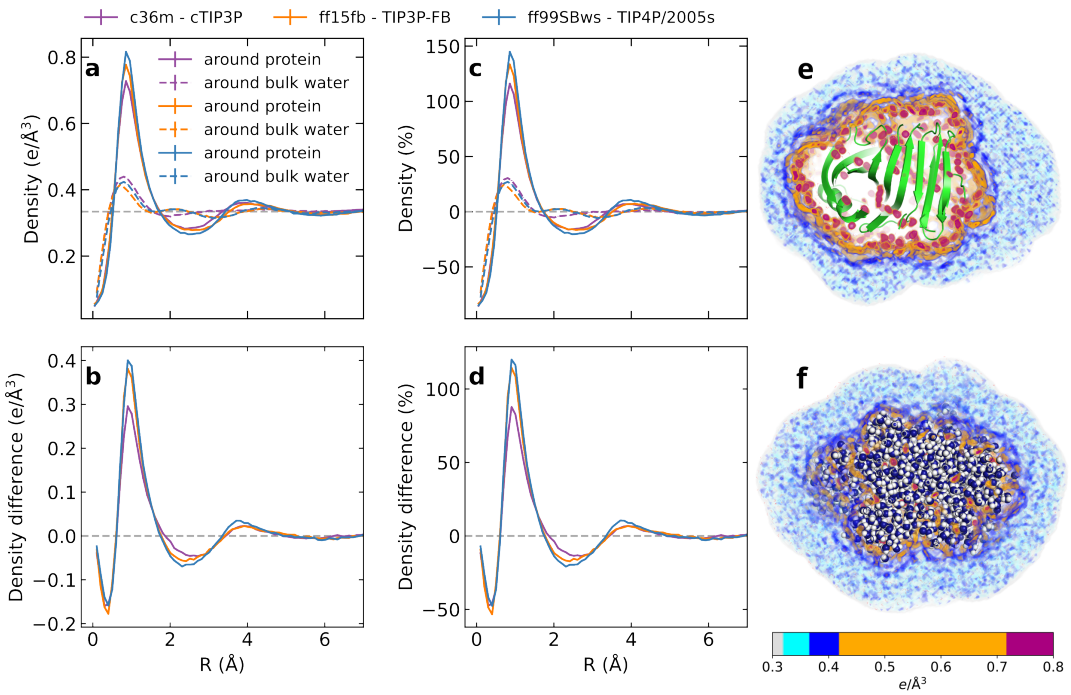


Figure S8: Solvent density modulations owing to water–protein interactions are by far larger as compared to density modulations owing to the internal structure of bulk water: (a) Solid lines: solvent density versus distance R from the Van-der-Waals surface of xylanase and averaged over the protein surface for c36–cTIP3P (purple), ff15fb–TIP3P-FB (orange), or ff99SBws–TIP4P/2005s (blue). Dashed colored lines: solvent density versus distance R from a volume of restrained bulk water atoms, taking approximately the shape of xylanase (see panels e and f). Dashed grey line: experimental bulk solvent density of $0.334\,e/\text{Å}^3$. (b) Density difference profiles of density around xylanase relative to density around bulk water, computed as the difference between the solid and dashed lines of panel a. The marked density difference manifests in an increased R_g as detected by SAXS. (c/d) Same analyses as in panels a/b, however, plotted as the relative difference of the density, similar to the analysis presented by Merzel and Smith.[11] (e) Electron density of solvent around xylanse with restrained heavy atoms (green cartoon) and (f) around bulk water with restrained oxygen atoms (blue/white spheres). Solvent densities are visualized in shades from light grey (bulk water) to blue to orange (see color bar), highlighting a by far more pronounced hydration shell around xylanase as compared to the solvent around bulk water. Simulations of the restrained solvent molecules inside the protein volume were performed as described in the Supplementary Methods. Densities were computed with the gmx genenv module of GROMACS-SWAXS.

Figure S9: Structures of RNaseA in simulations with CHARMM36m–OPC after equilibration (turquoise) and at ∼100 ns of unrestrained MD simulation (red). The N-terminal helix partly unfolded in the unrestrained simulation. In simulation with other force field combinations, the N-terminal helix exhibited some flexibility as well, however, the helix re-folded during the simulations. The conformational instability of RNaseA in CHARMM36m–OPC simulations may explain the the large ΔR_g^{SAS} value in unrestrained simulations of RNAseA (Fig. 3b, light turquoise bar labeled with c36), in contrast to all other unrestrained simulations with CHARMM36m (see Fig. 3) or all restrained simulations CHARMM36m (including CHARMM36m–OPC, see Fig. S6).

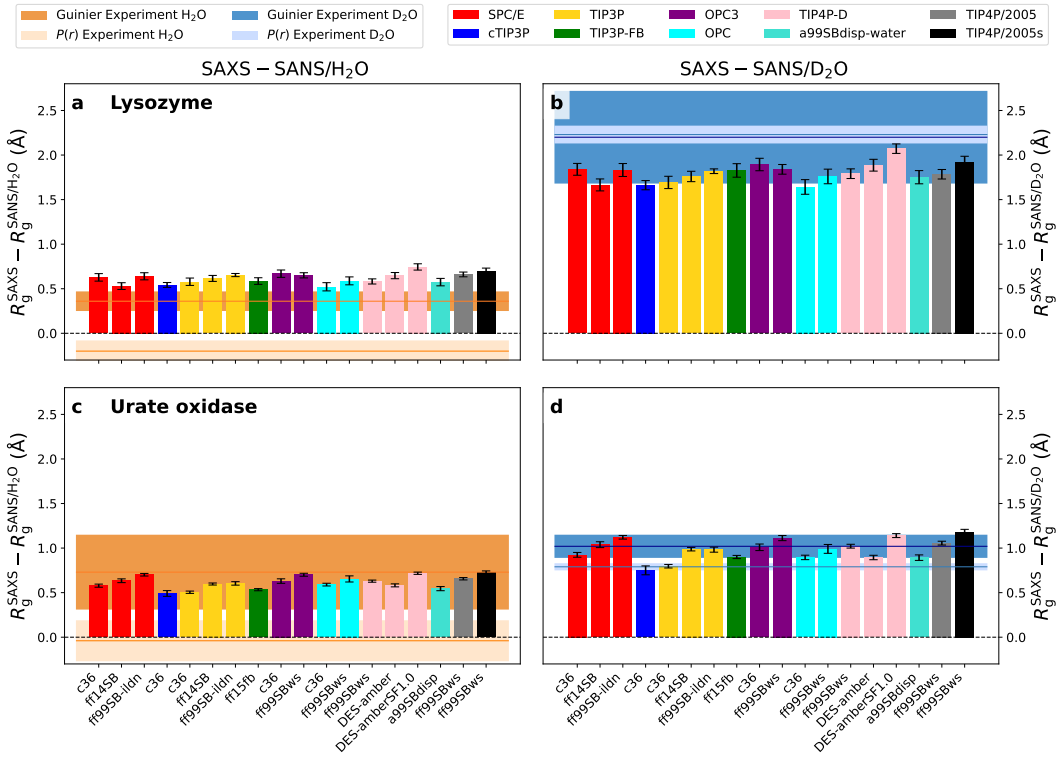


Figure S10: ΔR_g^{SAS} values of (a/b) lysozyme and (c/d) urate oxidase for SAXS relative to SANS/H_2O (left column) and SAXS relative to SANS/D_2O (right column). Experimental consensus data (horizontal lines) and reported uncertainties (shaded areas) from Guinier analysis (dark brown, dark blue) and from $P(r)$ analysis (light brown, light blue) are shown for reference. Experimental values for lysozyme and urate oxidase are subject to increased uncertainty, owing to problems with radiation damage and aggregation (see Supplementary Discussion).[1]

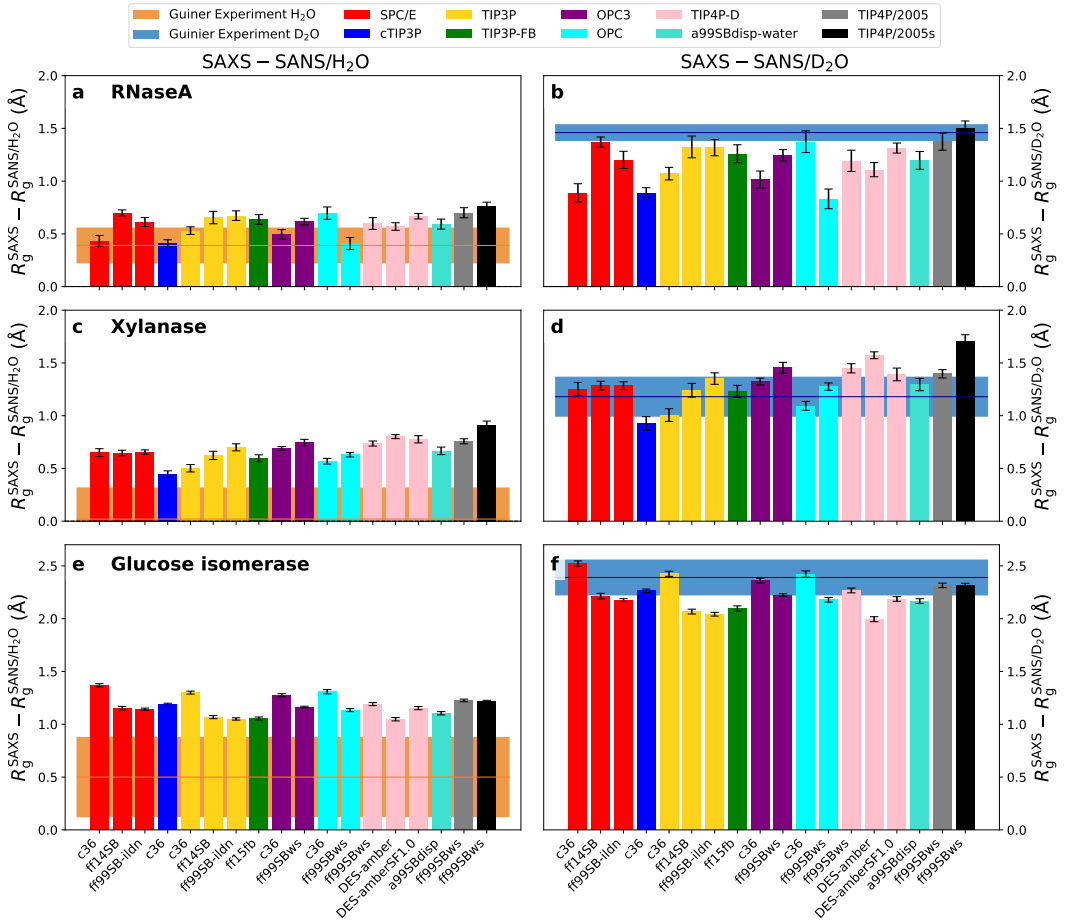


Figure S11: ΔR_g^{SAS} for (a/b) RNaseA, (c/d) xylanase, and (e/f) glucose isomerase, following the same labeling and color code as in Fig. 3. Experimental values obtained form Guinier analysis are shown as horizontal lines with uncertainties as shaded areas,[1] in contrast to experimental values from the $P(r)$ function in Fig. 3. The larger experimental errors as compared to the results from $P(r)$ analysis and the slightly poorer agreement between simulations and experiment (compared to Fig. 3) are rationalized by the fact that Guinier analysis is more sensitive to small amounts of undetected protein–protein aggregation.[1] Left column: ΔR_g^{SAS} from SAXS relative to SANS/H_2O revealing poor agreement between simulation and experiment, likely caused by poorer signal-to-noise ratio during SANS/H_2O experiments owing to increased incoherent scattering in H_2O as compared to D_2O and lower contrast.

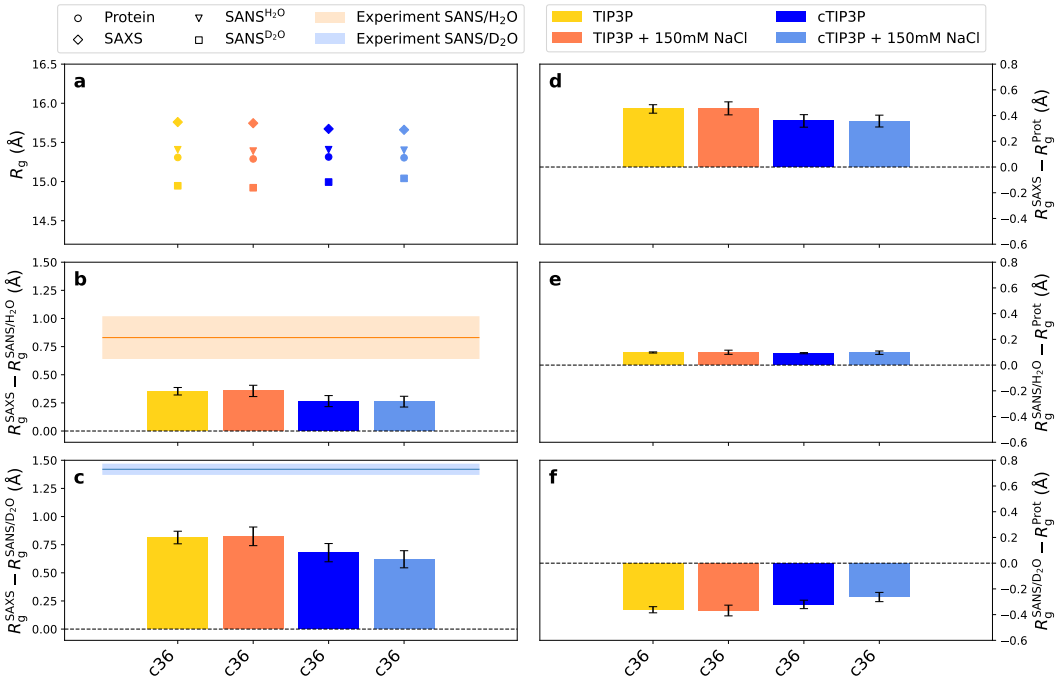


Figure S12: On the effect of 150 mM NaCl on SAS-derived R_g values of xylanase as compared to simulations with purely counter ions. Simulations were carried out with Charmm36m combined with three different water models using purely counter ions (SPC/E, TIP3P, or cTIP3P, see legend for color code) or with 150 mM NaCl (denoted SPC/E-ions, TIP3P-ions, cTIP3P-ions, see legend). (a) SAS-derived R_g^{SAS} computed from unrestrained MD simulations for SAXS (diamonds), SANS/H_2O (triangles), and SANS/D_2O (squares). R_g of the pure protein (R_g^{Prot}) shown as circles. Experimental R_g^{SAS} values from $P(r)$ analysis are shown has horizontal lines for SAXS (purple), SANS/H_2O (orange), and SANS/D_2O (blue). (b) Difference between R_g values from SAXS and SANS/H_2O as well as (c) between SAXS and SANS/D_2O. (d) Difference between SAS-derived R_g values from explicit-solvent MD and R_g^{Prot} values for SAXS,(e) for SANS/H_2O, and (f) for SANS/D_2O. Statistical errors (1 SE) were obtained from block averaging. Horizontal lines and shaded areas indicated experimental consensus values and uncertainties.[1] Adding 150 mM NaCl to the buffer has only a small effect on the ΔR_g values of the only slightly charged xylanase.

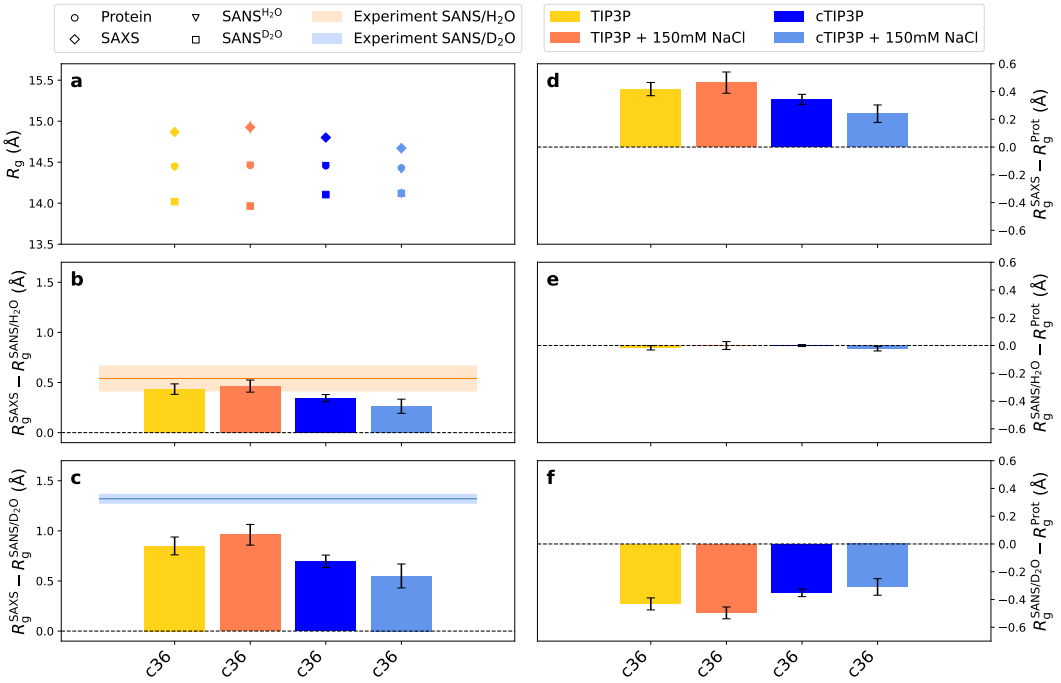


Figure S13: On the effect of 150 mM NaCl on SAS-derived R_g values of RNaseA. Same analysis and as shown for xylanase in Fig. S12.

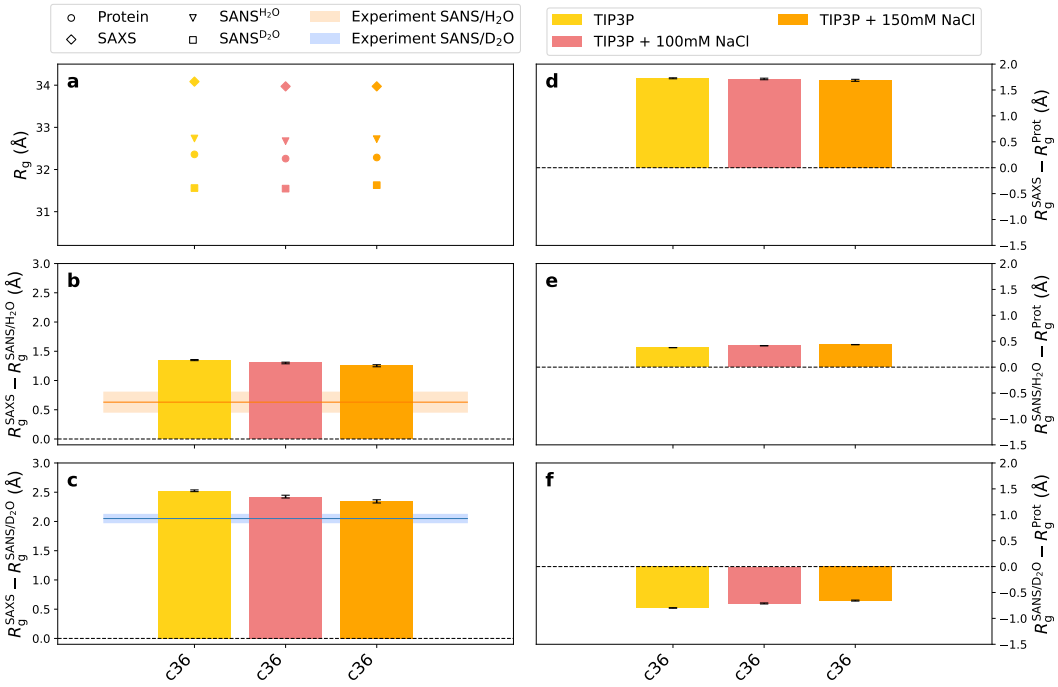


Figure S14: On the effect of 100 mM NaCl (pink) or 150 mM NaCl (orange) on SAS-derived R_g values of the highly anionic glucose isomerase ($-60\,e$), as compared to using only counter ions (yellow). Same analysis as shown for xylanase and RNaseA in Figs. S12 and S13, however restricted to TIP3P water.The lower the ion concentration in the buffer, the larger the Debye length and thus the effect of the hydration shell on the R_g.

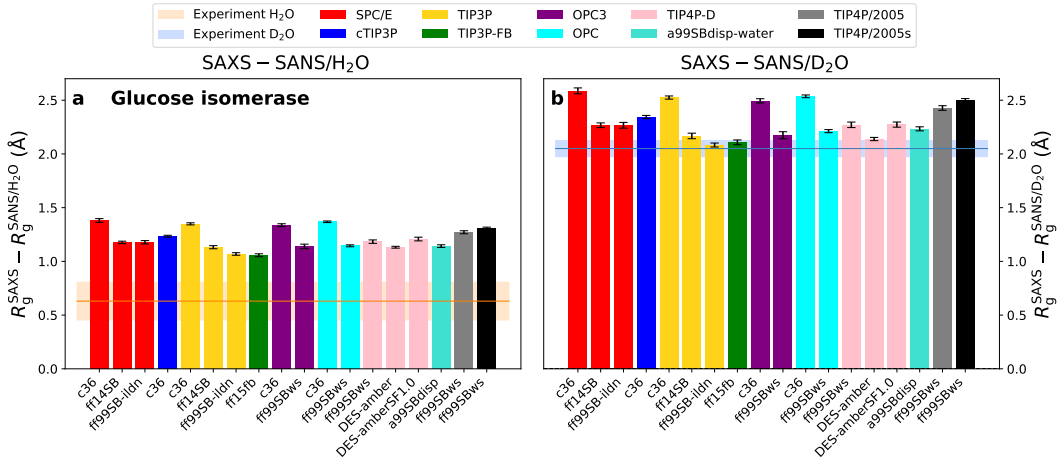


Figure S15: On the importance of salt for accurate R_g predictions of highly anionic glucose isomerase. (a) ΔR_g^{SAS} values from SAXS relative to SANS/H_2O and (b) of SAXS relative to SANS/D_2O from unrestrained simulations of glucose isomerase with only counter ions (60 Na^+) but no additional salt. Experimental consensus values and uncertainties from $P(r)$ analysis are shown as horizontal lines and shaded areas, respectively. By comparison with Fig. 3e/f, the lack of salt increases ΔR_g^{SAS} systematically, caused by an overly extended counter ion cloud. In turn, in presence of salt, the Debye length decreases, thus leading to a spatially more compact counter ion cloud and slightly smaller ΔR_g^{SAS} values (compare with Fig. 3).

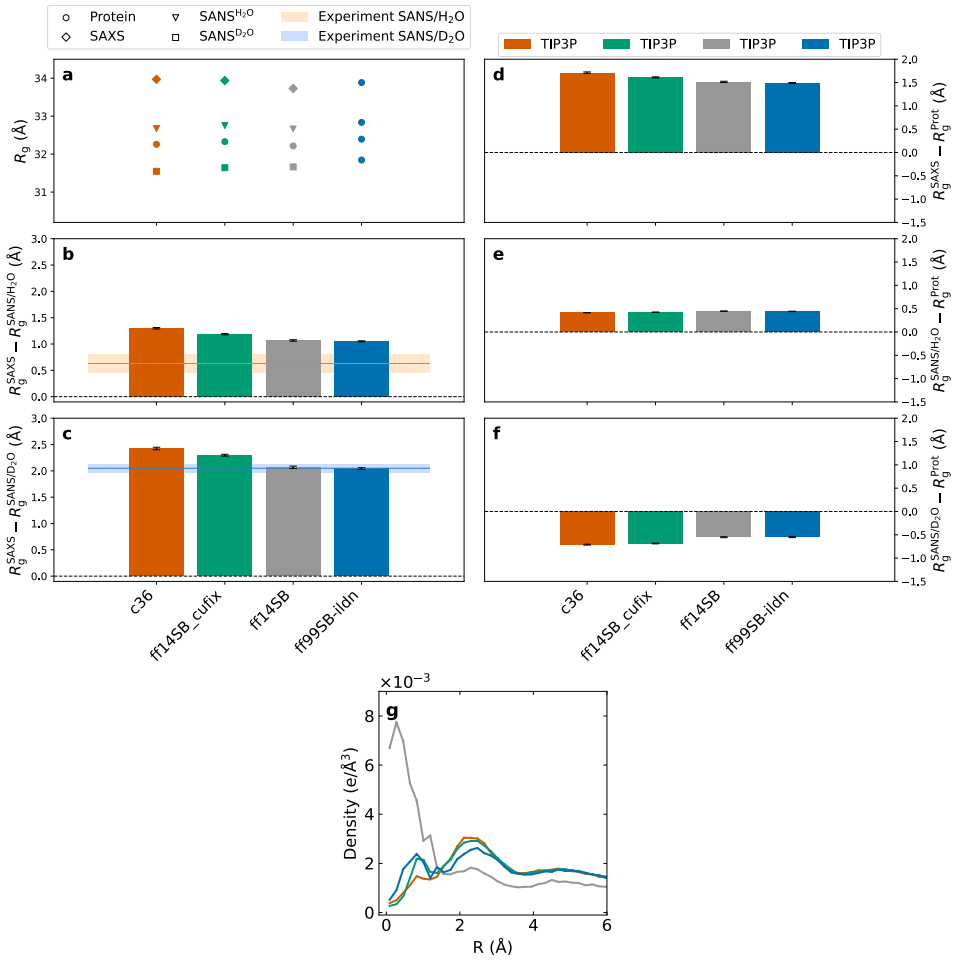


Figure S16: On the effect of non-bonded fix parameters (CUFIX, also called NBFIX) with the Amber14SB protein force field. CUFIX implements modified Lennard-Jones interactions between Na^+ ions and carboxylate oxygen atoms of aspartate and glutamate residues with the aim to avoid Na^+–COO^- overbinding.[7] CHARMM36m implements CUFIX by default in the version of July 2020 as used here. (a) R_g, (b/c) ΔR_g^{SAS}, and (d–f) ΔR_g values are reported for the highly anionic glucose isomerase (GI, $-60\,e$). Unrestrained simulations of GI were carried out with CHARMM36m (yellow bar), ff14SB with CUFIX (pink), as well as ff14SB or ff99SB-ildn without CUFIX (grey and blue, respectively). Experimental consensus values and uncertainties from $P(r)$ analysis are shown as horizontal lines and shaded areas, respectively. Statistical errors (1 SE) were obtained from block averaging. Upon refining Na^+–carboxylate interactions by ff14SB with CUFIX, ΔR_g^{SAS} increased by only ∼0.2 Å and ∼0.1Å for SANS in D_2O and H_2O, respectively. (g) Density of Na^+ as function of distance from the protein surface for four different force fields (color code according to panels a–f). CUFIX greatly reduces Na^+ overbinding in ff14SB simulations.

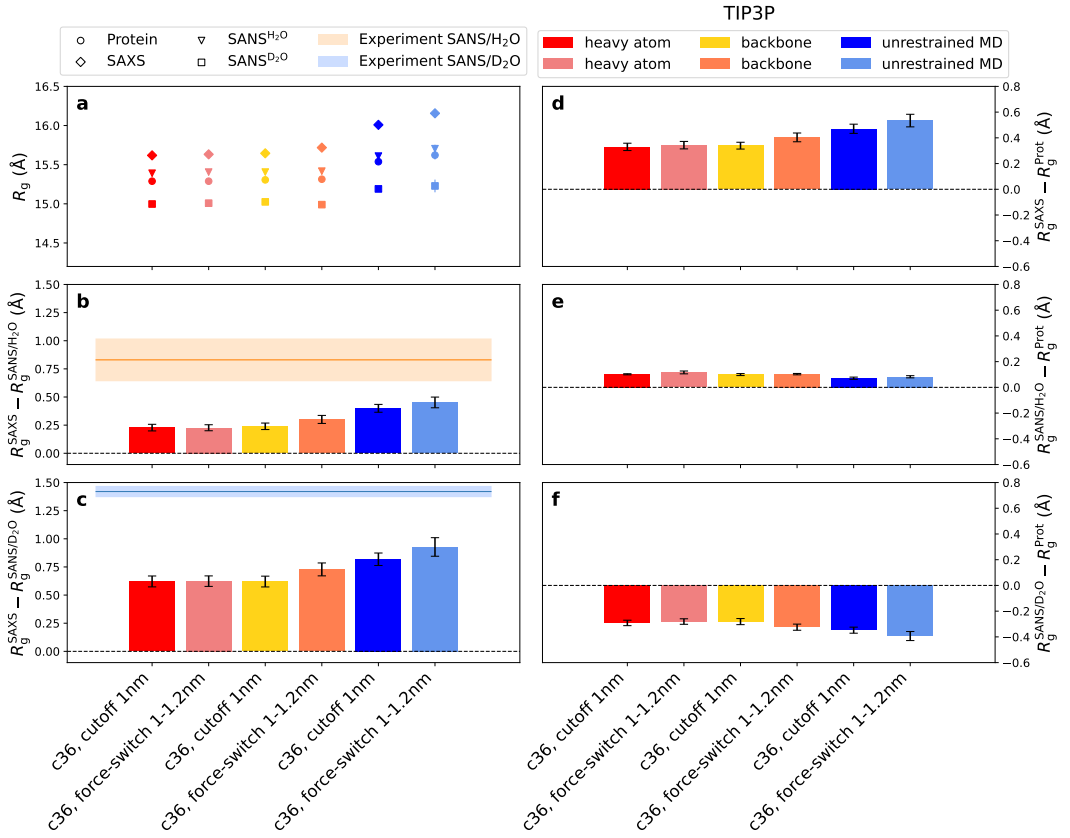


Figure S17: On the role of force field-specific Lennard-Jones (LJ) cutoff settings: (a) R_g, (b/c) ΔR_g^{SAS}, and (d–f) ΔR_g values for xylanase using the CHARMM36m–TIP3P force field combination. Simulations were carried out either using a plain LJ cutoff at 1.0 nm or by switching off the LJ forces between 1 nm and 1.2 nm, as recommended for the CHARMM36m force field (force-switch 1–1.2 nm). Results are shown for simulations with restrained heavy atoms (red, pink), restrained backbone (yellow, orange) or for unrestrained MD (dark and light blue). In simulations with flexible side chains (unrestrained or backbone-restrained MD), cutoff settings following CHARMM standards leads to larger ΔR_g^{SAS} values as compared to using a plain 1 nm cutoff (panels b/c, orange vs. yellow, light blue vs. dark blue). In simulations with restrained heavy atoms, using a different cutoff stetting has only a marginal effect (pink vs. red). Thus, the additional LJ interactions between 1 nm and 1.2 nm lead to increased ΔR_g^{SAS} values owing to more densely packed water structures around flexible protein side chains, and not owing to stronger attraction of water onto a (fixed) protein surface.

Table S1: Eighteen combinations of protein force field and water model used in this study. CHARMM36m was taken from version of July 2020.

Protein force field	Abbreviation	Water model	Refs.
CHARMM36m	c36	cTIP3P	18, 19
		TIP3P	20
		SPC/E	21
		OPC3	22
		OPC	23
Amber14SB	ff14SB	SPC/E	24 , 21
		TIP3P	20
Amber99SB-ildn	ff99SB-ildn	SPC/E	25, 21
		TIP3P	20
Amber15/force-balance	ff15fb	TIP3P-FB	26
Amber99SBws	ff99SBws	TIP4P/2005s	27, 28
		TIP4P/2005	29
		TIP4P-D	30
		OPC3	22
		OPC	23
DES-amber	DES-amber	TIP4P-D	31, 30
DES-amber without scaled charges	DES-amberSF1.0	TIP4P-D	31, 30
Amberr99SBdisp	a99SBdisp	a99SBdisp-water	32

Table S2: Experimental consensus R_g values for SAXS, SANS/H_2O, and SANS/D_2O from Trewhella et al.[1] R_g are shown as obtained from Guinier or $P(r)$ analysis. Two right columns: ΔR_g^{SAS} values from SAXS relative to SANS/H_2O and SAXS relative to SANS/D_2O, respectively.

Protein	Type of R_g	R_g^{SAXS} (Å)	R_g^{SANS/H_2O} (Å)	R_g^{SANS/D_2O} (Å)	$R_g^{SAXS} - R_g^{SANS/H_2O}$ (Å)	$R_g^{SAXS} - R_g^{SANS/D_2O}$ (Å)
RNaseA	Guinier	15.13 ± 0.02	14.74 ± 0.17	13.67 ± 0.08	0.39 ± 0.17	1.46 ± 0.08
	$P(r)$	15.04 ± 0.01	14.50 ± 0.13	13.72 ± 0.05	0.54 ± 0.13	1.32 ± 0.05
Xylanase	Guinier	16.05 ± 0.01	16.03 ± 0.30	14.87 ± 0.19	0.02 ± 0.3	1.18 ± 0.19
	$P(r)$	15.85 ± 0.01	15.02 ± 0.19	14.43 ± 0.05	0.83 ± 0.19	1.42 ± 0.05
Glucose isomerase	Guinier	33.11 ± 0.05	32.61 ± 0.38	30.72 ± 0.17	0.5 ± 0.38	2.39 ± 0.17
	$P(r)$	32.93 ± 0.01	32.30 ± 0.18	30.88 ± 0.08	0.63 ± 0.18	2.05 ± 0.08
Lysozyme	Guinier	14.64 ± 0.05	14.28 ± 0.12	12.44 ± 0.52	0.36 ± 0.11	2.2 ± 0.52
	$P(r)$	14.46 ± 0.01	14.66 ± 0.12	12.23 ± 0.10	-0.2 ± 0.12	2.23 ± 0.1
Urate Oxidase	Guinier	32.30 ± 0.06	31.57 ± 0.42	31.18 ± 0.12	0.73 ± 0.42	1.02 ± 0.13
	$P(r)$	31.63 ± 0.01	31.67 ± 0.23	30.84 ± 0.04	-0.4 ± 0.23	0.79 ± 0.04

force field	water model	R_g^{Prot} (Å)	R_g^{SAXS} (Å)	R_g^{SANS/H_2O} (Å)	R_g^{SANS/D_2O} (Å)	$R_g^{SAXS} - R_g^{Prot}$ (Å)	$R_g^{SANS/H_2O} - R_g^{Prot}$ (Å)	$R_g^{SANS/D_2O} - R_g^{Prot}$ (Å)	$R_g^{SAXS} - R_g^{SANS/H_2O}$ (Å)	$R_g^{SAXS} - R_g^{SANS/D_2O}$ (Å)
c36	SPC/E	14.69 ± 0.03	15.10 ± 0.07	14.67 ± 0.03	14.21 ± 0.04	0.42 ± 0.06	-0.02 ± 0.02	-0.47 ± 0.04	0.43 ± 0.06	0.89 ± 0.09
ff14SB	SPC/E	14.68 ± 0.03	15.36 ± 0.04	14.66 ± 0.03	13.99 ± 0.04	0.68 ± 0.03	-0.02 ± 0.01	-0.69 ± 0.03	0.70 ± 0.03	1.37 ± 0.05
ff99SB-ildn	SPC/E	14.51 ± 0.02	15.11 ± 0.06	14.50 ± 0.03	13.91 ± 0.05	0.60 ± 0.05	-0.01 ± 0.02	-0.60 ± 0.04	0.61 ± 0.05	1.20 ± 0.09
c36	cTIP3P	14.65 ± 0.05	15.08 ± 0.06	14.67 ± 0.05	14.19 ± 0.04	0.43 ± 0.04	0.02 ± 0.02	-0.46 ± 0.03	0.41 ± 0.04	0.89 ± 0.05
c36	TIP3P	14.76 ± 0.03	15.28 ± 0.05	14.75 ± 0.03	14.21 ± 0.04	0.52 ± 0.04	-0.01 ± 0.02	-0.55 ± 0.04	0.53 ± 0.04	1.07 ± 0.06
ff14SB	TIP3P	14.75 ± 0.02	15.37 ± 0.07	14.72 ± 0.02	14.05 ± 0.05	0.62 ± 0.06	-0.03 ± 0.01	-0.70 ± 0.05	0.65 ± 0.06	1.32 ± 0.11
ff99SB-ildn	TIP3P	14.73 ± 0.02	15.36 ± 0.05	14.69 ± 0.02	14.04 ± 0.04	0.63 ± 0.05	-0.04 ± 0.02	-0.69 ± 0.04	0.67 ± 0.05	1.32 ± 0.08
ff15fb	TIP3P-FB	14.51 ± 0.02	15.12 ± 0.06	14.48 ± 0.02	13.86 ± 0.04	0.61 ± 0.05	-0.03 ± 0.02	-0.65 ± 0.05	0.64 ± 0.05	1.26 ± 0.09
c36	OPC3	14.55 ± 0.01	15.03 ± 0.05	14.53 ± 0.02	14.01 ± 0.05	0.48 ± 0.05	-0.02 ± 0.02	-0.53 ± 0.04	0.50 ± 0.05	1.01 ± 0.09
ff99SBws	OPC3	14.74 ± 0.03	15.32 ± 0.04	14.70 ± 0.03	14.07 ± 0.03	0.58 ± 0.04	-0.03 ± 0.02	-0.66 ± 0.03	0.62 ± 0.04	1.25 ± 0.06
c36	OPC	14.65 ± 0.02	15.33 ± 0.07	14.63 ± 0.02	13.95 ± 0.05	0.68 ± 0.06	-0.02 ± 0.02	-0.69 ± 0.06	0.70 ± 0.06	1.37 ± 0.11
ff99SBws	OPC	14.58 ± 0.04	14.98 ± 0.08	14.57 ± 0.03	14.15 ± 0.04	0.40 ± 0.06	-0.01 ± 0.01	-0.44 ± 0.04	0.41 ± 0.06	0.83 ± 0.10
ff99SBws	TIP4P-D	14.90 ± 0.06	15.44 ± 0.08	14.84 ± 0.06	14.25 ± 0.08	0.55 ± 0.06	-0.05 ± 0.02	-0.65 ± 0.05	0.60 ± 0.06	1.19 ± 0.11
DES-amber	TIP4P-D	14.65 ± 0.02	15.17 ± 0.05	14.60 ± 0.02	14.06 ± 0.03	0.52 ± 0.05	-0.05 ± 0.02	-0.59 ± 0.03	0.57 ± 0.04	1.11 ± 0.07
DES-amberSF1.0	TIP4P-D	14.79 ± 0.03	15.41 ± 0.04	14.74 ± 0.02	14.09 ± 0.03	0.61 ± 0.04	-0.05 ± 0.01	-0.70 ± 0.02	0.67 ± 0.03	1.31 ± 0.05
a99SBdisp	a99SBdisp-water	14.72 ± 0.02	15.29 ± 0.06	14.70 ± 0.02	14.09 ± 0.04	0.57 ± 0.05	-0.03 ± 0.01	-0.63 ± 0.04	0.59 ± 0.05	1.20 ± 0.09
ff99SBws	TIP4P/2005	14.75 ± 0.02	15.42 ± 0.05	14.72 ± 0.02	14.04 ± 0.05	0.66 ± 0.05	-0.04 ± 0.02	-0.71 ± 0.05	0.70 ± 0.05	1.37 ± 0.09
ff99SBws	TIP4P/2005s	14.78 ± 0.03	15.49 ± 0.05	14.73 ± 0.03	13.99 ± 0.04	0.71 ± 0.05	-0.05 ± 0.02	-0.79 ± 0.03	0.76 ± 0.04	1.50 ± 0.07

Table S3: R_g, ΔR_g, and ΔR_g^{SAS} values from unrestrained MD simulations of 18 different combinations of protein force fields and water models for the protein RNaseA.

force field	water model	R_g^{Prot} (Å)	R_g^{SAXS} (Å)	R_g^{SANS/H_2O} (Å)	R_g^{SANS/D_2O} (Å)	$R_g^{SAXS} - R_g^{Prot}$ (Å)	$R_g^{SANS/H_2O} - R_g^{Prot}$ (Å)	$R_g^{SANS/D_2O} - R_g^{Prot}$ (Å)	$R_g^{SAXS} - R_g^{SANS/H_2O}$ (Å)	$R_g^{SAXS} - R_g^{SANS/D_2O}$ (Å)
c36	SPC/E	15.53 ± 0.02	16.26 ± 0.05	15.61 ± 0.02	15.00 ± 0.03	0.73 ± 0.04	0.08 ± 0.01	-0.53 ± 0.03	0.65 ± 0.04	1.25 ± 0.07
ff14SB	SPC/E	15.31 ± 0.01	16.04 ± 0.03	15.39 ± 0.02	14.75 ± 0.03	0.73 ± 0.03	0.08 ± 0.01	-0.56 ± 0.02	0.65 ± 0.03	1.29 ± 0.05
ff99SB-ildn	SPC/E	15.37 ± 0.01	16.09 ± 0.03	15.44 ± 0.01	14.81 ± 0.02	0.72 ± 0.03	0.07 ± 0.01	-0.56 ± 0.02	0.65 ± 0.03	1.28 ± 0.04
c36	cTIP3P	15.62 ± 0.06	16.15 ± 0.03	15.70 ± 0.05	15.22 ± 0.08	0.52 ± 0.05	0.08 ± 0.02	-0.40 ± 0.03	0.45 ± 0.04	0.93 ± 0.07
c36	TIP3P	15.47 ± 0.04	16.04 ± 0.06	15.54 ± 0.04	15.04 ± 0.04	0.58 ± 0.04	0.08 ± 0.01	-0.43 ± 0.03	0.50 ± 0.04	1.01 ± 0.06
ff14SB	TIP3P	15.32 ± 0.01	16.02 ± 0.04	15.40 ± 0.01	14.78 ± 0.03	0.71 ± 0.05	0.08 ± 0.01	-0.54 ± 0.03	0.62 ± 0.04	1.24 ± 0.07
ff99SB-ildn	TIP3P	15.52 ± 0.02	16.30 ± 0.04	15.60 ± 0.02	14.95 ± 0.03	0.78 ± 0.04	0.08 ± 0.01	-0.57 ± 0.03	0.70 ± 0.04	1.35 ± 0.06
ff15fb	TIP3P-FB	15.33 ± 0.02	15.99 ± 0.04	15.39 ± 0.02	14.76 ± 0.03	0.66 ± 0.04	0.06 ± 0.01	-0.57 ± 0.03	0.60 ± 0.04	1.23 ± 0.06
c36	OPC3	15.50 ± 0.02	16.27 ± 0.04	15.58 ± 0.03	14.94 ± 0.03	0.76 ± 0.03	0.07 ± 0.01	-0.56 ± 0.02	0.69 ± 0.02	1.32 ± 0.04
ff99SBws	OPC3	15.57 ± 0.04	16.40 ± 0.04	15.65 ± 0.03	14.95 ± 0.05	0.82 ± 0.03	0.08 ± 0.01	-0.63 ± 0.03	0.75 ± 0.03	1.45 ± 0.06
c36	OPC	15.49 ± 0.02	16.10 ± 0.04	15.53 ± 0.02	15.01 ± 0.03	0.61 ± 0.03	0.04 ± 0.01	-0.49 ± 0.02	0.57 ± 0.03	1.09 ± 0.05
ff99SBws	OPC	15.43 ± 0.02	16.14 ± 0.03	15.51 ± 0.02	14.87 ± 0.03	0.71 ± 0.03	0.08 ± 0.01	-0.56 ± 0.02	0.63 ± 0.03	1.28 ± 0.04
ff99SBws	TIP4P-D	15.42 ± 0.02	16.24 ± 0.04	15.50 ± 0.02	14.79 ± 0.03	0.82 ± 0.03	0.08 ± 0.01	-0.63 ± 0.03	0.74 ± 0.03	1.45 ± 0.05
DES-amber	TIP4P-D	15.35 ± 0.02	16.20 ± 0.03	15.40 ± 0.02	14.63 ± 0.03	0.85 ± 0.02	0.05 ± 0.01	-0.72 ± 0.02	0.80 ± 0.02	1.57 ± 0.04
DES-amberSF1.0	TIP4P-D	15.51 ± 0.02	16.30 ± 0.04	15.52 ± 0.02	14.91 ± 0.04	0.79 ± 0.04	0.01 ± 0.02	-0.60 ± 0.03	0.78 ± 0.04	1.39 ± 0.07
a99SBdisp	a99SBdisp-water	15.65 ± 0.04	16.38 ± 0.07	15.71 ± 0.04	15.08 ± 0.04	0.73 ± 0.04	0.06 ± 0.01	-0.57 ± 0.03	0.67 ± 0.04	1.30 ± 0.06
ff99SBws	TIP4P/2005	15.41 ± 0.03	16.20 ± 0.05	15.45 ± 0.03	14.81 ± 0.02	0.80 ± 0.03	0.04 ± 0.01	-0.60 ± 0.02	0.76 ± 0.03	1.40 ± 0.04
ff99SBws	TIP4P/2005s	15.57 ± 0.03	16.52 ± 0.05	15.61 ± 0.03	14.81 ± 0.05	0.95 ± 0.04	0.04 ± 0.01	-0.75 ± 0.03	0.91 ± 0.04	1.70 ± 0.07

Table S4: R_g, ΔR_g, and ΔR_g^{SAS} values from unrestrained MD simulations of 18 different combinations of protein force fields and water models for the protein xylanase.

force field	water model	R_g^{Prot} (Å)	R_g^{SAXS} (Å)	R_g^{SANS/H_2O} (Å)	R_g^{SANS/D_2O} (Å)	$R_g^{SAXS} - R_g^{Prot}$ (Å)	$R_g^{SANS/H_2O} - R_g^{Prot}$ (Å)	$R_g^{SANS/D_2O} - R_g^{Prot}$ (Å)	$R_g^{SAXS} - R_g^{SANS/H_2O}$ (Å)	$R_g^{SAXS} - R_g^{SANS/D_2O}$ (Å)
c36	SPC/E	32.22 ± 0.01	34.00 ± 0.01	32.63 ± 0.01	31.48 ± 0.02	1.78 ± 0.02	0.41 ± 0.01	-0.75 ± 0.02	1.37 ± 0.02	2.52 ± 0.03
ff14SB	SPC/E	32.15 ± 0.01	33.74 ± 0.02	32.59 ± 0.01	31.53 ± 0.02	1.59 ± 0.03	0.44 ± 0.01	-0.63 ± 0.02	1.15 ± 0.02	2.21 ± 0.03
ff99SB-ildn	SPC/E	32.28 ± 0.01	33.85 ± 0.02	32.71 ± 0.01	31.67 ± 0.01	1.57 ± 0.02	0.43 ± 0.01	-0.61 ± 0.01	1.14 ± 0.02	2.18 ± 0.02
c36	cTIP3P	32.46 ± 0.02	34.08 ± 0.02	32.90 ± 0.02	31.82 ± 0.02	1.62 ± 0.01	0.43 ± 0.01	-0.64 ± 0.01	1.19 ± 0.02	2.26 ± 0.02
c36	TIP3P	32.26 ± 0.01	33.97 ± 0.02	32.67 ± 0.01	31.55 ± 0.02	1.71 ± 0.02	0.41 ± 0.01	-0.71 ± 0.02	1.30 ± 0.02	2.43 ± 0.03
ff14SB	TIP3P	32.22 ± 0.01	33.73 ± 0.02	32.66 ± 0.01	31.66 ± 0.02	1.51 ± 0.02	0.45 ± 0.01	-0.55 ± 0.02	1.07 ± 0.02	2.07 ± 0.03
ff99SB-ildn	TIP3P	32.39 ± 0.01	33.89 ± 0.01	32.84 ± 0.01	31.84 ± 0.02	1.49 ± 0.01	0.44 ± 0.01	-0.55 ± 0.02	1.05 ± 0.02	2.04 ± 0.02
ff15fb	TIP3P-FB	32.06 ± 0.01	33.54 ± 0.02	32.48 ± 0.01	31.44 ± 0.01	1.47 ± 0.02	0.42 ± 0.01	-0.62 ± 0.02	1.05 ± 0.02	2.10 ± 0.03
c36	OPC3	32.24 ± 0.01	33.93 ± 0.02	32.65 ± 0.01	31.57 ± 0.01	1.69 ± 0.02	0.42 ± 0.01	-0.67 ± 0.01	1.27 ± 0.02	2.36 ± 0.03
ff99SBws	OPC3	32.33 ± 0.01	33.93 ± 0.01	32.76 ± 0.01	31.70 ± 0.01	1.60 ± 0.01	0.44 ± 0.01	-0.62 ± 0.01	1.16 ± 0.01	2.22 ± 0.02
c36	OPC	32.30 ± 0.01	34.02 ± 0.03	32.71 ± 0.01	31.60 ± 0.02	1.72 ± 0.03	0.41 ± 0.01	-0.70 ± 0.02	1.31 ± 0.03	2.42 ± 0.04
ff99SBws	OPC	32.44 ± 0.02	34.00 ± 0.02	32.87 ± 0.02	31.82 ± 0.03	1.56 ± 0.02	0.43 ± 0.01	-0.62 ± 0.02	1.13 ± 0.02	2.18 ± 0.03
ff99SBws	TIP4P-D	32.42 ± 0.02	34.04 ± 0.02	32.85 ± 0.03	31.77 ± 0.03	1.62 ± 0.02	0.43 ± 0.01	-0.65 ± 0.01	1.19 ± 0.02	2.27 ± 0.03
DES-amber	TIP4P-D	32.30 ± 0.01	33.77 ± 0.02	32.72 ± 0.01	31.78 ± 0.02	1.47 ± 0.02	0.42 ± 0.01	-0.53 ± 0.01	1.05 ± 0.02	2.00 ± 0.03
DES-amberSF1.0	TIP4P-D	32.35 ± 0.01	33.93 ± 0.02	32.78 ± 0.02	31.74 ± 0.02	1.58 ± 0.02	0.43 ± 0.01	-0.60 ± 0.01	1.15 ± 0.02	2.19 ± 0.03
a99SBdisp	a99SBdisp-water	32.29 ± 0.02	33.83 ± 0.03	32.73 ± 0.02	31.66 ± 0.02	1.54 ± 0.02	0.44 ± 0.01	-0.62 ± 0.02	1.10 ± 0.02	2.17 ± 0.03
ff99SBws	TIP4P/2005	32.36 ± 0.02	34.01 ± 0.02	32.78 ± 0.02	31.69 ± 0.03	1.65 ± 0.02	0.43 ± 0.01	-0.66 ± 0.02	1.23 ± 0.02	2.32 ± 0.03
ff99SBws	TIP4P/2005s	32.50 ± 0.03	34.16 ± 0.04	32.94 ± 0.03	31.84 ± 0.02	1.66 ± 0.02	0.44 ± 0.01	-0.66 ± 0.01	1.21 ± 0.02	2.31 ± 0.03

Table S5: R_g, ΔR_g, and ΔR_g^{SAS} values from unrestrained MD simulations of 18 different combinations of protein force fields and water models for the protein glucose isomerase.

force field	water model	R_g^{Prot} (Å)	R_g^{SAXS} (Å)	R_g^{SANS/H_2O} (Å)	R_g^{SANS/D_2O} (Å)	$R_g^{SAXS} - R_g^{Prot}$ (Å)	$R_g^{SANS/H_2O} - R_g^{Prot}$ (Å)	$R_g^{SANS/D_2O} - R_g^{Prot}$ (Å)	$R_g^{SAXS} - R_g^{SANS/H_2O}$ (Å)	$R_g^{SAXS} - R_g^{SANS/D_2O}$ (Å)
c36	SPC/E	14.17 ± 0.02	14.95 ± 0.04	14.32 ± 0.03	13.11 ± 0.04	0.78 ± 0.05	0.15 ± 0.02	-1.06 ± 0.03	0.63 ± 0.05	1.84 ± 0.07
ff14SB	SPC/E	14.13 ± 0.01	14.78 ± 0.04	14.25 ± 0.02	13.11 ± 0.04	0.65 ± 0.04	0.12 ± 0.02	-1.02 ± 0.04	0.53 ± 0.04	1.66 ± 0.07
ff99SB-ildn	SPC/E	14.06 ± 0.01	14.82 ± 0.04	14.18 ± 0.01	12.99 ± 0.04	0.76 ± 0.04	0.12 ± 0.01	-1.07 ± 0.04	0.64 ± 0.05	1.83 ± 0.08
c36	cTIP3P	14.31 ± 0.02	14.98 ± 0.03	14.43 ± 0.02	13.31 ± 0.03	0.66 ± 0.03	0.12 ± 0.01	-1.00 ± 0.03	0.54 ± 0.03	1.66 ± 0.06
c36	TIP3P	14.08 ± 0.02	14.80 ± 0.05	14.22 ± 0.03	13.11 ± 0.04	0.72 ± 0.04	0.14 ± 0.02	-0.97 ± 0.04	0.58 ± 0.05	1.69 ± 0.07
ff14SB	TIP3P	14.12 ± 0.01	14.86 ± 0.04	14.25 ± 0.01	13.10 ± 0.03	0.74 ± 0.04	0.12 ± 0.01	-1.02 ± 0.03	0.62 ± 0.04	1.76 ± 0.06
ff99SB-ildn	TIP3P	14.06 ± 0.02	14.84 ± 0.03	14.19 ± 0.03	13.02 ± 0.02	0.78 ± 0.03	0.12 ± 0.01	-1.04 ± 0.02	0.65 ± 0.02	1.82 ± 0.03
ff15fb	TIP3P-FB	14.02 ± 0.01	14.73 ± 0.05	14.14 ± 0.02	12.90 ± 0.04	0.71 ± 0.05	0.12 ± 0.02	-1.12 ± 0.04	0.59 ± 0.04	1.83 ± 0.08
c36	OPC3	14.13 ± 0.02	14.92 ± 0.04	14.25 ± 0.03	13.02 ± 0.05	0.78 ± 0.04	0.12 ± 0.02	-1.11 ± 0.04	0.67 ± 0.05	1.89 ± 0.07
ff99SBws	OPC3	14.01 ± 0.02	14.76 ± 0.04	14.11 ± 0.02	12.93 ± 0.04	0.76 ± 0.04	0.10 ± 0.01	-1.08 ± 0.03	0.65 ± 0.03	1.84 ± 0.06
c36	OPC	14.30 ± 0.03	14.94 ± 0.06	14.41 ± 0.03	13.29 ± 0.04	0.63 ± 0.05	0.11 ± 0.02	-1.01 ± 0.05	0.52 ± 0.05	1.64 ± 0.09
ff99SBws	OPC	14.12 ± 0.01	14.82 ± 0.05	14.23 ± 0.02	13.06 ± 0.04	0.69 ± 0.05	0.10 ± 0.02	-1.07 ± 0.04	0.59 ± 0.05	1.76 ± 0.09
ff99SBws	TIP4P-D	14.12 ± 0.02	14.84 ± 0.03	14.26 ± 0.02	13.05 ± 0.03	0.72 ± 0.03	0.14 ± 0.01	-1.07 ± 0.03	0.58 ± 0.03	1.79 ± 0.06
DES-amber	TIP4P-D	14.25 ± 0.01	15.02 ± 0.04	14.37 ± 0.01	13.14 ± 0.03	0.77 ± 0.04	0.12 ± 0.01	-1.12 ± 0.03	0.65 ± 0.04	1.89 ± 0.07
DES-amberSF1.0	TIP4P-D	14.22 ± 0.02	15.06 ± 0.04	14.31 ± 0.02	12.99 ± 0.03	0.84 ± 0.04	0.10 ± 0.02	-1.23 ± 0.03	0.74 ± 0.04	2.07 ± 0.06
a99SBdisp	a99SBdisp-water	14.10 ± 0.03	14.82 ± 0.06	14.24 ± 0.03	13.07 ± 0.04	0.71 ± 0.05	0.14 ± 0.01	-1.04 ± 0.04	0.57 ± 0.05	1.75 ± 0.08
ff99SBws	TIP4P/2005	13.97 ± 0.02	14.75 ± 0.04	14.09 ± 0.03	12.97 ± 0.03	0.78 ± 0.03	0.12 ± 0.01	-1.00 ± 0.04	0.66 ± 0.03	1.78 ± 0.06
ff99SBws	TIP4P/2005s	14.03 ± 0.03	14.83 ± 0.05	14.13 ± 0.03	12.91 ± 0.04	0.80 ± 0.04	0.10 ± 0.01	-1.12 ± 0.04	0.70 ± 0.04	1.92 ± 0.07

Table S6: R_g, ΔR_g, and ΔR_g^{SAS} values from unrestrained MD simulations of 18 different combinations of protein force fields and water models for the protein lysozyme.

force field	water model	R_g^{Prot} (Å)	R_g^{SAXS} (Å)	R_g^{SANS/H_2O} (Å)	R_g^{SANS/D_2O} (Å)	$R_g^{SAXS} - R_g^{Prot}$ (Å)	$R_g^{SANS/H_2O} - R_g^{Prot}$ (Å)	$R_g^{SANS/D_2O} - R_g^{Prot}$ (Å)	$R_g^{SAXS} - R_g^{SANS/H_2O}$ (Å)	$R_g^{SAXS} - R_g^{SANS/D_2O}$ (Å)
c36	SPC/E	31.19 ± 0.01	32.15 ± 0.02	31.57 ± 0.01	31.22 ± 0.02	0.96 ± 0.02	0.38 ± 0.01	0.03 ± 0.02	0.58 ± 0.02	0.92 ± 0.03
ff14SB	SPC/E	31.27 ± 0.02	32.28 ± 0.03	31.65 ± 0.01	31.25 ± 0.02	1.02 ± 0.03	0.38 ± 0.01	-0.02 ± 0.02	0.63 ± 0.03	1.04 ± 0.04
ff99SB-ildn	SPC/E	31.50 ± 0.02	32.58 ± 0.03	31.88 ± 0.02	31.46 ± 0.02	1.08 ± 0.02	0.37 ± 0.01	-0.05 ± 0.01	0.70 ± 0.02	1.12 ± 0.03
c36	cTIP3P	31.30 ± 0.02	32.14 ± 0.05	31.65 ± 0.02	31.39 ± 0.03	0.84 ± 0.04	0.35 ± 0.01	0.09 ± 0.02	0.49 ± 0.04	0.75 ± 0.05
c36	TIP3P	31.23 ± 0.02	32.10 ± 0.02	31.59 ± 0.01	31.30 ± 0.02	0.87 ± 0.02	0.36 ± 0.01	0.07 ± 0.01	0.50 ± 0.02	0.80 ± 0.02
ff14SB	TIP3P	31.16 ± 0.02	32.14 ± 0.02	31.55 ± 0.02	31.16 ± 0.03	0.99 ± 0.02	0.39 ± 0.01	-0.00 ± 0.01	0.60 ± 0.02	0.99 ± 0.02
ff99SB-ildn	TIP3P	31.31 ± 0.01	32.30 ± 0.02	31.69 ± 0.01	31.31 ± 0.02	0.99 ± 0.02	0.39 ± 0.01	0.01 ± 0.02	0.60 ± 0.02	0.98 ± 0.03
ff15fb	TIP3P-FB	31.10 ± 0.02	32.01 ± 0.02	31.48 ± 0.02	31.11 ± 0.02	0.91 ± 0.02	0.38 ± 0.01	0.01 ± 0.01	0.53 ± 0.02	0.90 ± 0.02
c36	OPC3	31.21 ± 0.03	32.21 ± 0.05	31.58 ± 0.03	31.20 ± 0.03	1.00 ± 0.03	0.37 ± 0.01	-0.01 ± 0.02	0.63 ± 0.03	1.01 ± 0.04
ff99SBws	OPC3	31.49 ± 0.02	32.56 ± 0.02	31.86 ± 0.02	31.45 ± 0.03	1.07 ± 0.02	0.37 ± 0.01	-0.04 ± 0.02	0.70 ± 0.02	1.11 ± 0.03
c36	OPC	31.35 ± 0.02	32.28 ± 0.03	31.69 ± 0.02	31.39 ± 0.03	0.93 ± 0.02	0.34 ± 0.01	0.04 ± 0.02	0.59 ± 0.02	0.90 ± 0.03
ff99SBws	OPC	31.63 ± 0.03	32.63 ± 0.05	31.97 ± 0.02	31.64 ± 0.03	1.00 ± 0.04	0.35 ± 0.01	0.01 ± 0.02	0.65 ± 0.04	0.99 ± 0.05
ff99SBws	TIP4P-D	31.54 ± 0.02	32.56 ± 0.02	31.93 ± 0.02	31.54 ± 0.02	1.02 ± 0.02	0.39 ± 0.01	-0.00 ± 0.01	0.63 ± 0.02	1.02 ± 0.03
DES-amber	TIP4P-D	31.35 ± 0.01	32.28 ± 0.03	31.70 ± 0.02	31.39 ± 0.02	0.94 ± 0.02	0.35 ± 0.01	0.04 ± 0.01	0.58 ± 0.02	0.89 ± 0.03
DES-amberSF1.0	TIP4P-D	31.32 ± 0.02	32.40 ± 0.03	31.68 ± 0.02	31.26 ± 0.03	1.08 ± 0.02	0.36 ± 0.01	-0.06 ± 0.01	0.72 ± 0.02	1.14 ± 0.03
a99SBdisp	a99SBdisp-water	31.38 ± 0.02	32.31 ± 0.03	31.76 ± 0.01	31.42 ± 0.01	0.93 ± 0.02	0.38 ± 0.01	0.03 ± 0.02	0.55 ± 0.03	0.89 ± 0.04
ff99SBws	TIP4P/2005	31.47 ± 0.01	32.50 ± 0.02	31.84 ± 0.01	31.44 ± 0.01	1.03 ± 0.02	0.37 ± 0.01	-0.03 ± 0.01	0.66 ± 0.02	1.06 ± 0.03
ff99SBws	TIP4P/2005s	31.57 ± 0.01	32.66 ± 0.02	31.94 ± 0.02	31.48 ± 0.02	1.09 ± 0.02	0.37 ± 0.01	-0.09 ± 0.02	0.72 ± 0.03	1.18 ± 0.04

Table S7: R_g, ΔR_g, and ΔR_g^{SAS} values from unrestrained MD simulations of 18 different combinations of protein force fields and water models for the protein urate oxidase.

Supplementary References

(1) Trewhella, J.; Vachette, P.; Bierma, J.; Blanchet, C.; Brookes, E.; Chakravarthy, S.; Chatzimagas, L.; Cleveland, T.; Cowieson, N.; Crossett, B. e. a. A Round Robin Approach Provides a Detailed Assessment of Biomolecular Small-Angle Scattering Data Reproducibility and Yields Consensus Curves for Benchmarking. *Acta Cryst. D78* **2022**,

(2) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267–2272.

(3) Grishaev, A.; Guo, L.; Irving, T.; Bax, A. Improved Fitting of Solution X-ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. *J. Am. Chem. Soc.* **2010**, *132*, 15484–15486.

(4) Ivanović, M. T.; Bruetzel, L. K.; Shevchuk, R.; Lipfert, J.; Hub, J. S. Quantifying the influence of the ion cloud on SAXS profiles of charged proteins. *Phys. Chem. Chem. Phys.* **2018**, *20*, 26351–26361.

(5) Holst, M. J. The Poisson-Boltzmann Equation: Analysis and Multilevel Numerical Solution. Ph.D. thesis, California institute of Technology, 1994.

(6) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *J. Phys. Chem. B* **2013**, *117*, 10183–10192.

(7) Yoo, J.; Aksimentiev, A. New Tricks for Old Dogs: Improving the Accuracy of Biomolecular Force Fields by Pair-Specific Corrections to Non-Bonded Interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8432–8449.

(8) Ivanović, M. T.; Bruetzel, L. K.; Shevchuk, R.; Lipfert, J.; Hub, J. S. Quantifying the Influence of the Ion Cloud on SAXS Profiles of Charged Proteins. *Physical Chemistry Chemical Physics* **2018**, *20*, 26351–26361.

(9) Chen, P.-c.; Hub, J. S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014**, *107*, 435–447.

(10) Chatzimagas, L.; Hub, J. S. In *Small Angle Scattering Part A: Methods for Structural Investigation*; Tainer, J. A., Ed.; Methods in Enzymology; Academic Press, 2022; Vol. 677; pp 433–456.

(11) Merzel, F.; Smith, J. C. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5378–5383.

(12) Chen, P.-c.; Hub, J. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014**, *107*, 435–447.

(13) Wang, J.; Dauter, M.; Alkire, R.; Joachimiak, A.; Dauter, Z. Triclinic lysozyme at 0.65 Å resolution. *Acta Cryst. D* **2007**, *63*, 1254–1268.

(14) Wlodawer, A.; Svensson, L. A.; Sjoelin, L.; Gilliland, G. L. Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochem.* **1988**, *27*, 2705–2717.

(15) Watanabe, N.; Akiba, T.; Kanai, R.; Harata, K. Structure of an orthorhombic form of xylanase II from *Trichoderma reesei* and analysis of thermal displacement. *Acta Cryst. D* **2006**, *62*, 784–792.

(16) Nowak, E.; Panjikar, S.; Tucker, P. Atomic structure of Glucose isomerase . *To be published*

(17) Gabison, L.; Chiadmi, M.; El Hajji, M.; Castro, B.; Colloc'h, N.; Prangé, T. Near-atomic resolution structures of urate oxidase complexed with its substrate and analogues: the protonation state of the ligand. *Acta Cryst. D* **2010**, *66*, 714–724.

(18) Huang, J.; Rauscher, S.; Nawrocki, G.; Ting, R.; Feig, M.; de Groot, B.; Grubmüller, H.; MacKerell, A. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.

(19) MacKerell, A. D. J. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(21) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(22) Izadi, S.; Onufriev, A. V. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **2016**, *145*, 074501.

(23) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(25) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.

(26) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023–4039.

(27) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(28) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

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

(30) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(31) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *J. Chem. Theory Comput.* **2020**, *16*, 2494–2507.

(32) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758–E4766.

Description of additional supplementary information

Johanna-Barbara Linse and Jochen S. Hub*

Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, 66123, Germany

E-mail: jochen.hub@uni-saarland.de

Supplementary Movie 1: Illustration of the hydration layers around the protein xylanase.

Three-dimensional solvent electron density taken from a simulation of xylanase obtained with ff99SBws and TIP4P/2005s water using position restraints on all heavy atoms. Only density inside the envelope is shown in shades from light grey (bulk water) to blue to orange, revealing the first (orange) and the second (mostly blue) hydration layers. For the color code, see the color bar in Fig. 1B.

Chapter 7

How protein hydration depends on amino acid composition, peptide conformation, and force fields

This chapter was originally published in Biophysical Journal, Volume 125, Issue 1, Pages 255-269 (2026) (Ref. 237). We thank the journal for permission to reprint it.

The authors of this paper are Johanna-Barbara Linse, Tobias M. Fischbach and Jochen S. Hub, with the following contributions:

J.S.H. designed research and contributed software.
T.M.F. and J.-B.L. performed research and analyzed data.
J.-B.L. and J.S.H. wrote the article.

Biophysical Journal
Article

How protein hydration depends on amino acid composition, peptide conformation, and force fields

Johanna-Barbara Linse,[1] Tobias M. Fischbach,[1] and Jochen S. Hub[1,*]
[1]Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, Germany

ABSTRACT The protein hydration shell is a key mediator of processes such as molecular recognition, protein folding, and proton transfer. How solvent-exposed amino acids shape the hydration shell structure is not well understood. We combine molecular dynamics simulations with explicit-solvent predictions of small-angle x-ray scattering (SAXS) curves to quantify the contributions of all 20 proteinogenic amino acids to the hydration shell of the globular GB3 domain and the intrinsically disordered protein (IDP) XAO. We focus on two quantities encoded by SAXS curves: the hydration shell effect on the radius of gyration and the electron density contrast between protein and solvent. We derive an amino acid-specific contrast score, revealing that acidic residues generate the strongest contrast with 1–1.5 excess water molecules relative to alanine, followed by cationic and polar residues. In contrast, apolar residues generate a water depletion layer. These trends are consistent across simulations with different water models. Around the XAO peptide, the hydration shell is generally far weaker compared with the globular GB3 domain, indicating unfavorable water-peptide packing at the IDP surface. The hydration shell effect on the radius of gyration of the IDP is strongly conformation-dependent. Together, the calculations show that the composition and spatial arrangement of solvent-exposed amino acids govern the hydration shell structure, with implications for a wide range of biological functions and for hydration-sensitive experimental techniques such as solution scattering.

SIGNIFICANCE Hydration shells of biomolecules constitute a large fraction of the water in crowded cellular environments and play key roles in biological functions such as enzymatic reactions and conformational transitions. Small-angle x-ray scattering (SAXS) has shown that hydration shells differ in density from bulk water, yet how solvent-exposed amino acids and protein surface geometry shape the hydration shell is not well understood. We combined molecular dynamics simulations with explicit-solvent SAXS predictions to quantify how solvent-exposed chemical moieties and protein geometry drive variations in hydration shell density. Notably, the hydration shell of a globular protein differs markedly from that of an intrinsically disordered protein. Our study offers a comprehensive characterization of protein hydration and informs the interpretation of hydration-sensitive experimental techniques.

INTRODUCTION

Proteins in solution are enveloped by a hydration shell, formed through electrostatic and dispersive interactions between water molecules and solvent-exposed protein moieties. The hydration shell actively participates in various biological functions, such as protein folding, molecular recognition, enzyme catalysis, proton transfer, or avoidance of unspecific aggregation, and is thus considered an integral part of proteins (1–5). The structure and dynamics of the hydration shell differ from those of bulk water, as revealed by nuclear magnetic resonance, terahertz spectroscopy, time-dependent fluorescence Stokes shift, inelastic neutron scattering, molecular dynamics (MD) simulations, and several other techniques (6–16). Consequently, the vibrational, rotational, and translational dynamics of water molecules in the hydration shell are slowed down by approximately two- to fivefold. Scattering experiments have revealed that the water density in the hydration shell of many proteins is increased relative to the density of bulk water, with the magnitude of this density increase likely being protein dependent (17–20). In the crowded cytoplasm of biological cells, up to 70% of water belongs to a hydration shell, indicating that water involved in life is predominantly non-bulk-like (3,21,22).

Submitted June 16, 2025, and accepted for publication November 24, 2025.
*Correspondence: jochen.hub@uni-saarland.de
Editor: Frank Gabel.
https://doi.org/10.1016/j.bpj.2025.11.2683

How the composition and relative arrangement of solvent-exposed amino acids control the properties of the protein hydration shell is not well understood. Small-angle scattering with x-rays and neutrons (SAXS/SANS) of highly charged proteins suggested that solvent-exposed anionic aspartate or glutamate residues increase the hydration shell density more than cationic lysine or arginine residues (20,23), which aligns with the large number of structured water molecules located at anionic residues in protein crystals (24). Furthermore, spectroscopic techniques revealed that the polarity of surfaces influences the properties of water at interfaces. At polar surfaces, water exhibits decreased internal water order and fewer internal hydrogen bonds. At apolar surfaces, in contrast, water exhibits increased internal order and more internal hydrogen bonds, and it may form low-density clathrate structures (25–35). Additionally, water has been shown to form a depletion layer with reduced density at hydrophobic surfaces (36–38). However, the quantitative influence of solvent-exposed amino acids or of specific chemical moieties on the hydration shell architecture remains largely unexplored.

Upon modifying solvent-exposed moieties of proteins by mutagenesis, the hydration shell may change via two distinct effects: first, due to altered residue-water interaction energies at fixed protein conformations (see previous paragraph). Second, mutations may shift the conformational ensemble of the protein, thereby replacing protein-water contacts with protein-protein contacts or vice versa. Such effects are especially pronounced in intrinsically disordered proteins (IDPs), as their shallow free energy landscapes are sensitive to mutations (39,40). Nevertheless, surface mutations may also alter the conformation of globular proteins, for example by triggering unfolding or a transition to a molten globule (41,42). By simulating proteins at preselected, fixed conformations, this study focuses on the first effect while noting that, under experimental conditions, both effects often play a role.

We recently validated the protein hydration shell from MD simulations by comparing results from explicit-solvent SAXS/SANS predictions (43–45) with consensus experimental data obtained from a worldwide community effort (20,46). SAXS and SANS data reflect the contrast of the, respectively, electron density or neutron scattering length density of the protein relative to bulk solvent, thereby including contributions of the hydration shell. We observed that many but not all combinations of protein force fields and water models accurately reproduce the hydration shell effect on the radius of gyration R_g. We furthermore found that the hydration shell effect on R_g depends on protein size, geometry, and surface composition, suggesting that the effect represents a protein-specific footprint of the hydration shell. In this study, we use MD simulations to quantify the influence of all proteinogenic amino acids on the hydration shell of a globular and an intrinsically disordered protein on two parameters that are encoded by SAXS curves, namely on the R_g and on the overall contrast between solute and solvent. We derive an amino acid-specific contrast score for solvent-exposed proteinogenic residues and show that the hydration shell structure and its effect on SAXS data depends not only on chemical composition but also on peptide conformation and water models. We rationalize the amino acid-specific effects on SAXS curves by analyzing three-dimensional solvent densities as well as radial distribution functions around solvent-exposed amino acid side chains.

MATERIALS AND METHODS

Simulation setup and parameters for the GB3 domain

The initial structure of the GB3 domain was taken from the protein data bank (PDB (47): 1IGD (48)). Ten amino acids on the surface of the GB3 domain were selected and mutated to one of the 21 proteinogenic amino acid, involving two protonation states of histidine, with the software Chimera (49), namely residues 15, 18, 20, 22, 24, 27, 33, 37, 47, and 51 (Fig. 1 *A*). Hydrogen atoms were added with pdb2gmx. MD simulations of GB3 were carried out with GROMACS, version 2020.3 (50). Interactions of the proteins were described with the following variants of the AMBER03 force field (51): ff03* (52), ff03w (53), and ff03ws (54). The starting structures were placed in a dodecahedral box, where the distance between the protein and the box edges was at least 2.0 nm, and solvated in TIP3P (55), TIP4P/2005 (56), or TIP4P/2005s (54) water. The simulation systems were neutralized by adding Na^+ or Cl^- ions. After 400 steps of minimization with the steepest decent algorithm, the systems were equilibrated for 100 ps with harmonic position restraints applied to the heavy atoms of the proteins (force constant 1000 kJ mol^{-1} nm^{-2}). Subsequently, production runs were started for 50 ns with harmonic position restraints applied to the backbone atoms of the proteins (force constant 2000 kJ mol^{-1} nm^{-2}). The equations of motion were integrated using the leapfrog algorithm (57). The temperature was controlled at 298.15 K, using velocity rescaling (τ = 1 ps) (58). The pressure was controlled at 1 bar with the Berendsen barostat (τ = 1 ps) (59) and with the Parrinello-Rahman barostat (τ = 5 ps) (60) during equilibration and production simulations, respectively. The geometry of the water molecules was constrained with the SETTLE algorithm (61), and LINCS (62) was used to constrain all other bond length. A time-step of 2 fs was used. Dispersive interactions and short-range repulsion were described by a Lennard-Jones potential with a cutoff at 1 nm. The pressure and the energy were corrected for missing dispersion corrections beyond the cut-off. Neighbor lists were updated with the Verlet scheme. Coulomb interactions were computed with the smooth particle-mesh Ewald (PME) method (63,64). We used a Fourier spacing of approx. 0.12 nm, which was optimized by the GROMACS mdrun module at the beginning of each simulation.

Simulation setup and parameters for the XAO peptide

To obtain an ensemble of XAO, we carried out maximum-entropy ensemble refinement (65) of XAO against experimental SAXS data. To this end, four parallel replicas of XAO simulations were coupled to SAXS data as described in the supporting material. From the ensemble, we obtained 20 frames that reasonably represent the confrontational space adopted by XAO, thus including compact and extended conformations. The simulation systems for the XAO conformations and mutants were set up as described above for the GB3 domain, except that the MD simulations were carried out with the GROMACS 2021.7 (50). Side chains of ornithin (Orn) and

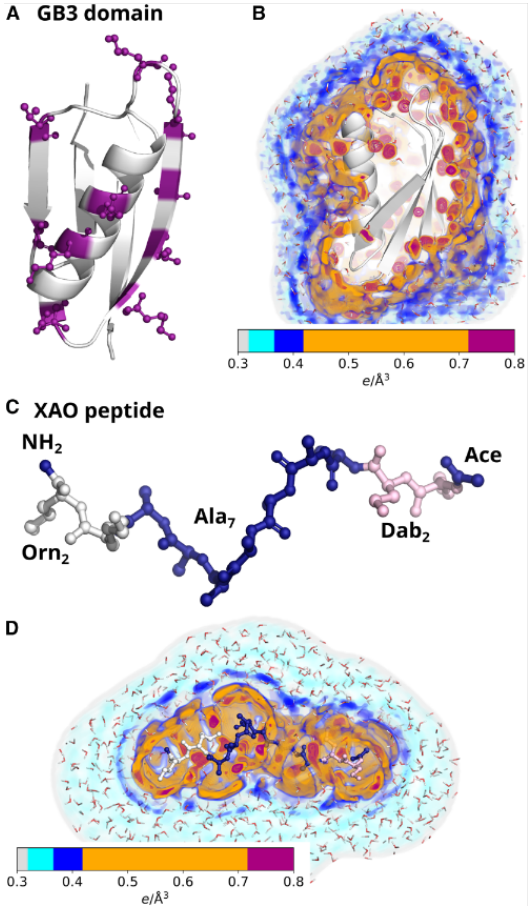


FIGURE 1 Structures and hydration shells of the GB3 domain and XAO peptide. (*A*) Cartoon representation of the GB3 domain. Ten surface amino acids shown in purple ball-and-stick representation were mutated into each of the 20 proteinogenic amino acids. (*B*) Three-dimensional density of the hydration shell around the wild-type GB3 domain (*for colors, see colorbar*). The first and second hydration layers appear as orange and blue densities, respectively. Red densities indicate a well-defined water position at the protein surface. (*C*) XAO peptide with four unnatural amino acids at the termini: two 2,4-diaminobutyric acid (Dab) and two ornithin (Orn) shown in pink and white, respectively. (*D*) Three-dimensional solvent density around the wild-type XAO peptide.

2,4-diaminobutyric acid (Dab) were modeled based on the parameters for lysine by removing either one or two CH2 groups from the lysine topology, respectively. Four residues of the XAO, the two Orn and the two Dab residues, were mutated to one of the 21 proteinogenic amino acids with the software Chimera (49).

Explicit-solvent SAXS calculations

2251 simulation frames from the time interval between 5 and 50 ns from MD simulations of the GB3 domain and 2501 simulation frames from the time interval between 0 and 50 ns from simulations of the XAO peptide were used for SAXS calculations. The SAXS calculations were performed with GROMACS-SWAXS, an in-house modification of GROMACS 2021.7, as also implemented by the web server WAXSiS (43,44,66). The code and tutorials are available at https://cbjh.gitlab.io/gromacs-swaxs-docs/. For more background on explicit-solvent SAXS calculations, we refer to recent reviews (45,67). Explicit-solvent SAXS calculations have previously been successfully compared with experimental data from diverse solutes including proteins (43,68), IDPs (69), protein-detergent complexes (70), or detergent micelles (71,72), suggesting that the predictions are robust and accurate. A spatial envelope was built around all solute frames from the proteins. Solvent atoms inside the envelope contributed to the calculated SAXS curves. The distance between the protein and the envelope surface was at least 9 Å, such that all water atoms of the hydration shell were included. The buffer subtraction was carried out using 2251 simulations frames of a pure-water simulation box, which was simulated for 50 ns and large enough to enclose the envelope. The orientational average was carried out using 200 **q**-vectors for the GB3 domain and 50 **q**-vectors for the XAO peptide for each absolute value of **q**, and the solvent electron density was corrected to the experimental value of 334 e/nm^3 as described previously (43).

Statistical errors were computed for simulations with the GB3 domain by binning the trajectory into 10 time blocks of 4.5 ns and computing the standard error. In the case of the XAO peptide, statistical errors were calculated from simulations of 20 independent conformations.

Calculation of the hydration shell contrast

The forward scattering intensity $I_0 = \Delta N_e^2$ of a SAXS curve is given by the square of the contrast ΔN_e^2 between solute (including the hydration shell) and the solvent in number of electrons. Thus, I_0 follows by

$$I_0 = \left[\Delta N_e^{\text{prot}} + \Delta N_e^{\text{hs}}\right]^2 \tag{1}$$

$$= \left[N_e^{\text{prot}} - \rho_{\text{solv}} V^{\text{prot}} + \Delta N_e^{\text{hs}}\right]^2, \tag{2}$$

where N_e^{prot} is the number of electrons of the protein, V^{prot} the protein volume, ΔN_e^{hs} the contrast imposed by the hydration shell, and ρ_{solv} the solvent electron density taken as 334 e/nm^3. Thus, we have

$$\Delta N_e^{\text{hs}} = \pm\sqrt{I_0} - \left(N_e^{\text{prot}} - \rho_{\text{solv}} V^{\text{prot}}\right). \tag{3}$$

Here, the plus and minus signs correspond to the cases where $\Delta N_e^{\text{prot}} + \Delta N_e^{\text{hs}}$ is positive or negative, respectively. In our implementation, N_e^{prot} is taken from atomic form factor at zero scattering angle as defined via the Cromer-Mann parameters of the atoms (73). The volume of the solute V^{prot} was defined as the cavity volume calculated with the 3V volume calculator (74) with a grid spacing of 0.16 Å and a probe radius of 1.4 Å corresponding to the van der Waals radius of a water molecule. Volumes were computed as an average over 20 simulations frames, which were randomly rotated before running the 3V software. Statistical uncertainties of volume calculations correspond to 1 SE obtained from the 20 frames. Volumes of GB3 and XAO variants are shown in Fig. S13. The influence of the probe radius and the grid spacing on the volume calculation is analyzed in Fig. S14.

RESULTS

To quantify the effects of amino acid composition on the hydration shell of proteins, we simulated the GB3 domain as a representative for globular proteins (Fig. 1 *A* and *B*) and the XAO peptide (75,76) as representative for IDPs (Fig. 1 *C* and *D*). The three-dimensional electron densities of solvent

around the GB3 domain or around XAO are shown in Fig. 1 *B* and *D*, here computed from simulations with position restraints on all heavy atoms leading to spatially well-defined densities from surface-bound water molecules. The densities reveal highly localized water molecules (red densities), the first hydration shell (orange/red densities), as well as the second hydration shell (dark blue densities). A highly shallow third shell is hardly visible in the three-dimensional density representation (Fig. 1 *D*, cyan density layer). The structure of the hydration shell involving a pronounced first shell, a shallow second, and a highly shallow third shell agrees with many previous MD studies (Ref. (77) and references therein).

The hydration shell of the globular GB3 domain strongly depends on the surface amino acid composition

SAXS experiments of proteins probe the electron density contrast between the protein and the bulk solvent, including the density contrast contributed by the hydration shell. To quantify the effects of different amino acids to the hydration shell, and to relate variations among different amino acids to putative solution scattering experiments, we computed SAXS curves of the wild-type and of 21 mutants of the globular GB3 domain. We used explicit-solvent SAXS calculations (43), as also implemented by the WAXSiS web server (66), thereby accounting for all non-bulk-like solvent molecules of the hydration shell and using an explicit representation of the excluded solvent. In contrast to implicit-solvent SAXS predictions, the method does not require any solvent-related fitting parameters or a choice for a thickness for the hydration shell (45). We selected 10 solvent-exposed residues (Fig. 1 *A*, pink ball-and-stick representation) and mutated these residues to each of the 20 proteinogenic amino acids while including histidine in the neutral form (δ-nitrogen protonated, His^0) and in the cationic form (δ- and ϵ-nitrogen protonated, His^+), resulting in 22 GB3 variants (wild-type and 21 mutants). We performed explicit-solvent MD simulations with restraints on the backbone atoms to maintain all GB3 variants in identical backbone conformation and to prevent unfolding of putatively unstable GB3 mutants such as mutants with many hydrophobic solvent-exposed residues. SAXS curves $I(q)$ computed for the 22 GB3 variants differ (Fig. 2). Since we computed the SAXS curves taking the solvent explicitly into account (43,45,66), the variations among the SAXS curves include effects owing to variations of the hydration shell contrast.

In this study, we used two quantities that provide a footprint for the hydration shell while being encoded by the SAXS curves: 1) the forward scattering intensity $I_0 = I(q = 0)$, which is related to the contrast between protein and solvent; and 2) the radius of gyration R_g, which quantifies the spatial extent of the protein. Focusing first on the former quantity, the forward scattering is given by the

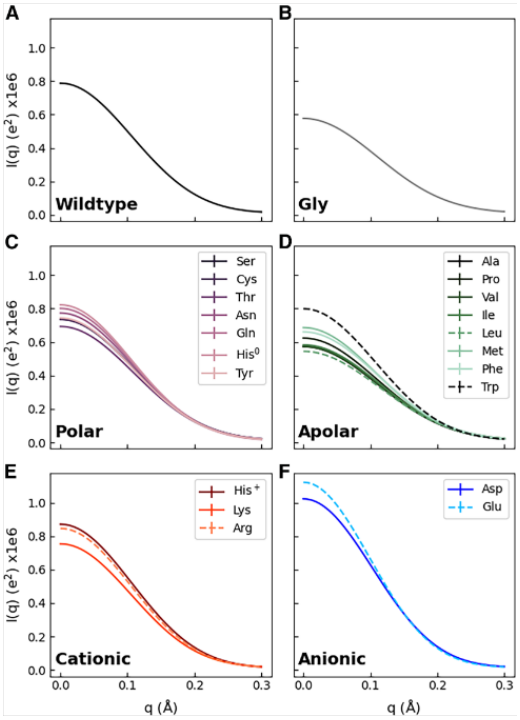


FIGURE 2 SAXS curves of the GB3 domain from explicit-solvent SAXS calculations with the TIP4P/2005 water model in combination with the ff03w protein force field (*A*) for GB3 wild-type or (*B–F*) for 21 GB3 variants with 10 mutated solvent-exposed amino acids each (*for color code and line style, see legends*). For clarity, SAXS curves are grouped by the amino acid property (glycine, polar, apolar, cationic, anionic) in (*B*)–(*F*). Guinier plots of these curves are shown in Fig. S2.

square of the total contrast ΔN_e in number of electrons between protein and solvent, i.e., $I_0 = \Delta N_e^2$. We decomposed the total contrast into contributions from the contrast of the bare protein $\Delta N_e^{\mathrm{prot}}$ and the contrast of the hydration shell ΔN_e^{hs}, via $\Delta N_e = \Delta N_e^{\mathrm{prot}} + \Delta N_e^{\mathrm{hs}}$ (see Materials and Methods). Fig. 3 *A* (yellow bars) presents the contrast of the hydration shell for 22 GB3 variants, here computed with the AMBER force field ff03w in conjunction with the TIP4P/2005 water model (53,56), which revealed excellent agreement with experimental SAXS/SANS data in our previous study and may, therefore, be taken as reference force field (20). The contrast is plotted in number of water molecules as $N_e^{\mathrm{hs}}/10$ given that each water molecule contains 10 electrons. Evidently, the contrast of the hydration shell differs greatly among GB3 variants with different solvent-exposed amino acid types. The hydration shell of wild-type GB3 exhibits a positive contrast of 4.1 water molecules implying the presence of additional 4.1 water molecules in the hydration shell relative to an equivalent volume of bulk water, in line with the well-known densely packed

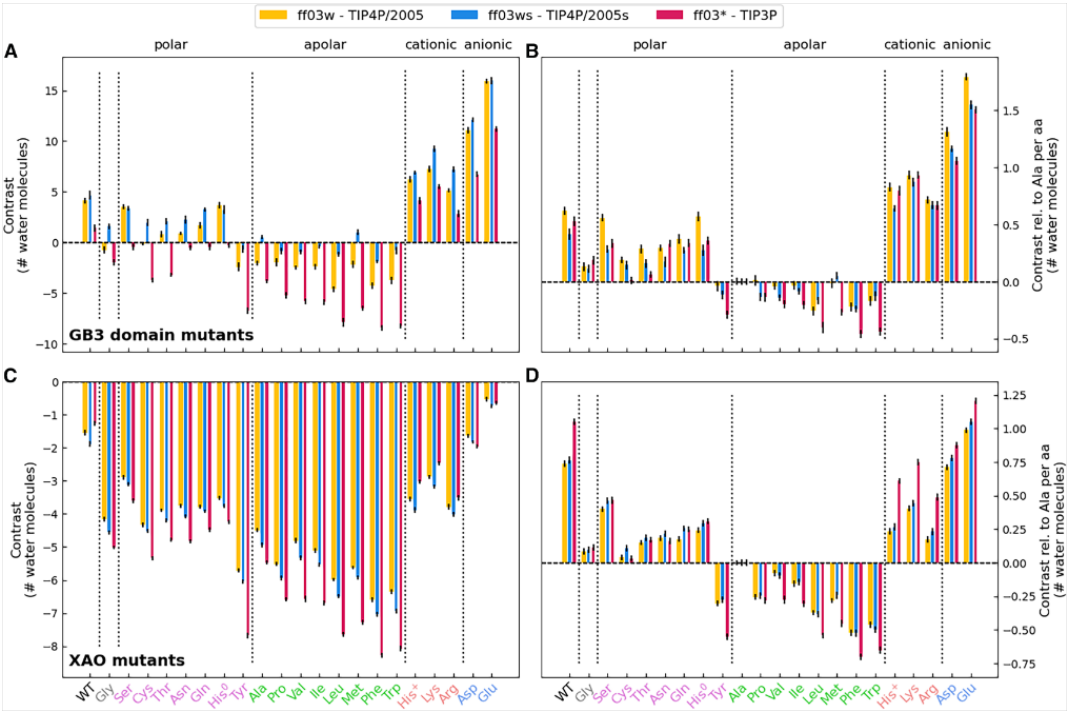


FIGURE 3 Analysis of hydration shell contrasts. (*A*) Contrast of the hydration shell in number of water molecules of GB3 wild-type and 21 GB3 mutants; see labels at abscissa colored by the property of the amino acid: Gly (*gray*), polar (*pink*), apolar (*green*), cationic (*orange*), and anionic (*blue*) residues. Contrast values are shown for three combinations of protein force field and water model: ff03w–TIP4P/2005 (*yellow*), ff03ws–TIP4P/2005s (*blue*), ff03*–TIP3P (*red*). (*B*) Contrast per amino acid for GB3 domain relative to alanine. (*C* and *D*) Same analysis as in (*A*) and (*B*) for the XAO peptide. Error bars denote 1 SE.

hydration shell documented by SAXS experiments of several proteins (17,18,20). Among the mutated GB3 variants, the variants with additional anionic residues (Asp/Glu) reveal the largest contrast, followed in decreasing order by GB3 variants with cationic (Lys/Arg/His$^+$) and polar charge-neutral amino acids (Fig. 3, pink labels at the abscissa). The marked hydration shell imposed by the anionic residues Glu/Asp aligns with previous SAXS/SANS experiments of super-charged variants of green fluorescent protein (78) and of the highly anionic glucose isomerase (20,46). In contrast, GB3 variants with many apolar solvent-exposed residues reveal a negative hydration shell contrast, indicating a water depletion layer in the vicinity of hydrophobic amino acids, in line with reports for other types of hydrophobic surfaces (36–38). An exception to the order anionic–cationic–polar–apolar is given by tyrosine, which displays a more negative contrast compared with all other polar amino acids, rationalized by the presence of the apolar six-membered aromatic ring.

To quantify the hydration shell contrast imposed by individual amino acids, Fig. 3 *B* presents the contrast per amino acid and relative to alanine as reference. Accordingly, the hydration shells of anionic residues exhibit approximately 1.5 additional water molecules relative to alanine, with glutamate standing out as the amino acid whose hydration shell imposes the largest contrast, indicative for a particularly densely packed hydration shell. The hydration shells of cationic residues contain roughly one additional water molecule relative to alanine, whereas the hydration shells of polar residues contain approximately 0.2–0.5 additional water molecules relative to alanine. Bulky apolar amino acids such as leucine or phenylalanine may contain up to 0.3 fewer water molecules relative to alanine. Tyrosine with its polar hydroxyl group and apolar aromatic ring represents an intermediate case between polar and apolar residues. These values provide an amino acid-resolved hydration layer contrast score for a common globular protein such as GB3, thus quantifying how chemical specificities of proteinogenic amino acids control the density of the protein hydration shell. Below, we further analyze these data to dissect how individual chemical moieties control the hydration shell.

Critically, the overall contrast ΔN_e for a given MD simulation is determined unambiguously from the forward scattering calculated using explicit-solvent SAXS calculations. In contrast, the decomposition of ΔN_e into contributions from the bare protein $\Delta N_e^{\mathrm{prot}}$ and the hydration shell ΔN_e^{hs} depends on how the protein volume is defined. Assigning a larger volume to the protein leads to a decrease in $\Delta N_e^{\mathrm{prot}}$ and a corresponding increase in ΔN_e^{hs}. This effect is illustrated in Fig. S4 *A* and *C*, where protein volumes were computed using three different probe radii: 1.3 Å, 1.4 Å, or 1.5 Å. However, the contrast per amino acid relative to alanine depends only marginally on the volume definition (Fig. S4 *B* and *D*), suggesting that ambiguities in our volume calculations cancel out upon comparing different amino acids. Consequently, although the choice of volume definition (see Materials and Methods) affects the absolute contrasts shown in Fig. 3 *A* and *C*, it has only a marginal impact on the relative contrasts in Fig. 3 *B* and *D*.

Among different water models, effects of amino acid properties on the hydration layer contrast agree qualitatively but differ quantitatively

To test the influence of water models, and to exclude that our key findings are not biased by the choice of the water model, we calculated the contrasts for three different water models, namely TIP3P (55), TIP4P/2005 (56), and TIP4P/2005s (54), in combination with their corresponding AMBER03 force field (51) variant. Among these water models, TIP3P is the most widely used model since the popular CHARMM and AMBER protein force fields have originally been parameterized in conjunction with TIP3P. However, TIP3P shows poor agreement with experimental data as it yields a too low density, a too high diffusion coefficient, and a too high isothermal compressibility (79). TIP4P/2005 reproduces water properties more accurately compared with TIP3P, and it furthermore captures the water density over a wide temperature range (56). TIP4P/2005s takes water-water interactions from TIP4P/2005; however, it implements increased water-protein dispersion interactions with the aim to balance water-water against water-protein interactions in protein simulations with AMBER force fields (54).

Fig. 3 *A* compares the total hydration shell contrast of 22 GB3 variants for simulations with these three water models (yellow, blue, red bars, see legend). Irrespective of the water model, the contrasts follow the order anionic–cationic–polar–apolar, suggesting that the effects of amino acid classes on the hydration shell contrast agree qualitatively among the water models. This finding is confirmed by comparing the contrast per amino acid relative to alanine (Fig. 3 *B*). However, the total contrasts shown in Fig. 3 *A* furthermore reveal considerable quantitative differences among water models. TIP3P yields the lowest contrast for all GB3 variants with approximately four water molecules fewer within the overall GB3 hydration shell compared with TIP4P/2005 (Fig. 3 *A*, red vs. yellow bars). In contrast, TIP4P/2005s yields an increased contrast for most GB3 variants with typically zero to three additional water molecules in the hydration shell compared with TIP4P/2005 (blue vs. yellow bars), rationalized by the increased protein-water dispersion interactions implemented by TIP4P/2005s (54). More specifically, TIP4P/2005s yields a similar contrast compared with TIP4P/2005 for the GB3 wild-type and for the Ser, His^0, His^+, Asp, Glu and an increased contrast for all the other variants. Thereby, the hydration shell contrast by TIP4P/2005s largely exceeds the contrast by TIP3P, in particular for the apolar variants, with up to 7.5 additional water molecules in the hydration shell for the Trp variant.

Hydration shell contrasts controlled by amino acid type and force field impose experimentally accessible variations of the radius of gyration

Analysis of the hydration shell based on I_0 and hydration shell contrasts involve two caveats. First, experimental I_0 values—or, equivalently, total electron density contrasts—are subject to relatively high uncertainty because obtaining I_0 from an experimental SAXS curve would require precise knowledge of the solute concentration. Because the solute concentration is typically only approximately known, quantitative comparisons of I_0 between MD simulations and experiments is difficult. In contrast, the radius of gyration R_g obtained by SAXS coupled to size exclusion chromatography (SEC-SAXS) enables R_g measurements with sub-Ångström accuracy (46), thereby enabling quantitative validation of MD simulations against experiments (20). Second, whereas the overall solute contrast is unambiguously defined in explicit-solvent SAXS calculations via $(I_0)^{1/2}$, its decomposition into contrast contributions from the protein and hydration shell depends on the definition of the protein volume or, equivalently, on the definition of the dividing surface at the protein-water interface (77). In contrast, computing the hydration shell effect on R_g does not require a definition of the dividing surface.

Thus, as a second indicator for the hydration shell, we analyzed hydration shell effects on the R_g values of 22 GB3 variants. We computed the change of R_g owing to the hydration shell, as given by $\Delta R_g = R_g^{\mathrm{SAXS}} - R_g^{\mathrm{Prot}}$, where R_g^{SAXS} denotes the R_g value obtained by Guinier analysis of the SAXS curve, thereby taking the hydration shell into account, and R_g^{Prot} denotes the R_g value of the bare proteins computed from the atomic coordinates of protein atoms. Fig. 4 *A* presents ΔR_g values for the 22 GB3 variants, whereas Fig. 4 *B* shows the ΔR_g values relative to alanine per mutated amino acid. In line with the contrasts discussed above, the presence of anionic residues (Asp/Glu) imposes by far the largest increase of ΔR_g values by ∼1.5 Å and an increase per amino acid relative to alanine by ∼0.1 Å.

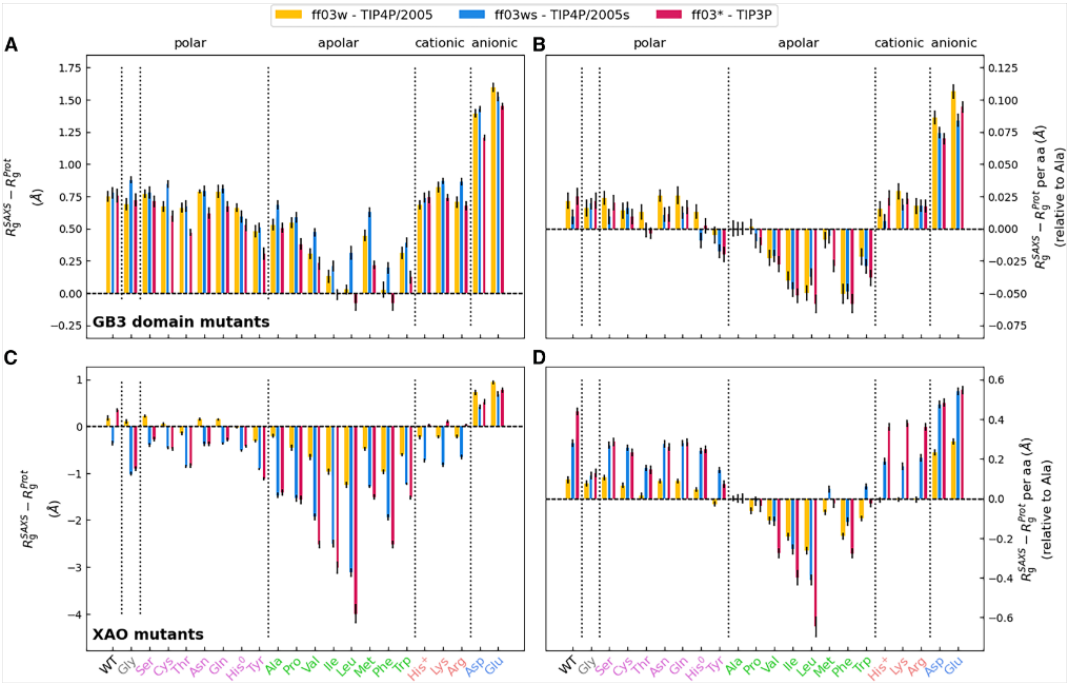


FIGURE 4 Effects of hydration shell on radii of gyration. (*A*) Hydration-induced shift in the radius of gyration ΔR_g from explicit-solvent SAXS calculation of 22 GB3 variants. ΔR_g values are shown for GB3 wild-type (WT) and 21 mutants; see labels at the abscissa colored by the property of the amino acid: Gly (*gray*), polar (*pink*), apolar (*green*), cationic (*orange*), and anionic (*blue*) residues. Results are shown for three combinations of protein force field and water model: ff03w–TIP4P/2005 (*yellow*), ff03ws–TIP4P/2005s (*blue*), ff03*–TIP3P (*red*). (*B*) ΔR_g values relative to the alanine mutant and per amino acid. (*C* and *D*) Same analysis as in (*A*) and (*B*) for the XAO peptide. Error bars denote 1 SE.

These results confirm that anionic residues impose a tightly packed hydration shell (20,24,78). In contrast, bulky hydrophobic residues such as Val, Ile, Leu, Phe, or Trp lead to small ΔR_g values and a decrease per amino acid relative to alanine by up to ∼0.05 Å, in line with a water depletion layer at the hydrophobic surface. Cationic and many polar residues lead to intermediate and similar ΔR_g values, which may be surprising considering that hydration shell contrasts imposed by cationic residues clearly exceed the contrasts imposed by polar residues (see Fig. 3 *A* and *B*). The findings might reflect that ΔR_g is sensitive to the spatial distributions of contrasts, which may lead to different amino acid-specific effects on ΔR_g compared with effects on the total contrast derived from I_0.

Comparing the ΔR_g values between simulations with TIP3, TIP4P/2005, and TIP4P/2005s reveals specific properties of water models, although the relative differences among ΔR_g values are overall smaller compared with the differences among the contrasts (compare Fig. 4 *A* and *B* with Fig. 3 *A* and *B*). Whereas the ΔR_g values of GB3 wild-type are in excellent agreement among the three water models (Fig. 4 *A*, left column), TIP3 yields lower ΔR_g for many GB3 mutants such as mutants with anionic residues (Asp/Glu), several apolar residues, as well as for several apolar residues such as Thr or Tyr. The largest variations of ΔR_g values among water models is found for the bulky apolar residues, where in particular TIP4P/2005s, but also TIP4P/2005, yields by far larger ΔR_g values compared with TIP3P. Thus, the increased protein-water dispersion interactions implemented by TIP4P/2005s lead to a partial loss of the water depletion layer at hydrophobic surfaces, with a footprint on the radii of gyration.

Hydration shell of the intrinsically disordered protein XAO exhibits a negative amino acid-dependent contrast

Compared with globular proteins, IDPs exhibit a larger surface/volume ratio, suggesting that—at a given protein density—IDPs exhibit more protein-water contacts and perturb a larger volume of water, rationalizing the tight coupling between IDP and water dynamics (80). Many IDPs carry out their function by partial folding on the surface of other proteins, thereby involving large rearrangements of protein-water

and water-water interaction networks. Nevertheless, the hydration shell of IDPs has attracted less attention compared with the hydration shell of globular proteins (81,82). A previous MD study suggested that accurate representation of the hydration shell density by explicit-solvent models is critical for predicting SAXS curves of IDPs since even small variations of the hydration shell density may strongly influence predicted SAXS curves (19,65). MD simulations using TIP4P/2005 or TIP4P/2005s revealed that the water structure in the hydration shell of an IDP is perturbed relative to the bulk as indicated by a loss of tetrahedrality; however, this perturbation has been weaker as compared with the hydration shell of a folded protein (83). How the amino acid composition of an IDP controls the density of its hydration shell has not been systematically addressed.

As a model IDP, we here consider the XAO peptide with the sequence Ace-(diaminobutyric acid)$_2$-(Ala)$_7$-(ornithine)$_2$-NH_2, whose conformational ensemble has been studied by SAXS as well as by NMR and circular dichroism spectroscopy (75,76). We used maximum-entropy ensemble refinement against SAXS data taken from Ref. (75) to obtain the heterogeneous ensemble of XAO. We randomly selected 20 frames from the ensemble, thereby representing the heterogeneous ensemble, involving compact and expanded XAO conformations (see Materials and Methods and Fig. S1). In follow-up simulations of XAO and its mutants, these conformations were maintained using backbone restraints.

To reveal how the amino acid composition influences the hydration shell of an IDP, we mutated the four terminal XAO residues to 20 proteinogenic amino acids, again including histidine in the neutral and cationic form (21 mutants). For each XAO variant, we carried out 20 MD simulations with backbone restraints to the 20 conformations taken from the XAO ensemble (see above) and computed the SAXS curve using explicit-solvent SAXS calculations (Figs. S3 and S5). Following the analysis described above for the GB3 domain, we obtained the contrast of the hydration shell of the XAO wild-type and 21 mutants (22 variants), using the three water models TIP4P/2005, TIP4P/2005s, and TIP3P. We computed the overall contrast of the hydration shell relative to an equivalent volume of bulk water (Fig. 3 *C*) as well as the contrast per amino acid relative to alanine (Fig. 3 *D*).

At variance with the analysis for GB3, we found a negative hydration shell contrast for all 22 XAO variants, implying that the hydration shell of XAO contains fewer water molecules compared with bulk water. Whereas XAO wild-type and anionic mutants reveal a contrast of up to approximately -2 water molecules, apolar variants may reveal contrast of up to -7 water molecules or even fewer (Fig. 3 *C*). We rationalize the negative contrasts with the high flexibility of XAO, leading to more loosely packed conformations compared with the structure of the globular GB3. Thereby, small voids between the XAO backbone and side chains may exclude water molecules or enable only incomplete packing of water around XAO moieties, leading to a lower water density around XAO compared with GB3.

The effects of amino acid classes on the hydration shells are consistent with the findings for GB3; namely the contrasts follow roughly the order anionic–cationic–polar–apolar (Fig. 3 *D*). However, contrasts per amino acid differ quantitatively for XAO compared with our results for GB3. For instance, arginine residues imposed smaller contrasts in XAO compared with arginine in GB3 (compare Fig. 3 *B* with Fig. 3 *D*), possibly because the hydrophobic C_β, C_γ, and C_δ atoms of arginine near the XAO termini are more solvent-exposed compared with arginines at the GB3 surface.

The hydration shell of an IDP may reduce the radius of gyration detected by SAXS

In sharp contrast to the findings for GB3 variants, for which ΔR_g values were mostly positive in the range of approximately 0–1.5 Å, ΔR_g values for XAO variants with different force fields are mostly negative and may take large absolute values (Fig. 4 *C*). The radius of gyration detected by SAXS is given by

$$R_g^2 = \left(\int \Delta\rho(\mathbf{r})\, d\mathbf{r} \right)^{-1} \int r^2 \Delta\rho(\mathbf{r})\, d\mathbf{r}, \qquad (4)$$

where $\Delta\rho(\mathbf{r})$ denotes the electron density contrast, and r is the distance from the (contrast-weighted) center of mass. The large negative ΔR_g values up to -4 Å may be rationalized by 1) the overall small contrast of XAO (small value in brackets in Eq. (4)), leading to a large impact of the hydration shell contrast on R_g, and 2) hydration shell contributions to the contrast close to the center of mass (at small r), as occurring for extended conformations for which moieties near the center of mass are solvent-exposed. Positive ΔR_g values are found only for the anionic mutants for which Asp/Glu residues may impose large contrast near the endpoints of XAO, but also for the XAO wild-type and few polar variants, mostly with the TIP4P/2005 water model (Fig. 4 *C*, yellow bars). Thus, whereas the hydration shell of globular proteins often leads to an increased R_g detected by SAXS (17,23,43), the hydration shell of IDPs may also lower the R_g value.

Three-dimensional densities around XAO reveal amino acid- and force field-specific hydration shell structures

To rationalize the variations of contrast and R_g values in structural terms, we computed the three-dimensional solvent densities with the TIP4P/2005 water model around extended conformations of four XAO variants, in which the four terminal residues were mutated to serine, leucine, lysine, or glutamate, providing one representative each for

a polar, apolar, cationic, or anionic variant (Fig. 5 *E–H*). In line with the findings from the XAO ensemble (Figs. 3 *C* and *D* and 4 *C* and *D*), these mutations lead to large variations of the forward scattering I_0 (Fig. 5 *A*), contrast of the hydration shell between approximately −1 and −6 water molecules (Fig. 5 *B*), R_g values obtained from SAXS between ∼7.5 Å and ∼11.5 Å (Fig. 5 *C*, blue squares), and ΔR_g values between −3 Å and +1 Å (Fig. 5 *D*). These values align qualitatively with the variations of solvent densities near the endpoints of XAO. For instance, the solvent structure around the anionic glutamate residues reveals several tightly bound water molecules (Fig. 5 *H*, red spots) and a pronounced second hydration shell. In contrast, the hydration shell around the apolar leucine lacks any structured water molecules, whereas the second hydration shell is blurred out (Fig. 5 *F*). The serine and lysine variants yield intermediate cases (Fig. 5 *E* and *G*). Thus, the variations of contrast and R_g values quantified above are footprints of amino acid-specific hydration shell structures revealed by three-dimensional solvent densities.

The varying solvent structures around different amino acid types are also evident from radial distribution functions (RDFs) of water oxygen atoms with respect to heavy atoms of the side chains of the mutated amino acids. The RDFs for anionic residues reveal a pronounced peak at 2.65 Å, which is not present for any other residue, illustrating their exceptionally tightly bound hydration shell (Figs. 6, blue; S6 *D*; S7 *D*). The first peak in the RDFs of cationic residues is much smaller than that of anionic residues and is located at a larger distance (Figs. 6, orange; S6 *D*; S7 *D*), rationalizing the weaker hydration shell contrast (Fig. 3). RDFs of polar residues depend strongly on the amino acid species, reflecting the relative abundances of polar and hydrophobic moieties; for instance, water oxygen atoms around serine, and to a lesser degree around threonine, glutamine, or asparagine, show considerable density at 2.8 Å, whereas water is on average located at larger distances from cysteine or tyrosine (Figs. 6, black; S6 *A/B*; S7 *A/B*). The water depletion layer at apolar residues is illustrated by RDFs that are nearly zero below 3 Å (Figs. 6, green curve; S6 *C*; S7 *C*).

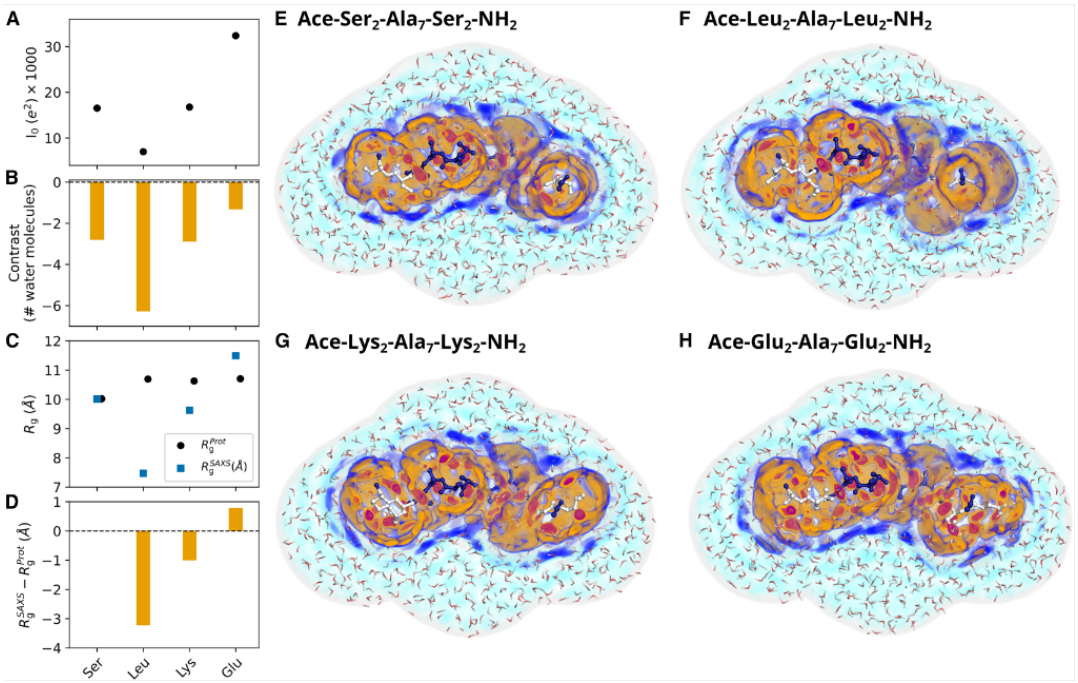


FIGURE 5 Comparison of hydration shells of XAO mutants with different types of mutated residues: Ser (polar), Leu (apolar), Lys (cationic), Glu (anionic). (*A*) Forward scattering I_0, (*B*) hydration shell contrast in number of water molecules, (*C*) radius of gyration R_g^{Prot} from the bare peptide (*black dots*) and R_g^{SAXS} from Guinier analysis (*blue squares*), and (*D*) $\Delta R_g = R_g^{\mathrm{SAXS}} - R_g^{\mathrm{Prot}}$. Values were calculated from simulations with the TIP4P/2005 water model in combination with the ff03w protein force field. (*E–H*) Shaded colors show the three-dimensional solvent density maps from 50 ns simulations around the mutants analyzed in (*A*)–(*D*) with color code taken from Fig. 1 *B* and *D*. MD simulations used to compute density maps were carried out with restraints on all heavy atoms, thereby yielding spatially well-defined hydration shells. The density is overlayed with one MD frame, showing the XAO in ball-and-stick representation and water molecules within the envelope as red/white lines. Color code according to Fig. 1 *B* and *D*. Example PyMOL scripts for visualizing the densities are provided in the supporting material.

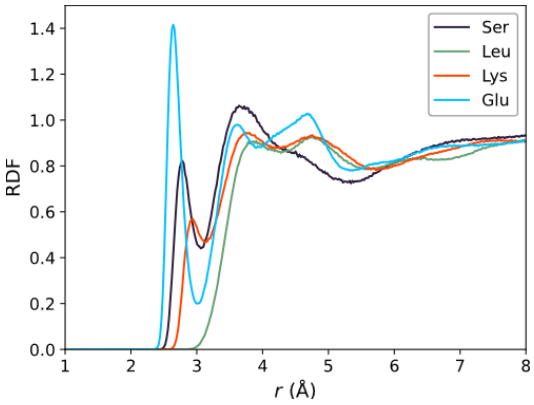


FIGURE 6 Radial distributions functions (RDFs) of water oxygens with respect to side chain heavy atoms of serine, leucine, lysine, or glutamate (*see legend*). RDFs were computed from simulations of XAO with TIP4P/2005, averaged over 20 XAO conformation. RDFs for other amino acids or computed from simulations of GB3 are shown in Figs. S6 and S7.

Moreover, the RDFs show that the hydration shell depends on the protonation state, as evident from more tightly bound water at doubly protonated histidine compared with singly protonated histidine (Fig. S8). Together, the RDFs demonstrate the highly distinct hydration shell structures around different amino acid types, and they align with the varying contrasts discussed above.

The solvent densities around the extended XAO peptide depend not only on amino acid composition but furthermore on the water model, as evident, for instance, from more pronounced solvent structures modeled by TIP4P/2005s compared with TIP3P (Fig. S9). Such force-field-dependent solvent densities rationalize variations of hydration shell contrasts by the ensembles of 22 XAO variants presented in Fig. 3 *C* and *D*, which qualitatively align with the findings for GB3 (Fig. 3 *A* and *B*). Specifically, TIP3P yields smaller (more negative) hydration shell contrasts relative to TIP4P/2005, whereas the increased protein-water dispersion interactions implemented by TIP4P/2005s yield larger (less negative) contrasts (Fig. 3 *C* and *D*, yellow, blue, or red bars). Notably, owing to the small contrast $\int \Delta\rho(\mathbf{r})\, d\mathbf{r}$ of the overall XAO, ΔR_g values are highly sensitive with respect to water model-imposed variations of the hydration shell, leading to large ΔR_g variations by up to 2 Å and even more (Fig. 4 *C* and *D*, yellow, blue, and red bars; cf. Eq. (4)).

Hydration shell effect on R_g strongly depends on IDP conformation

To test how the peptide conformation influences the hydration shell of XAO, we computed hydration shell contrasts and ΔR_g values for 20 representative conformations of the aspartate mutant of XAO (Fig. S10) and analyzed two example conformations—one compact and one extended conformation—in detail in Fig. 7. We found that peptide conformations have only a small effect on the contrast (Fig. 7 *A* and *B*) but may impose large variations of ΔR_g by up to ∼0.8 Å (Figs. 7 *C* and *D* and S10). Notably, ΔR_g values neither correlate significantly with R_g values nor with I_0, suggesting that subtle details of the peptide and hydration shell geometries determine the ΔR_g value.

How solvent-exposed chemical moieties determine the hydration shell contrast

The hydration shell contrasts computed for all proteinogenic amino acids (Fig. 3) enabled us to quantify how chemical modifications of solvent-exposed moieties alter the hydration shell contrast. Table 1 lists 12 chemical modifications, together with the accompanying changes of contrasts $\Delta\Delta N_e^{hs}$ in number of electrons, as taken from the GB3 or XAO simulations with TIP4P/2005. Since one water molecule contains 10 electrons, the $\Delta\Delta N_e^{hs}$ values may be translated to number of water molecules by dividing by 10. The values are qualitatively consistent among the analysis from GB3 and XAO. However, values differ quantitatively, likely reflecting that the degree to which these moieties are solvent-exposed differs between GB3 and XAO.

The analysis enables the following conclusions: 1) replacing a hydrogen (H) with a methyl group (CH_3) decreases the contrast by about one electron (Table 1, row 2). 2) Replacing a methyl (CH_3) with a hydroxyl group (OH) increases the contrast by about three electrons (rows 1, 2, 10). 3) Replacing carbamoyl group ($CONH_2$) as present in Asn or Gln with a carboxyl group (COO^-) as present in Asp or Glu leads a marked increase of the contrast between 5 and 14 electrons (rows 7, 8). 4) Extending a side chain by the addition of a CH_2 group increases the contrast for a polar side chain (e.g., Asp → Glu, row 5) and decreases the contrast for apolar side chains (row 4), rationalized by the fact that longer side chains are more solvent-exposed. In other words, chemical modifications at the tip of longer side chains have a larger effect on the hydration shell contrast compared with modifications at shorter side chains. Together, these values quantify how solvent-exposed chemical moieties modulate the electron density of the hydration shell of globular proteins or IDPs.

DISCUSSION

We quantified the effect of the 20 proteinogenic amino acids on the protein hydration shell density of the GB3 domain and the XAO peptide as representatives for the classes of globular or intrinsically disorderd proteins with focus on two quantities encoded by SAXS curves: 1) the forward scattering intensity I_0, which reports on the overall contrast between the protein—including its hydration shell—relative to the buffer; and 2) the radius of gyration R_g, which is modified by the hydration shell relative to the R_g of the bare protein.

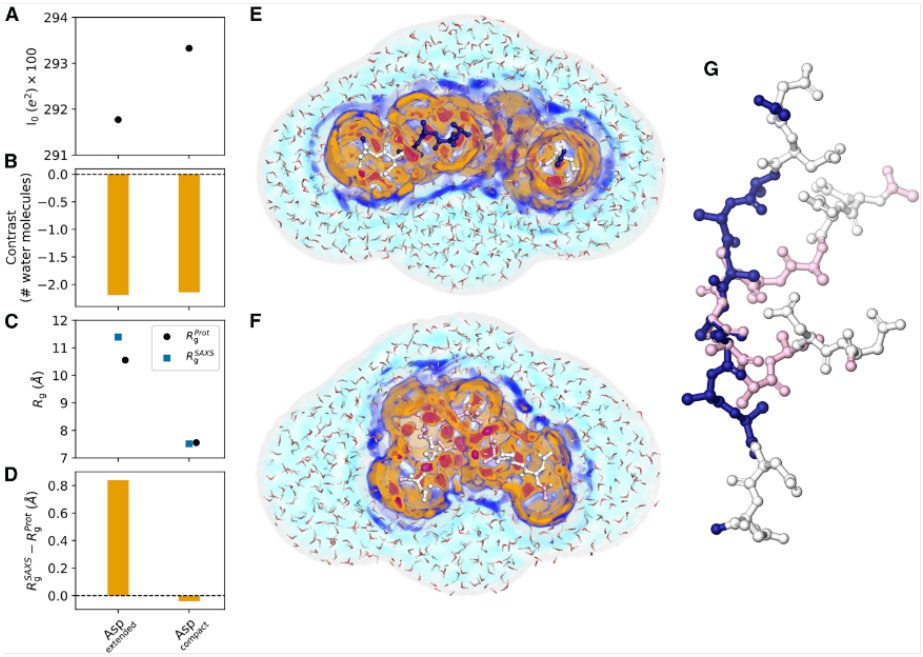


FIGURE 7 Effects of peptide conformation on the hydration shell. (*A*) Forward scattering I_0, (*B*) hydration shell contrast, (*C*) R_g, and (*D*) ΔR_g values of one extended and one compact conformation of the aspartate mutant of XAO obtained with TIP4P/2005 and ff03w. (*E* and *F*) Shaded colors show the three-dimensional solvent density maps from 50 ns MD simulations around the XAO mutant analyzed in (*A*)–(*D*) with color code taken from Fig. 1 *B* and *D*. The density is overlayed with one MD frame, showing the XAO as balls/sticks and water molecules within the envelope as red/white lines. MD simulations used to compute density maps were carried out with restraints on all heavy atoms, thereby yielding spatially well-defined hydration shells. (*G*) Extended (*blue/white*) and compact conformation (*pink/white*) of XAO with four terminal residues mutated to aspartate (*white*).

Across both proteins and three different water models, we observed a consistent trend in hydration shell density at solvent-exposed amino acids: anionic > cationic > polar > apolar residues. Substituting alanine with an anionic residue at the GB3 surface, the hydration shell density increased considerably and contained approximately 1–1.5 additional water molecules relative to the bulk density. Given that 1) water is nearly incompressible and 2) amino acids are in contact with only a few water molecules, these values demonstrate a highly condensed packing of water at anionic amino acids and rationalize the presence of a marked hydration shell around anionic proteins (20,78). In contrast, replacing alanine with bulkier hydrophobic residues led to a decreased hydration shell density, which supports the presence of a water depletion layer previously reported for hydrophobic surfaces (36–38).

Although the amino acid-specific effects on the hydration shell are qualitatively consistent between the GB3 domain and the XAO peptide, we also observed distinct differences in their hydration shells and their effects on SAXS curves. The hydration shell contrast was positive for the GB3 wild type and for most polar or charged GB3 mutants (Fig. 3 *A*), indicating tightly packed water on the protein surface. In contrast, the hydration shell of most XAO variants revealed a negative contrast, indicating the presence of water depletion layers at the XAO surface (Fig. 3 *C*). Thus, hydration shells of IDPs may differ substantially from hydration shells of globular proteins.

Furthermore, our calculations revealed that the effect of the hydration shell on R_g strongly depends on the conformation of XAO. This finding can be rationalized by the fact that the R_g of XAO is sensitive to the spatial distribution of hydration shell contrast: contrasts located farther from the XAO center of mass have greater effect on R_g compared with contrasts closer the center of mass. Thus, the structure of the hydration shell and its effects on SAXS curves is controlled by an interplay between amino acid composition and peptide conformation.

Computing the hydration shell contrast from I_0 values requires a definition of the protein volume or, equivalently, of the protein-water dividing surface. Indeed, different conventions for the dividing surface, for instance based on different Voronoi tessellation schemes, have led to slightly different protein volumes and different estimates for the hydration shell density (77,84–87). Accordingly, we observed that different choices for computing the dividing surface influence the absolute hydration shell contrasts; however, the

TABLE 1 Change of Hydration Shell Contrast in Number of Electrons upon Various Chemical Modifications of Solvent-Exposed Amino Acids

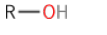
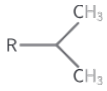
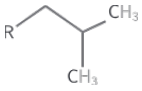
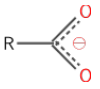
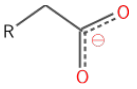
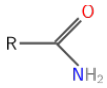
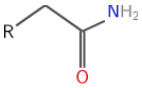
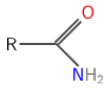
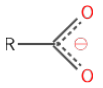
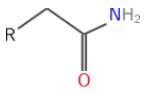
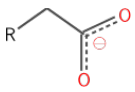
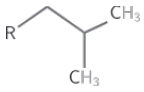
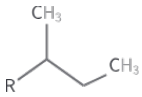
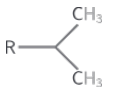
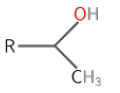
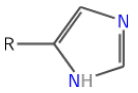
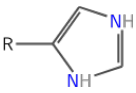
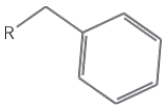
	Chemical Modification		GB3	XAO
	From	To	$\Delta\Delta N_e^{hs}$	$\Delta\Delta N_e^{hs}$
1			1.8 ± 0.5	2.2 ± 0.3
2			−1.3 ± 0.4	−0.8 ± 0.2
3			3.7 ± 0.3	3.6 ± 0.2
4			−2.2 ± 0.4	−2.9 ± 0.2
5			4.8 ± 0.4	2.8 ± 0.2
6			0.8 ± 0.4	−0.1 ± 0.2
7			10.2 ± 0.4	5.3 ± 0.2

Table 1. ***Continued***

	Chemical Modification		GB3	XAO
	From	To	$\Delta\Delta N_e^{hs}$	$\Delta\Delta N_e^{hs}$
8			14.2 ± 0.4	8.1 ± 0.2
9			2.3 ± 0.4	2.2 ± 0.2
10			3.3 ± 0.4	2.3 ± 0.2
11			2.6 ± 0.4	−0.1 ± 0.2
12			−2.2 ± 0.4	−2.1 ± 0.1

Errors denote 1 SE.

choices have only a marginal effect on the contrasts relative to alanine found in this study (Figs. S4 and S14). We speculate that different volume definitions might explain why a previous study found similar water volumes at anionic, cationic, and polar moieties (77), whereas our analysis suggested by far larger contrasts imposed by the anionic Asp/Glu residues. Critically, the pronounced hydration shells by Asp/Glu are confirmed by our ΔR_g calculation, which do not require assumptions on the dividing surface. In addition, they align with previous experimental SAXS/SANS studies of super-charged GFP variants and with large ΔR_g values found for the highly anionic glucose isomerase (46,68).

We previously found that MD simulations with certain but not all force fields accurately reproduce the difference in R_g between SAXS relative to SANS in D_2O, suggesting that they capture the hydration shell contrast of several globular proteins (20). To experimentally test our amino acid-specific contrast values reported here, future SAXS/SANS studies of proteins and their mutants would be desirable. Designed ultrastable proteins may be particularly suitable for such studies, as their conformations are relatively insensitive to surface substitutions (88). For IDPs, whose ensembles are strongly amino acid dependent, careful SAXS/SANS measurements in both H_2O and D_2O could provide a means to isolate amino acid-specific effects on the hydration shell.

We anticipate that the residue-resolved hydration shell contrast scores derived for all 20 proteinogenic amino acids will be useful for several future developments. Contrast scores may be used to parameterize computationally efficient SAXS curve predictions that account for residue-specific hydration while avoiding the need for explicit-solvent MD simulations for each protein conformation. Thereby, our calculations may bridge the gap between accurate yet computationally expensive explicit-solvent SAXS calculations (43,66,71,89–92) and simplified implicit-solvent methods that require fitting of the hydration shell against experimental data (93,94). In addition, quantifying residue-specific hydration will be key to understanding how targeted modulation of the water structure by protein-water interactions promotes biomolecular function, for instance in contexts of antifreeze proteins, molecular recognition, or biomolecular phase separation (95–97).

DATA AND CODE AVAILABILITY

A modified GROMACS version that implements the explicit-solvent SAXS calculations is available at https://gitlab.com/cbjh/gromacs-swaxs.

ACKNOWLEDGMENTS

We thank Jan Lipfert for stimulating discussions and for sharing SAXS data of XAO. This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via grants HU 1971/3-1 and INST 256/539-1.

AUTHOR CONTRIBUTIONS

J.-B.L. and T.M.F. performed research and analyzed data. J.S.H. designed research and contributed software. J.-B.L. and J.S.H. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

J.B.L. and J.S.H. acknowledge the assistance of ChatGPT-4 for providing proofreading support during the writing process.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2025.11.2683.

REFERENCES

1. Bellissent-Funel, M.-C., A. Hassanali, …, A. E. Garcia. 2016. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* 116:7673–7697.
2. Levy, Y., and J. N. Onuchic. 2006. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* 35:389–415.
3. Ball, P. 2008. Water as an Active Constituent in Cell Biology. *Chem. Rev.* 108:74–108.
4. Daidone, I., M. B. Ulmschneider, …, J. C. Smith. 2007. Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proc. Natl. Acad. Sci. USA.* 104:15230–15235.
5. Fogarty, A. C., and D. Laage. 2014. Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation. *J. Phys. Chem. B.* 118:7715–7729.
6. Mondal, S., S. Mukherjee, and B. Bagchi. 2017. Protein Hydration Dynamics: Much Ado about Nothing? *J. Phys. Chem. Lett.* 8:4878–4882.
7. Fogarty, A. C., E. Duboué-Dijon, …, D. Laage. 2013. Biomolecular hydration dynamics: a jump model perspective. *Chem. Soc. Rev.* 42:5672–5683.
8. Bagchi, B. 2005. Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem. Rev.* 105:3197–3219.
9. Wüthrich, K., M. Billeter, …, G. Wider. 1996. NMR studies of the hydration of biological macromolecules. *Faraday Discuss.* 103:245–253.
10. Crilly, C. J., J. E. Eicher, …, G. J. Pielak. 2021. Water's Variable Role in Protein Stability Uncovered by Liquid-Observed Vapor Exchange NMR. *Biochemist.* 60:3041–3045.
11. Laage, D., T. Elsaesser, and J. T. Hynes. 2017. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* 117:10694–10725.
12. Pal, S. K., J. Peon, …, A. H. Zewail. 2002. Biological Water: Femtosecond Dynamics of Macromolecular Hydration. *J. Phys. Chem. B.* 106:12376–12395.
13. Ebbinghaus, S., S. J. Kim, …, M. Havenith. 2007. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci. USA.* 104:20749–20752.
14. Born, B., S. J. Kim, …, M. Havenith. 2009. The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss.* 141:161–207.
15. Sushko, O., R. Dubrovka, and R. S. Donnan. 2015. Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected. *J. Chem. Phys.* 142:055101.
16. Li, T., A. A. Hassanali, …, S. J. Singer. 2007. Hydration Dynamics and Time Scales of Coupled Water-Protein Fluctuations. *J. Am. Chem. Soc.* 129:3376–3382.
17. Svergun, D. I., S. Richard, …, G. Zaccai. 1998. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95:2267–2272.
18. Merzel, F., and J. C. Smith. 2002. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci. USA.* 99:5378–5383.
19. Henriques, J., L. Arleth, …, M. Skepö. 2018. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* 430:2521–2539.
20. Linse, J.-B., and J. S. Hub. 2023. Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data. *Commun. Chem.* 6:272.
21. Zimmerman, S. B., and S. O. Trach. 1991. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* 222:599–620.
22. Harada, R., Y. Sugita, and M. Feig. 2012. Protein Crowding Affects Hydration Structure and Dynamics. *J. Am. Chem. Soc.* 134:4842–4849.
23. Kim, H. S., and F. Gabel. 2015. Uniqueness of models from small-angle scattering data: the impact of a hydration shell and complementary NMR restraints. *Acta Crystallogr. D.* 71:57–66.
24. Biedermannová, L., and B. Schneider. 2015. Structure of the Ordered Hydration of Amino Acids in Proteins: Analysis of Crystal Structures. *Acta Crystallogr. D.* 71:2192–2202.
25. Hecht, D., L. Tadesse, and L. Walters. 1993. Correlating Hydration Shell Structure with Amino Acid Hydrophobicity. *J. Am. Chem. Soc.* 115:3336–3337.
26. Ide, M., Y. Maeda, and H. Kitano. 1997. Effect of Hydrophobicity of Amino Acids on the Structure of Water. *J. Phys. Chem. B.* 101:7022–7026.
27. Djikaev, Y. S., and E. Ruckenstein. 2011. The variation of the number of hydrogen bonds per water molecule in the vicinity of a hydrophobic surface and its effect on hydrophobic interactions. *Curr. Opin. Colloid Interface Sci.* 16:272–284.
28. Meral, D., S. Toal, …, B. Urbanc. 2015. Water-Centered Interpretation of Intrinsic pPII Propensities of Amino Acid Residues: In Vitro-Driven Molecular Dynamics Study. *J. Phys. Chem. B.* 119:13237–13251.
29. Lanza, G., and M. A. Chiacchio. 2020. The water molecule arrangement over the side chain of some aliphatic amino acids: A quantum chemical and bottom-up investigation. *Int. J. Quant. Chem.* 120:e26161.
30. Robinson, G. W., and C. H. Cho. 1999. Role of Hydration Water in Protein Unfolding. *Biophys. J.* 77:3311–3318.
31. Collins, K. D., and M. W. Washabaugh. 1985. The Hofmeister effect and the behaviour of water at interfaces. *Q. Rev. Biophys.* 18:323–422.
32. Walrafen, G. E., and Y.-C. Chu. 2000. Nature of collagen–water hydration forces: a problem in water structure. *Chem. Phys.* 258:427–446.

33. Esser, A., H. Forbert, …, D. Marx. 2018. Hydrophilic Solvation Dominates the Terahertz Fingerprint of Amino Acids in Water. *J. Phys. Chem. B.* 122:1453–1459.
34. Pertsemlidis, A., A. M. Saxena, …, R. M. Glaeser. 1996. Direct evidence for modified solvent structure within the hydration shell of a hydrophobic amino acid. *Proc. Natl. Acad. Sci. USA.* 93:10769–10774.
35. Grdadolnik, J., F. Merzel, and F. Avbelj. 2017. Origin of hydrophobicity and enhanced water hydrogen bond strength near purely hydrophobic solutes. *Proc. Natl. Acad. Sci. USA.* 114:322–327.
36. Doshi, D. A., E. B. Watkins, …, J. Majewski. 2005. Reduced Water Density at Hydrophobic Surfaces: Effect of Dissolved Gases. *Proc. Natl. Acad. Sci. USA.* 102:9458–9462.
37. Poynor, A., L. Hong, …, P. A. Fenter. 2006. How Water Meets a Hydrophobic Surface. *Phys. Rev. Lett.* 97:266101.
38. Janeček, J., and R. R. Netz. 2007. Interfacial Water at Hydrophobic and Hydrophilic Surfaces: Depletion versus Adsorption. *Langmuir.* 23:8417–8429.
39. Schrag, L. G., X. Liu, …, J. Chen. 2021. Cancer-Associated Mutations Perturb the Disordered Ensemble and Interactions of the Intrinsically Disordered P53 Transactivation Domain. *J. Mol. Biol.* 433:167048.
40. Flores, E., N. Acharya, …, S. Sukenik. 2025. Single-Point Mutations in Disordered Proteins: Linking Sequence, Ensemble, and Function. *Curr. Opin. Struct. Biol.* 91:102987.
41. Gu, H., N. Doshi, …, D. Baker. 1999. Robustness of Protein Folding Kinetics to Surface Hydrophobic Substitutions. *Protein Sci.* 8:2734–2741.
42. Hill, R. B., and W. F. DeGrado. 2000. A Polar, Solvent-Exposed Residue Can Be Essential for Native Protein Structure. *Structure.* 8:471–479.
43. Chen, P.-c., and J. S. Hub. 2014. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* 107:435–447.
44. Chen, P.-C., R. Shevchuk, …, J. S. Hub. 2019. Combined Small-Angle X-ray and Neutron Scattering Restraints in Molecular Dynamics Simulations. *J. Chem. Theor. Comput.* 15:4687–4698.
45. Chatzimagas, L., and J. S. Hub. 2022. Predicting solution scattering patterns with explicit-solvent molecular simulations. *Methods Enzymol.* 677:433–456.
46. Trewhella, J., P. Vachette, and X. Zuo. 2022. A Round-Robin Approach Provides a Detailed Assessment of Biomolecular Small-Angle Scattering Data Reproducibility and Yields Consensus Curves for Benchmarking. *Acta Crystallogr. D.* 78:1315–1336.
47. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
48. Derrick, J. P., and D. B. Wigley. 1994. The Third IgG-Binding Domain from Streptococcal Protein G: An Analysis by X-ray Crystallography of the Structure Alone and in a Complex with Fab. *J. Mol. Biol.* 243:906–918.
49. Pettersen, E. F., T. D. Goddard, …, T. E. Ferrin. 2004. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.
50. Abraham, M. J., T. Murtola, …, E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 1–2:19–25.
51. Duan, Y., C. Wu, …, P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.
52. Best, R. B., and G. Hummer. 2009. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B.* 113:9004–9015.
53. Best, R. B., and J. Mittal. 2010. Balance between α and β Structures in Ab Initio Protein Folding. *J. Phys. Chem. B.* 114:8790–8798.
54. Best, R. B., W. Zheng, and J. Mittal. 2014. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theor. Comput.* 10:5113–5124.
55. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79:926–935.
56. Abascal, J. L. F., and C. Vega. 2005. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* 123:234505.
57. Hockney, R. W., S. P. Goel, and J. W. Eastwood. 1974. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* 14:148–158.
58. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
59. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
60. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
61. Miyamoto, S., and P. A. Kollman. 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.
62. Hess, B. 2008. A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theor. Comput.* 4:116–122.
63. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
64. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
65. Hermann, M. R., and J. S. Hub. 2019. SAXS-Restrained Ensemble Simulations of Intrinsically Disordered Proteins with Commitment to the Principle of Maximum Entropy. *J. Chem. Theor. Comput.* 15:5103–5115.
66. Knight, C. J., and J. S. Hub. 2015. WAXSiS: A Web Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics. *Nucleic Acids Res.* 43:W225–W230.
67. Chatzimagas, L., and J. S. Hub. 2023. Structure and ensemble refinement against SAXS data: Combining MD simulations with Bayesian inference or with the maximum entropy principle. *Methods Enzymol.* 678:23–54.
68. Linse, J.-B., and J. S. Hub. 2023. Scrutinizing the Protein Hydration Shell from Molecular Dynamics Simulations against Consensus Small-Angle Scattering Data. *Commun. Chem.* 6:272.
69. Cordeiro, T. N., P.-c. Chen, …, P. Bernadó. 2017. Disentangling Polydispersity in the PCNA-p15PAF Complex, a Disordered, Transient and Multivalent Macromolecular Assembly. *Nucleic Acids Res.* 45:1501–1515.
70. Chen, P.-c., and J. S. Hub. 2015. Structural Properties of Protein-Detergent Complexes from SAXS and MD Simulations. *J. Phys. Chem. Lett.* 6:5116–5121.
71. Ivanović, M. T., M. R. Hermann, …, J. S. Hub. 2020. Small-Angle X-ray Scattering Curves of Detergent Micelles: Effects of Asymmetry, Shape Fluctuations, Disorder, and Atomic Details. *J. Phys. Chem. Lett.* 11:945–951.
72. Ivanović, M. T., L. K. Bruetzel, …, J. S. Hub. 2018. Temperature-Dependent Atomic Models of Detergent Micelles Refined against Small-Angle X-ray Scattering Data. *Angew. Chem. Int. Ed.* 57:5635–5639.
73. Cromer, D. T., and J. B. Mann. 1968. X-ray scattering factors computed from numerical Hartree–Fock wave functions. *Acta Crystallogr. A.* 24:321–324.
74. Voss, N. R., and M. Gerstein. 2010. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* 38:W555–W562.
75. Zagrovic, B., J. Lipfert, …, V. S. Pande. 2005. Unusual Compactness of a Polyproline Type II Structure. *Proc. Natl. Acad. Sci. USA.* 102:11698–11703.
76. Makowska, J., S. Rodziewicz-Motowidło, …, H. A. Scheraga. 2006. Polyproline II Conformation Is One of Many Local Conformational States and Is Not an Overall Conformation of Unfolded Peptides and Proteins. *Proc. Natl. Acad. Sci. USA.* 103:1744–1749.

77. Persson, F., P. Söderhjelm, and B. Halle. 2018. The Geometry of Protein Hydration. *J. Chem. Phys.* 148:215101.

78. Kim, H. S., A. Martel, …, F. Gabel. 2016. SAXS/SANS on Supercharged Proteins Reveals Residue-Specific Modifications of the Hydration Shell. *Biophys. J.* 110:2185–2194.

79. Linse, J.-B., and J. S. Hub. 2021. Three- and Four-Site Models for Heavy Water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW. *J. Chem. Phys.* 154:194501.

80. Gallat, F.-X., A. Laganowsky, …, M. Weik. 2012. Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. *Biophys. J.* 103:129–136.

81. Taricska, N., M. Bokor, …, A. Perczel. 2019. Hydration Shell Differentiates Folded and Disordered States of a Trp-cage Miniprotein, Allowing Characterization of Structural Heterogeneity by Wide-Line NMR Measurements. *Sci. Rep.* 9:2947.

82. Waszkiewicz, R., A. Michaś, …, A. Niedzwiecka. 2024. Hydrodynamic Radii of Intrinsically Disordered Proteins: Fast Prediction by Minimum Dissipation Approximation and Experimental Validation. *J. Phys. Chem. Lett.* 15:5024–5033.

83. Vural, D., U. R. Shrestha, …, J. C. Smith. 2023. Water Molecule Ordering on the Surface of an Intrinsically Disordered Protein. *Biophys. J.* 122:4326–4335.

84. Merzel, F., and J. C. Smith. 2002. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? *Proc. Natl. Acad. Sci. USA.* 99:5378–5383.

85. Neumayr, G., T. Rudas, and O. Steinhauser. 2010. Global and Local Voronoi Analysis of Solvation Shells of Proteins. *J. Chem. Phys.* 133:084108.

86. Voloshin, V. P., N. N. Medvedev, …, A. Geiger. 2011. Volumetric Properties of Hydrated Peptides: Voronoi–Delaunay Analysis of Molecular Simulation Runs. *J. Phys. Chem. B.* 115:14217–14228.

87. Smolin, N., and R. Winter. 2004. Molecular Dynamics Simulations of Staphylococcal Nuclease: Properties of Water at the Protein Surface. *J. Phys. Chem. B.* 108:15928–15937.

88. Kuhlman, B., G. Dantas, …, D. Baker. 2003. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science.* 302:1364–1368.

89. Merzel, F., and J. C. Smith. 2002. SASSIM: A Method for Calculating Small-Angle X-ray and Neutron Scattering and the Associated Molecular Envelope from Explicit-Atom Models of Solvated Proteins. *Acta Crystallogr. D.* 58:242–249.

90. Oroguchi, T., and M. Ikeguchi. 2012. MD–SAXS Method with Nonspherical Boundaries. *Chem. Phys. Lett.* 541:117–121.

91. Park, S., J. P. Bardhan, …, L. Makowski. 2009. Simulated X-Ray Scattering of Protein Solutions Using Explicit-Solvent Models. *J. Chem. Phys.* 130:134114.

92. Köfinger, J., and G. Hummer. 2013. Atomic-Resolution Structural Information from Scattering Experiments on Macromolecules in Solution. *Phys. Rev. E.* 87:052712.

93. Svergun, D., C. Barberato, and M. H. J. Koch. 1995. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* 28:768–773.

94. Schneidman-Duhovny, D., M. Hammel, and A. Sali. 2010. A Web Server for Rapid Computation and Fitting of SAXS Profiles. *Nucleic Acids Res.* 38:W540–W544.

95. Xu, Y., A. Bäumer, …, M. Havenith. 2016. Protein–Water Dynamics in Antifreeze Protein III Activity. *Chem. Phys. Lett.* 647:1–6.

96. Ribeiro, S. S., N. Samanta, …, J. C. Marcos. 2019. The Synergic Effect of Water and Biomolecules in Intracellular Phase Separation. *Nat. Rev. Chem.* 3:552–561.

97. Maurer, M., and C. Oostenbrink. 2019. Water in Protein Hydration and Ligand Recognition. *J. Mol. Recogn.* 32:e2810.

Biophysical Journal, Volume *125*

Supplemental information

How protein hydration depends on amino acid composition, peptide conformation, and force fields

Johanna-Barbara Linse, Tobias M. Fischbach, and Jochen S. Hub

Supporting Information for: How protein hydration depends on amino acid composition, peptide conformation, and force fields

Johanna-Barbara Linse, Tobias M. Fischbach, and Jochen S. Hub*

Theoretical Physics and Center for Biophysics, Saarland University, 66123 Saarbrücken, Germany

E-mail: jochen.hub@uni-saarland.de

Supporting Information Methods

Maximum-entropy refinement with the XAO peptide ensemble against SAXS data

To obtain a set of 20 XAO conformations that are representative for the XAO solution ensemble, we carried out SAXS-restrained ensemble simulations with commitment to the maximum entropy principle.[1] Four parallel XAO simulation replicas were coupled on-the-fly to the SAXS curve taken from Ref. 2. Simulations we carried out with GROMACS-SWAXS, version 2021.5, as freely available at https://gitlab.com/cbjh/gromacs-swaxs. Documentation for GROMACS-SWAXS is available at https://cbjh.gitlab.io/gromacs-swaxs-docs.

Four starting structures for the SAXS-restrained simulations were taken from the XAO ensemble refined against NMR data by Makowska et al.[3] and set up for simulations as

described in the Methods. To couple the simulations to the SAXS data, SAXS curves were computed from the simulations on-the-fly using explicit-solvent SAXS calculations, thereby taking scattering contributions from the hydration shell into account.[4–6] SAXS curves were averaged on-the-fly using a memory kernel that decays exponentially into to the past using a memory time of 100 ps [molecular dynamics parameter (mdp) option `waxs-tau`]. A q range from 0.065 Å^{-1} to 0.58 Å^{-1} with 30 equally-spaced q-point was used (mdp options `waxs-startq, waxs-endq, waxs-nq`). The SAXS curve was updated every 125 ps (mdp options `waxs-nstcalc` together with `dt`). A force constant of unity was applied, and the restraints were turned on gradually over 10 ns (mdp options `waxs-fc, waxs-t-target`). During the simulations, and prior to computing SAXS-derived forces, the experimental SAXS curve was fitted to the calculated curve via $I_{exp,fit}(q) = f \cdot I_{exp} + c$, by minimizing χ^2 with respect to the calculated curve. Here the factor f accounts for the overall scale, and the offset c accounts for a putative uncertainty from the buffer subtraction. No fitting parameters owing to the hydration layer or excluded solvent were used, implying that also the radius of gyration was not adjusted by the fitting parameters. The agreement of the SAXS curve obtained from the refined XAO ensemble with the experimental data is shown in Fig. S1.

SAXS-restrained simulations were carried out for 150 ns. The spatial envelope was built at a distance of 12 Å from all XAO atoms during free simulation that started from four different structures. Solvent atoms within the envelope contributed to the calculated SAXS curve as described previously.[5] The temperature was controlled at 298.15 K using a stochastic dynamics integrator.[7] All other simulation parameters were chosen as described in the Methods.

From the four trajectories collected from the four parallel replicas, five configurations each were taken from the simulation times 30 ns, 60 ns, 90 ns, 120 ns and 150 ns, thereby providing 20 independent conformations. These conformations were mutated as described in the main text and used for follow-up SAXS calculations. During follow-up simulations, these 20 conformations were maintained by applying positions restraints either to the backbone or

to heavy atoms, as described above.

On the calculation of the radius of gyration $R_{\mathrm{g}}^{\mathrm{Prot}}$ of the bare protein

In this study, the radius of gyration $R_{\mathrm{g}}^{\mathrm{Prot}}$ was computed from the coordinates of the atoms $\mathbf{r}_i$, weighted by the number of electrons of the atoms n_i:

$$(R_{\mathrm{g}}^{\mathrm{Prot}})^2 = N^{-1} \sum_i n_i (\mathbf{r}_i - \mathbf{r}_{\mathrm{com}})^2,$$

where $\mathbf{r}_{\mathrm{com}} = N^{-1} \sum_i n_i \mathbf{r}_i$ is the respective electron-weighted center of mass and $N = \sum_i n_i$ the total number of electrons. The sum runs over all protein atoms. This value differs marginally from the R_{g} that would be obtained from the protein electron density,

$$(R_{\mathrm{g}}^{\mathrm{Prot},\rho})^2 = N^{-1} \int \rho(\mathbf{r}) (\mathbf{r} - \mathbf{r}_{\mathrm{com}})^2 \, \mathrm{d}\mathbf{r},$$

because the electrons are spatially distributed around the nuclei. The difference between these two values may be estimated with a simple model: (a) Let a set of electrons be positioned at $\mathbf{R}_0 = (R_0, 0, 0)$, implying that the radius of gyration $R_{\mathrm{g,m}}$ with respect to the origin is R_0. (b) Alternatively, let the electrons be distributed as a three-dimensional Gaussian of width σ centered at $\mathbf{R}_0$, where $\sigma \ll R_0$. Then, the radius of gyration is given via

$$(R'_{\mathrm{g,m}})^2 = \frac{1}{(2\pi)^{(3/2)} \sigma^3} \int \mathbf{r}^2 \, \mathrm{e}^{-(\mathbf{r}-\mathbf{R}_0)^2/2\sigma^2} \mathrm{d}^3\mathbf{r} = R_0^2 + 3\sigma^2$$

and the difference in radius of gyration between cases (a) and (b) is $R'_{\mathrm{g,m}} - R_{\mathrm{g,m}} \approx 3\sigma^2/(2R_0)$. Thus, upon smearing out electrons around the position $\mathbf{R}_0$, the R_{g} value is slightly increased.

As a numerical example, let $R_0 = 11\,\text{Å}$, corresponding approximately to the R_{g} value of the GB3 domain, and $\sigma = 0.4\,\text{Å}$, modeling an electron density distribution that, very approximately, resembles the electron density distribution of carbon. Then, $R'_{\mathrm{g,m}} - R_{\mathrm{g,m}} = 0.024\,\text{Å}$. (A more precise calculation using electron densities from Slater-type orbitals yields

similar values.) To test these estimates, we computed $R_g^{Prot,\rho}$ from a simulation of GB3, obtained with a Guinier analysis of a SAXS curve computed after setting the form factors of all solvent atoms to zero. We obtained $R_g^{Prot,\rho} = 11.032\,\text{Å}$, whereas the value from the atomic coordinates was $R_g^{Prot} = 11.002\,\text{Å}$. The difference is in reasonable agreement with the simple model, suggesting that the model provides a reasonable order-of-magnitude estimate for the increase of $R_g^{Prot,\rho}$ relative to R_g^{Prot}.

Together, this analysis suggests that R_g^{Prot} provides a good approximation to $R_g^{Prot,\rho}$. Thus, in this study, for the sake of simplicity, we used R_g^{Prot} to quantify the R_g value of the protein.

Supporting Information Figures

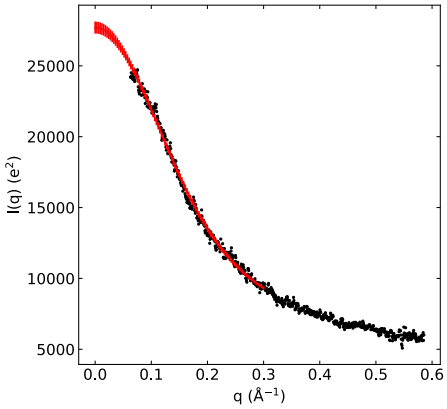


Figure S1: Experimental SAXS data by Zagrovic *et al.*[2] (black dots) and SAXS curve of XAO ensemble obtained by maximum-entropy ensemble refinement (red curve).[1] From the refined XAO ensemble, 20 frames were selected as representative conformations of the heterogeneous XAO ensemble and used subsequently for computing SAXS curves of XAO mutant. The experimental SAXS curve represents the extrapolation to infinite dilution from a set of experiments in 100 mM acetate buffer at 15, 10, 5, 2.5, and 1.25 mg/ml. SAXS data kindly provided by Jan Lipfert.[2]

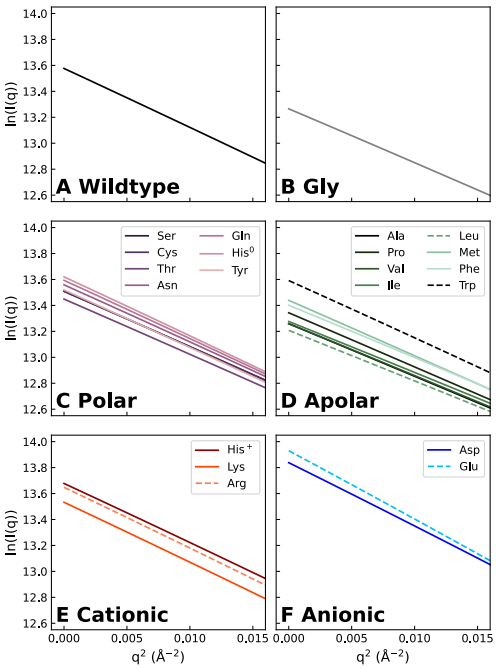


Figure S2: Guinier plots of SAXS curves of the GB3 domain from explicit-solvent SAXS calculations with the TIP4P/2005 water model in combination with the ff03w protein force field. Same SAXS curves as shown in Fig. 2A–F, however plotted as $\ln(I(q))$ vs. q^2.

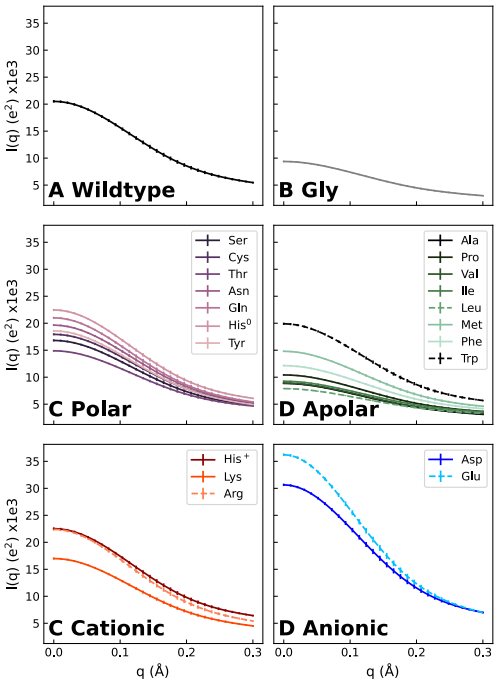


Figure S3: SAXS curves of the heterogeneous ensemble of the XAO peptide from explicit-solvent SAXS calculations with the TIP4P/2005 water model in combination with the ff03w protein force field. Backbone positions were restrained in simulations for all XAO mutants to the backbone positions of the XAO wild type ensemble refined against experimental SAXS data (see Fig. S1), suggesting that variations among the computed SAXS curves are purely caused by presence of four different amino acids (at fixed backbone positions) and by variations of the hydration shell. SAXS curves are shown (A) for the XAO wild type and (B–F) for for 21 mutants with four mutated surface-exposed amino acids each (for color code and line style, see legends). For clarity, SAXS curves are grouped by the amino acid property (glycine, polar, apolar, cationic, anionic) in panels B–F.

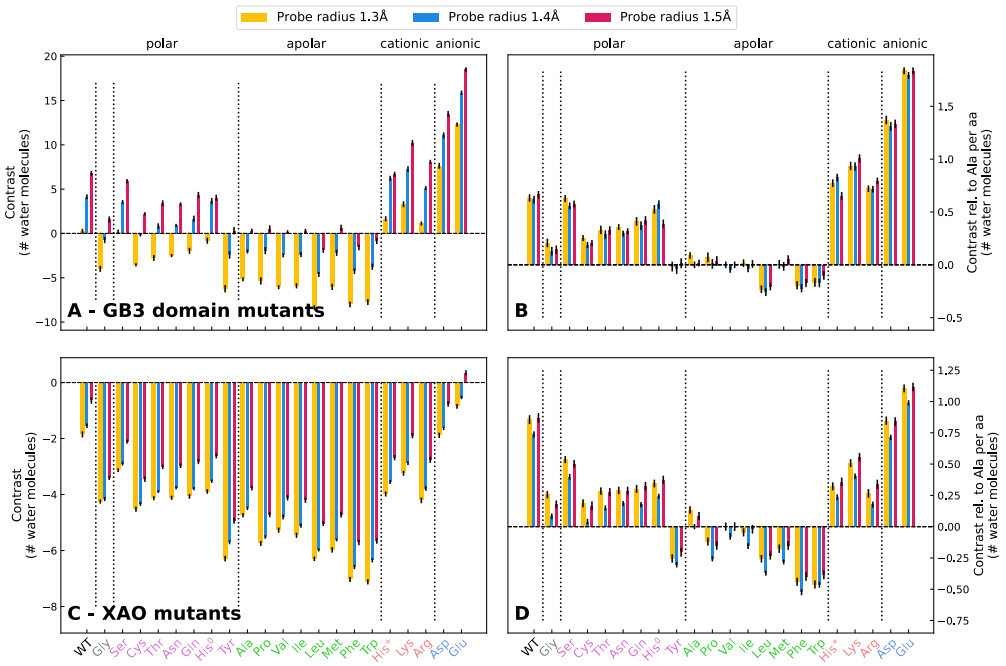


Figure S4: On the effect of using different probe radii for computing the protein volume on computed hydration shell contrasts. (A) Contrast of the hydration shell in number of water molecules of GB3 wild type and 21 GB3 mutants, see labels at abscissa colored by the property of the amino acid: Gly (grey), polar (pink), apolar (green), cationic (orange), anionic (blue) residues. Contrast values were obtained using three different probe radii for computing the protein volume with the 3V volume calculator:[8] 1.3 Å (yellow), 1.4 Å (blue), or 1.5 Å (red). A grid spacing of 0.16Å was used. (B) Contrast per amino acid for GB3 domain relative to alanine. (C/D) Same analysis as in panels (A/B) for the XAO peptide. From a given MD simulation, the total contrast is unambiguously obtained via the forward scattering $I(0)$. However, using larger probe radii lead to larger computed protein volumes, thereby to smaller density contrasts assigned to the bare protein, and, consequently, larger contrasts assigned to the hydration shell as shown in panels (A/C). In contrasts relative to alanine hardly depend on the choice of the probe radius, as shown in panels (B/D).

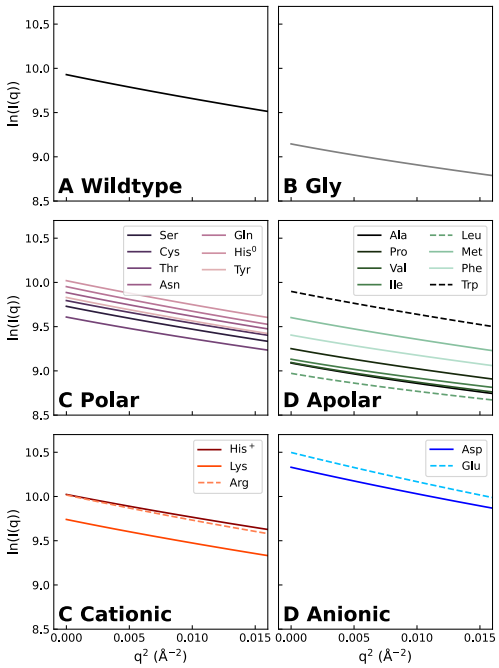


Figure S5: Guinier plots of SAXS curves of the heterogeneous ensemble of the XAO peptide from explicit-solvent SAXS calculations with the TIP4P/2005 water model in combination with the ff03w protein force field. Same SAXS curves as shown in Fig. S3A–F, however plotted as $\ln(I(q))$ vs. q^2.

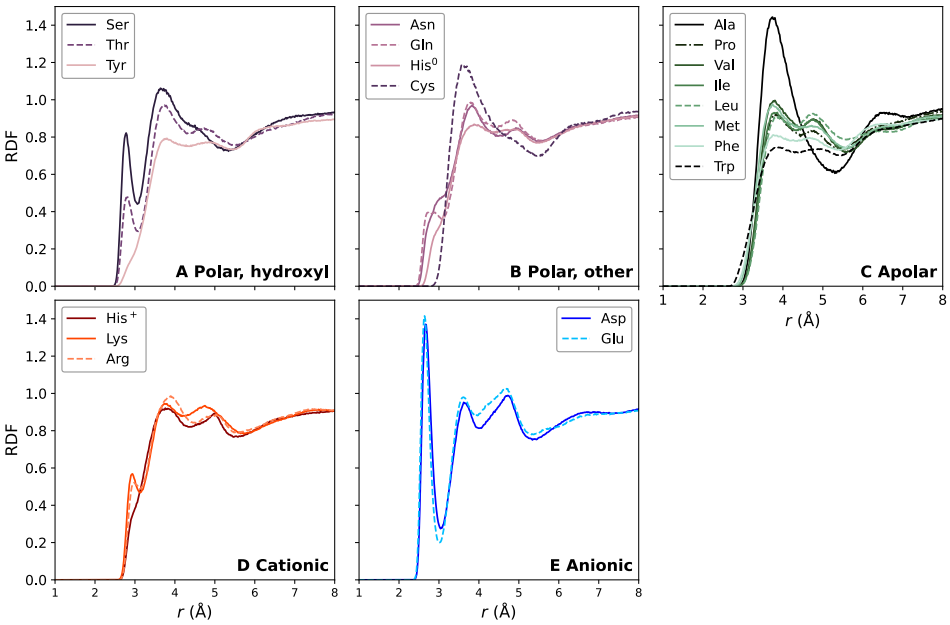


Figure S6: Radial distributions functions (RDFs) of water oxygen atoms with respect to side chain heavy atoms of amino acids computed from simulations of mutated XAO variants. For color code, see legends. RDFs were obtained from simulations with TIP4P/2005 and averaged over 20 XAO conformations. Results from different amino acids are grouped by (A) polar residues containing a hydroxyl group, (B) other polar residues, (C) apolar, (D) cationic, and (E) anionic residues.

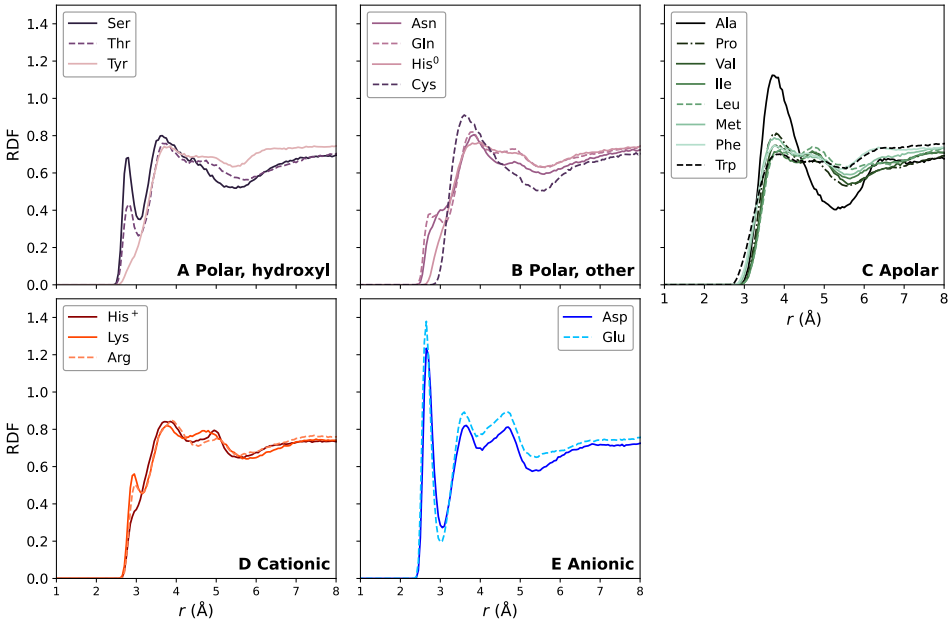


Figure S7: RDFs of water oxygen atoms with respect to side chain heavy atoms of the ten mutated residues of mutated GB3 variants. Presentation analogous to Fig. S6.

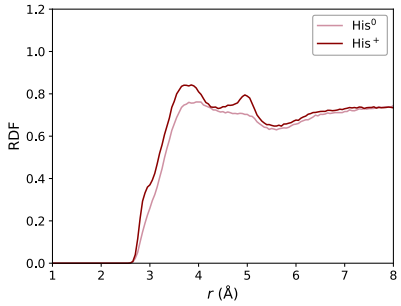


Figure S8: RDFs of water oxygen with respect to side chain heavy atoms of neutral δ-protonated (His^0) and cationic double-protonated histidine (His^+). RDFs were computed from simulations of mutated GB3 variants. Water is more tightly packed around His^+ compared to His^0.

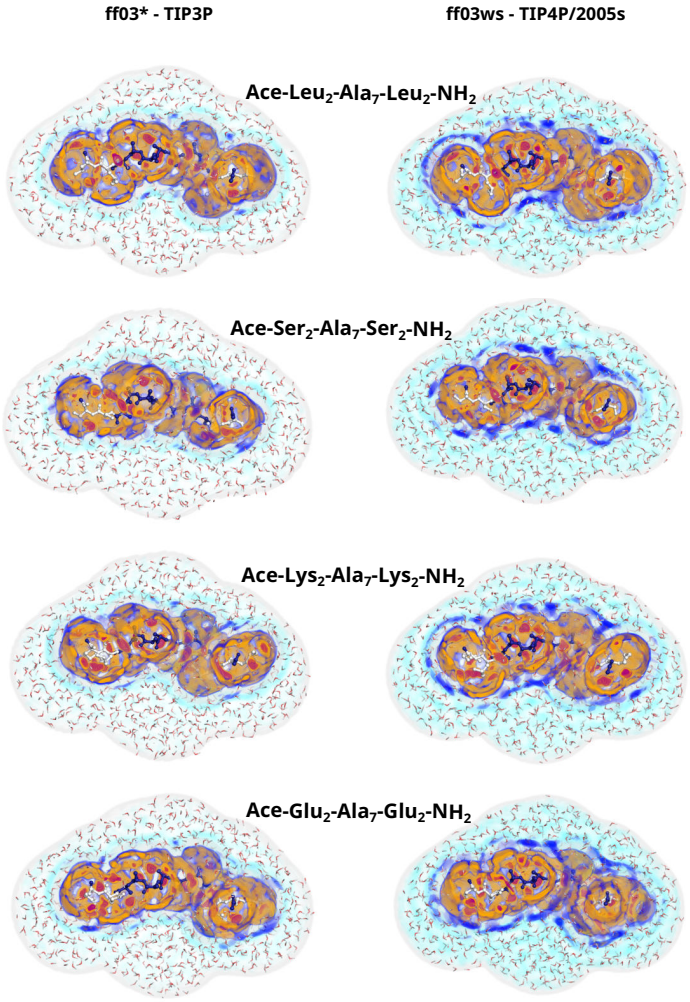


Figure S9: Three-dimensional densities of the hydration shell around the XAO mutants with four leucine, serine, lysine, or glutamate residues at the termini (see labels). The densities were calculated from simulations using ff03* in conjunction with TIP3P (left column) or using ff03ws in conjunction with TIP4P/2005s (right column). Color code is taken from Fig. 1B/D. The solvent densities depend on amino acid type and on the force field.

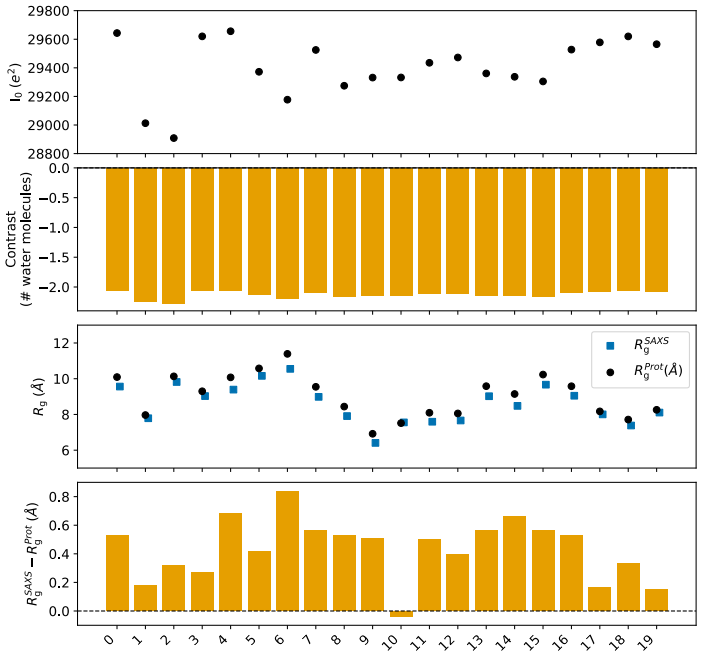


Figure S10: Forward scattering I_0, hydration shell contrast in number of water molecules, R_g, and ΔR_g values for 20 conformations of the aspartate mutant of XAO, obtained with TIP4P/2005 and ff03w. The different conformations impose similar hydration shell contrasts, yet lead to greatly different ΔR_g values.

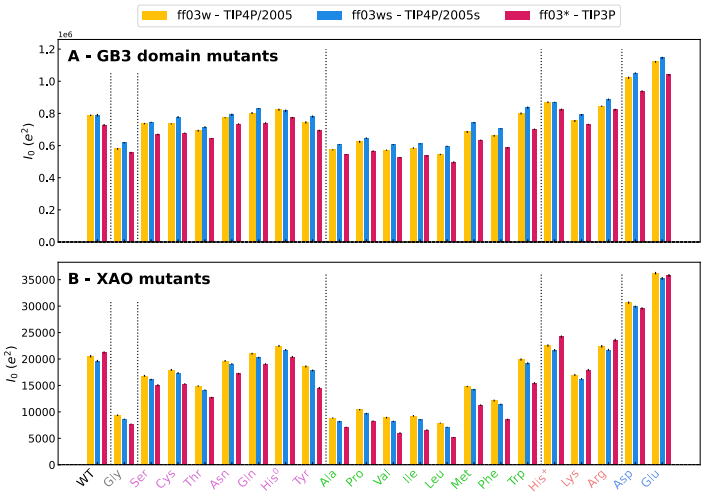


Figure S11: Forward scattering intensity I_0 from SAXS curves for wild type and 21 mutants (see labels on abscissa) of (A) GB3 domain and (B) XAO peptide from simulations with three different combinations of protein force field and water model (for color code, see legend).

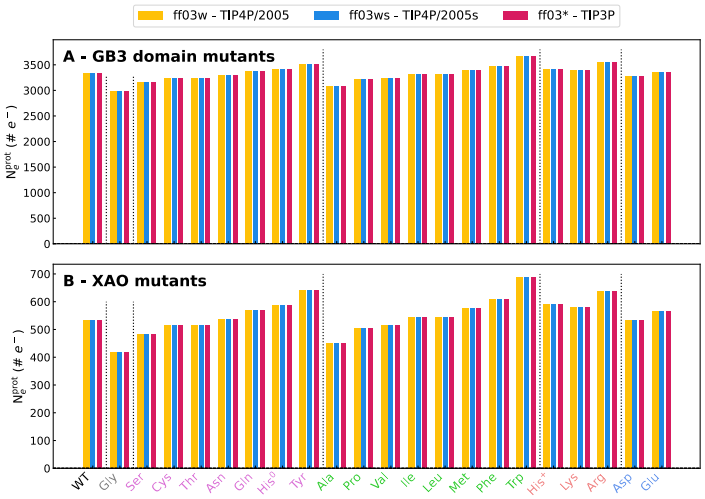


Figure S12: Number electrons (# e^-) of the solute for WT and 21 mutants of (A) GB3 domaind and (B) XAO peptide from simulations with three different combinations of protein force field and water model (for color code, see legend).

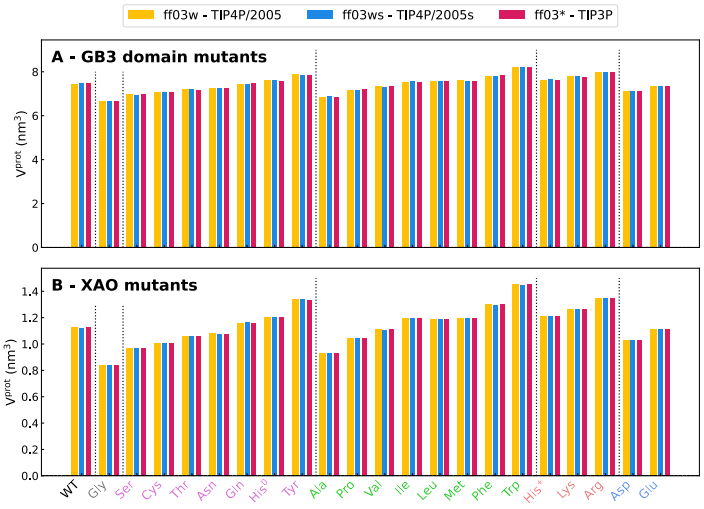


Figure S13: Volumes of WT and 21 mutants (see labels on abscissa) of (A) GB3 domain and (B) XAO peptide from simulations with three different combinations of protein force field and water model (for color code, see legend).

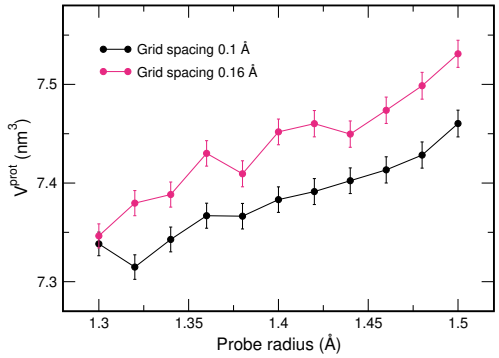


Figure S14: On the effect of using different probe radii or grid spacing for computing the protein volume. GB3 volumes were computed with the 3V volume calculator[8] using different probe radii (abscissa) or using a grid spacing of either 0.1 Å (black) or 0.16 Å (magenta). Error bars denote 1 SE obtained from the average of 20 MD simulation frames. The choices of both probe radius and grid spacing influence the computed protein volumes and, thereby, the absolute contrasts assigned to the hydration shell. In contrast, the contrasts relative to alanine depend only marginally on the choice of the volume definition (see Fig. S4).

References

(1) Hermann, M. R.; Hub, J. S. SAXS-Restrained Ensemble Simulations of Intrinsically Disordered Proteins with Commitment to the Principle of Maximum Entropy. *J. Chem. Theory Comput.* **2019**, *15*, 5103–5115.

(2) Zagrovic, B.; Lipfert, J.; Sorin, E. J.; Millett, I. S.; Van Gunsteren, W. F.; Doniach, S.; Pande, V. S. Unusual Compactness of a Polyproline Type II Structure. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11698–11703.

(3) Makowska, J.; Rodziewicz-Motowidło, S.; Bagińska, K.; Vila, J. A.; Liwo, A.; Chmurzyński, L.; Scheraga, H. A. Polyproline II Conformation Is One of Many Local Conformational States and Is Not an Overall Conformation of Unfolded Peptides and Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1744–1749.

(4) Knight, C. J.; Hub, J. S. WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res.* **2015**, *43*, W225–W230.

(5) Chen, P.-c.; Hub, J. S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014**, *107*, 435–447.

(6) Chen, P.-c.; Hub, J. S. Interpretation of Solution X-ray Scattering by Explicit-Solvent Molecular Dynamics. *Biophys. J.* **2015**, *108*, 2573–2584.

(7) van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Sim.* **1988**, *1*, 173–185.

(8) Voss, N. R.; Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555–W562.

Chapter 8

Depletion of the protein hydration shell with increasing temperature observed by small-angle X-ray scattering and molecular simulations

The authors of this paper are Johanna-Barbara Linse, Hyun Sun Cho, Friedrich Schotte, Philip A. Anfinrud and Jochen S. Hub, with the following contributions:

J.S.H. motivated the project.
J.-B.L. performed and analyzed MD simulations.
H.S.C., F.S. and P.A.A. performed the SAXS experiments and analysis.
J.B.L., H.S.C., F.S., P.A.A. and J.S.H. designed the project.
J.-B.L., H.S.C., P.A.A. and J.S.H. wrote the article.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS

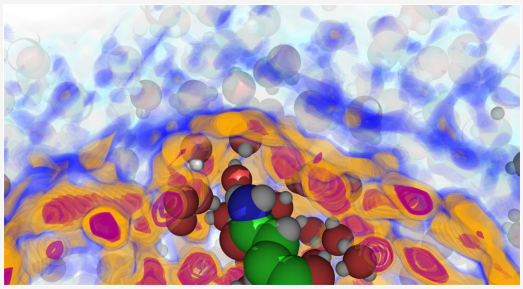
Article

Depletion of the Protein Hydration Shell with Increasing Temperature Observed by Small-Angle X-ray Scattering and Molecular Simulations

Johanna-Barbara Linse, Hyun Sun Cho, Friedrich Schotte, Philip A. Anfinrud,* and Jochen S. Hub*

Cite This: J. Am. Chem. Soc. 2025, 147, 47117−47125

ABSTRACT: The hydration shell is an integral part of proteins since it plays key roles in conformational transitions, molecular recognition, and enzymatic activity. While the dynamics of the hydration shell have been described by spectroscopic techniques, the structure of the hydration shell remains less understood due to the lack of hydration shell-sensitive structural probes with high spatial resolution. We combined temperature-ramp small-angle X-ray scattering (*T*-ramp SAXS) from 255 to 335 K with molecular simulations to demonstrate that the hydration shells of the IgG-binding domain of Protein G (GB3) and the villin headpiece are remarkably temperature-sensitive. For proteins in the folded state, *T*-ramp SAXS data and explicit-solvent SAXS predictions consistently demonstrate decays of protein contrasts and radii of gyration with increasing temperature, which are shown to reflect predominantly temperature-sensitive, depleting hydration shells. The depletion is caused not merely by enhanced disorder within the hydration shells but also by partial displacements of surface-coordinated water molecules. Together, *T*-ramp SAXS and explicit-solvent SAXS calculations provide a novel structural view of the protein hydration shell, which underlies temperature-dependent processes such as cold denaturation, thermophoresis, or biomolecular phase separation.



INTRODUCTION

The hydration shell is an integral part of biomolecules since it mediates a wide range of biological functions. Water in the hydration shell of enzymes actively participates in hydrolytic enzymatic reactions, acid−base reactions, or proton transfer via the Grotthus mechanism.[1−3] The hydration shell also orchestrates protein folding, ligand binding, or protein−protein recognition because such processes involve extensive rearrangements of protein−water and water−water interaction networks. Hydration-shell water exhibits different properties compared to bulk water. Techniques such as NMR, terahertz, sum-frequency generation, or inelastic neutron scattering spectroscopy have shown that the vibrational, rotational, and translational dynamics of hydration shell water are slowed down approximately two- to fivefold relative to those of bulk water.[4−18]

While the dynamics of the hydration shell have been extensively studied, its structural properties remain less well understood. X-ray and neutron crystallography provide insight into highly localized water molecules that coordinate with biomolecular surface residues;[19,20] however, these techniques are largely blind to more dynamic water or to water in the second and third hydration layers. Small-angle scattering with X-rays or neutrons (SAXS/SANS) is sensitive to the hydration shell, but it yields data with low information content and low spatial resolution. Accordingly, SAXS/SANS has shown that many proteins exhibit hydration shells with an overall increased density relative to the bulk, which manifests in an increased radius of gyration (R_g) relative to the R_g of the bare protein.[21,22] Molecular dynamics (MD) simulations with explicit water corroborated these findings and explained the modified R_g values by an increase in hydration shell density of ∼6%.[23−28] However, the influence of protein geometry, surface composition, pH, and temperature on the hydration shell remains poorly characterized, largely due to the paucity of experimental methods capable of probing the hydration-shell structure at atomic resolution.

Solvation of biomolecules is governed by a delicate balance of significant enthalpic and entropic contributions, which often mostly compensate to yield moderate solvation free energies. Because entropic effects are amplified at high temperatures, solvation is inherently temperature-dependent, with implica-

Received: August 5, 2025
Revised: October 29, 2025
Accepted: December 3, 2025
Published: December 11, 2025

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

Journal of the American Chemical Society pubs.acs.org/JACS Article

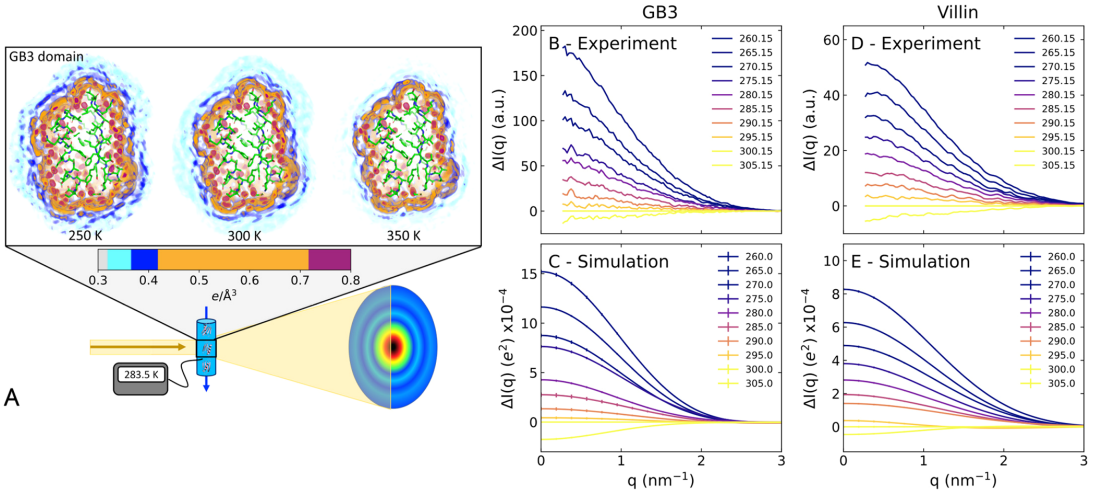


Figure 1. (A) Illustration of temperature-ramp SAXS experiments. Upper row: three-dimensional electron density around the GB3 domain at 250, 300, and 350 K from MD simulations with restrained heavy atoms. The protein is shown as sticks, and the solvent electron density is represented in colors ranging from cyan to purple, as indicated by the color bar, revealing the depletion of the hydration shell structure with increasing temperature. Scattering intensity difference relative to 300 K, $\Delta I(q)$, of the GB3 domain at temperatures spanning ~260 to ~305 K (see the legend) (B) from experiments or (C) from simulations. $\Delta I(q)$ for the villin headpiece (D) from experiments or (E) from simulations. Absolute SAXS curves are shown in Figures S1 and S2.

tions for biomolecules and soft-matter systems. For instance, the hydrophobic effect, which drives protein folding and micelle and lipid membrane assembly, is temperature-dependent and entropy-driven at 22 °C yet enthalpy-driven at 113 °C.[27−29] Temperature-dependent solvation drives the unfolding of proteins at low temperatures (cold denaturation)[30,31] as well as the collapse of intrinsically disordered proteins[32] or liquid−liquid phase separation at high temperatures.[33] Such effects would be at odds with the naive expectation that high temperatures would generally favor polymer disorder, thus highlighting the importance of hydration shell effects. The drift of molecules or beads along temperature gradients, an effect known as thermophoresis, thermodiffusion, or the Soret effect, is driven by the temperature dependence of the solvation entropy.[34,35] The mechanisms by which such effects contribute to the response of biological systems to changing temperatures, for instance, during thermosensing by membrane channels or thermotaxis by bacteria, are not known. Additionally, the size and shape of detergent or polymer micelles are temperature-sensitive, which has likewise been associated with temperature-dependent solvation.[36,37] Thus, structural insight into temperature-dependent solvation is essential to rationalizing a plethora of biological or soft-matter phenomena.

We combine temperature-ramp (T-ramp) SAXS experiments and MD simulations to reveal how temperature controls the hydration shell structure of two proteins in their folded state, the third IgG-binding domain of Protein G (GB3) and chicken villin headpiece (also termed HP35), which have been used extensively as model proteins for biophysical studies. We developed infrastructure on the BioCARS beamline at the Advanced Photon Source for acquiring SAXS data over a broad temperature range, spanning 257−393 K, which covers supercooled conditions up to unfolded proteins. To rationalize the temperature effects on our SAXS data by atomic means, we use all-atom MD simulations combined with explicit-solvent SAXS calculations.[25,38] Our strategy is supported by recent findings that the overall hydration shell density obtained by MD simulations with many protein and water force fields aligns accurately with consensus SAXS/SANS data.[39,40] Our T-ramp SAXS data and MD simulations are quantitatively consistent and demonstrate that, in the folded state, the protein hydration shell is remarkably temperature-sensitive, as shown by partial displacements of surface-coordinated water molecules.

RESULTS

Figure 1A (upper row) presents the three-dimensional electron density of the solvent around the GB3 domain, computed from an MD simulation with the TIP4P/2005 water model, which has been parametrized to reproduce water properties across a broad temperature range.[41] By using position restraints on all heavy atoms, conformational fluctuations of solvent-exposed protein atoms were excluded, thereby yielding a spatially well-defined hydration shell with a first and a second hydration layer (orange and blue densities, respectively). The density maps reveal that increasing the temperature from 250 to 350 K leads to a partial loss of the hydration shell structure, as evidenced by the decreased densities of the first and second hydration layers. These simulations provide the atomistic rationale for interpreting the temperature-dependent scattering intensities observed in SAXS experiments.

SAXS data for the GB3 domain and villin headpiece were acquired on the BioCARS beamline at the Advanced Photon Source using a T-ramp protocol that generates scattering images while repeatedly ramping the sample temperature between 257 K and up to 393 K. The SAXS curves of both proteins exhibited clear temperature dependence (Figures S1A and S2A). Thanks in part to the small volume of protein solution in the HF-etched, temperature-controlled capillary, the solution could be repeatably supercooled to 257 K without freezing. The temperature dependence of the SAXS curves is highlighted by the difference

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

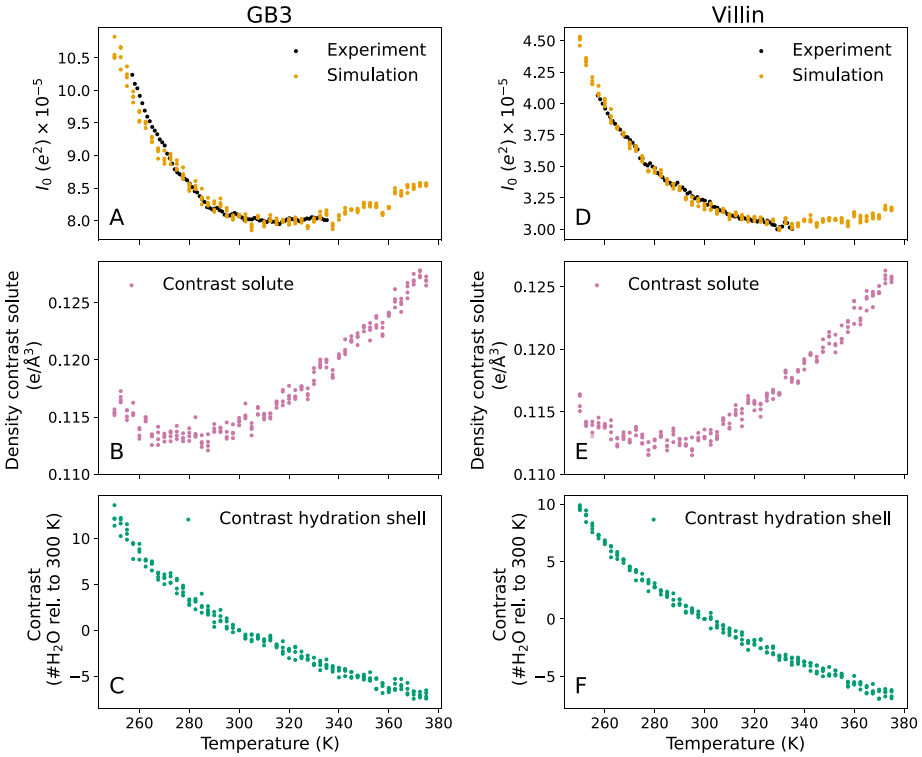


Figure 2. Forward scattering I_0 and decomposition of the solute−solvent contrast into contributions from the bare protein and the hydration shell for the (A−C) GB3 domain and (D−F) villin headpiece. (A, D) I_0 versus temperature from experiments (black) and backbone-restrained MD simulations with explicit-solvent SAXS calculations (orange). The experimental curves were scaled by a constant factor of the simulation data in the temperature range below 303 K. (B, E) Density contrast of the backbone-restrained bare protein from MD simulations, reflecting the temperature dependence of the solvent density. (C, F) Contrast of the hydration shell in the number of water molecules relative to 300 K. In all panels, colored dots indicate simulation results from four independent simulations per temperature. The analysis of panels (A−F) visualized as contrasts in the number of electrons is shown in Figure S5.

intensities $\Delta I(q)$ relative to the experimental reference temperature of 300.15 K (Figure 1B,D). In this study, we focus on two key parameters that are frequently obtained from the small-angle regime: (i) the forward scattering intensity, $I_0 = I(q = 0)$, which is given by the square of the contrast between the solute and the solvent (in number of electrons), scaled in the experiment by a concentration-dependent factor; and (ii) the radius of gyration, R_g. These parameters were obtained via Guinier analysis: $\ln(I(q)/I_0) \approx q^2R_g^2/3$, where $I(q)$ is the SAXS curve and q is the momentum transfer. Specifically, Guinier analysis was carried out after extrapolating a series of concentration-dependent intensities to the infinite dilution limit, thereby removing modest effects from the structure factor at low q (Figure S3).

Our SAXS experiments revealed that the forward scattering I_0 decreased by more than 20% for both the GB3 domain and the villin headpiece across the measured temperature range, demonstrating a decreasing electron density contrast that plateaued at ∼320 K for the GB3 domain and at ∼335 K for the villin headpiece (Figure 2A,D, black dots). The R_g values decreased by 0.4 Å for the GB3 domain and by 0.3 Å for the villin headpiece between 257.15 and ∼300 K (Figure 3A,C, black dots). This trend is followed by a slight increase in R_g up to ∼330 K, likely indicating enhanced conformational fluctuations, followed by sharp increases in R_g, indicating protein unfolding. The temperatures of the sharp R_g increase are compatible with previously reported melting temperatures of the GB3 domain and villin headpiece.[42−44]

We hypothesized that the decreases in I_0 and R_g between 257 and 300 K, where the proteins remain folded, reflect gradual depletions of the protein hydration shells. To test this, we carried out MD simulations of the GB3 domain and villin headpiece (Figure S4) over a temperature range of 250−375 K in steps of 2.5 K. SAXS curves were computed from MD simulations according to the WAXSiS method, thereby taking the explicit solvent into account.[25,45] To isolate the effect of the solvent and to prevent unfolding, the simulations were carried out with position restraints on backbone atoms; thus, variations in the computed SAXS curve were purely caused by temperature-dependent variations in the hydration shell and excluded solvent densities.

In line with our SAXS experiments, the computed SAXS curves were temperature-dependent, as indicated by the absolute scattering curves (Figures S1B and S2B) and by the difference curves relative to 300 K (Figure 1C,E). The I_0 and R_g values taken from the calculated curves via Guinier analysis are in good agreement with the experimental values in the ≲300 K

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

Journal of the American Chemical Society pubs.acs.org/JACS Article

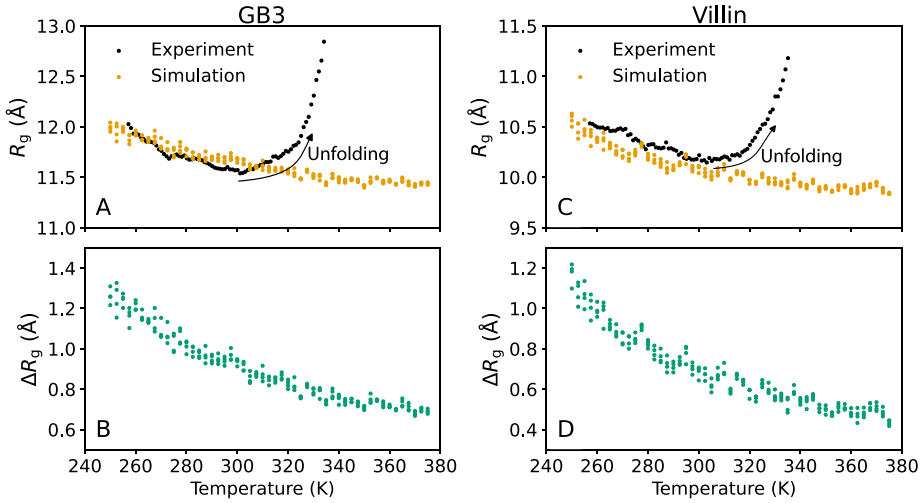


Figure 3. Radius of gyration R_g versus temperature for the (A, B) GB3 domain and (C, D) villin headpiece. (A, C) R_g from Guinier analysis from experiments (black) or backbone-restrained MD simulations (orange). (B, D) From MD simulations and explicit-solvent SAXS calculations, the difference ΔR_g between the R_g from Guinier analysis (including hydration shell contributions) and the R_g of the bare protein. A decrease in R_g with temperature in backbone-restrained simulations demonstrates the depletion of the hydration shell. Colored dots indicate simulation results from four independent simulations per temperature.

regime, where the proteins remain fully folded (Figures 2A,D and 3A,C, yellow dots). The agreement is remarkable considering that, in our explicit-solvent SAXS calculations, neither the hydration shell nor the excluded solvent density was fitted to the experimental data. The agreement implies that (i) thermal expansion of the protein, which is excluded in our MD simulations with backbone restraints, has only a small effect on I_0 and R_g in this temperature range; and (ii) the MD simulations with the TIP4P/2005 water model provide a realistic representation of the hydration shell structure across a broad temperature range.

To isolate the effect of temperature on the hydration shell, we decomposed the total contrast ΔN_e (in the number of electrons) into contributions from the protein and the hydration shell. The forward scattering follows $I_0 = \Delta N_e^2$, and the total contrast is

$$\Delta N_e = \Delta N_e^{prot} + \Delta N_e^{hs}$$

where ΔN_e^{hs} is the contrast imposed by the hydration shell and $\Delta N_e^{prot} = N_e^{prot} - \rho_{solv} V^{prot}$ is the contrast of the bare protein with the temperature-dependent solvent density ρ_{solv}, V^{prot} is the protein volume, and N_e^{prot} is the number of electrons in the protein. Figure 2B,E shows the electron density contrast of the bare proteins, $\Delta N_e^{prot}/V^{prot}$, reflecting purely the temperature dependence of the water density as the protein volumes are nearly fixed. Here, the convex shapes of the density contrast curves reflect the well-known maximum of the water density at 4 °C, with the density decreasing at both lower and higher temperatures. Critically, the contrast between the hydration shell and bulk water ΔN_e^{hs} greatly decreases with increasing temperature, demonstrating a depleting hydration shell structure (Figure 2C,F). Specifically, relative to 300 K, the hydration shell of both the GB3 domain and the villin headpiece contains approximately ten additional water molecules at 250 K and five fewer water molecules at 360 K (Figure 2C,F). Over the entire temperature range simulated here, the hydration shells of the GB3 domain and villin headpiece lose 19 and 17 water molecules, respectively. We note that the quantitative decomposition of the ΔN_e into contributions from protein and hydration shell depends on the definition of the protein volume; thus, a different definition may lead to a constant shift between ΔN_e^{prot} and ΔN_e^{hs},[26] which would, however, not alter the trends as a function of temperature. Together, this analysis supports our hypothesis that the experimentally and computationally observed decay of I_0 reflects temperature-sensitive depletion of the hydration shell.

As a second indicator of the hydration shell, we computed from the simulations the increase of the radius of gyration due to the hydration shell, $\Delta R_g = R_g - R_g^{prot}$, defined as the R_g from Guinier analysis relative to the R_g^{prot} of the bare protein (Figure 3B,D). Recently, we found excellent agreement for ΔR_g values at room temperature obtained from consensus SAXS/SANS data and MD simulations using various protein force fields and water models, including the TIP4P/2005 water model employed here.[39,40] According to the simulations, ΔR_g decays with increasing temperature by ~0.6 Å, reflecting the diminishing hydration shell in line with the decreasing I_0 discussed above. Considering that (i) R_g^{prot} was fixed in simulations due to backbone restraints and (ii) R_g values from simulations agreed with experimental values in the $T \leq 300$ K regime where the proteins remain folded (Figure 3A,C), this analysis suggests that the R_g decays in experiments up to 300 K are likewise signatures of depleting hydration shells with increasing temperature.

The results described above were obtained from simulations with restrained backbone atoms, thereby avoiding conformational transitions and isolating the effect of the hydration shell on the contrast and R_g. To exclude the possibility that backbone fluctuations might influence our key findings, we carried out an additional series of simulations without restraints for the GB3 domain (Figure S9, red triangles). Whereas the results reveal slightly increased variability among different temperatures, as expected from enhanced protein fluctuations, GB3 remained

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

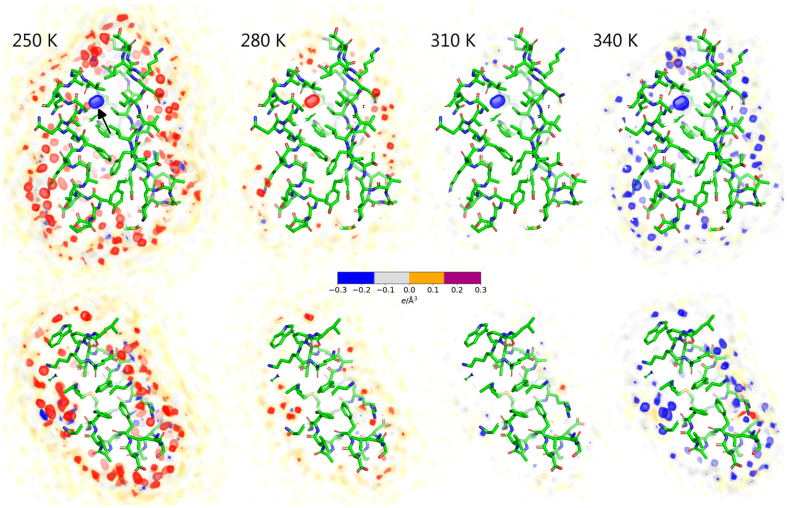


Figure 4. Electron density difference of the solvent relative to 300 K for the GB3 domain (upper row) and villin headpiece (lower row) for temperatures of (from left to right) 250, 280, 310, or 340 K. Densities were computed from MD simulations with restraints on all heavy protein atoms. Increased and decreased densities relative to 300 K are represented by red and blue densities, respectively (see the color bar), revealing the depletion of surface-bound densities with increasing temperature. The localized density (black arrow) refers to a randomly placed water molecule within a GB3 cavity, which is not exchanged with the bulk within the simulation time. Proteins are shown as sticks.

folded within the simulation time, even at high temperatures, and the trends are conserved across the entire temperature range.

Having established that the temperature-dependent hydration shell contrast agrees between simulations and experiments, we used the simulations to gain structural and energetic insights into the hydration shell and its interaction with the protein. Visual inspection of the hydration shell densities confirms that the hydration shells are gradually depleted as the temperature increases (Movies S1 and S2 and Figure 1A). To resolve contributions from individual surface-bound water molecules, we computed the density differences relative to 300 K from an additional series of simulations with restraints on all heavy atoms, thereby suppressing side-chain fluctuations and leading to a spatially well-resolved hydration shell structure (Figure 4 and Movies S3 and S4). In addition to the overall depletion of the first and second hydration layers, these density difference maps reveal the loss of numerous localized solvent densities. Visual inspection showed that most of these localized density differences originate from molecules that are coordinated via hydrogen bonds with the protein, whereas a few density differences arise from water molecules trapped in small hydrophobic pockets (Figure 4, red and blue spots). Critically, localized densities from surface-bound water at low temperatures are not merely smeared out at higher temperatures but are partially lost. This finding is supported by the hydration shell densities as a function of distance from the van der Waals surface of the proteins, which reveal a partial loss of the first and second hydration shell density peaks (Figures S6A and S7A). These structural changes are accompanied by a ∼13% decrease in both protein−water interaction energies and the number of hydrogen bonds over the simulated temperature range (Figures S6B,C and S7B,C). Nevertheless, even at 350 K, the overall hydration shell structure with its pronounced first and shallow second peaks remains intact (Figures S6A and S7A), suggesting that the decrease of I_0 and R_g reflects a partial but not complete loss of the protein−water coordination with increasing temperature.

In addition to analyzing the protein−water coordination discussed above, we investigated how temperature affects the internal structure of the hydration shell. To this end, we computed the number of water−water hydrogen bonds within a distance of 9 Å from the protein surfaces of the GB3 domain and villin headpiece (Figures S6D and S7D), revealing a ∼20% decrease over the simulated temperature range. This trend, together with an increasingly smeared-out water−water radial distribution function within the hydration shell (Figure S8), demonstrates a considerable loss of internal water structure within the hydration shell, consistent with previous computational and spectroscopic studies.[46−49] At high temperatures, the decrease in the number of hydrogen bonds is more pronounced in the hydration shell compared to the bulk solvent, suggesting that the hydration shell is more temperature-sensitive than bulk water (Figures S6D and S7D). Together, these analyses suggest that increasing temperature leads not only to a loss of coordinated water densities at the protein surface but also to a generally less structured hydration shell, as demonstrated by fewer water−protein and water−water hydrogen bonds, more dispersed protein−water and water−water correlations, and reduced enthalpic protein−water interactions. These findings provide the structural rationale for the decreasing contrast and ΔR_g values observed by our SAXS experiments and explicit-solvent SAXS calculations.

In summary, using a novel *T*-ramp SAXS setup at BioCARS, we investigated the GB3 domain and villin headpiece across a broad temperature range, spanning supercooled conditions to protein unfolding. When warming from cold temperatures, the SAXS data revealed a systematic decrease in the protein's electron density contrasts and radii of gyration. MD simulations combined with explicit-solvent SAXS calculations showed excellent agreement with the experimental data, attributing

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

these trends primarily to the temperature-dependent depletion of the hydration shell. The depletion is not solely caused by increased water disorder, as expected at increasing temperatures, but involves the partial displacement of surface-coordinated water molecules. Together, our SAXS experiments and simulations provide detailed structural insight into the protein hydration shell, highlighting its remarkable sensitivity to temperature and its potential influence on biological processes, such as cold denaturation, thermophoresis, and biomolecular phase separation.

METHODS

***T*-ramp SAXS Experiments.** The 35-residue villin headpiece subdomain (LSDED FKAVF GMTRS AFANL PLWKQ QHLKK EKGLF) was obtained from California Peptide Research Inc. The peptide was dialyzed against 20 mM acetate buffer (pH 4.9) with 150 mM NaCl. The GB3 domain (MQYKL VINGK TLKGE TTTKA VDAET AEKAF KQYAN DNGVD GVWTY DDATK TFTVTE) was expressed and purified as described previously[50] and dissolved in 40 mM acetate buffer with 150 mM NaCl, 0.05% NaN_3, and 5 mM DTT at pH 5.5.

Temperature-dependent small- and wide-angle X-ray scattering (SAXS-WAXS) data were acquired on the BioCARS 14IDB beamline at the Advanced Photon Source.[51−54] Briefly, a peristaltic pump circulates in a closed-loop sample solution through a 560 mm long capillary (Polymicro TSP250350) that is supported on a home-built high-speed XYZ stage. To minimize scattering from the capillary walls, the region where X-rays pass through was HF-etched to a wall thickness of approximately 15−20 μm. X-rays passing through the capillary are scattered and detected by a large-area Rayonix MS340-HS detector positioned 186 mm downstream from the capillary. Thanks to the small, 0.51 mm diameter beamstop located near the midpoint between the sample and detector, the range of q accessible with 12 keV photons spans from 0.2 to the far-WAXS regime of 52 nm^{-1}, albeit this study focused on the small-angle regime up to 3 nm^{-1} (Figure 1). The high-speed stage translates the sample capillary at a constant velocity over a ∼20 mm range, during which 40 X-ray shots are transmitted through the sample at ∼40 Hz with a separation of 0.5 mm along the capillary, thereby distributing the X-ray dosage over a large volume of the sample. During the return stroke, the X-ray scattering image is saved, and a fresh aliquot of solution is drawn from a ∼120 μL sample reservoir and pushed into the capillary. A homemade temperature controller ramped the sample temperature up and down between −16 and 120 °C at a rate of nearly 1 °C/s, repeating the ramp three times for each data set. Scattering data were acquired at each of three different concentrations (2.2, 6.9, 20 mg/mL for the villin headpiece and 2.7, 8.4, 24.7 mg/mL for GB3). To prevent boiling when ramping the capillary temperature to 120 °C, the sample reservoir was pressured with helium at 3 atm. Since we focused in this study on the hydration shell of folded proteins, we restricted further analysis to temperatures below 63 °C. They were averaged, extrapolated to the infinite dilution limit, and analyzed using the Guinier method to generate temperature-dependent I_0 and R_g curves.

Simulation Setup and Explicit-Solvent SAXS Calculations. Structures of the villin headpiece and the GB3 domain were taken from the protein data bank (PDB; codes 1yrf[55] and 2oed,[42] respectively). Crystal water was kept in the structures, and hydrogen atoms were added using pdb2gmx software.[56] MD simulations were carried out using the GROMACS software, version 2021.7.[56] Interactions of the proteins were described with the amber99SB-ildn force field (ff99SB-ildn).[57] The starting structures were placed in a dodecahedral box, where the distance between the protein and box edges was at least 2.0 nm, and solvated in TIP4P/2005[41] water. The charges of the proteins were neutralized by adding Na^+ counterions. Additional salt was not added because (i) effects of NaCl on SAXS curves were small[39] and (ii) force fields for ion−protein interactions were not well validated for wide temperature ranges and would, therefore, add uncertainty. After 400 steps of minimization with the steepest descent algorithm, the systems were equilibrated for 100 ps with harmonic position restraints applied to the heavy atoms of the proteins (force constant, 2000 kJ mol^{-1} nm^{-2}). Subsequently, the production runs were carried out for 50 ns with harmonic position restraints applied to backbone atoms (force constant 2000 kJ mol^{-1} nm^{-2}) at temperatures between 250 and 375 K in steps of 2.5 K. For each protein and temperature, four independent simulations were carried out. The equations of motion were integrated using the leapfrog algorithm.[58] The temperature was controlled using velocity rescaling (τ = 1 ps).[59] The pressure was controlled at 1 bar with the Berendsen barostat (τ = 1 ps)[60] and the Parrinello−Rahman barostat (τ = 5 ps)[61] during equilibration and production simulations, respectively. Since the experiments were carried out at 3 atm, we performed an additional series of simulations at the same pressure and found that the increased pressure had only a small effect on the results (Figure S9). The geometry of the water molecules was constrained with the SETTLE algorithm,[62] and LINCS[63] was used to constrain all other bond lengths. A time step of 2 fs was used. Dispersive interactions and short-range repulsion were described by a Lennard−Jones potential, which had a cutoff at 1 nm. The pressure and energy were corrected for missing dispersion corrections beyond the cutoff. Neighbor lists were updated with the Verlet scheme. Coulomb interactions were computed with the smooth particle-mesh Ewald method.[64,65] We used a Fourier spacing of approximately 0.12 nm, which was optimized by the GROMACS mdrun module at the beginning of each simulation.

To compute the SAXS curves, 2500 simulation frames were used from the time interval between 0 and 50 ns. The SAXS calculations were performed using GROMACS-SWAXS, as also implemented by our web server WAXSiS.[25,45,66] The implementation, documentation, and tutorials are available at https://gitlab.com/cbjh/gromacs-swaxs. A spatial envelope was built around the protein atoms in all frames. The distance between the protein and the envelope surface was at least 12 Å, such that all water atoms of the hydration shell were within the envelope. One envelope was generated for GB3, and one envelope was generated for villin; each was used throughout this study. Solvent atoms inside the envelope contributed to the calculated SAXS curves. Critically, the size of the envelope is not a fitting parameter, as it is not adjusted to match experimental data. Instead, the envelope is chosen to be large enough such that correlations between solvent densities inside and outside the envelope are due to bulk water.[25,67] The buffer subtraction was carried out using 2498 simulation frames of a pure-water simulation box, which was simulated for 50 ns at the same temperature, as the protein simulation, and large enough to enclose the envelope. The orientational average was calculated by using 150 **q**-vectors for each absolute value of **q**. The solvent electron density was corrected to the temperature-dependent experimental value at the respective temperature such as 334 e/nm^3 for 298.15 K, as described previously,[66] using the empirical equation for water density as a function of temperature by Kell.[68] In simulations at a pressure of 3 atm, the solvent density was, in addition to the equation by Kell, corrected by the isothermal compressibility of ambient water. No fitting parameters due to the hydration layer or excluded solvent were used, implying that the R_g values were not adjusted by fitting parameters. Nevertheless, the SAXS curves for GB3 and villin revealed reasonable agreement with the experimental curves, also over a wide q range (Figure S10). I_0 and R_g values obtained via Guinier analysis from four independent replicas per temperature showed excellent agreement, as reflected by low standard errors relative to the corresponding mean values (I_0: GB3 0.27%, villin 0.26%; R_g: GB3 0.11%, villin 0.13%), as visualized by the good agreement from simulation results from individual simulations (Figures 2 and 3, colored dots).

Three-dimensional solvent densities around the GB3 domain and villin headpiece (Figures 1A and 4 and Movies S1−S4) were computed with the mdrun module of GROMACS-SWAXS using the environment variable GMX_WAXS_GRID_DENSITY and visualized with PyMOL (see Supporting Methods and Supporting ZIP folder).[69] One-dimensional solvent densities around the protein (Figures S6A and S7A) were computed with the gmx genenv module of GROMACS-SWAXS. Here, three- and one-dimensional densities were computed from 12,500 frames extracted from a 25 ns simulation with restrained heavy atoms, which prevents the smearing of solvent densities near the

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

protein surface due to side-chain fluctuations. The total number of excess electrons implied by the density profiles aligns with the literature (see Supporting Results). Protein−water interaction energies (Figures S6B and S7B) were computed as the sum of protein−water Lennard−Jones and short-range Coulomb interactions. Hydrogen bonds were computed with the gmx hbond module using standard settings.

RDFs between water oxygen atoms within the hydration shell were computed using an in-house modification of gmx genenv by using simulations with restraints on all backbone atoms. To this end, envelopes were constructed at distances of 0, 3, 5, or 7 Å from the van der Waals surface of the protein. For each MD frame of the production simulations, water oxygen atoms between (i) the envelope at 0 Å distance and (ii) the envelope at x Å distance were selected, where $x \in \{3, 5, 7\}$. Thereby, oxygen atoms within a distance of x Å from the protein were selected.

To decompose the total contrast into the contrast of the bare protein and the hydration shell, we defined the protein volume V_{prot} as the cavity volume calculated using the 3V volume calculator[70] with a grid size of 0.16 Å and a probe radius of 1.4 Å, as used previously.[71,72] Each volume was computed from 20 independent MD frames. The number of solute electrons N_e^{prot} was taken from the atomic form factors at zero scattering angle, as defined by the Cromer−Mann parameters of the respective atoms.

■ ASSOCIATED CONTENT

Data Availability Statement

A Gromacs variant GROMACS-SWAXS that implements explicit-solvent SAXS calculations is freely available at https://gitlab.com/cbjh/gromacs-swaxs. Documentation and tutorials are available at https://cbjh.gitlab.io/gromacs-swaxs-docs/.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.5c13497.

Absolute SAXS curves, Guinier analysis, simulation systems, analysis of solvent density correlations, hydrogen bonds, interaction energies, control simulations, and SAXS curve comparison to wider angles (Figures S1−S10) (PDF)

Three-dimensional electron density maps for all temperatures (Movie S1) (MP4)

Three-dimensional electron density maps for all temperatures (Movie S2) (MP4)

Three-dimensional electron density maps for all temperatures (Movie S3) (MP4)

Three-dimensional electron density maps for all temperatures (Movie S4) (MP4)

Electron densities and visualization scripts (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Philip A. Anfinrud − *Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States*; Email: philipa@intra.niddk.nih.gov

Jochen S. Hub − *Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken 66123, Germany*; orcid.org/0000-0001-7716-1767; Email: jochen.hub@uni-saarland.de

Authors

Johanna-Barbara Linse − *Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken 66123, Germany*

Hyun Sun Cho − *Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States*

Friedrich Schotte − *Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.5c13497

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.-B.L. and J.S.H. were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via grants HU 1971/3-1 and INST 256/539-1. The temperature-dependent SAXS studies were performed on the BioCARS 14IDB beamline at the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The use of BioCARS was supported by the National Institute of General Medical Sciences of the National Institutes of Health under grant number P41 GM118217. This research was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health. This research was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) within the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

■ REFERENCES

(1) Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108*, 74−108.

(2) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673−7697.

(3) de Grotthuss, C. J. T. Sur La Décomposition de l'eau et Des Corps Qu'elle Tient En Dissolution à l'aide de l'électricité Galvanique. *Ann. Chim.* **1806**, *LVIII*, 54−74.

(4) Mondal, S.; Mukherjee, S.; Bagchi, B. Protein Hydration Dynamics: Much Ado about Nothing? *J. Phys. Chem. Lett.* **2017**, *8*, 4878−4882.

(5) Fogarty, A. C.; Duboué-Dijon, E.; Sterpone, F.; Hynes, J. T.; Laage, D. Biomolecular hydration dynamics: a jump model perspective. *Chem. Soc. Rev.* **2013**, *42*, 5672−5683.

(6) Bagchi, B. Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem. Rev.* **2005**, *105*, 3197−3219.

(7) Wüthrich, K.; Billeter, M.; Güntert, P.; Luginbühl, P.; Riek, R.; Wider, G. NMR studies of the hydration of biological macromolecules. *Faraday Discuss.* **1996**, *103*, 245−253.

(8) Crilly, C. J.; Eicher, J. E.; Warmuth, O.; Atkin, J. M.; Pielak, G. J. Water's Variable Role in Protein Stability Uncovered by Liquid-Observed Vapor Exchange NMR. *Biochemistry* **2021**, *60*, 3041−3045.

(9) Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017**, *117*, 10694−10725.

(10) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical

hydration shell around proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20749−20752.
(11) Born, B.; Kim, S. J.; Ebbinghaus, S.; Gruebele, M.; Havenith, M. The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss.* **2009**, *141*, 161−173.
(12) Sushko, O.; Dubrovka, R.; Donnan, R. S. Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected. *J. Chem. Phys.* **2015**, *142*, No. 055101.
(13) Li, T.; Hassanali, A. A.; Kao, Y.-T.; Zhong, D.; Singer, S. J. Hydration Dynamics and Time Scales of Coupled Water-Protein Fluctuations. *J. Am. Chem. Soc.* **2007**, *129*, 3376−3382.
(14) Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Y. Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Cent. Sci.* **2022**, *8*, 1404−1414.
(15) Furse, K. E.; Corcelli, S. A. The Dynamics of Water at DNA Interfaces: Computational Studies of Hoechst 33258 Bound to DNA. *J. Am. Chem. Soc.* **2008**, *130*, 13103−13109.
(16) Halle, B.; Nilsson, L. Does the Dynamic Stokes Shift Report on Slow Protein Hydration Dynamics? *J. Phys. Chem. B* **2009**, *113*, 8210−8213.
(17) Yang, M.; Szyc, Ł.; Elsaesser, T. Decelerated Water Dynamics and Vibrational Couplings of Hydrated DNA Mapped by Two-Dimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2011**, *115*, 13093−13100.
(18) Petersen, C.; Tielrooij, K.-J.; Bakker, H. J. Strong temperature dependence of water reorientation in hydrophobic hydration shells. *J. Chem. Phys.* **2009**, *130*, No. 214511.
(19) Chatake, T.; Ostermann, A.; Kurihara, K.; Parak, F. G.; Niimura, N. Hydration in Proteins Observed by High-resolution Neutron Crystallography. *Proteins: Struct., Funct., Bioinf.* **2003**, *50*, 516−523.
(20) Nakasako, M. Water-Protein Interactions from High-Resolution Protein Crystallography. *Philos. Trans. R. Soc., B* **2004**, *359*, 1191−1206.
(21) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein Hydration in Solution: Experimental Observation by x-Ray and Neutron Scattering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267−2272.
(22) Kim, H. S.; Martel, A.; Girard, E.; Moulin, M.; Härtlein, M.; Madern, D.; Blackledge, M.; Franzetti, B.; Gabel, F. SAXS/SANS on Supercharged Proteins Reveals Residue-Specific Modifications of the Hydration Shell. *Biophys. J.* **2016**, *110*, 2185−2194.
(23) Merzel, F.; Smith, J. C. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378−5383.
(24) Köfinger, J.; Hummer, G. Atomic-Resolution Structural Information from Scattering Experiments on Macromolecules in Solution. *Phys. Rev. E* **2013**, *87*, No. 052712.
(25) Chen, P.-c.; Hub, J. S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014**, *107*, 435−447.
(26) Persson, F.; Söderhjelm, P.; Halle, B. The Geometry of Protein Hydration. *J. Chem. Phys.* **2018**, *148*, No. 215101.
(27) Baldwin, R. L. Temperature Dependence of the Hydrophobic Interaction in Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8069−8072.
(28) Huang, D. M.; Chandler, D. Temperature and Length Scale Dependence of Hydrophobic Effects and Their Possible Implications for Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324−8327.
(29) Southall, N. T.; Dill, K.; Haymet, A. D. J. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **2002**, *106*, 521−533.
(30) Agashe, V. R.; Udgaonkar, J. B. Thermodynamics of Denaturation of Barstar: Evidence for Cold Denaturation and Evaluation of the Interaction with Guanidine Hydrochloride. *Biochemistry* **1995**, *34*, 3286−3299.
(31) Dias, C. L.; Ala-Nissila, T.; Wong-Ekkabut, J.; Vattulainen, I.; Grant, M.; Karttunen, M. The Hydrophobic Effect and Its Role in Cold Denaturation. *Cryobiology* **2010**, *60*, 91−99.
(32) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. Temperature-Dependent Solvation Modulates the Dimensions of Disordered Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 5213−5218.
(33) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J. Temperature-Controlled Liquid-Liquid Phase Separation of Disordered Proteins. *ACS Cent. Sci.* **2019**, *5*, 821−830.
(34) Ludwig, C. Diffusion Zwischen Ungleich Erwärmten Orten Gleich Zusammengesetzter Lösungen. *Sitzungber. Bayer. Akad. Wiss., Wien Math.-Naturwiss. Kl.* **1856**, *20*, 539.
(35) Duhr, S.; Braun, D. Why Molecules Move along a Temperature Gradient. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19678−19682.
(36) Kumbhakar, M.; Goel, T.; Mukherjee, T.; Pal, H. Role of Micellar Size and Hydration on Solvation Dynamics: A Temperature Dependent Study in Triton-X-100 and Brij-35 Micelles. *J. Phys. Chem. B* **2004**, *108*, 19246−19254.
(37) Ivanović, M. T.; Bruetzel, L. K.; Lipfert, J.; Hub, J. S. Temperature-Dependent Atomic Models of Detergent Micelles Refined against Small-Angle X-ray Scattering Data. *Angew. Chem., Int. Ed.* **2018**, *57*, 5635−5639.
(38) Hub, J. S. Interpreting Solution X-ray Scattering Data Using Molecular Simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 18−26.
(39) Linse, J.-B.; Hub, J. S. Scrutinizing the Protein Hydration Shell from Molecular Dynamics Simulations against Consensus Small-Angle Scattering Data. *Commun. Chem.* **2023**, *6*, No. 272.
(40) Trewhella, J.; Vachette, P.; Bierma, J.; et al. A Round-Robin Approach Provides a Detailed Assessment of Biomolecular Small-Angle Scattering Data Reproducibility and Yields Consensus Curves for Benchmarking. *Acta Crystallogr., Sect. D: Struct. Biol.* **2022**, *78*, 1315−1336.
(41) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, No. 234505.
(42) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy. *J. Am. Chem. Soc.* **2003**, *125*, 9179−9191.
(43) Kubelka, J.; Eaton, W. A.; Hofrichter, J. Experimental Tests of Villin Subdomain Folding Simulations. *J. Mol. Biol.* **2003**, *329*, 625−630.
(44) Lei, H.; Deng, X.; Wang, Z.; Duan, Y. The Fast-Folding HP35 Double Mutant Has a Substantially Reduced Primary Folding Free Energy Barrier. *J. Chem. Phys.* **2008**, *129*, No. 155104.
(45) Knight, C. J.; Hub, J. S. WAXSiS: AWeb Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics. *Nucleic Acids Res.* **2015**, *43*, W225−W230.
(46) Brovchenko, I.; Krukau, A.; Smolin, N.; Oleinikova, A.; Geiger, A.; Winter, R. Thermal breaking of spanning water networks in the hydration shell of proteins. *J. Chem. Phys.* **2005**, *123*, No. 224905.
(47) Smolin, N.; Winter, R. Effect of Temperature, Pressure, and Cosolvents on Structural and Dynamic Properties of the Hydration Shell of SNase: A Molecular Dynamics Computer Simulation Study. *J. Phys. Chem. B* **2008**, *112*, 997−1006.
(48) Tielrooij, K.-J.; Hunger, J.; Buchner, R.; Bonn, M.; Bakker, H. J. Influence of Concentration and Temperature on the Dynamics of Water in the Hydrophobic Hydration Shell of Tetramethylurea. *J. Am. Chem. Soc.* **2010**, *132*, 15671−15678.
(49) Ben-Amotz, D. Hydration-Shell Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2019**, *141*, 10569−10580.
(50) Tian, P.; Louis, J. M.; Baber, J. L.; Aniana, A.; Best, R. B. Co-Evolutionary Fitness Landscapes for Sequence Design. *Angew. Chem. Int. Ed.* **2018**, *57*, 5674−5678.
(51) Graber, T.; Anderson, S.; Brewer, H.; et al. BioCARS: A Synchrotron Resource for Time-Resolved X-ray Science. *J. Synchrotron Radiat.* **2011**, *18*, 658−670.
(52) Cho, H. S.; Schotte, F.; Stadnytskyi, V.; DiChiara, A.; Henning, R.; Anfinrud, P. Dynamics of Quaternary Structure Transitions in R-

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

State Carbonmonoxyhemoglobin Unveiled in Time-Resolved X-ray Scattering Patterns Following a Temperature Jump. *J. Phys. Chem. B* **2018**, *122*, 11488−11496.
(53) Cho, H. S.; Schotte, F.; Stadnytskyi, V.; Anfinrud, P. Time-Resolved X-ray Scattering Studies of Proteins. *Curr. Opin. Struct. Biol.* **2021**, *70*, 99−107.
(54) Henning, R. W.; Kosheleva, I.; Šrajer, V.; Kim, I.-S.; Zoellner, E.; Ranganathan, R. BioCARS: Synchrotron Facility for Probing Structural Dynamics of Biological Macromolecules. *Struct. Dyn.* **2024**, *11*, No. 014301.
(55) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7517−7522.
(56) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(57) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.
(58) Hockney, R.; Goel, S.; Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974**, *14*, 148−158.
(59) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(60) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(61) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(62) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(63) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(64) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(65) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(66) Chen, P.-c.; Shevchuk, R.; Strnad, F. M.; Lorenz, C.; Karge, L.; Gilles, R.; Stadler, A. M.; Hennig, J.; Hub, J. S. Combined Small-Angle X-ray and Neutron Scattering Restraints in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 4687−4698.
(67) Park, S.; Bardhan, J. P.; Roux, B.; Makowski, L. Simulated X-Ray Scattering of Protein Solutions Using Explicit-Solvent Models. *J. Chem. Phys.* **2009**, *130*, No. 134114.
(68) Kell, G. S. Density, Thermal Expansivity, and Compressibility of Liquid Water from 0. Deg. to 150. Deg.. Correlations and Tables for Atmospheric Pressure and Saturation Reviewed and Expressed on 1968 Temperature Scale. *J. Chem. Eng. Data* **1975**, *20*, 97.
(69) Schrödinger, LLC. *The PyMOL Molecular Graphics System*.
(70) Voss, N. R.; Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555−W562.
(71) Shumilina, A. In *A Fast Method for Determination of Solvent-Exposed Atoms and Its Possible Applications for Implicit Solvent Models*, Computational Science and Its Applications − ICCSA 2005; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 1075−1082.
(72) Lee, B.; Richards, F. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.

https://doi.org/10.1021/jacs.5c13497
J. Am. Chem. Soc. 2025, 147, 47117−47125

Supplementary material for: Depletion of the protein hydration shell with increasing temperature observed by small-angle X-ray scattering and molecular simulations

Johanna-Barbara Linse,[†] Hyun Sun Cho,[‡] Friedrich Schotte,[‡] Philip A. Anfinrud,[*,‡,¶] and Jochen S. Hub[*,†,¶]

†*Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, 66123, Germany*

‡*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA*

¶*Corresponding author*

E-mail: philipa@intra.niddk.nih.gov; jochen.hub@uni-saarland.de

Supporting Results

Solvent density profiles as function of distance from the protein surface

To illustrate the solvent structure within the hydration layer, we computed the solvent density as a function of the distance from the van der Waals surface of the protein, taken from simulations with restrained heavy atoms (Fig. S6A, S7A). Here, the restraints on heavy atoms were used to exclude spatial fluctuations of protein side chains, which would lead to a smearing of the water density and, thereby, to a smearing of the density peaks in Figs. S6A and S7A. The density profiles were computed with the GROMACS-SWAXS module gmx genenv with option -od (Fig. S6A, S7A). To this end, gmx genenv constructs a series of spatial envelopes around the protein surface at distances up to 12Å in steps of 0.1Å and counts the number of solvent electrons within the 0.1Å thick concentric volume elements, as averaged over 12,500 simulation frames taken from a 25 ns simulation for each temperature.

Critically, since such removal of side chain fluctuations likely influences the packing of water on the protein surface, the density profiles obtained from such simulations with restrained heavy atoms are not suitable for obtaining the overall excess number of water molecules for comparison with our SAXS experiments. Nevertheless, to test whether the density profiles align qualitatively with excess densities reported in the literature, we computed the excess number of electrons implied by the profiles via

$$\Delta N_e = \sum_i (\rho(R_i) - \rho_{\text{bulk}}) \Delta V_i,$$

where R_i is the binned distance from the van der Waals surface of the protein, $\rho(R_i)$ the density, ρ_{bulk} the bulk solvent density defined as the average of $\rho(R_i)$ in the $R_i > 9$ Å region, and ΔV_i denotes volume elements between concentric pairs of envelopes constructed around the protein by gmx genenv, as also used to compute the densities. Taking the profile for GB3 at 300 K as an example (Fig. S6A, orange curve), this leads to $\Delta N_e = 401 \pm 4$ excess electrons in the hydration shell, or, taking the approximate surface area of the GB3 domain

of 3800 Å^2 obtained with the gmx sasa module,[1] an excess of $0.105 \pm 0.1\,e$ per Å^2 of protein surface.

Assuming that these excess electrons would be uniformly distributed in a layer of a thickness of 3 Å, the excess of $0.105\,e/$Å^2 would correspond to uniform excess density of $0.035\,e/$Å^3 or ∼10% of the bulk density of $0.334\,e/$Å^3. This value aligns with common assumptions made by implicit-solvent SAXS predictors that model the hydration shell as a uniform 3Å layer with an excess density of 10–15% of the bulk density.[2]

Supporting Methods

Solvent density calculations

Three-dimensional (3D) solvent densities were computed with the rerun functionality of the GROMACS-SWAXS mdrun module (gmx mdrun -rerun), as activated with the environment variable GMX_WAXS_GRID_DENSITY. See https://cbjh.gitlab.io/gromacs-swaxs-docs/ for documentation. In brief, mdrun places a cubic grid with a 1Å spacing across the volume of the envelope. Then, electron densities are assigned to the grid based on the number of electrons of atoms, here taken from the Cromer-Mann parameters. Densities are assigned based on nuclei position. Densities are averaged over MD frames. With the environment variable GMX_WAXS_GRID_DENSITY_MODE=2, only solvent densities (but not solute densities) are computed, as used here. The mdrun module writes the resulting density to a CUBE file to allow visualization with PyMOL.[70]

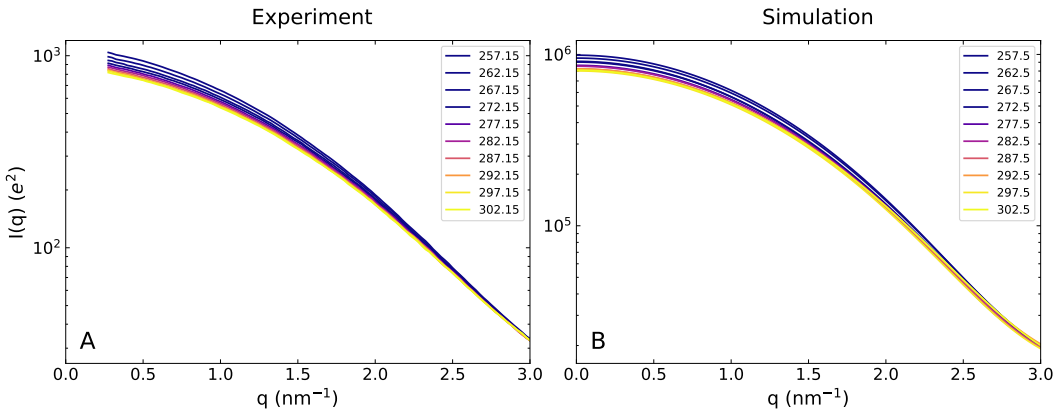


Figure S1: SAXS curves of the GB3 domain from (A) experiment, extrapolated to infinite dilution, and (B) MD simulation over the temperature range shown in the figure legend.

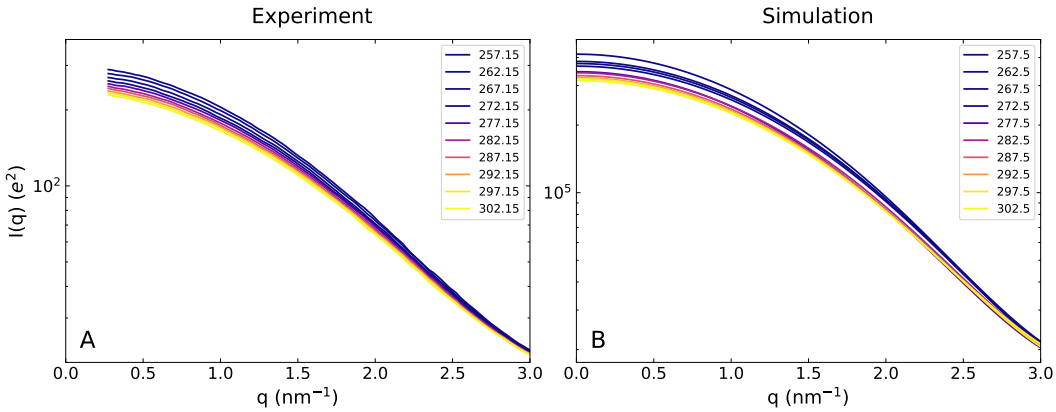


Figure S2: SAXS curves of villin head headpiece from (A) experiment, extrapolated to infinite dilution, and (B) MD simulation over the temperature range shown in the figure legend.

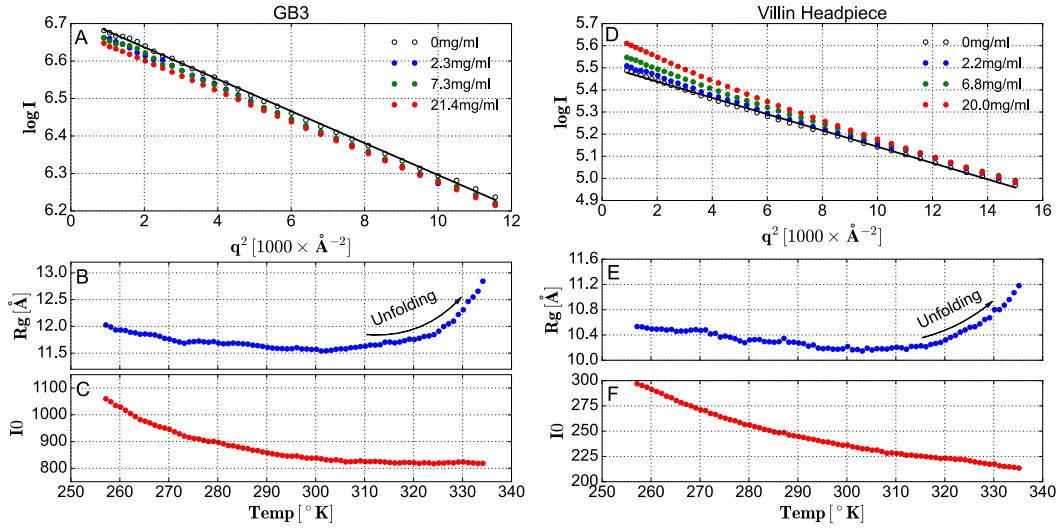


Figure S3: Guinier analysis of GB3 and Villin SAXS scattering. Guinier analysis of GB3 and Villin SAXS scattering. (A) Concentration-dependent SAXS scattering of GB3 at 295 K. The concentration-dependent, buffer-subtracted scattering amplitudes I_i were divided by their relative concentration rc_i to put them on the same scale, i.e., rc_i is approximately [1 (red), 1/3 (green), 1/9 (blue)]. Differences between the normalized scattering curves arise from the concentration-dependent packing structure factor, whose effects can be mitigated by extrapolating the experimental curves to the limit of infinite dilution (open circles). This extrapolated curve corresponds to the infinite dilution limit of a statistically-weighted linear fit of (I_i/rc_i) as a function of concentration at each value of q. The solid line is a Guinier fit of the infinite dilution limit. (B) Radius of gyration R_g from Guinier fits of the infinite dilution limit calculated from concentration-dependent scattering data acquired in 1 deg steps from 257–335 K. (C) I_0 from corresponding Guinier fits. (D–F) Same as (A–C) but for villin headpiece. Due to attractive interparticle interactions, the scattering intensities of Villin headpiece exhibit opposite concentration behavior compared to GB3. The decrease of R_g and I_0 with increasing temperature (while folded) in both GB3 and Villin headpiece correlates with depletion of waters in their hydration shell.

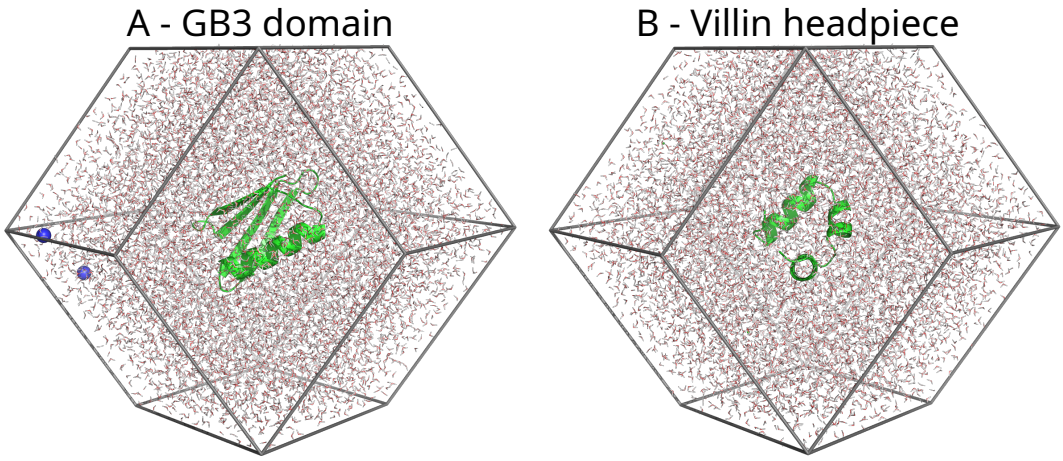


Figure S4: Simulation systems of (A) the GB3 domain and (B) villin headpiece in dodecaheral simulation boxes. Proteins are shown in green cartoon representation, water as red/white sticks, and counter ions as blue spheres.

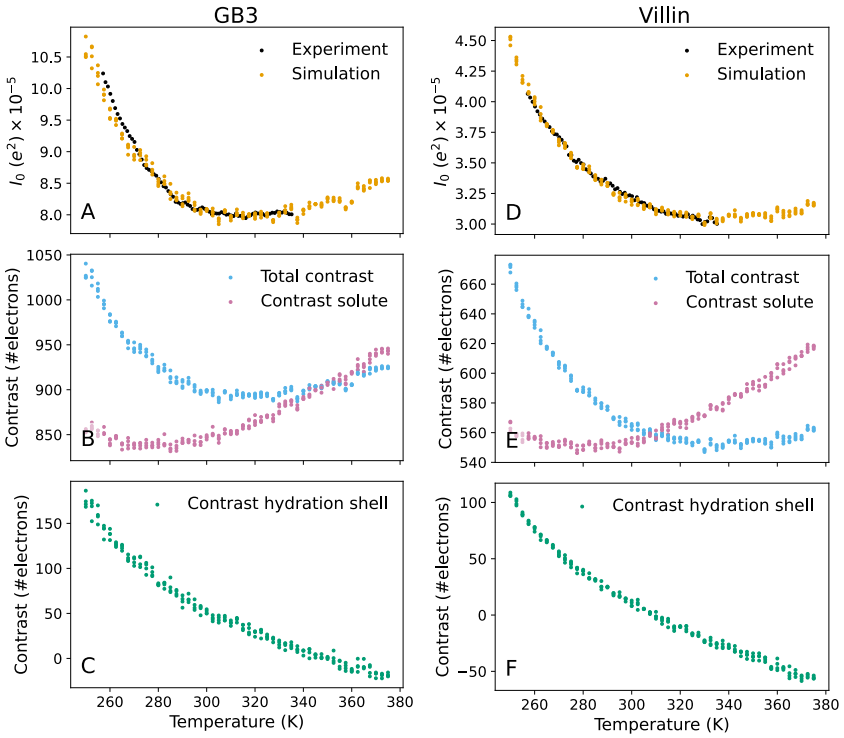


Figure S5: Forward scattering I_0 and analysis of contrast in number of electrons versus temperature for (A–C) the GB3 domain and (D–F) villin headpiece. Same data as presented in Fig. 2, however presented to directly illustrate our analysis. (A/D) I_0 from experiment (black) and backbone-restrained MD simulations (orange) versus temperature. The experimental data was scaled by a constant factor to the simulation data in the temperature range below 303 K. (B/E) Total contrast in number of electrons computed as $I_0^{1/2}$ from MD simulation (blue) and contrast ΔN_e^{prot} due to the solute (pink). (C/F) Contrast of the hydration shell ΔN_e^{hs} given as the difference between the total contrast (B/E, blue) and solute contrast (B/E, pink). Colored dots in panels (A–F) indicate simulation results from four independent simulation replicates per temperature.

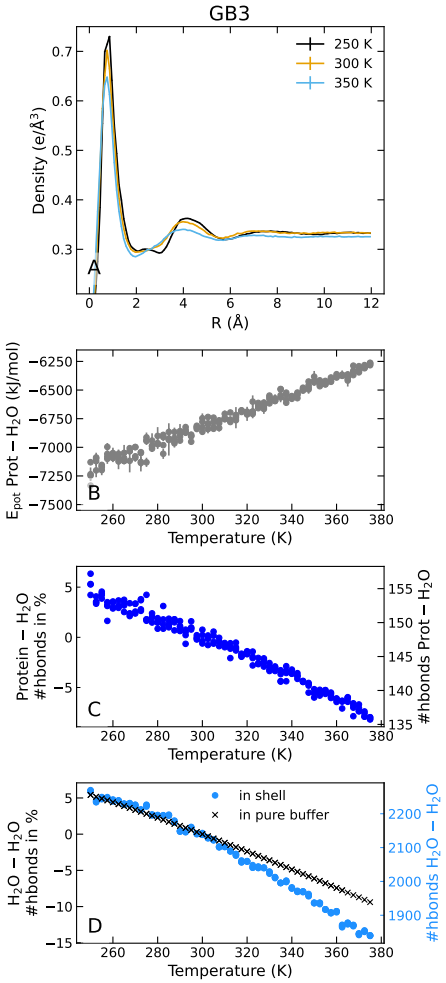


Figure S6: Analysis of the hydration shell of the GB3 domain. (A) Solvent density versus distance from the Van-der-Waals surface of the protein as averaged over the protein surface at three temperatures (see legend), computed from simulations with restrained heavy atoms. (B) Protein–water interaction energy versus temperature computed as sum of Lennard-Jones and short-range Coulomb energies. (C) Number of protein–water hydrogen bonds, plotted either as total number of hydrogen bonds (right ordinate) or as change of number of hydrogen bonds relative to 300 K (left ordinate). (D) Number of water–water hydrogen bonds in the hydration shell (blue dots), defined as water within a distance of 9 Å from the protein surface. Plotted as number of hydrogen bonds (right ordinate) and relative to 300 K (left ordinate). For reference, the relative change of the number of water–water hydrogen bonds is shown for bulk water (black crosses, left black ordinate). Dots in panels B/C indicate simulation results from four independent simulation replicates per temperature.

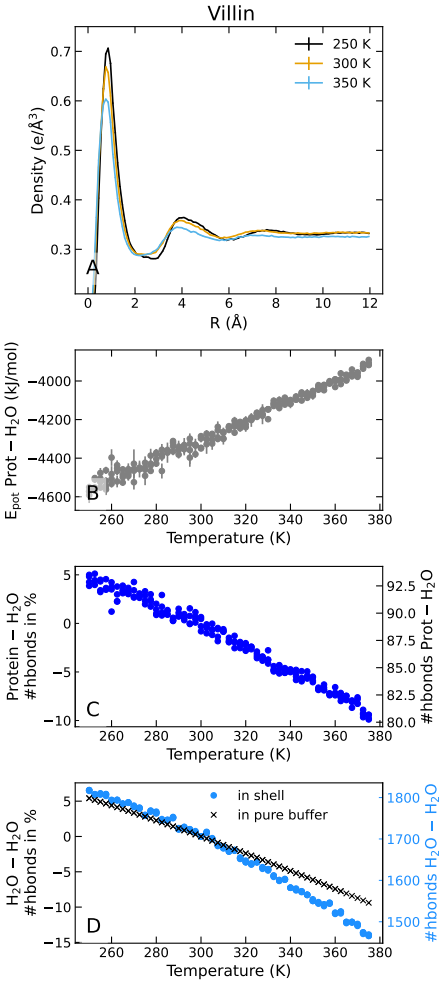


Figure S7: Analysis of the hydration shell of villin. (A) Solvent density versus distance from the Van-der-Waals surface of the protein as averaged over the protein surface at three temperatures (see legend), computed from simulations with restrained heavy atoms. (B) Protein–water interaction energy versus temperature computed as sum of Lennard-Jones and short-range Coulomb energies. (C) Number of protein–water hydrogen bonds, plotted either as total number of hydrogen bonds (right ordinate) or as change of number of hydrogen bonds relative to 300 K (left ordinate). (D) Number of water–water hydrogen bonds in the hydration shell (blue dots), defined as water within a distance of 9 Å from the protein surface. Plotted as number of hydrogen bonds (right ordinate) and relative to 300 K (left ordinate). For reference, the relative change of the number of water–water hydrogen bonds is shown for bulk water (black crosses, left black ordinate). Dots in panels B/C indicate simulation results from four independent simulation replicates per temperature.

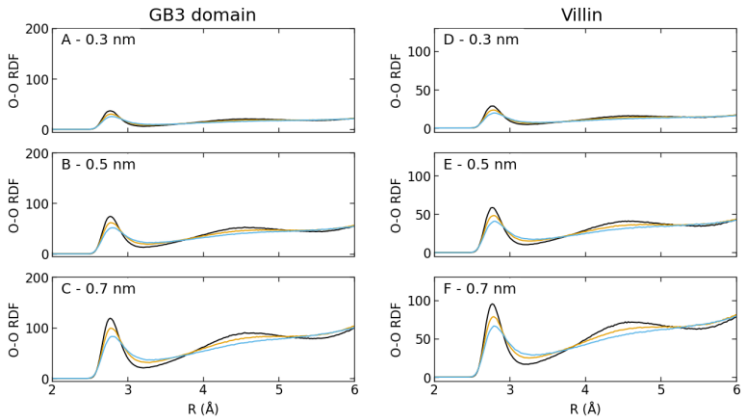


Figure S8: Analysis of the internal water structure within the hydration shell. (A) Non-normalized radial distribution functions (RDFs) between pairs of water oxygen atoms (O–O) for (A–C) the GB3 domain or (D–F) villin headpiece. RDFs are shown for temperatures of 250 K (black), 300 K (orange), and 350 K (blue). RDFs were computed for O–O pairs within distances of (A/D) 0.3 nm, (B/E) 0.5 nm, or (C/F) 0.7 nm from the protein surface. Peaks of the RDFs decay with increasing temperature, indicating a loss of water structure within the hydration shell.

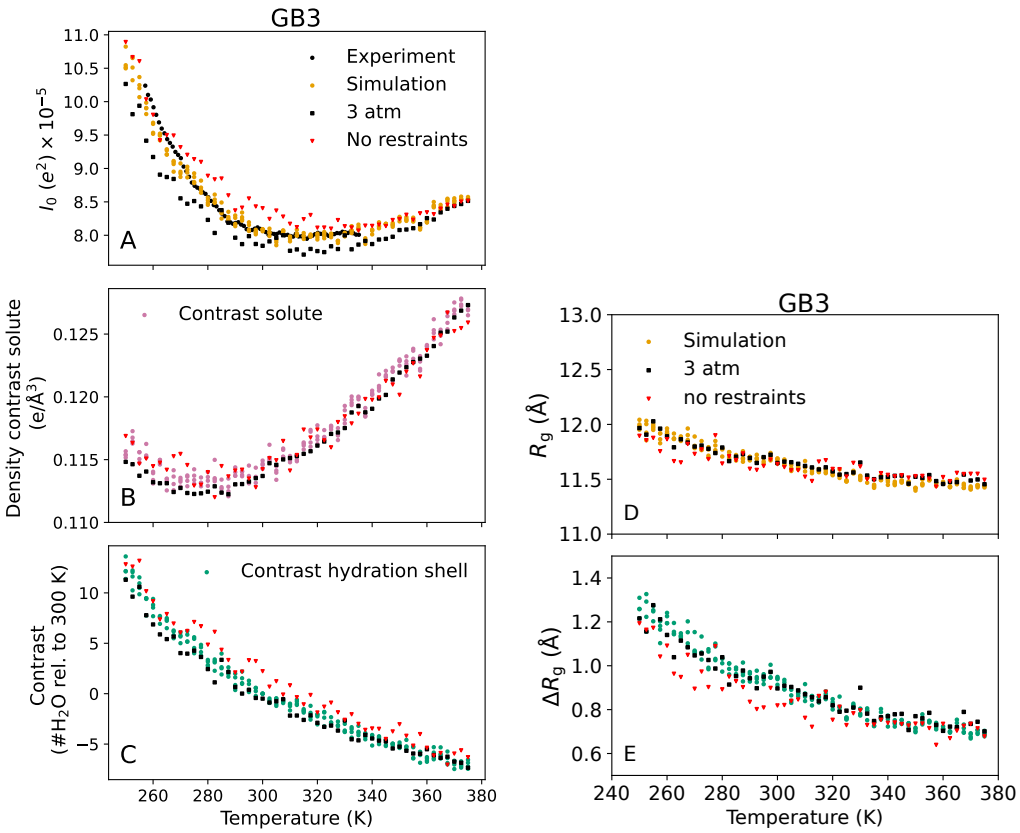


Figure S9: Control simulations with the GB3 domain of 25 ns each for testing the effects of simulating at an increased pressure of 3 atm (black squares) or simulating without positions restraints (red triangles down) instead of simulating with ambient pressure and with position restraints on backbone atoms (yellow, purple, or green circles). (A) Forward scattering I_0 from MD simulations versus temperature, (B) density contrast of the bare protein, and (C) contrast of the hydration shell, according to the analysis in Fig. 2A–C. (D) Radius of gyration R_g from MD simulations and (E) difference ΔR_g between R_g from Guinier analysis (including hydration shell contributions) and R_g of the bare protein, according to the analysis in Fig. 3A/B. Simulating at 3 atm instead of 1 atm or simulating without position restraints has only a small effect on the results. Notably, in unrestrained simulations, the GB3 domain did not unfold at temperatures within simulation time of 25 ns, rationalizing the good agreement with the restrained simulations across the entire temperature range.

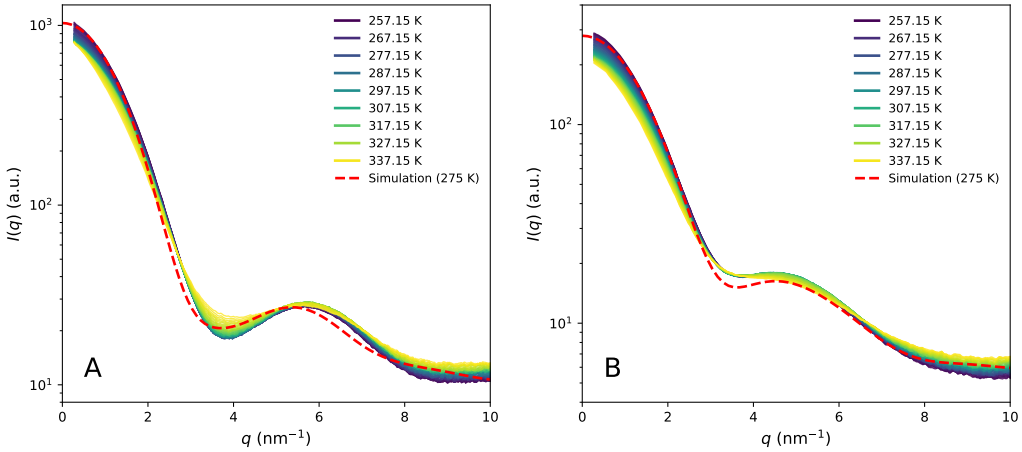


Figure S10: Comparison of SAXS curves from experiment (color gradient) with an example curve from simulations at 275 K (red dashed) for (A) the GB3 domain and (B) villin headpiece to wider angles up to $q = 10\,\mathrm{nm}^{-1}$. The legend shows the color for a few representative temperatures. Reasonable agreement is found, despite the fact that the hydration shell and, thereby, the radius of gyration is not fitted against the data. Instead, for this figure only, the calculated curve was fitted to the experimental curve at the closest temperature via $I_{\mathrm{fitted}}(q) = fI_{\mathrm{exp}}(q) + c$ by adjusting the absolute scale f and a constant offset c. While the absolute scale was likewise adjusted for comparison of I_0 in Figs. 2A/D and S9A, a constant offset was not adjusted for other analyses of this study.

References

(1) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19 – 25.

(2) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.* **1995**, *28*, 768–773.

Chapter 9

Conclusion and Outlook

> "What are the biological implications of your research?" - "Well, I simulate water."
>
> Petter Johansson

The present thesis provides a comprehensive and detailed analysis of the hydration shells of proteins. By combining state-of-the-art MD simulations with experimental SAS techniques the four studies of this thesis addresses a continuing challenge in structural biology and computational biophysics: the complex nature of the hydration shell. Hence, the four studies address fundamental questions concerning how proteins interact with their solvent surroundings and how precisely these interactions can be modeled computationally. The four studies also explore how various physical and chemical parameters influence the behavior of protein hydration shells. This conclusion summarizes the main results of these four studies. More detailed results of the individual projects are provided in the respective chapters.

Development of new refined heavy water pair potentials. In our first study, we addressed a methodological gap by developing customized models for heavy water (D_2O), a solvent commonly used in different biophysical experiments, e.g. neutron scattering experiments. Traditional water models developed primarily for H_2O often fail to capture the peculiarities introduced by the isotopic substitution, including changes in hydrogen bonds, viscosity, and diffusion behavior. By carefully parameterizing and validating three- and four-sided models for D_2O, we introduced three heavy-water models – SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW –, developed on the basis of widely used H_2O models to ensure compatibility with established simulation frameworks. These models have been validated against experimental data – including density, radial distribution functions and diffusion coefficients – demonstrating that these models successfully reproduce the key experimental differences between H_2O and D_2O. This ensures their applicability in biologically relevant situations. Therefore, new models provide a powerful set of tools for simulations

that require accurate modeling of solvent conditions. This methodological advance is crucial for simulations that aim to bridge the gap between theoretical predictions and experimental observations. Accurate models for heavy water increase the reliability of various biomolecular simulations, such as protein folding, conformational changes and interactions in D_2O-rich environments. The models developed in this study provide the basis for a broader application in computational studies that require high accuracy in solvent behavior.

Validation of the hydration shell from molecular dynamics against experimental SAXS/SANS data. The overall structure of the hydration shell, including the density contrast relative to bulk solvent, how many layers, how the shell depends on protein shape and charge, was poorly understood because appropriate high-precision experimental observables were lacking. It was also unclear how well current protein force fields and water models reproduce the hydration shell contrast in quantitative agreement with experiment. Therefore, in the second study of this thesis, a detailed and critical assessment is made of how well MD simulations reproduce the hydration shell of proteins that is also observed in SAXS and SANS experiments.

We performed explicit-solvent MD simulations for five globular proteins, namely RNaseA, lysozyme, xylanase, urate oxidase and glucose isomerase. Overall, we considered 18 different protein force field and water model combinations, including variants from the AMBER and CHARMM families. For each explicit-solvent MD simulation we computed simulated SAS intensity curves. From each of these SAS curves we extracted the radius of gyration R_g of the solute and the hydration shell, and compared them to the R_g of the "bare" protein. The difference radius of gyration, $\Delta R_g = R_g^{SAS} - R_g^{Prot}$, is sensitive to the hydration shell contrast. This means that the hydration shell modulates the measured radius of gyration in SAS. With SAXS an increase in R_g is observed, because the hydration shell has a positive electron-density contrast relative to the bulk solvent. While with SANS in D_2O a decrease in R_g is observed, because the neutron scattering contrast of the protein is negative and the one of the hydration shell is positive. We also found that the solvent density around the protein surface in simulations shows distinct hydration layers: a first – quite dense – layer, a second weaker, and a third very weak. Finally, we compared the simulation results to high-precision consensus experimental SAS data obtained from a worldwide round-robin study for all five proteins [124]. By comparing the simulation results with the experimental data, discrepancies are revealed

and the force fields are tested for their ability to reproduce the thickness, density and spacial structure of the hydration shells. Among the 18 force-field/water-model combinations we tested, many—but not all—produced hydration-shell contrasts in remarkable agreement with experiments. For near-neutral proteins – e.g., RnaseA and xylanase –, simulations using the CHARMM36m force field often underestimated the hydration shell contrast compared to the experiment, especially when used with the standard TIP3P water. For the highly anionic protein glucose isomerase, with a net charge of roughly −60e, the hydration shell contrast was large in comparison to the near-neutral proteins, while some of the charms now overestimate hydration shell contrasts.
Overall, we were able to show that the effect of the hydration shell on R_g is protein-specific: dependent on charge, surface composition and size and shape of the protein, so one cannot assume a "one-size-fits-all" hydration shell. The study emphasizes the need to compare computational models with reliable empirical data to ensure the reliability of simulation-based conclusions. By integrating all-atom MD simulations with simulated and experimental SAS data, this work offers a new way to overcome the long-standing difficulty of accurately characterizing the structure of the protein hydration shell and provide a quantitative observable of the hydration shell: namely ΔR_g (or the difference in R_g between SAXS vs SANS) that is sensitive to how the hydration shell packs onto the protein surface.

Influence of surface amino acids on the protein hydration shell. The hydration shell significantly influences protein folding, binding, and proton transfer, but it has remained unclear to what extent individual amino acid types and protein conformation quantitatively and generally influence this hydration shell. Therefore, in our third study, we investigated how specific protein properties, such as surface amino acids and protein conformations, influence the hydration behavior.
We performed all-atom explicit-solvent MD simulations of two model systems: a globular GB3 domain and an intrinsically disordered peptide, the XAO peptide. In a next step, we computed SAXS curves from the computed explicit solvent trajectories and used SAXS-sensitive quantities – the hydration-induced change in radius of gyration, and the electron-density contrast – to probe the hydration-shell effects. From these data, we derived amino-acid-specific contrast scores describing the influence of each residue type on local water density, either increasing or decreasing it, relative to alanine as a reference. Acidic residues produce the largest positive contrasts. Cationic and polar residues follow; apolar residues often create a deple-

tion layer, which results in a negative contrast. We tested several water models and force-field combinations to check robustness and we found that the amino-acid ranking of the contrasts and ΔR_g values are consistent across the tested water model and protein force field combinations [anionic > cationic (>) polar > apolar], indicating that the relative hydration preferences of amino acids – describing which residues attract or repel water – are a stable conclusion for all simulated configurations tested. In contrast to globular proteins, IDPs, which are characterized by the lack of a fixed tertiary structure, exhibit hydration patterns that are more heterogeneous and sensitive to the sequence composition and conformation. These differences are further modulated by the amino acid composition and the conformational ensemble captured during the simulation. For the disordered XAO peptide, the hydration-shell effect on SAXS-derived R_g is strongly conformation-dependent. And also the comparative analysis between globular proteins and IDPs underlines the fundamental finding that the hydration behaviour is highly context-dependent and cannot be generalized for all protein types. This showed again, that the hydration structure is not universal. It depends on residue chemistry and on the local conformation of the peptide or protein. Despite that the ΔR_g and contrast values are consistent across all tested combinations of water model and protein force field, the study also highlights the sensitivity of hydration properties to force-field selection and demonstrates that methodological choices in MD simulations can (substantially) influence the conclusions drawn.

The amino-acid contrast score – introduced in this paper – links local chemistry to an experimentally accessible, SAXS-sensitive quantity, namely the hydration effect on R_g and on the electron-density contrast, enabling direct comparison between simulation and experiment. The functional roles of IDPs are inseparably linked to their interaction with the solvent. Therefore, understanding hydration layers is not only a scientific question, but a step towards elucidating mechanisms of molecular recognition, phase separation and signal transduction. This has also far-reaching implications not only for fundamental biophysical understanding, but also for the interpretation of hydration-sensitive experimental techniques, the design of biomolecular simulations and the development of new proteins and drugs.

Temperature-dependent behavior of protein hydration shells. It has been elusive how temperature affects the density, thickness, and packing of water near the protein surface because of the lack of high-resolution structural probes of the hydration shell in solution. Therefore, the most recent study investigates how tem-

perature changes affect the protein-solvent interface. Temperature is a critical factor affecting protein conformations, flexibility and solubility, and this study shows how the temperature also modulates the protein hydration shells.

Our collaborators performed T-ramp SAXS experiments spanning from around 255 K up to 335 K on two folded model proteins: the GB3 and the villin headpiece. We performed all-atom MD simulations with explicit solvent across a wide temperature range (250 K up to 375 K) for the same proteins, using backbone-restrained simulations to keep the proteins folded and computing explicit-solvent SAXS profiles from the trajectories. In a following step we calculated key observables from both experiment and simulation: (i) the forward scattering intensity $I(0)$ (which is proportional to contrast), (ii) the radius of gyration R_g of the solute and hydration shell, and (iii) a decomposition of the contrast into contributions from the "bare" protein and from the hydration shell in number of electrons or equivalent water-molecule count. We used the simulation data to map changes of the water density inside the hydration shell, water-protein hydrogen bonds, as well as water-water hydrogen bonds within the shell, and visualized solvent density maps near the protein surface at increasing temperatures. The experimental data reveal, that the forward scattering intensity decreased by more than 20% for both the GB3 domain and the villin headpiece across the measured temperature range, indicating a decreasing solute–solvent contrast. The data also revealed an decrease of the experimentally determined R_g values by approximately 0.4 Å for GB3 domain and around 0.3 Å for the villin headpiece between 257 K and 300 K, before increasing at higher temperature likely due to unfolding of the proteins. The explicit-solvent SAXS profiles calculated from the MD simulations reproduced the temperature trends in both $I(0)$ and R_g below unfolding. The measurable global observables from SAXS – $I(0)$ and R_g – can reflect subtle changes in the hydration shell. In combination with explicit-solvent MD simulations, these changes can be decomposed into a reduction of the hydration shell and other factors such as protein expansion and changes in solvent density. From the decomposition of the simulated data, we were able to show that the hydration shell lost roughly 15 water molecules at 360 K relative to 250 K for both proteins. We could also show that the loss of hydration shell contrast arises not only from increased disorder of the shell – water–water H-bonds decrease by approximately 20% – but also by partially displacing water molecules on the protein surface—these water molecules are more strongly bound at lower temperatures and are lost from the shell at higher temperatures. Solvent-density maps show the first and second hydration layers are decreasing with increasing temperature.

Overall, the hydration shell of folded proteins is remarkably temperature-sensitive: as temperature increases (within the folded regime), the hydration shell is gradually depleted – fewer water molecules remain tightly associated and the shell density drops. A depletion of the hydration layer around proteins, characterized by a decrease in hydration layer density and thickness, can in turn can have profound effects on the protein dynamics and functions. For example, a less structured hydration layer at high(er) temperatures could influence the stability of temporary protein conformations or promote local unfolding processes. These findings are particularly important for understanding thermal denaturation, enzyme catalysis under changing conditions and the behavior of extremophilic proteins.
The study supports the idea that hydration is not a static feature, but a dynamic, temperature-dependent property that needs to be interpreted context-dependent. Furthermore, the agreement between experiment and simulation allows us to use these data to (i) validate temperature-dependent changes in the hydration shell by assessing both the overall contrast and the internal structure, and to (ii) assess whether current water models accurately reproduce these temperature-dependent effects. This study is therefore another example of the synergistic power of combining MD simulations and experiments.

Conclusions and future implications Taken together, these studies provide a multidimensional understanding of protein hydration shells, including methodological development, theoretical validation and biological relevance. Several key aspects emerge from this integrated perspective:

- *Model accuracy and parametrization:* Fitting solvent models, especially for D_2O, increases the applicability of simulations to experimental conditions and improves the interpretation of scattering data.
- *Experimental and computational symbiosis:* Integrating experimental SAS data with MD simulations proves to be a powerful strategy for refining models and validating hypotheses about hydration behavior.
- *Protein-specific behavior:* The hydration shell is not uniform across protein classes; IDPs and globular proteins show different hydration patterns reflecting their structural and functional differences.
- *Dynamic nature of hydration:* Temperature and protein conformations jointly influence the hydration shell, highlighting the need for contextual modeling approaches.

From a broader scientific point of view, this work enriches the conceptual and prac-

tical toolkit for studying protein-solvent interactions. It provides information for the development of more accurate simulation protocols, assists in the interpretation of experimental data, and forms the basis for future studies of how hydration affects the behavior of biomolecules in health and disease. The work thus represents a significant advance in the field of computational and structural biophysics.

Outlook: Challenges and future directions in biomolecular simulations. Current MD methods often rely on classical, non-polarizable force fields that treat atomic charges as fixed quantities. This simplification neglects the dynamic nature of electronic distributions [239], as found in heterogeneous and polar environments such as the hydration shells of proteins. As a result, the simulations may misrepresent the subtle electrostatic interactions between water molecules and protein surfaces, especially in charged or highly flexible regions. Future studies should investigate the implementation of polarizable force fields such as AMOEBA [240–242] or Drude-based [243–245] models that provide a more accurate representation of solvent polarization effects. These models, while computationally more demanding, could provide a more realistic representation of protein-solvent interactions and are expected to improve the quantitative accuracy of hydration shell features.

Although polarizable force fields represent a promising advance, they are not without significant problems. Their implementation is computationally intensive and often requires an order of magnitude more resources than fixed charge models [246, 247]. This limits their routine application to small systems and short simulation periods [239, 248]. Furthermore, the parametrization of polarizable models is complex and still under development, especially for proteins with different side-chain chemistries and dynamic conformations [249]. There are also concerns about transferability, stability of simulations and compatibility with long-established tools and workflows.

Moreover, the refinement and comparison of force fields remains an ongoing challenge, especially for IDPs [250]. As this thesis demonstrates, IDPs exhibit fundamentally different hydration behaviors compared to globular proteins, often requiring specialized force fields that can adequately represent their extended conformational ensembles. The continued development of IDP-specific force fields supported by experimental data such as SAXS and NMR will be crucial for further progress in this field.

Additionally, the application of ensemble refinement approaches - where experimental data guide the simulation of structural ensembles - provides a powerful method

for overcoming the ensemble and sampling problems inherent in MD [251]. Such approaches are particularly relevant for systems characterized by high flexibility and dynamic disorder, where simulations with a single trajectory are not sufficient. Looking further ahead, advances in computational power and the advent of exascale computing will enable the routinized simulation of increasingly complex and large biomolecular systems such as a whole cell, ribosomes or membrane protein complexes over biologically meaningful time periods [252, 253]. Combined with improved (polarizable) force fields and refined sampling algorithms, these simulations will provide unprecedented insight into the dynamic behavior of macromolecular hydration in realistic cellular environments [254].

In conclusion, the future of MD simulations in the field of protein hydration studies lies in overcoming current limitations through more accurate physical modeling, better integration with experimental data, and the use of next generation computational tools. This thesis sets the stage for these developments by providing a rigorous and detailed framework for characterizing hydration phenomena with theoretical accuracy and experimental precision.

Declaration of generative AI and AI-assisted technologies in the writing process

I used chatGPT-4, an AI-assisted editing tool, to improve spelling, grammar, clarity, and readability during preparation. After using this tool, I carefully reviewed and edited the text and take full responsibility for the final content.

Bibliography

[1] J. Watson and F. Crick. "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid." In: *Clin. Orthop. Relat. Res.* 462 (2007), pp. 3–5.

[2] R. M. Daniel, J. L. Finney, M. Stoneham, R. V. Dunn, and R. M. Daniel. "The use of gas–phase substrates to study enzyme catalysis at low hydration". In: *Philos. Trans. R. Soc. Lond. B* 359.1448 (2004), pp. 1309–1320.

[3] M. Levitt and B. H. Park. "Water: now you see it, now you don't". In: *Structure* 1.4 (1993), pp. 223–226.

[4] T. M. Raschke. "Water structure and interactions with protein surfaces". In: *Curr. Opin. Struct. Biol.* 16.2 (2006). Theory and simulation/Macromolecular assemblages, pp. 152–159.

[5] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, and A. E. Garcia. "Water Determines the Structure and Dynamics of Proteins". In: *Chem. Rev.* 116.13 (2016), pp. 7673–7697.

[6] R. Metzler. "The Dance of Water Molecules around Proteins". In: *Physics* 11.59 (2018).

[7] P. Ball. "Water as an Active Constituent in Cell Biology". In: *Chem. Rev.* 108.1 (2008), pp. 74–108.

[8] Y. Levy and J. N. Onuchic. "Water and proteins: A love–hate relationship". In: *Proc. Natl. Acad. Sci. U.S.A.* 101.10 (2004), pp. 3325–3326.

[9] G. Schirò and M. Weik. "Role of hydration water in the onset of protein structural dynamics". In: *J. Phys. Condens. Matter* 31 (2019).

[10] M. Chaplin. "Opinion – Do we underestimate the importance of water in cell biology?" In: *Nat. Rev. Mol. Cell Biol.* 7 (2006), pp. 861–6.

[11] F. R. Salsbury. "Molecular dynamics simulations of protein dynamics and their relevance to drug discovery". In: *Curr. Opin. Pharmacol.* 10.6 (2010). Endocrine and metabolic diseases/New technologies - the importance of protein dynamics, pp. 738–744.

[12] A. Yazhini, S. Chakraborti, and N. Srinivasan. "Protein Structure, Dynamics and Assembly: Implications for Drug Discovery". In: *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design*. Ed. by S. K. Singh. Singapore: Springer Singapore, 2021, pp. 91–122.

[13] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli. “Role of Molecular Dynamics and Related Methods in Drug Discovery”. In: *J. Med. Chem.* 59.9 (2016), pp. 4035–4061.

[14] J. D. Durrant and J. A. McCammon. “Molecular dynamics simulations and drug discovery”. In: *BMC Biol.* 9.71 (2011).

[15] Y. Shi. “A Glimpse of Structural Biology through X-Ray Crystallography”. In: *Cell* 159.5 (2014), pp. 995–1014.

[16] W. Kühlbrandt. “The Resolution Revolution”. In: *Sci.* 343.6178 (2014), pp. 1443–1444.

[17] K. Wüthrich. “The way to NMR structures of proteins”. In: *Nat. Struct. Biol.* 8 (2001), pp. 923–5.

[18] A. H. Kwan, M. Mobli, P. R. Gooley, G. F. King, and J. P. Mackay. “Macromolecular NMR spectroscopy for the non-spectroscopist”. In: *FEBS J.* 278.5 (2011), pp. 687–703.

[19] M. Bieri, A. H. Kwan, M. Mobli, G. F. King, J. P. Mackay, and P. R. Gooley. “Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination”. In: *FEBS J.* 278.5 (2011), pp. 704–715.

[20] R. C. Oliver, W. Potrzebowski, S. M. Najibi, M. N. Pedersen, L. Arleth, N. Mahmoudi, and I. André. “Assembly of Capsids from Hepatitis B Virus Core Protein Progresses through Highly Populated Intermediates in the Presence and Absence of RNA”. In: *ACS Nano* 14.8 (2020), pp. 10226–10238.

[21] E. Choi, J. G. Kim, J. Kim, H. Ki, Y. Lee, S. Lee, K. Yoon, J. Kim, J. Kim, and H. Ihee. “Filming ultrafast roaming-mediated isomerization of bismuth triiodide in solution”. In: *Nat. Commun.* 12 (2021).

[22] C. E. Blanchet and D. I. Svergun. “Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution”. In: *Annu. Rev. Phys. Chem.* 64.1 (2013). PMID: 23216378, pp. 37–54.

[23] J. Trewhella. “Small-angle scattering and 3D structure interpretation”. In: *Curr. Opin. Struct. Biol.* 40 (2016), pp. 1–7.

[24] D. I. Svergun, M. H. J. Koch, P. A. Timmins, and R. P. May. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. Oxford University Press, 2013.

[25] D. Svergun. “Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing”. In: *Biophys. J.* 76.6 (1999), pp. 2879–2886.

[26] D. Franke and D. I. Svergun. "*DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering". In: *J. Appl. Crystallogr.* 42.2 (2009), pp. 342–346.

[27] D. I. Svergun, M. V. Petoukhov, and M. H. Koch. "Determination of Domain Structure of Proteins from X-Ray Solution Scattering". In: *Biophys. J.* 80.6 (2001), pp. 2946–2953.

[28] P.-c. Chen and J. S. Hub. "Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data". In: *Biophys. J.* 107.2 (2014), pp. 435–447.

[29] J. S. Hub. "Interpreting solution X-ray scattering data using molecular simulations". In: *Curr. Opin. Struct. Biol.* 49 (2018), pp. 18–26.

[30] A. Berthaud, J. Manzi, J. Pérez, and S. Mangenot. "Modeling Detergent Organization around Aquaporin-0 Using Small-Angle X-ray Scattering". In: *J. Am. Chem. Soc.* 134.24 (2012), pp. 10080–10088.

[31] N. T. Johansen, M. C. Pedersen, L. Porcar, A. Martel, and L. Arleth. "Introducing SEC–SANS for studies of complex self-organized biological systems". In: *Acta Cryst. D.* 74.12 (2018), pp. 1178–1191.

[32] M. R. Chance, E. R. Farquhar, S. Yang, D. T. Lodowski, and J. Kiselar. "Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution". In: *J. Mol. Biol.* 432.9 (2020), pp. 2973–2984.

[33] *RCSB PDB*. URL: `http://www.rcsb.org/`.

[34] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. 2nd ed. Cambridge University Press, 2004.

[35] B. J. Alder and T. E. Wainwright. "Phase Transition for a Hard Sphere System". In: *J. Chem. Phys.* 27.5 (1957), pp. 1208–1209.

[36] F. H. Stillinger and A. Rahman. "Improved Simulation of Liquid Water by Molecular Dynamics". In: *J. Chem. Phys.* 60.4 (1974), pp. 1545–1557.

[37] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. Phillips. "A three-dimensional model of the myoglobin molecule obtained by X-ray analysis". In: *Nature* 181.4610 (1958), pp. 662–666.

[38] J. A. McCammon, B. R. Gelin, and M. Karplus. "Dynamics of Folded Proteins". In: *Nature* 267 (1977), pp. 585–590.

[39] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, and K. Schulten. "Discovery Through the Computational Microscope". In: *Structure* 17.10 (2009), pp. 1295–1306.

[40] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw. “Biomolecular Simulation: A Computational Microscope for Molecular Biology”. In: *Annu. Rev. Biophys.* 41.1 (2012). PMID: 22577825, pp. 429–452.

[41] B. Z. Zsidó and C. Hetényi. “The role of water in ligand binding”. In: *Curr. Opin. Struct. Biol.* 67 (2021), pp. 1–8.

[42] D. Laage, T. Elsaesser, and J. T. Hynes. “Water Dynamics in the Hydration Shells of Biomolecules”. In: *Chem. Rev.* 117.16 (2017), pp. 10694–10725.

[43] M. K. Gilson and T. Kurtzman. “Free Energy Density of a Fluid and Its Role in Solvation and Binding”. In: *J. Chem. Theory Comput.* 20.7 (2024), pp. 2871–2887.

[44] J. A. Lemkul and D. R. Bevan. “The Role of Molecular Simulations in the Development of Inhibitors of Amyloid β -Peptide Aggregation for the Treatment of Alzheimer’s Disease”. In: *ACS Chem. Neurosci.* 3.11 (2012), pp. 845–856.

[45] O. Coskuner-Weber and V. N. Uversky. “Insights into the Molecular Mechanisms of Alzheimer’s and Parkinson’s Diseases with Molecular Simulations: Understanding the Roles of Artificial and Pathological Missense Mutations in Intrinsically Disordered Proteins Related to Pathology”. In: *Int. J. Mol. Sci.* 19 (2018), pp. 336–401.

[46] G. Torrie and J. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *J. Comput. Phys.* 23.2 (1977), pp. 187–199.

[47] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proc. Natl. Acad. Sci. U. S. A.* 99.20 (2002), pp. 12562–12566.

[48] P. G. Bolhuis, C. Dellago, and D. Chandler. “Sampling ensembles of deterministic transition pathways”. In: *Faraday Discuss.* 110 (1998), pp. 421–436.

[49] C. Dellago, P. G. Bolhuis, and D. Chandler. “Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements”. In: *J. Chem. Phys.* 108.22 (1998), pp. 9236–9245.

[50] P. G. Bolhuis and C. Dellago. “Trajectory-Based Rare Event Simulations”. In: *Rev. Comput. Chem.* John Wiley & Sons, Ltd, 2010. Chap. 3, pp. 111–210.

[51] C. Dellago, P. G. Bolhuis, and P. L. Geissler. “Transition Path Sampling”. In: *Adv. Chem. Phys.* John Wiley & Sons, Ltd, 2002. Chap. 1, pp. 1–78.

[52] R. Rambo and J. Tainer. “Accurate assessment of mass, models and resolution by small-angle scattering”. In: *Nature* 496 (2013), pp. 477–81.

[53] G. Hummer and J. Köfinger. "Bayesian ensemble refinement by replica simulations and reweighting". In: *J. Chem. Phys.* 143.24 (2015).

[54] S. Bottaro, T. Bengtsen, and K. Lindorff-Larsen. "Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy reweighting approach". In: *bioRxiv* (2018).

[55] W. Rieping, M. Habeck, and M. Nilges. "Inferential Structure Determination". In: *Sci.* 309.5732 (2005), pp. 303–306.

[56] A. B. Ward, A. Sali, and I. A. Wilson. "Integrative Structural Biology". In: *Sci.* 339.6122 (2013), pp. 913–915.

[57] S. Bottaro and K. Lindorff-Larsen. "Biophysical experiments and biomolecular simulations: A perfect match?" In: *Sci.* 361.6400 (2018), pp. 355–360.

[58] M. P. Rout and A. Sali. "Principles for Integrative Structural Biology Studies". In: *Cell* 177.6 (2019), pp. 1384–1403.

[59] P. R. L. Markwick, T. Malliavin, and M. Nilges. "Structural Biology by NMR: Structure, Dynamics, and Interactions". In: *PLOS Comput. Biol.* 4.9 (2008), pp. 1–7.

[60] C. J. Knight and J. S. Hub. "WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics". In: *Nucleic Acids Res.* 43.W1 (2015), W225–W230.

[61] P.-c. Chen and J. Hub. "Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics". In: *Biophys. J.* 108 (2015), pp. 2573–84.

[62] F. Persson, P. Söderhjelm, and B. Halle. "The geometry of protein hydration". In: *J. Chem. Phys.* 148.21 (2018), p. 215101.

[63] T. Begušić and G. A. Blake. "Two-dimensional infrared-Raman spectroscopy as a probe of water's tetrahedrality". In: *Nat. Commun.* 14.1 (2023), p. 1950.

[64] K. Ramasesha, S. T. Roberts, R. A. Nicodemus, A. Mandal, and A. Tokmakoff. "Ultrafast 2D IR anisotropy of water reveals reorientation during hydrogen-bond switching". In: *J. Chem. Phys.* 135.5 (2011), p. 054509.

[65] R. A. Nicodemus, S. A. Corcelli, J. L. Skinner, and A. Tokmakoff. "Collective Hydrogen Bond Reorganization in Water Studied with Temperature-Dependent Ultrafast Infrared Spectroscopy". In: *J. Phys. Chem. B* 115.18 (2011). PMID: 21417373, pp. 5604–5616.

[66] D. E. Moilanen, D. Wong, D. E. Rosenfeld, E. E. Fenn, and M. D. Fayer. "Ion–water hydrogen-bond switching observed with 2D IR vibrational echo

chemical exchange spectroscopy". In: *Proc. Natl. Acad. Sci. U.S.A.* 106.2 (2009), pp. 375–380.

[67] A. C. Fogarty and D. Laage. "Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation". In: *J. Phys. Chem. B* 118.28 (2014), pp. 7715–7729.

[68] J. T. King and K. J. Kubarych. "Site-Specific Coupling of Hydration Water and Protein Flexibility Studied in Solution with Ultrafast 2D-IR Spectroscopy". In: *J. Am. Chem. Soc.* 134.45 (2012). PMID: 23101613, pp. 18705–18712.

[69] A. Ghosh, J. S. Ostrander, and M. T. Zanni. "Watching Proteins Wiggle: Mapping Structures with Two-Dimensional Infrared Spectroscopy". In: *Chem. Rev.* 117.16 (2017), pp. 10726–10759.

[70] S. Hume, G. M. Greetham, P. M. Donaldson, M. Towrie, A. W. Parker, M. J. Baker, and N. T. Hunt. "2D-Infrared Spectroscopy of Proteins in Water: Using the Solvent Thermal Response as an Internal Standard". In: *Anal. Chem.* 92.4 (2020), pp. 3463–3469.

[71] N. V. Penkov. "Terahertz spectroscopy as a method for investigation of hydration shells of biomolecules". In: *Biophys. Rev.* 15.5 (2023), pp. 833–849.

[72] D. Ojha, N. K. Kaliannan, and T. D. Kühne. "Time-dependent vibrational sum-frequency generation spectroscopy of the air-water interface". In: *Commun. Chem.* 2.1 (2019), p. 116.

[73] C.-M. Saak and E. H. G. Backus. "The Role of Sum-Frequency Generation Spectroscopy in Understanding On-Surface Reactions and Dynamics in Atmospheric Model-Systems". In: *J. Phys. Chem. Lett.* 15.17 (2024), pp. 4546–4559.

[74] T. Seki, C.-C. Yu, X. Yu, T. Ohto, S. Sun, K. Meister, E. H. G. Backus, M. Bonn, and Y. Nagata. "Decoding the molecular water structure at complex interfaces through surface-specific spectroscopy of the water bending mode". In: *Phys. Chem. Chem. Phys.* 22 (19 2020), pp. 10934–10940.

[75] M. Lukas, R. Schwidetzky, A. T. Kunert, E. H. Backus, U. Pöschl, J. Fröhlich-Nowoisky, M. Bonn, and K. Meister. "Interfacial Water Ordering Is Insufficient to Explain Ice-Nucleating Protein Activity". In: *J. Phys. Chem. Lett.* 12.1 (2021), pp. 218–223.

[76] Y.-R. Qiu, W. Peng, R. Zhang, Y.-H. Wang, W. Guo, and J.-F. Li. "Frontier interfacial water properties characterization and applications". In: *Natl. Sci. Rev.* 12.11 (2025), nwaf284.

[77] T. Yamada and H. Seto. “Quasi-Elastic Neutron Scattering Studies on Hydration Water in Phospholipid Membranes”. In: *Front. Chem.* Volume 8 - 2020 (2020).

[78] N. A. Penkova, M. G. Sharapov, and N. V. Penkov. “Hydration Shells of DNA from the Point of View of Terahertz Time-Domain Spectroscopy”. In: *Int. J. Mol. Sci.* 22.20 (2021).

[79] S. Friesen, M. V. Fedotova, S. E. Kruchinin, and R. Buchner. “Hydration and dynamics of l-glutamate ion in aqueous solution”. In: *Phys. Chem. Chem. Phys.* 23 (2 2021), pp. 1590–1600.

[80] W. Li and Z. Sun. “Second Hydration Shell of Mg2+: Competition between Ion–Water Interaction and Hydrogen Bonding Interaction”. In: *J. Phys. Chem. Lett.* 16.2 (2025), pp. 503–509.

[81] D. M. Leitner, M. Gruebele, and M. Havenith. “Solvation dynamics of biomolecules: Modeling and terahertz experiments”. In: *HFSP Journal* 2.6 (2008), pp. 314–323.

[82] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. “Protein hydration in solution: Experimental observation by x-ray and neutron scattering”. In: *Proc. Natl. Acad. Sci. U.S.A.* 95.5 (1998), pp. 2267–2272.

[83] K. Amann-Winkel, M.-C. Bellissent-Funel, L. E. Bove, T. Loerting, A. Nilsson, A. Paciaroni, D. Schlesinger, and L. Skinner. “X-ray and Neutron Scattering of Water”. In: *Chem. Rev.* 116.13 (2016), pp. 7570–7589.

[84] K. Tompa, P. Bánki, M. Bokor, P. Kamasa, G. Lasanda, and P. Tompa. “Interfacial Water at Protein Surfaces: Wide-Line NMR and DSC Characterization of Hydration in Ubiquitin Solutions”. In: *Biophys. J.* 96.7 (2009), pp. 2789–2798.

[85] V. P. Denisov and B. Halle. “Protein Hydration Dynamics in Aqueous Solution: A Comparison of Bovine Pancreatic Trypsin Inhibitor and Ubiquitin by Oxygen-17 Spin Relaxation Dispersion”. In: *J. Mol. Biol.* 245.5 (1995), pp. 682–697.

[86] R. M. Daniel, J. L. Finney, M. Stoneham, and B. Halle. “Protein hydration dynamics in solution: a critical survey”. In: *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359.1448 (2004), pp. 1207–1224.

[87] D. Laage, T. Elsaesser, and J. T. Hynes. “Perspective: Structure and ultrafast dynamics of biomolecular hydration shells”. In: *Structural Dynamics* 4.4 (2017), p. 044018.

[88] A. C. Fogarty, R. Potestio, and K. Kremer. “Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties”. In: *J. Chem. Phys.* 142.19 (2015), p. 195101.

[89] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. “The missing term in effective pair potentials”. In: *J. Phys. Chem.* 91.24 (1987), pp. 6269–6271.

[90] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. “Comparison of simple potential functions for simulating liquid water”. In: *J. Chem. Phys.* 79.2 (1983), pp. 926–935.

[91] F. H. Stillinger. “Theory and Molecular Models for Water”. In: *Advances in Chemical Physics*. John Wiley & Sons, Ltd, 1975, pp. 1–101.

[92] P. Loche, D. J. Bonthuis, and R. R. Netz. “Molecular dynamics simulations of the evaporation of hydrated ions from aqueous solution”. In: *Commun. Chem.* 5.1 (2022), p. 55.

[93] J. Huang, P. E. M. Lopes, B. Roux, and A. D. J. MacKerell. “Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model”. In: *J. Phys. Chem. Lett.* 5.18 (2014), pp. 3144–3150.

[94] J. M. Kirsh, J. B. Weaver, S. G. Boxer, and J. Kozuch. “Critical Evaluation of Polarizable and Nonpolarizable Force Fields for Proteins Using Experimentally Derived Nitrile Electric Fields”. In: *J. Am. Chem. Soc.* 146.10 (2024), pp. 6983–6991.

[95] J. Schmidt, S. Roberts, J. Loparo, A. Tokmakoff, M. Fayer, and J. Skinner. “Are water simulation models consistent with steady-state and ultrafast vibrational spectroscopy experiments?” In: *Chem. Phys.* 341.1 (2007). Ultrafast Dynamics of Molecules in the Condensed Phase: Photon Echoes and Coupled Excitations, pp. 143–157.

[96] R. Weldon and F. Wang. “Simulating a flexible water model as rigid: Best practices and lessons learned”. In: *J. Chem. Phys.* 158.13 (2023), p. 134506.

[97] J. L. F. Abascal and C. Vega. “A general purpose model for the condensed phases of water: TIP4P/2005”. In: *J. Chem. Phys.* 123.23 (2005), p. 234505.

[98] P. R. L. Markwick and J. A. McCammon. “Studying functional dynamics in bio-molecules using accelerated molecular dynamics”. In: *Phys. Chem. Chem. Phys.* 13 (45 2011), pp. 20053–20065.

[99] J. D. Nickels, H. O’Neill, L. Hong, M. Tyagi, G. Ehlers, K. L. Weiss, Q. Zhang, Z. Yi, E. Mamontov, J. C. Smith, and A. P. Sokolov. “Dynamics of Protein

and its Hydration Water: Neutron Scattering Studies on Fully Deuterated GFP". In: *Biophys. J.l* 103.7 (2012), pp. 1566–1575.

[100] H. Nakagawa and M. Kataoka. "How can we derive hydration water dynamics with incoherent neutron scattering and molecular dynamics simulation?" In: *Biophysics and Physicobiology* 16 (2019), pp. 213–219.

[101] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith. "An extended dynamical hydration shell around proteins". In: *Proc. Natl. Acad. Sci. U.S.A.* 104.52 (2007), pp. 20749–20752.

[102] Y. Ye, X. Chen, J. Huang, L. Zheng, Q. Tang, L. Long, T. Yamada, M. Tyagi, V. G. Sakai, H. O'Neill, Q. Zhang, N. R. de Souza, X. Xiao, W. Zhao, L. Hong, and Z. Liu. "Dynamic entity formed by protein and its hydration water". In: *Phys. Rev. Res.* 6 (3 2024), p. 033316.

[103] R. L. Baldwin. "Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding". In: *Proc. Natl. Acad. Sci. U.S.A.* 111.36 (2014), pp. 13052–13056.

[104] M. Sarter, J. R. Stewart, G. J. Nilsen, M. Devonport, and K. Nemkovski. "Data Analysis of Dynamics in Protein Solutions Using Quasi-Elastic Neutron Scattering Important Insights from Polarized Neutrons". In: *J. Am. Chem. Soc.* 146.41 (2024), pp. 28023–28033.

[105] B. L. Dargaville and D. W. Hutmacher. "Water as the often neglected medium at the interface between materials and biology". In: *Nat. Commun.* 13.1 (2022), p. 4222.

[106] J. Whittaker and L. Delle Site. "Investigation of the hydration shell of a membrane in an open system molecular dynamics simulation". In: *Phys. Rev. Res.* 1 (3 2019), p. 033099.

[107] G. Czakó, E. Mátyus, and A. G. Császár. "Bridging Theory with Experiment: A Benchmark Study of Thermally Averaged Structural and Effective Spectroscopic Parameters of the Water Molecule". In: *J. Phys. Chem. A* 113.43 (2009), pp. 11665–11678.

[108] M. M. Castellanos, A. McAuley, and J. E. Curtis. "Investigating Structure and Dynamics of Proteins in Amorphous Phases Using Neutron Scattering". In: *Comput. Struct. Biotechnol. J.* 15 (2017), pp. 117–130.

[109] O. Dunne, M. Weidenhaupt, P. Callow, A. Martel, M. Moulin, S. Perkins, M. Haertlein, and V. Forsyth. "Matchout deuterium labelling of proteins for small-angle neutron scattering studies using prokaryotic and eukaryotic

expression systems and high cell-density cultures". In: *Eur. Biophys. J.* 46 (2017), pp. 425–432.

[110] F. Gabel. "Chapter Thirteen - Small-Angle Neutron Scattering for Structural Biology of Protein–RNA Complexes". In: *Structures of Large RNA Molecules and Their Complexes*. Ed. by S. A. Woodson and F. H. Allain. Vol. 558. Methods in Enzymology. Academic Press, 2015, pp. 391–415.

[111] K. E. Haslauer, D. Hemmler, P. Schmitt-Kopplin, and S. S. Heinzmann. "Guidelines for the Use of Deuterium Oxide (D_2O) in ^1H NMR Metabolomics". In: *Anal. Chem.* 91.17 (2019), pp. 11063–11069.

[112] G. Zuber, S. Prestrelski, and K. Benedek. "Application of Fourier transform infrared spectroscopy to studies of aqueous protein solutions". In: *Anal. Biochem.* 207.1 (1992), pp. 150–156.

[113] J. Grigera. "An effective pair potential for heavy water". In: *J. Chem. Phys.* 114 (2001), pp. 8064–8067.

[114] S. K. Pal and A. H. Zewail. "Dynamics of Water in Biological Recognition". In: *Chem. Rev.* 104.4 (2004), pp. 2099–2124.

[115] Y. Levy and J. N. Onuchic. "Water mediation in protein folding and molecular recognition". In: *Annu. rev. biophys. biomol. struct.* 35.1 (2006), pp. 389–415.

[116] S. Mondal, S. Mukherjee, and B. Bagchi. "Protein Hydration Dynamics: Much Ado about Nothing?" In: *J. Phys. Chem. Lett.* 8.19 (2017), pp. 4878–4882.

[117] T. Li, A. A. Hassanali, Y.-T. Kao, D. Zhong, and S. J. Singer. "Hydration Dynamics and Time Scales of Coupled Water-Protein Fluctuations". In: *J. Am. Chem. Soc.* 129.11 (2007), pp. 3376–3382.

[118] A. C. Fogarty, E. Duboué-Dijon, F. Sterpone, J. T. Hynes, and D. Laage. "Biomolecular hydration dynamics: a jump model perspective". In: *Chem. Soc. Rev.* 42 (13 2013), pp. 5672–5683.

[119] B. Bagchi. "Water Dynamics in the Hydration Layer around Proteins and Micelles". In: *Chem. Rev.* 105.9 (2005), pp. 3197–3219.

[120] K. Wüthrich, M. Billeter, P. Güntert, P. Luginbühl, R. Riek, and G. Wider. "NMR studies of the hydration of biological macromolecules". In: *Faraday Discuss.* 103 (0 1996), pp. 245–253.

[121] C. J. Crilly, J. E. Eicher, O. Warmuth, J. M. Atkin, and G. J. Pielak. "Water's Variable Role in Protein Stability Uncovered by Liquid-Observed Vapor Exchange NMR". In: *Biochem.* 60.41 (2021), pp. 3041–3045.

[122] B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebele, and M. Havenith. "The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin". In: *Faraday Discuss.* 141 (0 2009), pp. 161–173.

[123] O. Sushko, R. Dubrovka, and R. S. Donnan. "Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected". In: *J. Chem. Phys.* 142.5 (2015), p. 055101.

[124] J. Trewhella, P. Vachette, J. Bierma, C. Blanchet, E. Brookes, S. Chakravarthy, L. Chatzimagas, T. Cleveland, N. Cowieson, and B. e. a. Crossett. "A Round Robin Approach Provides a Detailed Assessment of Biomolecular Small-Angle Scattering Data Reproducibility and Yields Consensus Curves for Benchmarking". In: *Acta Cryst. D.* (2022).

[125] D. Hecht, L. Tadesse, and L. Walters. "Correlating Hydration Shell Structure with Amino Acid Hydrophobicity". In: *J. Am. Chem. Soc.* 115.8 (1993), pp. 3336–3337.

[126] A. Pertsemlidis, A. M. Saxena, A. K. Soper, T. Head-Gordon, and R. M. Glaeser. "Direct evidence for modified solvent structure within the hydration shell of a hydrophobic amino acid." In: *Proc. Natl. Acad. Sci. U.S. A.* 93.20 (1996), pp. 10769–10774.

[127] J. Grdadolnik, F. Merzel, and F. Avbelj. "Origin of hydrophobicity and enhanced water hydrogen bond strength near purely hydrophobic solutes". In: *Proc. Natl. Acad. Sci. U.S. A.* 114.2 (2017), pp. 322–327.

[128] D. A. Doshi, E. B. Watkins, J. N. Israelachvili, and J. Majewski. "Reduced Water Density at Hydrophobic Surfaces: Effect of Dissolved Gases". In: *Proc. Natl. Acad. Sci. U.S.A.* 102.27 (2005), pp. 9458–9462.

[129] A. Poynor, L. Hong, I. K. Robinson, S. Granick, Z. Zhang, and P. A. Fenter. "How Water Meets a Hydrophobic Surface". In: *Phys. Rev. Lett.* 97.26 (2006), p. 266101.

[130] J. Janeček and R. R. Netz. "Interfacial Water at Hydrophobic and Hydrophilic Surfaces: Depletion versus Adsorption". In: *Langmuir* 23.16 (2007), pp. 8417–8429.

[131] M. Ide, Y. Maeda, and H. Kitano. "Effect of Hydrophobicity of Amino Acids on the Structure of Water". In: *J. Phys. Chem. B* 101.35 (1997), pp. 7022–7026.

[132] Y. Djikaev and E. Ruckenstein. "The variation of the number of hydrogen bonds per water molecule in the vicinity of a hydrophobic surface and its

effect on hydrophobic interactions". In: *Curr. Opin. Colloid Interface Sci.* 16.4 (2011), pp. 272–284.

[133] D. Meral, S. Toal, R. Schweitzer-Stenner, and B. Urbanc. "Water-Centered Interpretation of Intrinsic pPII Propensities of Amino Acid Residues: In Vitro-Driven Molecular Dynamics Study". In: *J. Phys. Chem. B* 119.42 (2015). PMID: 26418575, pp. 13237–13251.

[134] G. Lanza and M. A. Chiacchio. "The water molecule arrangement over the side chain of some aliphatic amino acids: A quantum chemical and bottom-up investigation". In: *Int. J. Quantum Chem.* 120.9 (2020), e26161.

[135] G. W. Robinson and C. Cho. "Role of Hydration Water in Protein Unfolding". In: *Biophys. J.* 77.6 (1999), pp. 3311–3318.

[136] K. D. Collins and M. W. Washabaugh. "The Hofmeister effect and the behaviour of water at interfaces". In: *Quart. Rev. Biophys.* 18.4 (1985), pp. 323–422.

[137] G. E. Walrafen and Y.-C. Chu. "Nature of collagen–water hydration forces: a problem in water structure". In: *Chem. Phys.* 258.2 (2000), pp. 427–446.

[138] A. Esser, H. Forbert, F. Sebastiani, G. Schwaab, M. Havenith, and D. Marx. "Hydrophilic Solvation Dominates the Terahertz Fingerprint of Amino Acids in Water". In: *J. Phys. Chem. B* 122.4 (2018), pp. 1453–1459.

[139] R. L. Baldwin. "Temperature Dependence of the Hydrophobic Interaction in Protein Folding". In: *Proc. Natl. Acad. Sci. U.S.A.* 83.21 (1986), pp. 8069–8072.

[140] D. M. Huang and D. Chandler. "Temperature and Length Scale Dependence of Hydrophobic Effects and Their Possible Implications for Protein Folding". In: *Proc. Natl. Acad. Sci. U.S.A.* 97.15 (2000), pp. 8324–8327.

[141] V. R. Agashe and J. B. Udgaonkar. "Thermodynamics of Denaturation of Barstar: Evidence for Cold Denaturation and Evaluation of the Interaction with Guanidine Hydrochloride". In: *Biochem.* 34.10 (1995), pp. 3286–3299.

[142] C. L. Dias, T. Ala-Nissila, J. Wong-ekkabut, I. Vattulainen, M. Grant, and M. Karttunen. "The Hydrophobic Effect and Its Role in Cold Denaturation". In: *Cryobiology* 60.1 (2010), pp. 91–99.

[143] G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal. "Temperature-Controlled Liquid–Liquid Phase Separation of Disordered Proteins". In: *ACS Cent. Sci.* 5.5 (2019), pp. 821–830.

[144] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. 6th edition. New York: Garland Science, 2015.

[145] J.-B. Linse. "Molecular dynamic simulations of a protein-micelle complex". Master thesis. Ulm University, Ulm University of Applied Sciences, University of Gothenburg, 2019.

[146] P. Christen, R. Jaussi, and R. Benoit. *Biochemie und Molekularbiologie*. Berlin, Heidelberg: Springer-Verlag, 2016.

[147] J. W. Kadereit, C. Körner, B. Kost, and U. Sonnewald. *Lehrbuch der Pflanzenwissenschaften*. Berlin, Heidelberg: Springer Spektrum, 2014.

[148] S. Freeman, M. Black, E. Taylor, J. Monroe, L. Allison, G. Podgorski, and K. Quillin. *Biological Science*. 5th edition. Boston: Pearson, 2014.

[149] G. J. Mulder. "Sur la composition de quelques substances animales". In: *Bull. Sci. Phys. Nat. Neerl.* 104 (1938).

[150] H. S. Hartley. "Origin of the Word 'Protein'". In: *Nature* 168 (1951), pp. 244–244.

[151] J. F. Engelhart. "Commentatio de vera materiæ sanguini purpureum colorem impertientis natura, &c". In: *Edinb. Med. Surg. J.* 27.90 (1827), pp. 95–102.

[152] J. F. Engelhart. *Commentatio de vera materia sanguini purpureum colorem impertientis natura*. Göttingen: Dietrich, 1825.

[153] F. L. Hünefeld. *Der Chemismus in der thierischen Organisation: physiologisch-chemische Untersuchungen der materiellen Veränderungen oder des Bildungslebens im thierischen Organismus, insbesondere des Blutbildungsprocesses, der Natur der Blutkörperchen und ihrer Kernchen ; ein Beitrag zur Physiologie und Heilmittellehre ; gekrönte Preisschrift ; mit einer lithographirten Tafel*. Leipzig: F. A. Brockhaus, 1840.

[154] A. E. Blaurock and W. Stoeckenius. "Structure of the Purple Membrane". In: *Nat. New Biol.* 233.39 (1971), pp. 152–155.

[155] D. Oesterhelt and W. Stoeckenius. "Rhodopsin-like Protein from the Purple Membrane of Halobacterium halobium". In: *Nat. New Biol.* 233.39 (1971), pp. 149–152.

[156] D. Oesterhelt and W. Stoeckenius. "Functions of a New Photoreceptor Membrane". In: *Proc. Natl. Acad. Sci. U.S.A.* 70.10 (1973), pp. 2853–2857.

[157] W. Stoeckenius. "From membrane structure to bacteriorhodopsin". In: *J. Membr. Biol.* 139.3 (1994), pp. 139–148.

[158] O. Shimomura, F. H. Johnson, and Y. Saiga. “Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea”. In: *J. Cell. Comp. Physiol.* 59.3 (1962), pp. 223–239.

[159] O. Shimomura. “The discovery of aequorin and green fluorescent protein”. In: *J. Microsc.* 217.1 (2005), pp. 3–15.

[160] D. Nelson and M. Cox. *Lehninger Principles of Biochemistry*. 6th ed. New York and Basingstoke: W.H. Freeman and Company, 2013.

[161] E. Schrödinger. “Quantisierung als Eigenwertproblem”. In: *Ann. Phys.* Vierte Folge, Band 79 (1926), pp. 361–376, 489–527, 734–756.

[162] E. Schrödinger. “Quantisierung als Eigenwertproblem”. In: *Ann. Phys.* Vierte Folge, Band 80 (1926), pp. 438–490.

[163] E. Schrödinger. “Quantisierung als Eigenwertproblem”. In: *Ann. Phys.* Vierte Folge, Band 81 (1926), pp. 109–139.

[164] M. Born and R. Oppenheimer. “Zur Quantentheorie der Molekeln”. In: *Ann. Phys.* Vierte Folge, Band 84 (1927), pp. 457–484.

[165] P. Ehrenfest. “Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik”. In: *Z. Phys.* 45.7 (1927), pp. 455–457.

[166] E. Joos, H. D. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I.-O. Stamatescu. *Decoherence and the Appearance of a Classical World in Quantum Theory*. Springer Berlin, Heidelberg, 2010.

[167] T. Darden, D. York, and L. Pedersen. “Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems”. In: *J. Chem. Phys.* 98.12 (1993), pp. 10089–10092.

[168] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. “A smooth particle mesh Ewald method”. In: *J. Chem. Phys.* 103.19 (1995), pp. 8577–8593.

[169] P. P. Ewald. “Die Berechnung optischer und elektrostatischer Gitterpotentiale”. In: *Ann. Phys.* 369.3 (1921), pp. 253–287.

[170] C. L. Wennberg, T. Murtola, S. Páll, M. J. Abraham, B. Hess, and E. Lindahl. “Direct-Space Corrections Enable Fast and Accurate Lorentz–Berthelot Combination Rule Lennard-Jones Lattice Summation”. In: *J. Chem. Theory Comput.* 11.12 (2015), pp. 5737–5746.

[171] P. Robustelli, S. Piana, and D. E. Shaw. "Developing a molecular dynamics force field for both folded and disordered protein states". In: *Proc. Natl. Acad. Sci. U.S.A.* 115.21 (2018), E4758–E4766.

[172] S. Piana, P. Robustelli, D. Tan, S. Chen, and D. E. Shaw. "Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes". In: *J. Chem. Theory Comput.* 16.4 (2020), pp. 2494–2507.

[173] R. B. Best, W. Zheng, and J. Mittal. "Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association". In: *J. Chem. Theory Comput.* 10.11 (2014), pp. 5113–5124.

[174] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". In: *Proteins* 78.8 (2010), pp. 1950–1958.

[175] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". In: *J. Am. Chem. Soc.* 117.19 (1995), pp. 5179–5197.

[176] C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, and C. Simmerling. "ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution". In: *J. Chem. Theory Comput.* 16.1 (2020), pp. 528–552.

[177] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations". In: *J. Comput. Chem.* 24.16 (2003), pp. 1999–2012.

[178] M. C. Lee and Y. Duan. "Distinguish protein decoys by Using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model". In: *Proteins: Struct., Funct., Bioinf.* 55.3 (2004), pp. 620–634.

[179] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB". In: *J. Chem. Theory Comput.* 11.8 (2015), pp. 3696–3713.

[180] L.-P. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martínez, and V. S. Pande. "Building a More

Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15". In: *J. Phys. Chem. B* 121.16 (2017), pp. 4023–4039.

[181] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations". In: *J. Comput. Chem.* 4.2 (1983), pp. 187–217.

[182] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. J. MacKerell. "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles". In: *J. Chem. Theory Comput.* 8.9 (2012), pp. 3257–3273.

[183] A. D. J. MacKerell, D. Bashford, M. Bellott, R. L. J. 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. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins". In: *J. Phys. Chem. B* 102.18 (1998). PMID: 24889800, pp. 3586–3616.

[184] J. Huang, S. Rauscher, G. Nawrocki, R. Ting, M. Feig, B. de Groot, H. Grubmüller, and A. MacKerell. "CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins". In: *Nat. Methods* 14 (2017), pp. 71–73.

[185] M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen. "Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field". In: *J. Chem. Theory Comput.* 11.7 (2015), pp. 3499–3509.

[186] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations". In: *J. Phys. Chem. B* 111.27 (2007), pp. 7812–7824.

[187] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink. "The MARTINI Coarse-Grained Force Field: Extension to Proteins". In: *J. Chem. Theory Comput.* 4.5 (2008), pp. 819–834.

[188] S. Izadi, R. Anandakrishnan, and A. V. Onufriev. "Building Water Models: A Different Approach". In: *J. Phys. Chem. Lett.* 5.21 (2014), pp. 3863–3871.

[189] F. Naughton. *What is this MD thing anyway?* 2016. URL: https://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway (visited on 08/01/2026).

[190] G. Pomberger and H. Dobler. *Algorithmen und Datenstrukturen: Eine systematische Einführung in die Programmierung*. München: Pearson Studium, 2008.

[191] M. Griebel, S. Knapek, G. Zumbusch, and A. Caglar. *Numerische Simulation in der Moleküldynamik. Numerik, Algorithmen, Parallelisierung, Anwendungen*. Berlin, Heidelberg: Springer, 2004.

[192] S. Gerlach. *Computerphysik. Einführung, Beispiele und Anwendungen*. Berlin, Heidelberg: Springer Spektrum, Springer-Verlag, 2016.

[193] M. Tuckerman, B. J. Berne, and G. J. Martyna. "Reversible multiple time scale molecular dynamics". In: *J. Chem. Phys.* 97.3 (1992), pp. 1990–2001.

[194] K. A. Feenstra, B. Hess, and H. Berendsen. "Improving Efficiency of Large Time-scale Molecular Dynamics Simulations of Hydrogen-rich Systems". In: *J. Comput. Chem.* 20 (1999), pp. 786–798.

[195] B. Hess, H. Bekker, H. Berendsen, and J. G. E. M. Fraaije. "LINCS: A Linear Constraint Solver for molecular simulations". In: *J. Comput. Chem.* 18.12 (1998), pp. 1463–1472.

[196] B. Hess. "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation". In: *J. Chem. Theory Comput.* 4.1 (2008), pp. 116–122.

[197] M. P. Allen and D. Tildesley. *Computer simulation of liquids*. 1st edition. Oxford: Clarendon Pr., 1987.

[198] R. W. Hockney. "The Potential Calculation and Some Applications". In: *Methods Comp. Phys.* 9 (1970), pp. 135–211.

[199] R. W. Hockney and J. W. Eastwood. *Computer simulation using particles*. 1st edition. New York: Taylor & Francis Group, 1988.

[200] G. Bussi, D. Donadio, and M. Parrinello. "Canonical sampling through velocity rescaling". In: *J. Chem. Phys.* 126.1 (2007), p. 014101.

[201] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular dynamics with coupling to an external bath". In: *J. Chem. Phys.* 81.8 (1984), pp. 3684–3690.

[202] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190.

[203] M. P. Allen. “Introduction to Molecular Dynamics Simulation”. In: *NIC Series. Computational Soft Matter: From Synthetic Polymers to Proteins* 23 (2004). Ed. by N. Attig, K. Binder, H. Grubmüller, and K. Kremer, pp. 1–28.

[204] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. “The Protein Data Bank”. In: *Nucl. Acids Res.* 28.1 (2000), pp. 235–242.

[205] H. Berman, K. Henrick, and H. Nakamura. “Berman, H, Henrick, K and Nakamura, H. Announcing the worldwide Protein Data Bank. Nat Struct Biol 10: 980”. In: *Nat. Struct. Biol.* 10 (2004), p. 980.

[206] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, and D. Hassabis. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596 (2021), pp. 1–11.

[207] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar. “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models”. In: *Nucl. Acids Res.* 50.D1 (2021), pp. D439–D444.

[208] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, and D. Baker. “Accurate prediction of protein structures and interactions using a three-track neural network”. In: *Science* 373.6557 (2021), pp. 871–876.

[209] P. Bryant, G. Pozzati, W. Zhu, A. Shenoy, P. Kundrotas, and A. Elofsson. “Predicting the structure of large protein complexes using AlphaFold and Monte Carlo tree search”. In: *Nat. Commun.* 13 (2022).

[210] L. M. F. Bertoline, A. N. Lima, J. E. Krieger, and S. K. Teixeira. “Before and after AlphaFold2: An overview of protein structure prediction”. In: *Front. Bioinform.* 3 (2023).

[211] A. Perrakis and T. K. Sixma. “AI revolutions in biology”. In: *EMBO Rep.* 22.11 (2021), e54046.

[212] J. Trewhella, P. Vachette, J. Bierma, C. Blanchet, E. Brookes, S. Chakravarthy, L. Chatzimagas, T. E. Cleveland IV, N. Cowieson, B. Crossett, A. P. Duff, D. Franke, F. Gabel, R. E. Gillilan, M. Graewert, A. Grishaev, J. M. Guss, M. Hammel, J. Hopkins, Q. Huang, J. S. Hub, G. L. Hura, T. C. Irving, C. M. Jeffries, C. Jeong, N. Kirby, S. Krueger, A. Martel, T. Matsui, N. Li, J. Pérez, L. Porcar, T. Prangé, I. Rajkovic, M. Rocco, D. J. Rosenberg, T. M. Ryan, S. Seifert, H. Sekiguchi, D. Svergun, S. Teixeira, A. Thureau, T. M. Weiss, A. E. Whitten, K. Wood, and X. Zuo. “A round-robin approach provides a detailed assessment of biomolecular small-angle scattering data reproducibility and yields consensus curves for benchmarking”. In: *Acta Cryst. D.* 78.11 (2022), pp. 1315–1336.

[213] D. A. Jacques and J. Trewhella. “Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls”. In: *Protein Sci.* 19.4 (2010), pp. 642–657.

[214] E. Valentini, A. G. Kikhney, G. Previtali, C. M. Jeffries, and D. I. Svergun. “SASBDB, a repository for biological small-angle scattering data”. In: *Nucl. Acids Res.* 43.D1 (2014), pp. D357–D363.

[215] M. Soundararajan, F. Willard, A. Kimple, A. Turnbull, L. Ball, G. Schoch, C. Gileadi, O. Fedorov, E. Dowler, V. Higman, S. Hutsell, M. Sundström, D. Doyle, and D. Siderovski. “Structural diversity in the RGS domain and its interaction with heterotrimeric G protein -subunits”. In: *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008), pp. 6457–62.

[216] O. Glatter. “A new method for the evaluation of small-angle scattering data”. In: *J. Appl. Crystallogr.* 10.5 (1977), pp. 415–421.

[217] A. Guinier. “La diffraction des rayons X aux tres petits angles : applications a l’etude de phenomenes ultramicroscopiques”. In: *Ann. Phys.* 11. Sér. 12 (2010), pp. 161–237.

[218] A. Guinier and G. Fournet. “Small Angle Scattering of X-rays”. In: *Wiley-VCH Verlag GmbH & Co. KGaA, New York* (1955).

[219] R. P. Hjelm Jnr. “The small-angle approximation of X-ray and neutron scatter from rigid rods of non-uniform cross section of finite length”. In: *J. Appl. Crystallogr.* 18.6 (1985), pp. 452–460.

[220] T. Harroun, G. Wignall, J. Fitter, T. Gutberlet, and J. Katsaras. “Neutron Scattering for Biology”. In: 2006, pp. 1–18.

[221] G. H. Lathe and C. R. J. Ruthven. “The separation of substances on the basis of their molecular weights, using columns of starch and water”. In: *Biochem. J.* 60 (4 1955).

[222] G. H. Lathe and C. R. J. Ruthven. “The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water”. In: *Biochem. J.* 62 (4 1956), pp. 665–674.

[223] W. T. Heller. “Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures”. In: *Acta Cryst. D.* 66.11 (2010), pp. 1213–1217.

[224] S. Park, J. P. Bardhan, B. Roux, and L. Makowski. “Simulated x-ray scattering of protein solutions using explicit-solvent models”. In: *J. Chem. Phys.* 130.13 (2009), p. 134114.

[225] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. “*ATSAS 2.8*: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions”. In: *J. Appl. Crystallogr.* 50.4 (2017), pp. 1212–1225.

[226] D. Svergun, C. Barberato, and M. Koch. “CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates”. In: *J. Appl. Cryst.* 28 (1995), pp. 768–773.

[227] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. “Protein hydration in solution: Experimental observation by x-ray and neutron scattering”. In: *Proc. Natl. Acad. Sci. U.S.A.* 95.5 (1998), pp. 2267–2272.

[228] S. Grudinin, M. Garkavenko, and A. Kazennov. “*Pepsi-SAXS*: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles”. In: *Acta Cryst. D* 73.5 (2017), pp. 449–464.

[229] J. Bardhan, S. Park, and L. Makowski. “*SoftWAXS*: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules”. In: *J. Appl. Crystallogr.* 42.5 (2009), pp. 932–943.

[230] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, and A. Sali. “FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles”. In: *Nucl. Acids. Res.* 44.W1 (2016), W424–W429.

[231] D. Schneidman-Duhovny, M. Hammel, and A. Sali. “FoXS: a web server for rapid computation and fitting of SAXS profiles”. In: *Nucl. Acids. Res.* 38.suppl_2 (2010), W540–W544.

[232] F. Poitevin, H. Orland, S. Doniach, P. Koehl, and M. Delarue. “AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models”. In: *Nucl. Acids Res.* 39.suppl_2 (2011), W184–W189.

[233] M. R. Hermann and J. S. Hub. “SAXS-Restrained Ensemble Simulations of Intrinsically Disordered Proteins with Commitment to the Principle of Maximum Entropy”. In: *J. Chem. Theory Comput.* 15.9 (2019), pp. 5103–5115.

[234] E. T. Jaynes. “Information Theory and Statistical Mechanics”. In: *Phys. Rev.* 106 (4 1957), pp. 620–630.

[235] J.-B. Linse and J. S. Hub. “Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW”. In: *J. Chem. Phys.* 154.19 (2021), p. 194501.

[236] J.-B. Linse and J. S. Hub. “Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data”. In: *Commun. Chem.* 6.1 (2023), p. 272.

[237] J.-B. Linse, T. M. Fischbach, and J. S. Hub. “How protein hydration depends on amino acid composition, peptide conformation, and force fields”. In: *Biophys. J.* 125.1 (2026), pp. 255–269.

[238] J.-B. Linse, H. S. Cho, F. Schotte, P. A. Anfinrud, and J. S. Hub. “Depletion of the Protein Hydration Shell with Increasing Temperature Observed by Small-Angle X-ray Scattering and Molecular Simulations”. In: *J. Am. Chem. Soc.* 147.51 (2025), pp. 47117–47125.

[239] J. Melcr and J.-P. Piquemal. “Accurate Biomolecular Simulations Account for Electronic Polarization”. In: *Front. Mol. Biosci.* Volume 6 - 2019 (2019).

[240] C. Liu, J.-P. Piquemal, and P. Ren. “Implementation of Geometry-Dependent Charge Flux into the Polarizable AMOEBA+ Potential”. In: *J. Phys. Chem. Lett.* 11.2 (2020), pp. 419–426.

[241] C. Liu, J.-P. Piquemal, and P. Ren. “AMOEBA+ Classical Potential for Modeling Molecular Interactions”. In: *J. Chem. Theory Comput.* 15.7 (2019), pp. 4122–4139.

[242] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren, and J. W. Ponder. “Tinker 8: Software Tools for

Molecular Design". In: *J. Chem. Theory Comput.* 14.10 (2018), pp. 5273–5289.

[243] G. Lamoureux and B. Roux. "Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm". In: *J. Chem. Phys.* 119.6 (2003), pp. 3025–3039.

[244] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, and A. D. MacKerell. "A polarizable model of water for molecular dynamics simulations of biomolecules". In: *Chem. Phys. Lett.* 418.1 (2006), pp. 245–249.

[245] G. Lamoureux, J. MacKerell Alexander D., and B. Roux. "A simple polarizable model of water based on classical Drude oscillators". In: *J. Chem. Phys.* 119.10 (2003), pp. 5185–5197.

[246] A. D. Mackerell Jr. "Empirical force fields for biological macromolecules: Overview and issues". In: *J. Comput. Chem.* 25.13 (2004), pp. 1584–1604.

[247] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren. "Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications". In: *Annu. Rev. Biophys.* 48.Volume 48 (2019), pp. 371–394.

[248] A. S. Kamenik, P. H. Handle, F. Hofer, U. Kahler, J. Kraml, and K. R. Liedl. "Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding". In: *J. Chem. Phys.* 153.18 (2020), p. 185102.

[249] A. Li, A. Voronin, A. T. Fenley, and M. K. Gilson. "Evaluation of Representations and Response Models for Polarizable Force Fields". In: *J. Phys. Chem. B* 120.33 (2016), pp. 8668–8684.

[250] P. Robustelli, S. Piana, and D. E. Shaw. "Developing a molecular dynamics force field for both folded and disordered protein states". In: *Proc. Natl. Acad. Sci. U.S.A.* 115.21 (2018), E4758–E4766.

[251] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu. "Biomolecular Modeling: Goals, Problems, Perspectives". In: *Angew. Chem. Int. Ed.* 45.25 (2006), pp. 4064–4092.

[252] M. Vendruscolo and C. M. Dobson. "Protein Dynamics: Moore's Law in Molecular Biology". In: *Curr. Biol.* 21.2 (2011), R68–R70.

[253] W. F. van Gunsteren and J. Dolenc. "Biomolecular simulation: historical picture and future perspectives". In: *Biochem. Soc. Trans.* 36.1 (2008), pp. 11–15.

[254] M. Levitt. “Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture)”. In: *Angew. Chem. Int. Ed.* 53.38 (2014), pp. 10006–10018.