Interpretation of In-Solution Small-Angle Scattering Data Using Molecular Dynamics Simulations

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

von

Miloš T. Ivanović

Saarbrücken

2020

Tag des Kolloquiums:	12. Februar 2020
Dekan:	Prof. Dr. Guido Kickelbick
Berichterstatter:	Prof. Dr. Jochen Hub
	Prof. Dr. Volkhard Helms
Vorsitz:	Prof. Dr. Martin Müser
Akad. Mitarbeiter:	Dr. Diego Andrada

"For every complex problem there is an answer that is clear, simple, and wrong." -H. L. Mencken

Acknowledgements

I extend my sincere gratitude to

Prof. Dr. Jochen Hub for his patient and generous guidance, continuous support and a contagious passion for science.

Prof. Dr. Sarah Köster and Prof. Dr. Jörg Enderlein from the Göttingen University (where I spent the first two years of my PhD) for theirs precious time and valuable insights during our TAC meetings.

Prof. Dr. Volkhard Helms for taking the role of Begleitperson when I continued my PhD studies at the Saarland University.

My current and former colleagues from the Computational Biophysics Group for being kind, supportive and open to discuss any topic.

GWDG for the computational time.

Markus Hermann, Stefan Ivanović and Miroslav Anđelković for redaction, comments and suggestions that made the thesis better.

Leonhard Starke for translating the thesis Abstract from English to German.

This thesis would not exist without the constant support from my family and my friends.

Thank You all!

Zusammenfassung

Die akkurate Bestimmung makromolekularer Strukturen erfordert häufig experimentelle Daten mit rechnerischen Methoden zu kombinieren. In dieser Arbeit werden MD Simulationen zur Interpretation von experimentellen Daten, die mittels Kleinwinkel-Röntgenstreuung (SAXS) und Kleinwinkel-Neutronenstreuung (SANS) aufgenommen wurden, genutzt. SAXS- und SANS-Experimente werden unter nahezu natürlichen Bedingungen durchgeführt, liefern jedoch nur ein beschränktes Maß an struktureller Information, welche sich zudem auch nur schwer interpretieren lässt. MD Simulationen eignen sich besonders gut zur Kombination mit Kleinwinkelstreuexperimenten, da die Simulationen verwendet werden können, um experimentelle Daten zu interpretieren. Umgekehrt werden experimentelle Daten benutzt, um Simulationen zu validieren und sie nötigenfalls zu lenken.

Im Folgenden werden vier verschiedene jedoch verbundene Fragestellungen behandelt. Im ersten Abschnitt wird der Einfluss einer lonenwolke auf die Interpretation von SAXS-Daten geladener Proteine untersucht. Danach wird die Form und Größe tensidbasierter Mizellen untersucht. Dies ist ein Ausgangspunkt, um die Stabilität von Protein-Detergenz Komplexen während des Lösens von Membranproteinen zu erhöhen. Im nächsten Schritt wird ein Ensemble tensidbasierter Mizellen bestimmt, welches mit expirmentellen Daten übereinstimmt und es ermöglicht, den Einfluss verschiedener Effekte auf die SAXS Kurve zu studieren. Dies stellt einen wichtigen Schritt dar, um experimentelle SAXS Daten besser zu verstehen. Als vierten Punkt demonstrieren wir, wie die Kombination von SAXS- und SANS-Daten zusammen mit MD-Simulation eine genaue Strukturbestimmung von Protein-Detergenz-Komplexen ermöglicht.

Abstract

The accurate determination of macromolecular structures often necessitates joint experimental and computational efforts. In this thesis, MD simulations are engaged to interpret small-angle scattering data of X-rays (SAXS) and neutrons (SANS). SAXS and SANS experiments are performed under near-native conditions, but provide only limited amount of structural information that are, in addition, difficult to interpret. Therefore, MD simulations are highly compatible with small-angle scattering experiments - simulations are used to interpret experimental data and in turn, experimental data are used to validate, and if necessary to guide simulations.

Here, four different yet related questions are addressed. First, we quantify the influence of the ion cloud on interpreting SAXS data of charged proteins. Secondly, we study the size and the shape of detergent micelles, as this represents a starting point in improving the stability of protein-detergent complexes during the solubilization of membrane proteins. In the next step, we derive an ensemble of detergent micelles in agreement with experimental data, enabling us to study the influence of various effects of SAXS curves and thereby making an important step towards the better understanding of the SAXS experimental data. Finally, we demonstrate how SAXS and SANS data can jointly be combined with MD simulations, allowing for fine structural characterization of protein-detergent complexes.

Content

Chapter 1

Introduction

A fundamental goal of modern structural biology and biophysics is to determine how biomolecules dynamically reorganize to perform their functions. Obtaining biomolecular structures and ensembles often necessitates combining the data from multiple different experiments. Even then, the information content of experimental data may be too low to define all relevant degrees of freedom of the examined biomolecule. Therefore, to derive the ensemble of biomolecular structures while avoiding overfitting, experimental data are complemented with additional physico-chemical knowledge [1–18].

Three experimental techniques enable biomolecular structure determination with atomic resolution: (i) X-ray crystallography, (ii) nuclear magnetic resonance spectroscopy (NMR) and (iii) cryogenic electron microscopy (cryo-EM). These three techniques are often complemented with low-resolution techniques, either to harvest additional information about the investigated biomolecule or to decrease the risk of overfitting. Among the spectrum of available techniques, the application of small angle scattering of X-rays (SAXS) or neutrons (SANS), as well as the range of spectroscopic techniques were shown to be particularly beneficial. A brief overview of advantages and disadvantages of the mentioned techniques with the focus on SAXS and SANS is given in Section 1.1. The main reason for the usage of MD simulation to interpret SAXS and SANS data is given in Section 1.2. Finally, the aim of the present thesis is outlined in Section 1.3. A detailed introduction of specific projects discussed in this thesis is given at the beginning of Chapters 3-6.

1.1 SAXS and SANS among other experimental methods

X-ray crystallography To apply this method, a biomolecule has to be purified and crystallized. Subsequently, the crystallized sample is irradiated by X-rays and the diffraction pattern is recorded. Based on the diffraction pattern and prior knowledge (e.g bond lengths and angles, sequences of amino acids), the electron density distribution can be determined and correlated with the mean positions of the atoms [8, 19, 20].
Ever since the first protein structures were determined in the mid-twentieth century [21, 22], X-ray crystallography has been the cornerstone of biomolecular three-dimensional structure determination [3, 19, 23]. About 89% of the biomolecular structures deposited in the Protein Data Bank (PDB) were determined using X-ray crystallography [8]. While this experimental technique allows for the three-dimensional structure determination down to an angstrom resolution, the preparation of the crystal sample may be a daunting task, especially in case of membrane proteins. Additionally, conformational heterogeneity is restricted by the crystal lattice [3–5, 17, 24, 25].

NMR In a typical NMR experiment, a biomolecule is placed in a magnetic field and probed with radio-frequency signals. The absorption of the radio-frequency signals is measured and used to determine the distance between atomic nuclei and dihedral angles. With this information, a set of three-dimensional macromolecular models compatible with the NMR data can be delivered [8, 19, 26].
In contrast to crystallography, the NMR experiment is performed in solution (i.e. under near-native conditions), and the set of structures is obtained, which is particularly useful when studying flexible proteins. On the other hand, NMR is limited to relatively small biomolecules or parts of larger biomolecules, up to about 50 kDa, as the signals in the NMR spectra of larger biomolecules overlap, complicating the structure reconstruction. [3, 17, 27–30]. This limit can in some cases be increased up to 900 kDa by pairing NMR with transverse relaxation optimized spectroscopy and deuteration [31, 32]. Another disadvantage of NMR is the high cost of the experimental setup, as it requires strong magnets. Currently, about 8% of the biomolecular structures deposited in the PDB were determined using NMR [8].

Cryo-EM Electron microscopy uses a beam of electrons and a system of lenses to directly investigate a biomolecule. Nowadays, it has become a standard to rapidly freeze a sample to cryogenic temperature, obtaining a thin layer of non-crystalline ice. In this manner, a large number of single particles are preserved in different, biologically relevant

conformations, while the demanding procedure of crystallization is avoided [8, 19, 33]. In the early 1990s, cryo-EM was a low-resolution technique, allowing for the three-dimensional structure determination with the resolution of about 40 Å. That imroved to about 10 Å at the beginning of this century [34–36]. Owing to the development of a new generation of electron detectors and improvements in image processing procedures, the progress in cryo-EM over the last few years is so swift that it is often called a revolution. Nowadays, the resolution of a structure determined using cryo-EM is approaching the resolution obtained using crystallography [18, 36, 37]. The exponential growth of the cryo-EM usage can be illustrated by the number of structures deposited in the PDB: while the first cryo-EM structure was deposited in 1997, this number in 2013 was only 520. By the end of 2019, this number increased to 4145, including 1423 structures deposited during the year 2019 [8]. In contrast to NMR, cryo-EM is not limited to small biomolecules, and the amount of necessary sample is considerably lower. Despite these advantages, sample preparation and reproducibility, image processing as well as high costs remain a challenge [11, 38–40]. While cryo-EM is argued to allow for sample preparation in near-native conditions [11, 41], the possible discrepancy to native conditions is still to be evaluated [42].

Spectroscopic methods While spectroscopic methods usually provide only a limited amount of structural information, they often serve as a significant supplement, either to validate or to complement high-resolution structures. The fact that spectroscopic methods commonly provide high time resolution at low cost rationalizes their broad range of applications [18, 43–45]. For example, Förster resonance energy transfer (FRET) is often used to measure distances between domains of a protein [46–49]. In a wider biophysical context, FRET is used to detect interactions between proteins [50, 51] and to investigate lipid rafts in cell membranes [52]. The main disadvantage of this method is that it requires labeling, which can cause changes in the local environment [53]. Other examples include the usage of infrared spectroscopy to probe protein conformational changes associated with protein interaction with metal ions or other macromolecules [54], and the usage of fluorescent and Raman spectroscopy to characterize the secondary structure of the protein [55, 56].

SAXS In a typical SAXS experiment, a solution of macromolecules is irradiated by a beam of collimated monochromatic X-rays and a scattering pattern is recorded on a detector. In order to obtain structural information about the macromolecule, a scattering pattern of the pure solvent (buffer) has to be recorded and subsequently subtracted from the scattering pattern of the solution. The resulting scattering pattern (excess scattering intensity) encodes the structural information about the macromolecule under

investigation [24].

It is important to note that the density of the water around the biomolecule differs from the density of bulk water (the so-called hydration layer) [57]. Therefore, the excess scattering intensity encodes both the information about the macromolecule *and* the hydration layer.

The utilization of SAXS to study the structure of biomolecules originates in the first decades of the twentieth century [29, 58]. SAXS experiments can be performed under a great variety of solution conditions, including near-native conditions, without any limitation to the size of the biomolecule, and at a reasonable cost, which is why it rapidly became an important tool for studying the size and the shape of biomolecules. However, the applicability of SAXS was limited over the next decades, mainly due to two challenges: (i) low scattering contrast between the biomolecule and the solvent; (ii) unclear data content of a typical SAXS curve. Both topics will be further discussed later on.

The applicability of SAXS flourished over the latest two decades, owing to the improvements of light sources, detectors, sample preparation methods, as well as the methods and software for data collection and analysis. Nowadays, SAXS experiments alone are routinely used to obtain information about the biomolecular size and flexibility, as well as to construct a three-dimensional envelope of a biomolecule with a resolution higher than 1.5 nm, which is often sufficient for addressing important biological questions [2, 3, 6, 15, 17, 29, 59–68]. Additionally, although generally considered a low-resolution technique, even small differences in macromolecular conformations are detectable in a scattering profile. This is often used in conjugation with computational methods that allow for SAXS profile prediction, to validate or disprove structural models [17, 69–72]. Recently established standards for quality control, and data deposition [73, 74] further led to reduced risk of experimental error. Additionally, newly developed SEC-SAXS approach [75, 76] greatly improves data quality [15, 17]. Advances in automated sample handling and preparation brought SAXS to the level of a high-throughput method [3, 17]. A particularly promising direction in further development of SAXS as a method lies in time-resolved SAXS (TR-SAXS). When performed at synchrotrons, these experiments capture transient and evolving macromolecular conformations occurring on 100 ps timescales [15, 17, 77]. When paired with free-electron lasers, such experiments reach sub-picosecond time resolution, offering invaluable insights into ultra-fast protein dynamics [15, 78, 79].

Beside the above applications, an additional value of SAXS lies in its compatibility with other experimental techniques. SAXS is often combined with crystallography, NMR and cryo-EM, as the overall structure and flexibility of the biomolecule determined by SAXS complement or validate high-resolution structures obtained using these techniques. Additionally, SAXS data are often compared with intramolecular distances obtained using

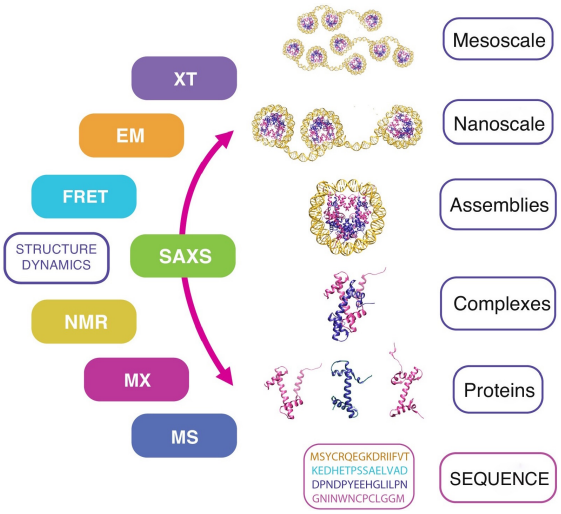

FIGURE 1.1: Compatibility of SAXS and other experimental techniques to study the biomolecular structure. XT denotes X-ray tomography, MX macromolecular crystallography, and MS mass spectroscopy. Figure adopted from Ref. 17.

FRET [2, 5, 29, 49, 80–86]. Being applicable on any biomolecular size and being compatible with variety of experimental techniques (Fig. 1.1), SAXS is expected to play a significant role in the further integrative modeling studies [17].

In spite of all the improvements over the last decades, interpretation of experimental SAXS data remains challenging. Namely, since the information content of data is low, and since it is generally unclear how the structural information is distributed over the scattering profile, interpretation of SAXS data comes with a substantial risk of overfitting [15]. Therefore, new methods to back-calculate the SAXS curve, as well as to combine SAXS data with other experimental and computational techniques, are constantly developing.

The focus of this thesis is to interpret SAXS (and SANS) experimental data using all-atom MD simulations. Namely, the detailed physico-chemichal knowledge that MD simulations contain greatly complements the information content of SAXS experiments. In turn, SAXS data can be used to validate, and if necessary, refine MD simulations [15, 87–94]. This will be further discussed in Section 1.2 and Chapter 2. The theoretical basis of SAXS, as well as methods to interpret SAXS experiments, are discussed in Chapter 2. Four projects related to the combination of SAXS data and MD simulations are given in Chapters 3-6.

SANS In a typical SANS experiment, a solution of macromolecules is irradiated by a collimated beam of neutrons. While the experimental setup and output, as well as the mathematical basis of scattering, are similar to SAXS, there is one important difference between SAXS and SANS: the beam of X-rays is mainly scattered by the electrons

of the atoms in a sample, and the electron scattering length density (SLD) is always positive; in contrast, the beam of the neutrons is mainly scattered by the nuclei, and the nuclear SLD can take both positive or negative values, depending on the nucleus, and often isotope type. For example, the scattering length of hydrogen is negative, and the scattering length of deuterium positive. This feature can be used to study components of biomolecular complexes or the subunits of larger biomolecules. Namely, one component of a biomolecular complex or a subunit of the biomolecule can be made "invisible" to scattering, by varying the H_2O to D_2O ratio.
The main disadvantage of SANS compared to SAXS is the considerably higher cost of the experiment. In addition, SANS is limited to smaller angles than SAXS, due to the low signal-to-noise ratio and incoherent scattering at wider scattering angles [4, 70, 94–97].
Here, it should be noted that the term "in-solution scattering" is commonly used in literature to refer to SAXS, WAXS (wide-angle scattering, which differs from SAXS only by the measured angle of scattering) and SANS. When referring to both SAXS and SANS, the term SAS (small-angle scattering) is commonly used.
More details about SANS theory and the methods used to interpret SANS experiments are given in Chapter 2. The project that combine SAXS and SANS data with MD simulations is described in Chapter 6.

1.2 Motivation for all-atom molecular dynamics simulations

In general, a wide range of physico-chemical models are used to complement experimental data during biomolcular structure determination. The resolution and the reliability of the obtained structures are determined by the information content of the experimental data and the level of physico-chemical knowledge contained in the used model. Since the information content of SAS data is low, very detailed physico-chemical models have to be used to derive reliable atomic macromolecular structures. As MD simulations contain detailed and accurate physico-chemical knowledge [15], they have been used throughout this thesis to derive atomic structures of different macromolecules.
Compatibility of MD simulations and SAS, as well as the strategies to combine SAS data with MD simulations will be discussed in Chapter 2, upon the introduction of the theoretical basis of MD and SAS.

1.3 Aim of the present thesis

This thesis aims to interpret SAXS data of detergent micelles and proteins, as well as SAS data of a protein-detergent complex. In Chapter 3 we quantify the influence of the ion cloud on SAXS profiles of charged proteins. Chapter 4 is dedicated to deriving atomic models of detergent micelles by incorporating SAXS data as an energetic restraint into MD simulations. In Chapter 5 an ensemble of detergent micelles is refined against SAXS data. Having the first reliable ensemble of the soft matter system in agreement with the experimental data enabled us to investigate contributions of different effects (model asymmetry, shape fluctuations, disorder and atomic details) on the SAXS curve. Finally, Chapter 6 demonstrate how SAXS and SANS data can be jointly used with MD simulations in order to precisely derive the structure of membrane-protein detergent complexes. Although these 4 chapters pursue relatively different inquires, the common denominator for all of them is the usage of all-atom MD simulations to predict SAS profiles. It is important to note that the commonly used implicit-solvent methods to predict SAS curves would not provide satisfactory precision in these studies. Additionally, in cases when profiles predicted from unbiased simulations do not perfectly agree with experimental profiles (Chapters 4 and 5), state of the art methods are used to bias MD simulations with SAXS data, thus overcoming simulation imperfections and driving the simulations into conformations that satisfy the experimental data.
A brief motivation and introduction to the four studies is given in the following paragraphs, while more detailed introductions are given at the beginning of each chapter.

Quantifying the influence of the ion cloud on SAXS profiles of charged proteins The interpretation of SAXS data is complicated by scattering contributions from the hydration layer and, in the case of charged macromolecules, from the ion cloud. The influence of the hydration layer on SAXS curves has been extensively studied previously [57, 88–90, 98–104]; in contrast, the influence of the ion cloud on SAXS curves, although nearly equally relevant, remained poorly understood, potentially leading to significant uncertainties during the interpretation of the data. To close this gap, we quantified the effect of the ion cloud on the radius of gyration (R_g) of the charged proteins using three computational models with decreasing complexity. Namely, we used (i) all-atom MD simulations in conjunction with explicit-solvent SAXS predictions; (ii) non-linear Poisson-Boltzmann calculations; and (iii) linearized Poisson-Boltzmann calculations in conjunction with a simple spherical protein model. The calculations are validated against experimental SAXS data. We find remarkable agreement between the three methods, suggesting that the ion cloud effect on R_g may be predicted even with a greatly simplified protein model. As a result of this study, we provide a computationally

efficient alternative to MD simulations for SAXS users interested in the effect of the ion cloud on R_g of charged proteins.

Temperature-dependent atomic models of detergent micelles refined against small-angle X-ray scattering data Detergent micelles play an important role in structural biology and biophysics. Namely, they are often used as membrane mimics to solubilize membrane proteins for biochemical or structural studies. Recent studies suggested that protein-detergent complexes are stable if the hydrophobic cross-section of the detergent micelle matches the membrane protein [105, 106]; however, since micellar structures are still uncertain, finding a suitable detergent remains a matter of trial and error. SAS experiments have been frequently used to study micelles, but the data content of such experiments is insufficient to derive atomic models [107–112]. Alternatively, MD simulations have been employed, but it remained unclear to which extent force field imperfections biased the derived models [113–118]. Therefore, we used SAXS-driven simulations [91] to derive the first accurate atomic models of two maltoside detergent micelles, n-dodecyl-β-D-maltoside (DDM) and n-decyl-β-D-maltoside (DM), in a range of experimentally relevant temperatures. By comparing calculated and experimental SAXS profiles, we found that the aggregation numbers of both micelles decrease with a temperature increase. These results are in remarkable agreement with independent estimates of the aggregation number from the experimental data and from an analytic model [119]. Further, in contrast to previously assumed two-axial ellipsoid micellar shapes [107, 108, 120–122], we determined that the shape of the DDM and DM micelles is a general tri-axial ellipsoid. Finally, we found that the minor maltoside micelle axes closely mimic lipid bilayers. While the aggregation numbers of the DDM and DM micelles decrease with increasing temperature, the thickness of the minor micelle axis is largely temperature-invariant, rationalizing why protein-detergent complexes may be stable over a wide temperature range.

SAXS Curves of Detergent Micelles: Effects of Asymmetry, Shape Fluctuations, Disorder, and Atomic Detail SAXS data of detergent micelles are often interpreted by fitting simple symmetrized continuum models, such as prolate or oblate ellipsoids, to the data, thereby neglecting shape fluctuations, disorder, possible asymmetry, and atomic details [107, 108, 121, 122]. It is not surprising that such models may explain only the low-angle scattering data whereas they fail at medium to wider angles. In order to derive an accurate, atomic, heterogeneous *ensemble* of a detergent micelle, we coupled MD simulations to experimental SAXS data, measured up to wider scattering angles [123]. Based on a recently developed method for ensemble refinement with commitment to the principle of maximum entropy [93], we ensure that the refined

ensemble is statistically well defined. Having an atomic ensemble in excellent agreement with the data as an accurate reference, we resort to simplified continuum models to decipher the effects of shape asymmetry, shape fluctuations, disorder and atomic details on the SAXS curve of the micelle. Our analysis shows that the previously assumed shape symmetry and lack of atomic details had been the main sources of discrepancy between continuum models and experiment, whereas disorder and shape fluctuations had been less problematic. Notably, we find remarkable agreement between the overall micelle shape determined using ensemble-refined MD simulations, and using a continuum model, if the symmetry of such a model is dropped. However, atomic and molecular details, as captured by the MD simulation, are necessary to quantitatively explain the SAXS curve at wider scattering angles.

Merging in-solution X-ray and neutron scattering data allows for fine structural analysis of membrane-protein detergent complexes Chen and Hub demonstrated previously that the number of detergent molecules forming a protein-detergent complex can be estimated by comparing the experimental SAXS profile with the SAXS profile calculated from a series of all-atom MD simulations with different number of detergent molecules [71]. Here, we followed the same approach to structurally characterize the ammonium transporter AmtB embedded in a DDM detergent belt. We found that the number of detergent molecules forming the protein-detergent complex determined by comparing experimental profiles with the profiles calculated from all-atom MD simulations is in an excellent agreement with the number determined using analytical ultracentrifugation. To further support this finding, we calculated the SANS profiles at different H_2O to D_2O ratios from the same sets of MD simulations and found good agreement with experimental SANS profiles. Finally, we demonstrated that the overall shape of the protein-detergent complex obtained using MD simulations is in a reasonable agreement with the *ab initio* models obtained using SANS data together with the SAXS data.

Chapter 2

Theory and Methods

This chapter provides a brief introduction to the underlying theory of the methods applied in this thesis. MD simulations are introduced in Section 2.1. SAXS is discussed in Section 2.2, and SANS in Section 2.3. Finally, direct coupling of MD simulations with SAXS data is described in Section 2.4.

2.1 MD simulations

MD simulations are an irreplaceable and powerful tool to investigate and visualize the dynamic behavior of macromolecules in solution [125, 126]. In a conventional MD simulation, atoms are described as spheres that carry partial charges, and the chemical bonds among atoms are modeled as harmonic springs (Fig. 2.1). The time evolution of the system (trajectory) is modeled by iteratively solving Newtons equation of motion

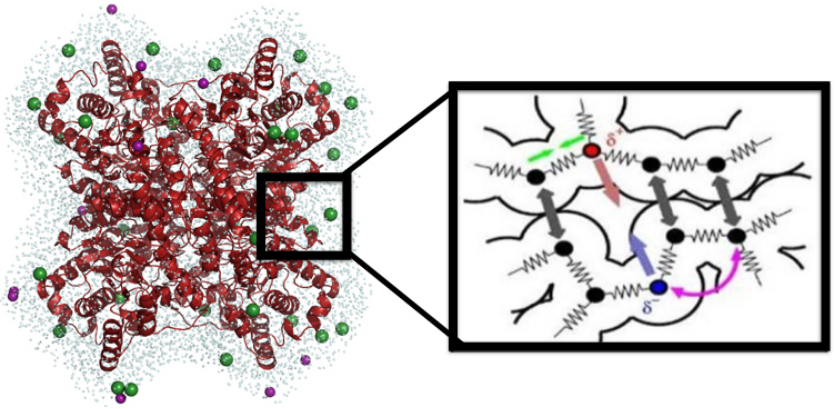

Figure 2.1: Left: snapshot of the typical MD simulation. Protein is visualized in red cartoon representation, and water and ions as colored spheres. For the simplicity, only the solvent atoms at distances smaller than 1 nm from the protein surface are shown. Right (adopted from Ref. 124): Illustration of the underlying principle of MD simulations - atoms are modeled as charged spheres, and chemical bonds as harmonic springs.

[127]. Underlying approximations, fundamentals of the MD algorithm as well as challenges of conventional MD simulations are briefly discussed in the following subsections. Additional details and applied parameters are given in the method sections of Chapters 3-6.

2.1.1 Three approximations of MD simulations

Conventional MD simulations rely on tree approximations [128]: (i) motions of the atomic nuclei can be decoupled from the motions of electrons (Born-Oppenheimer approximation); (ii) nuclei motions are described classically; (iii) the potential energy surface is truthfully described by an empirical force field.

Born-Oppenheimer approximation The time evolution of a system of particles is described by the time-dependent Schrödinger equation:

$$\mathcal{H} = i\hbar \frac{\partial \psi}{\partial t}, \tag{2.1}$$

where $\mathcal{H}$ denotes the Hamiltonian of the system and $\hbar$ denotes the Planck constant h divided by 2π. The wave function ψ at a given time t depends both on the positions of the atomic nuclei $\mathbf{R}$ and the positions of electrons $\mathbf{r}$. Here, $\mathbf{R}$ denotes positions of the k nuclei, $\mathbf{R} = \{\mathbf{R}_1, ..., \mathbf{R}_k\}$, $\mathbf{r}$ denotes the positions of the m electrons, $\mathbf{r} = \{\mathbf{r}_1, ..., \mathbf{r}_m\}$, and t stands for the time. The Born-Oppenheimer approximation assumes that electronic structure instantaneously adapts to the given nuclei positions [129]. In such a case, the motion of the electrons occur in a field of fixed nuclei, and the wave function in Eq. 2.1 can be replaced by a wave function of nuclei (ψ_n) and a wave function of electrons (ψ_e):

$$\mathcal{H}(\mathbf{R}, \mathbf{r}, t) = \psi_n(\mathbf{R})\, \psi_e(\mathbf{r}; \mathbf{R}). \tag{2.2}$$

Here, $\psi_e(\mathbf{r}; \mathbf{R})$ only parametrically depends on the nuclei positions. Therefore, the time-independent Schrödinger equation of the electrons can be written as:

$$\mathcal{H}_e(\mathbf{R})\, \psi_e(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R})\, \psi_e(\mathbf{r}; \mathbf{R}). \tag{2.3}$$

In the above equation, electronic Hamiltonian equals the difference between the kinetic energy of the nuclei T_n and the Hamiltonian of the system $\mathcal{H}$. $E_e(\mathbf{R})$ denotes the Born-Oppenheimer potential-energy surface and physically represents the potential that nuclei "feel" as they move. Finally, the time-dependent Schrödinger equation of nuclei motion can be written as:

$$(T_n + E_e(\mathbf{R}))\ \psi_n(\mathbf{R}, t) = i\hbar \frac{\partial \psi_n(\mathbf{R}, t)}{\partial t}. \tag{2.4}$$

Eq. 2.4 is valid if the eigenvalues of Eq. 2.3 differ significantly, meaning the potential energy surfaces of distinct excited states do not approach each other. For molecules in the ground state this is almost always the case.

Classical motion of nuclei In MD simulations, the number of atoms typically range from tens of thousands up to a few millions, making the solution of the time-dependent Schrödinger equation practically impossible. Therefore, in MD simulations, atomic nuclei are modeled as classical particles, whose motions follow Newton's second law:

$$\begin{aligned} m_i \frac{\partial^2 \mathbf{R}_i}{\partial t^2} &= \nabla_{\mathbf{R}_i} V(\mathbf{R}) \\ m_i \, \mathbf{a}_i &= \mathbf{F}_i, \end{aligned} \tag{2.5}$$

where m_i and $\mathbf{a}_i$ denotes mass and the acceleration of atom i, under the force $\mathbf{F}_i$, and $V(\mathbf{R}) = E_e^0(\mathbf{R})$ denotes the potential energy surface of the ground state, as given by Eq. 2.3. Since biomolecules in solution have been shown to behave classically, this approximation is well-justified [130–132].

Force fields Solving Eq. 2.3 is prohibitively expensive for the systems that contain a large number of electrons, hence the potential energy surface $E_e(\mathbf{R})$ of the biomolecules in solutions is modeled as a sum of simplified empirical expressions that describe $E_e(\mathbf{R})$ satisfactorily accurate. Such a sum (termed force field) can be evaluated in a computationally efficient manner. Within the force field, a macromolecule is modeled by a ball-and-spring-like model (Fig. 2.1). A typical force field includes the sum of bonded and non-bonded interactions between the atoms [127, 133]:

$$\begin{aligned} V(\mathbf{R}) &= V_{bonds} + V_{angles} + V_{dih.} + V_{imp.\,dih.} + V_{LJ} + V_{Coul} \\ &= \sum_{bonds\, i} \frac{k_i}{2} \left(l_i - l_{i,0}\right)^2 \\ &+ \sum_{angles\, i} \frac{f_i}{2} \left(\varphi_i - \varphi_{i,0}\right)^2 \\ &+ \sum_{dih.\, i} \frac{V_i}{2} \left[1 + \cos(n\phi_i - \phi_{i,0})\right] \\ &+ \sum_{im.\, dih.\, i} \kappa_i \left(\xi_i - \xi_{i,0}\right)^2 \\ &+ \sum_{pairs\ i,j} 4\,\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i\, q_j}{4\pi\, \varepsilon_0\, \varepsilon_r\, \varepsilon_{ij}} \end{aligned} \tag{2.6}$$

Here, the bond stretching potential V_{bonds}, the bond angle potential V_{angles} and the improper dihedral potential $V_{imp.\,dih.}$ (which describes out-of-plane bending modes) are

modeled as harmonic potentials, while the dihedral potentials $V_{dih.}$ are modeled by a cosine with periodicity n and potential barriers V_i. The last two terms in Eq. 2.6 describe non-bonded interactions. Non-bonded interactions are evaluated pairwise, neglecting many-particle effects. The short-range repulsive and attractive dispersion interactions are modeled by a Lennard-Jones (LJ) potential, where the parameters σ_{ij} and ε_{ij} represent width and the strength of the potential, respectively. The electrostatic interactions are represented by the Coulomb term, where q_i denotes the partial charge of particle i. The relative dielectric constant ε_r is typically set to 1.

There are several widely used families of force fields. Proteins are usually modeled using AMBER [134, 135], CHARMM [136, 137], GROMOS [138, 139] and OPLS [140, 141]. These force fields (or their modifications) can also be applied to study other macromolecules, such as detergents and lipids. In addition, more general force fields, such as GAFF and CGenFF are used to parameterize small molecules [142, 143]. It is important to note that terms of Eq. 2.6 may slightly differ between force fields.

2.1.2 MD algorithms

In this thesis, all MD simulations are performed using the Gromacs software package [144–146]. In the following paragraphs, key features of the MD algorithms applied to carry out computationally efficient, yet realistic simulations are briefly described.

Time integration Time integration in Gromacs is implemented via the *leap-frog* version of the Verlet algorithm [147], in which the velocities and the positions of the particles are updated as:

$$\begin{aligned} \mathbf{v}(t + \Delta t/2) &= \mathbf{v}(t - \Delta t/2) + \mathbf{F}(t)\,\Delta t/m, \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t + \Delta t/2)\,\Delta t. \end{aligned} \tag{2.7}$$

In order to prevent integration errors in the fastest motions, the time step Δt has to be sufficiently small. An explicit simulation of bond vibrations would require a time-step of 1 fs. Since the bond vibrations hardly couple to the global motion of the macromolecule, they are often constrained using the SETTLE algorithm for water molecules [148] and the LINCS algorithm [149] for the rest of the system, enabling a time step of 2 fs. The next fastest motion is given by bond angle oscillations of hydrogen atoms. In many cases hydrogen atoms can be modeled as virtual interaction sites, allowing a time step of 4 fs [150].

LJ and electrostatic interactions The calculation of non-bonded interactions requires a sum of pairs of atoms. If all possible interactions are explicitly evaluated, the

non-bonded interactions scale quadratically with the number of particles N in the system. The explicit calculation of all LJ and Coulomb interactions becomes prohibitively expensive for the systems whose number of atoms is larger than 10^4. Therefore, the LJ interactions are commonly cut off beyond a distance of 1.0 - 1.4 nm [151]. However, due to the long-range nature (r^{-1}) of decay of the Coulomb potential, cutting it of leads to significant artifacts [152, 153]. Therefore, long-range Coulomb interactions are commonly modeled using the Particle-Mesh Ewald (PME) method, which models the long-range electrostatics by assigning the charges to a grid and calculating the potential as a simple sum in reciprocal space. PME takes advantage of fast Fourier transformation calculations in reciprocal space and scales as $N \log N$ [154, 155]. This scaling is more favorable than the N^2 scaling of previously used Ewald summation (which has to be performed in a real space) [156].

Temperature and pressure coupling To model experimental conditions, MD simulations are commonly performed in an NPT ensemble (isobaric-isothermal conditions). Accordingly, simulations are coupled to temperature and pressure baths. Typically, the temperature coupling is performed via the Berendsen or Nosé-Hoover algorithms or velocity rescaling [157, 158]. Velocity rescaling is executed in a manner that the temperature decays exponentially to the target temperature [159]. Pressure coupling can be performed following different strategies. Through this thesis, we conducted a system equilibration using the Berendsen barostat, as it leads to a stable convergence to the target pressure [159]. Production simulations were performed using the Parrinello-Rahman barostat, as it guarantees the correct NPT ensemble [160, 161].

2.1.3 Main challenges in MD

The two main challenges of each MD simulation are force field accuracy and sampling [15, 16, 126]. These challenges as well as the range of applicability of conventional MD simulations will be discussed in following paragraphs.

Force field accuracy Parameters of force fields are constantly optimized and cross-validated using a range of experimental data and *ab initio* QM calculations [16, 162–166]. Careful optimizations over the last decades made modern force fields highly accurate and applicable to a wide variety of macromolecules. As a consequence, results obtained using different families of force fields are becoming more and more similar, and the agreement to the experimental data is constantly improving [16, 163, 166, 167]. Nonetheless, the choice of the most suitable force field to study a given property of a system of interest along with the accuracy of the applied force field is often far from obvious. Therefore, the

validity of the simulation results for any given application should be verified by comparing properties calculated from simulations with experimental data. Further development of force fields follows three routes: (i) including new experimental and quantum-chemical data for optimization and cross-validation [16, 168]; (ii) finding new strategies to optimize force fields - for example by using the machine learning [16, 169–172] and (iii) replacing simplified force field expressions (Eq. 2.6) with more realistic expressions. For example, force fields with additional terms, that explicitly take polarization into account already demonstrated promising results [173–175].

Sampling Sampling problem refers to the inability of a simulation to "visit" all relevant macromolecular conformations and to determine their relative populations [16, 176]. In studies where comparable experimental data are not available, checking for the sampling completeness is particularly difficult because it is difficult to predict the existence of a specific (important) conformation before it is "visited" by the MD simulation. Sampling problems in a given application may arise both from force field imperfections and from time-scales that are inaccessible using conventional MD simulations [176]. For example, some proteins undergo large, slow conformational changes, even on a time-scale of seconds, while state-of-the-art all-atom MD simulations reach up to milliseconds [16, 177–179]. Therefore, a variety of strategies to improve sampling are constantly developed. For example, the speed of MD simulations greatly improved over the last decade owing to new strategies to off-load calculations to graphic process units [180]. Additionally, hardware designed specifically for MD simulations, such as the massively parallel supercomputer Anton, significantly increase the simulation speed [181]. However relevant time-scales are still sometimes out of reach of unbiased all-atom simulations, either due to the system size or the slowness of relevant motions. In such cases, coarse-grained (CG) simulations may be employed [182–187]. The idea behind CG simulations is to represent multiple atoms as a single particle. While the force field expression, as shown in Eq. 2.6, usually remains unchanged, a new force field parametrization is required [184]. CG simulations are significantly faster compared to all-atom simulations, but the accuracy is reduced.

An alternative route to access longer time scales is to run many short simulations and extract information by, for example, constructing a Markov state model [188–190]. A detailed overview on the joint usage of Markov state models and all-atom MD simulations is given in Ref. 190. Additionally, the outcomes of a large set of short simulations may be used to start additional simulations from underpopulated biomolecular conformations. This approach is termed adaptive sampling [191–193].

Finally, long timescales may be reached by changing the simulation parameters, for example by employing enhanced sampling along one or more specific collective variables [194–197]. More details on enhanced sampling methods may be found in Refs. 16 and 198.

The range of applicability Relying on the approximations described in the Subsection 2.1.1 leaves the chemical reactions (involving the breaking of formation of covalent bonds), charge transfer processes, as well as excited electron states, out of the scope of the classical MD simulations. If any of these processes play an important role in a given study, MD simulations are often combined with the quantum-mechanical calculations [199–202].

2.1.4 How to combine experimental data with MD simulations

The compatibility of experimental data and MD simulations is predetermined by their nature [16, 203]: experimental data are synthesized into a coherent model by solving the inverse problem; on the other hand, simulations deal with the forward problem - building a model that can be compared with observed data. Accordingly, a variety of strategies have been proposed to combine experimental data with MD simulations. The simplest approach is to compare experimental data with data backcalculated from a free (unbiased) simulation [15, 16, 90, 93, 204]. If the back-calculated data agree with the experimental data, there is reason to believe that the experimental conditions are truthfully resembled in a simulation. In such cases, the analysis of the MD simulation provides additional information about the studied macromolecule. This approach is followed in Chapters 3 and 6 of this thesis. Although intuitive and conceptually simple, this approach may encounter many difficulties. First of all, experimental data and MD simulations often do not agree, owing either to the imperfection of simulations (force field and sampling problems), or to an unknown experimental error. Second, even if the back-calculated data agree with the experimental data, there is no guarantee that the simulation resembles experimental conditions. This challenge is particularly pronounced in the case of experiments that provide time and ensemble-averaged information (SAXS, NMR, and FRET, for example).

A more advanced approach is to reweight the simulation ensemble, such that it agrees with the available experimental data. However, this approach relies on exhaustive sampling - if relevant biomolecular conformations are not visited by a simulation, the derived ensamble will be overfitted and often dominated by only a few structures with high weights [15, 16, 93, 205].

Finally, experimental data may be directly integrated (as a restraint) in a simulation.

This route was followed in Chapters 4 and 5 of this thesis. Refining MD simulations against ensemble-averaged experimental data may also lead to overfitting [206], especially in the cases of flexible biomolecules [15, 16]. In such cases, a statistically well-founded framework is required. One possible approach is based on introducing the smallest possible perturbation to the simulation to match the experimental data. As the bias is chosen to be as small as possible, this approach follows the maximum-entropy principle [93, 207–209]. Although being powerful and mathematically rigorous, the maximum-entropy formalism does not explicitly take sources of errors into accout [16]. This is often no problem, since the force fields contain accurate physico-chemical knowledge. For example, the ensemble of the RS peptide that was refined against SAXS data following the maximum-entropy approach, was successfully cross-validated against orthogonal NMR data [93]. Nonetheless, when needed, a Bayesian framework can be employed, as it allows for the integration of data from multiple sources, while simultaneously estimating uncertainties [16, 92, 204, 206].

2.2 SAXS

As already outlined in the Introduction, SAXS is increasingly popular method to study biomolecules in solution, that is under near-native conditions [61]. In the next subsections, the basis of X-ray scattering, experimental set-up and output, as well as the theoretical framework to calculate SAXS curve from a given structural models are briefly discussed.

2.2.1 Basis of X-ray scattering

To illustrate the scattering process, we first consider the interaction of a monochromatic wave with a single particle, as shown in Fig. 2.2. The incident wave is described by the

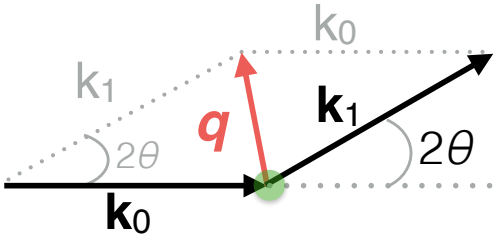

FIGURE 2.2: Illustration of the scattering of a wave on a single particle. $\mathbf{k}_0$ denotes the wave vector of the incident wave, and $\mathbf{k}_1$ denotes the wave vector after the elastic scattering on a single particle (green transparent sphere). The momentum transfer vector is denoted by $\mathbf{q}$, and 2θ denotes the scattering angle. Waves are represented by straight lines for simplicity.

wave vector $\mathbf{k}_0$ with the length $k_0 = 2\pi/\lambda$, where λ denotes the wavelength. As a result of the interaction of the wave with the particle, the particle starts to oscillate and emits a wave, whose wave vector is $\mathbf{k}_1$. The angle between the incident and scattered waves is denoted by $2\,\theta$.

In SAXS experiments, the incident X-rays are scattered at the electrons of the atomic shell. We confine to the coherent scattering because incoherent scattering is negligibly weak at small scattering angles [95, 210]. At the wavelengths used in SAXS, scattering is always elastic. Therefore, the wavelength of the incident wave does not change after the scattering, and the length of the scattered wave vector is given by $k_1 = 2\pi/\lambda$. The scattering intensity is commonly represented as a function of the momentum transfer vector length. The momentum transfer vector $\mathbf{q}$ has the same direction as $\mathbf{k}_1 - \mathbf{k}_0$ (Fig. 2.2), and the length of $\mathbf{q}$ is given as $q = 4\pi \sin(\theta)/\lambda$ [210].

Now we consider a scattering of the beam of collimated monochromatic X-ray waves on a macromolecule. Since the incident beam of monochromatic X-ray waves is linearly polarized and the scattered waves are coherent, the resulting amplitudes are simply added [210]:

$$A = \sum_{i=1}^{n} A_i, \tag{2.8}$$

where A_i denotes the scattering amplitude of the ith wave, and n denotes the total number of scattered waves. The total scattering intensity is given by the absolute square of the resulting amplitude [95, 210]:

$$I = \left|A^2\right| \tag{2.9}$$

The scattering curve of a macromolecule that consists of N molecules can be calculated using the Debye formula [95, 210]:

$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \, \frac{\sin\left(q \cdot r_{ij}\right)}{q \cdot r_{ij}} \tag{2.10}$$

where $r_{ij} = |\mathbf{r_i} - \mathbf{r_i}|$ and r_i and r_j are Cartesian coordinates of atoms i and j. The form factors of atoms i and j are denoted by $f_i(q)$ and $f_j(q)$, respectively. Physically, the form factor represents the Fourier transform of an electron distribution around the nucleus of a given atom [95, 210]. The form factors are accurately approximated via [90, 95, 210]:

$$f_j(q) = \sum_{k=1}^{4} a_k \, e^{-b_k (q/4\pi)^2} + c\,, \tag{2.11}$$

where a_k, b_k and c denote Cromer-Mann parameters [211].

For more detailed derivations and a description of the scattering process we refer to

Ref. 95 and Ref. 210. Details of SAXS experiments, information content of the experimental scattering curve and methods to calculate the scattering curve from a given structural model are discussed in the following paragraphs.

2.2.2 Experimental set-up and output

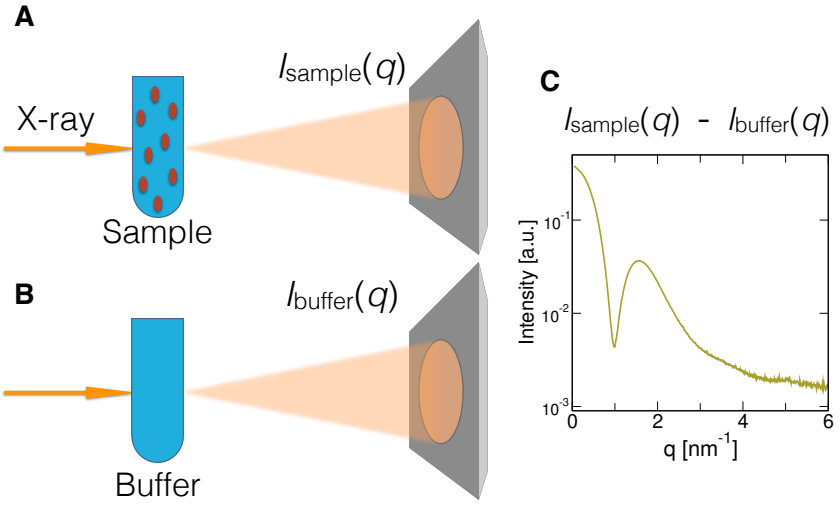

FIGURE 2.3: Illustration of the SAXS experiment: sample in a solution (A) and the pure buffer (B) are irradiated with X-ray. Scattering patterns are detected, orientationally averaged and subsequently subtracted, resulting in one-dimensional intensity curve shown in C.

Work-flow The work-flow of a typical SAXS experiment is as follows [24, 212]: the solvated biomolecules (sample) are placed in a quartz capillary and irradiated by a collimated monochromatic X-ray beam. The scattering pattern is recorded at the X-ray detector (Fig. 2.3A). A second quartz capillary is filled with pure solvent (buffer) that should be identical to the sample solvent. The scattering pattern of the buffer is recorded in the same manner as the scattering pattern of the biomolecular solution (Fig. 2.3B). Owing to the random orientation of macromolecules in solution, the scattering pattern is isotropic. Scattering patterns of the sample and the buffer are radially averaged and subsequently subtracted. In this way, the scattering intensity of the biomolecule (I) is determined as a function of momentum transfer $q = 4\pi\sin(\theta)/\lambda$ (Fig. 2.3C). Here λ denotes the wavelength of the X-ray beam, and 2θ denotes the scattering angle relative to the path of the beam. It is essential to note that, since the water around the biomolecule has the density that is different compared to the density of bulk water (so-called hydration layer), $I(q)$ records both the information about the biomolecule *and* the water density modification in the biomolecular vicinity.

Technical notes Nowadays, structural biology and biophysics communities take advantage of high brilliance of synchrotron sources with beamlines optimized for biological SAXS experiments. These include DESY (Hamburg, Germany), ESRF (Grenoble, France) Diamond (Oxford, Great Britain), ANL (Argonne, USA), SSRL (Stanford, USA) and Spring-8 (Himeji, Japan) [24, 213–215]. Nonetheless, reasonable results may be obtained using in-house SAXS beamlines, which are also employed to fine-tune samples before the beamtime at the synchrotron [15, 24, 216].
In a typical SAXS experiment, multiple concentrations of the same sample are prepared and measured. Subsequently, the highest sample concentration unaffected by the radiation damage, sample aggregation or biomolecule-biomolecule interactions is used for the analysis. This procedure requires $10 - 100\,\mu$l of a sample (meaning $1 - 2\,$mg of a purified protein) [24]. At the dedicated synchrotron beamlines, multiple parameters of the investigated sample, such as molecular weight, excluded particle volume, maximum dimension and the radius of gyration of the particle, are available immediately after the irradiation [3, 24].

Information content in experimental data In this paragraph, we discuss the information content of a typical SAXS curve, using examples given in Ref. 17: (i) folded, oligomeric PCNA (PDB code 1AXC [217]), (ii) modular GbpA (PDB code 2XWX [218]) and (iii) disordered elF3g (PDB code 4U1E [219]). The following analysis can be performed using a variety of available software packages, among which ATSAS [220] is probably the most popular one.
Often, the analysis of the SAXS data begins with the simple visual inspection of the scattering profile, as it may indicate the molecular geometry and dynamics. For example, the SAXS curve of well-folded, globular proteins contain several elevations and dips (Fig. 2.4b, blue line), in contrast to the featureless SAXS curve(Fig. 2.4b, green line) of intrinsically disordered proteins (IDPs).
In the next step, the well-understood Guinier region is analyzed (Figure 2.4b, insets). Namely, at very low q-values (chosen such that $q \cdot R_g < 1.3$), plotting the natural logarithm of I as a function of q^2 exhibits linear behavior. The slope of the fitted line is connected to R_g of the macromolecule. The extrapolation of the fitted line gives the forward angle scattering (I_0), which is used to determine the molecular mass of the protein [221]. Eventual non-linearity of the Guinier region is a clear indication of the sample aggregation, particle-particle repulsion or the radiation damage.
In the next step of the data analysis, the volume of the macromolecule can be estimated using the Porod law and Porod-Debye plot (Fig. 2.4c, top plot) [222]. Additionally, the decay of the pair-distance distribution function ($P(r)$) provides the estimate of the maximum diameter of the particle ($D_{\max}$ in the middle plot in the Fig. 2.4c). Information

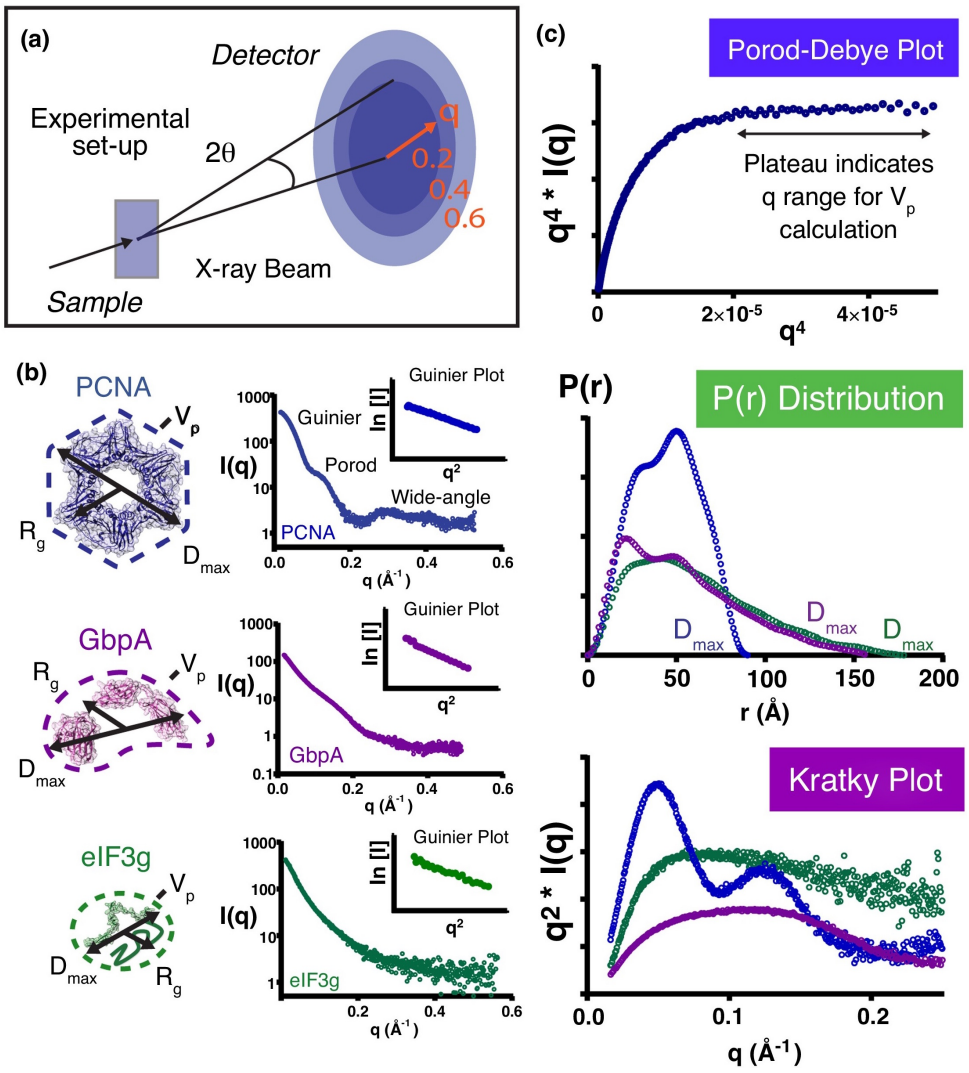

FIGURE 2.4: Scattering curves and various SAXS parameters determined for three proteins: PCNA (blue color-coding), GbpA (violet color coding) and eIF3g (green color-coding). Figure adopted from Ref. 17.

about the R_g, $D_{\max}$ and volume is subsequently used to construct the low-resolution envelope of the molecule (Fig. 2.4b, dotted lines).
Finally, the Kratky plot ($q^2 \cdot I(q)$ plotted as a function of q, Fig. 2.4c, bottom plot) offers a qualitative assessment of the compactness of the investigated molecule [6, 17, 24]: globular molecules exhibit parabolic Kratky curves, converging at high-q values (blue curve), while unfolded systems exhibit hyperbolic behavior (green curve). The non-parabolic profile of modular GbpA is the consequence of the flexibility of its linked domains (violet curve).

Beside the information about macromolecule size, low-resolution shape and R_g, there is no straight-forward procedure to obtain more structural information directly from the experimental curve. Moreover, the exact number of independent data points or how they are distributed over q-range is not fully understood [15]. Commonly, the number of

data points in a SAXS curve is estimated by the number of Shannon channels [2, 223]: $N_c = q_{\max} D_{\max}/\pi$, where $q_{\max}$ is the maximal q recorded in the experiment. For a typical SAXS setup, and an average-size protein, N_c is found to be between 10 and 30 [2, 15, 60]. The number of the backbone angles of a protein already exceed N_c by the two orders of magnitude, hindering any straight-forward procedure to obtain the high resolution structural model from the experimental data alone. Therefore, methods to back-calculate SAXS curve from a given structural model have been constantly developing [15].

2.2.3 Calculating SAXS curve from a given structural model

The physical basis of the SAXS curve is known - it is given by the orientationally averaged Fourier transform of the correlation function of the electron density contrast between solvent and solution [15, 95]. Nonetheless, as mentioned in Chapter 1, methods to predict SAXS curve vary greatly. Main differences stem from the treatment of the excluded solvent and hydration layer (implicit or explicit), shape fluctuations (included or not) and the resolution (coarse grained or atomistic). Detailed lists of methods to predict SAXS curve, outlining their differences can be found in Refs. 5, 15, 210. While implicit-solvent methods are considerably faster, they, in contrast to explicit-solvent methods, require at least two fitting parameters related to solvent, one to account for the hydration layer and one to account for the excluded solvent. Since the density of the hydration layer is biomolecule-dependent, and since the straight-forward procedure to predict the density of the hydration layer does not exist, implicit-solvent software like CRYSOL [224] or FoXS [225] treat the hydration layer as a parameter that is fitted in order to improve the agreement between the calculated and experimental SAXS curves. While this procedure is justified in many applications, it (like any other fitting procedure) increases the risk of overfitting, and may reduce the amount of extracted information [15, 71, 90]. For example, in a work by Chen and Hub (Ref. 71) it has been demonstrated that the number of detergent molecules forming a protein-detergent complex can be determined by comparing experimental SAXS curve with SAXS curves calculated from all-atom MD simulations of different number of detergents forming the protein-detergent complex. In contrast, CRYSOL and FoXS gave no clear answer, owing to the fitting of the hydration layer density. Additionally, implicit-solvent methods commonly use reduced atomic form factors [57, 87] to account for the displaced solvent. Since the determination of atomic form factors require accurate information on atomic volumes, which are known only approximately, the uncertainty of SAXS curve calculation is additionally increased [15, 94].

2.2.3.1 Calculating SAXS curve from all-atom MD simulation

As all-atom MD simulations are used through this thesis, the physical basis of the SAXS curve prediction will be explained following the approach that Chen and Hub (Ref. [90]) developed (and termed WAXiS [226]) building-up on the method proposed by Park *et al.* (Ref. [88]). This method requires two sets of simulations, one of biomolecule in a solution, and the second one of a pure solvent. In order to resemble the experiment, the simulations have to be long enough to sample all relevant bimolecular conformations, as well as solvent fluctuations. After the simulations are conducted, a fixed spatial envelope is constructed around the biomolecule (Fig. 3.5A). In each specific application, the envelope size is chosen carefully, such that all non-bulk-like solvent atoms are enclosed by the envelope at each simulation frame. Once constructed, exactly the same envelope is transferred in the pure solvent system, and the SAXS curve is predicted by taking into the account position of atoms inside the envelopes at each simulation frame. In this manner, the hydration layer, as well as the excluded solvent, are naturally included in the SAXS curve prediction, and bulk-like solvent atoms that do not contribute to the SAXS curve are excluded, making the calculations faster and reducing the noise.

Mathematical expressions used to compute SAXS curve In the following, we refer to all atoms inside the envelope in solvated biomolecule simulation as system A, and to all atoms inside the envelope in the pure solvent simulation as system B. The SAXS curve represents the scattering intensity difference (excess scattering intensity) between systems A and B:

$$I(q) = I_A(q) - I_B(q). \tag{2.12}$$

Scattering intensity I_A can be written as:

$$I_A = \left\langle \left| \tilde{A}(\mathbf{q}) \right|^2 \right\rangle^{'}, \tag{2.13}$$

where $< \cdots >^{'}$ denotes the ensemble average over all solvent and solute degrees of freedom, and $\tilde{A}(\mathbf{q})$ denotes Fourier transform of the instantaneous electron density $A(\mathbf{r})$ of the system A. An analogous expression is used to compute I_B. $D(\mathbf{q})$ is defined as:

$$D(\mathbf{q}) := \left\langle \left| \tilde{A}(\mathbf{q}) \right|^2 \right\rangle^{(\omega)} - \left\langle \left| \tilde{B}(\mathbf{q}) \right|^2 \right\rangle^{(\omega)}, \tag{2.14}$$

where $< \cdots >^{(\omega)}$ denotes average over solute and solvent fluctuations at fixed solute orientation ω. $I(q)$ is given by:

$$I(q) = \langle D(\mathbf{q}) \rangle_{\Omega}. \tag{2.15}$$

In the last equation, $< \cdots >_\Omega$ denotes the average over orientations of the solute. The scattering amplitude for an individual simulation frame i in system A is given as:

$$\tilde{A}_i(\mathbf{q}) = \sum_{j=1}^{N_A} f_i(q)\, e^{-i\mathbf{q}\cdot\mathbf{r}_j}\,, \tag{2.16}$$

where N_A denotes the number of atoms within the envelope, $f_j(q)$ are the atomic form factors, and $\mathbf{r}_j$ is the coordinate of atom j. The analogous expression is used to compute $\tilde{B}_i(\mathbf{q})$ (scattering amplitude for an individual simulation frame i in system B). The form factors are computed using the expression given in Eq. 2.11. Knowing $\tilde{A}_i(\mathbf{q})$ and $\tilde{B}_i(\mathbf{q})$, $D(\mathbf{q})$ can be computed as:

$$\begin{aligned} D(\mathbf{q}) = &\left\langle \left|\tilde{A}_i(\mathbf{q})\right|^2 \right\rangle^{(\omega)} - \left\langle \left|\tilde{B}_i(\mathbf{q})\right|^2 \right\rangle^{(\omega)} \\ &+ 2\mathrm{Re}\left[- \left\langle \tilde{B}_i^*(\mathbf{q}) \right\rangle^{(\omega)} \left\langle \tilde{A}(\mathbf{q}) \right\rangle^{(\omega)} - \left\langle \tilde{B}(\mathbf{q}) \right\rangle^{(\omega)} \right], \end{aligned} \tag{2.17}$$

where $*$ denotes the complex cojugate. Finally, $I(q)$ is computed using Eq. 2.15. Detailed derivation, as well as software details and validation can be found in Ref. [90]. Tutorials on SAXS (as well as SANS and SAXS-driven, see below) calculations using this method, written by the author of the thesis can be found at
https://biophys.uni-saarland.de/SANS_tut.html and
https://biophys.uni-saarland.de/grenoble-tut/.

2.3 SANS

Conceptually, SANS is very similar to SAXS, as already outlined in Chapter 1. In contrast to SAXS, due the significant incoherent scattering, which decreases the signal-to-noise ratio, SANS is restricted to lower q values [95, 227]. But SANS offers a great advantage - the usage of contrast variation experiments. These experiments are performed at different D_2O concentrations and at different perdeuteration conditions of the large macromoluecule or macromolecular complex (Fig. 2.5). Contrast variation experiments provide subunit-specific structural information, by making the domain of the large bioomolecule, or part of the complex "invisible" to the neutron scattering by making the contrast zero [4, 94, 97, 227–230]. Additionally, combining multiple SANS data-sets with SAXS data offers the possibility for cross-validation in order to decrease the risk of overfitting, as it is still a major challenge in SAS community [15, 72, 94].

Owing to the similarity between SAXS and SANS, procedures to calculate SANS curves from a given structural model are very similar to procedures to calculate SAXS curves

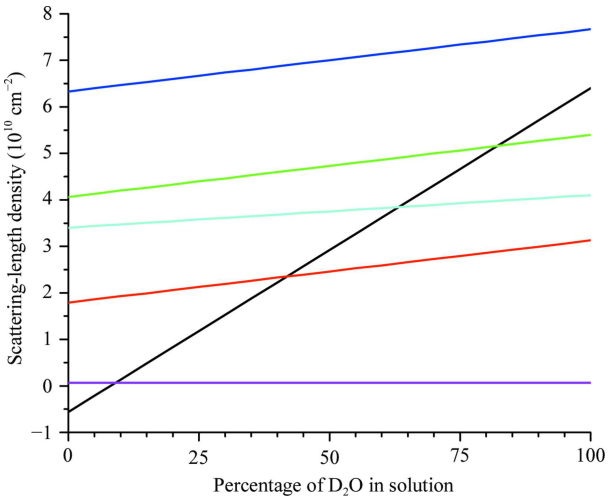

FIGURE 2.5: Scattering-length densities as a function of D_2O for water (black), hydrogenated protein (red), protein with 50% of the protons that do not exchange substituted by deuterium (green), protein with 100% of the protons that do not exchange substituted by deuterium (blue), DNA with an equal distribution of base pairs (cyan) and the lipid dimyristoyl phosphatidylcholine (violet). Adopted from Ref. 4.

that are described in the previous section. Namely, the mathematical basis of neutron scattering is identical to X-ray scattering - only X-ray atomic factors have to be replaced by neutron scattering lengths [94, 95, 231]. For the purpose of this thesis, an extension of WAXSiS software [90] was used. Software details are given in Ref. [94]. The project described in Chapter 6 represents the first application of this software.

2.4 SAXS-driven simulations

So far, we discussed the usage of MD simulations to back-calculate SAXS curves. If the curve back-calculated from the MD simulation agrees with the experimental curve, there is reason to believe that the simulation provides an accurate atomic model of the studied macromolecule. However, experimental and calculated curves often do not agree, sometimes due to experimental problems, but more often due to simulation imperfections. In order to overcome simulation imperfections, multiple methods that integrate SAXS data as an energetic restraint into MD simulations have been developed [91, 93, 232, 233]. The methods developed by Chen and Hub (Ref. 91) and Hermann and Hub (Ref. 93) have been employed in this thesis. The first method (in further text referred as single-replica simulations) is valid for structurally well-defined macromolecules that are truthfully described by moderate fluctuations of a single structure. However, in the case of heterogeneous ensembles (as adopted by IDPs) it is pointless to compare the scattering curve of a single structure with the ensemble-averaged experimental data that encodes

information on a diverse range of structures [93]. In such cases, multi-replica simulations offer a route to avoid overfitting.

2.4.1 Single-replica simulations

Details about this method, as well as its validation can be found in Ref. 91. In short, the coupling of the MD simulation to a target curve I_{exp} is achieved by a hybrid energy

$$E_{\text{hybrid}}(\mathbf{R}; I_{\text{exp}}) = E_{\text{MD}}(\mathbf{R}) + E^{(1)}_{\text{exp}}(\mathbf{R}; I_{\text{exp}}), \tag{2.18}$$

where $E_{\text{MD}}(\mathbf{R})$ denotes the energy derived from the force field for the biomolecule conformation $\mathbf{R}$. $E^{(1)}_{\text{exp}}(\mathbf{R}; I_{\text{exp}})$ denotes an experimentally derived energetic bias calculated via

$$E^{(1)}_{\text{exp}}(\mathbf{R}; I_{\text{exp}}) = \frac{k_r\, k_B T}{n_q} \sum_{i=1}^{n_q} \frac{[I_c(q_i, \mathbf{R}) - I_{\text{exp}}(q_i)]^2}{\sigma_i^2}. \tag{2.19}$$

Here, k_r denotes the force constant, k_B denotes the Boltzmann constant, and T denotes the temperature. The number of intensity points spread over the q-range is denoted by n_q, and $I_c(q_i, \mathbf{R})$ denotes the scattering intensity back-calculated (using the approach described in the previous section) from simulation coordinates $\mathbf{R}$. The uncertainty that accounts for experimental errors, statistical calculated errors and systematic error that originates from the uncertainty of the buffer density is denoted by σ_i.

2.4.2 Multi-replica simulations

In this method, multiple parallel simulations are refined against the experimental data by coupling a replica-averaged SAXS curve to the experiment. If the number of replica is high enough, this approach follows the maximum entropy principle [208, 234]. The idea follows Jaynes' maximum entropy principle: an unbiased ensemble distribution should be modified as minimally as possible into a biased distribution that explains the experimental data. Any bias that is not supported with experimental data should be avoided [93, 204, 235]. Therefore, multi-replica simulations that follow the principle of maximum entropy represent the statistically founded procedure to interpret SAXS data of heterogeneous ensembles of structures. Details about the method used in this thesis, as well as its cross-validation against NMR data can be found in Ref. 93. In short, in contrast to the single-replica refinement, the back-calculated scattering curve of N parallel replica simulation is first averaged among the replicas,

$$\bar{I}_c(q_i, \mathbf{R}_1, \ldots, \mathbf{R}_N) = N^{-1} \sum_{\alpha=1}^{N} I_c(q_i, \mathbf{R}_\alpha), \tag{2.20}$$

where α denotes the replica index. Subsequently, the energetic bias is calculated via

$$E_{\mathrm{exp}}(\mathbf{R}_1, \ldots, \mathbf{R}_N; I_{\mathrm{exp}}) = \frac{k_r N\, k_B T}{n_q} \sum_{i=1}^{n_q} \frac{\left[\bar{I}_c(q_i, \mathbf{R}_1, \ldots, \mathbf{R}_N) - I_{\mathrm{exp}}(q_i)\right]^2}{\sigma_i^2}. \tag{2.21}$$

Following this procedure, only the minimal bias necessary to match the experimental data is added, ensuring that if the simulations are long enough and the number of replicas is large enough, Jaynes' maximum entropy principle is fulfilled [207, 208].

Chapter 3

Quantifying the Influence of the Ion Cloud on SAXS Profiles of Charged Proteins

THIS chapter, under its title is published in Ref. 236. We acknowledge the permission from the *Physical Chemistry Chemical Physics* to reproduce the article.

Authors of the paper are Miloš T. Ivanović, Linda K. Bruetzel, Roman Shevchuk, Jan Lipfert and Jochen S. Hub, with the following contributions:
JSH motivated the project;
MTI performed and analyzed MD simulations;
RS, MTI and JSH wrote Python scripts for analytical calculations;
JL and LB performed the SAXS experiment;
all authors checked and discussed results;
JSH and MTI wrote the article.

3.1 Abstract

Small-angle X-ray scattering (SAXS) is a popular experimental technique used to obtain structural information on biomolecules in solution. SAXS is sensitive to the overall electron density contrast between the biomolecule and the buffer, including contrast contributions from the hydration layer and the ion cloud. This property may be used advantageously to probe the properties of the ion cloud around charged biomolecules. However, in turn, contributions from the hydration layer and ion cloud may complicate the interpretation of the data, because these contributions must be modelled during structure validation and refinement. In this work, we quantified the influence of the ion

cloud on SAXS curves of two charged proteins, bovine serum albumin (BSA) and glucose isomerase (GI), solvated in five different alkali chloride buffers of 100 mM or 500 mM concentrations. We compare three computational methods of varying physical detail, for deriving the ion cloud effect on the radius of gyration R_g of the proteins, namely (i) atomistic molecular dynamics simulations in conjunction with explicit-solvent SAXS calculations, (ii) non-linear Poisson-Boltzmann calculations, and (iii) a simple spherical model in conjunction with linearized Poisson-Boltzmann theory. The calculations for BSA are validated against experimental data. We find favorable agreement between the three computational methods and the experiment, suggesting that the influence of the ion cloud on R_g, as detected by SAXS, may be predicted with nearly analytic calculations. Our analysis further suggests that the ion cloud effect on R_g is dominated by the long-range distribution of the ions around the proteins, as described by Debye-Hückel theory, whereas the local salt structure near the protein surface plays a minor role.

3.2 Introduction

The ion cloud is an integral part of charged biomolecules, since the ions may influence the biomolecules' stability, structure, aggregation, and function [237–239]. Hence, major efforts have been invested over the last decades to understand the structure and the determinants of the ion cloud of biomolecules. Most studies have focused on the ion cloud of nucleic acids, owing to their pronounced negative charge [237, 239–254], whereas less work has focused on the ion cloud of charged proteins [102, 255].

Investigating the ion cloud is complicated by its fluctuating and dynamic nature; hence, combined experimental, theoretical, and simulation-based approaches are required to develop quantitative and atomic-level understanding of the ion clouds of biomolecules. Experimentally, the ion cloud has been probed by methods such as atomic emission spectroscopy, small-angle X-ray scattering (SAXS), as well as anomalous SAXS (ASAXS) [255–261]. Theoretically, Poisson-Boltzmann (PB) theory provides a framework for modeling ion distributions; however, unmodified or "native" PB theory neglects effects from the finite size of ions, ion-ion correlations, and from specific salt bridges formed between ions solution with charged groups on the biomolecule's surface. Hence, several extensions to the native PB theory have been proposed [262–265]. In principle, molecular dynamics (MD) simulations likewise overcome such limitations, hence they have been routinely used to model the distribution of ions and water around biomolecules [246–250, 252–254]. Results from MD simulations are compatible with other theoretical predictions and experiments [241, 242, 245, 251, 255], suggesting that MD simulations predict a

reasonably accurate ion distribution, despite some well-known limitations of certain ion force fields [266–269]. Complementary, the 3D-RISM method has been suggested as computationally efficient method for modeling the solvation shell of biomolecules [241].

SAXS is an increasingly popular method for obtaining low-resolution structural information of biomolecules in solution, described in a number of excellent reviews [2, 6, 29, 59–61, 66–68, 95]. Whereas SAXS has originally been used to obtain the approximate size and shape of biomolecules, it has developed into an increasingly quantitative probe thanks to technical developments in light sources and detectors, setups coupled with size-exclusion chromatography (SEC-SAXS) [270], and established standards for sample preparation and validation [73, 213]. Such developments reduced not only the statistical noise in the data, but, more critically, also the risk of systematic errors, for instance owing to aggregation, poor buffer matching, or radiation damage. However, to harvest the increasingly accurate and reliable structural information contained in the data, by means of validating or refining structural models against the SAXS data, increasingly accurate computational methods are required for predicting SAXS curves from a given structural model.

SAXS detects the electron density contrast, $\Delta\rho(\mathbf{r})$, between solute and solvent, including the contributions from the hydration layer and the ion cloud [2]. Hence, in order to draw structural conclusions form the data, understanding of the influence of the hydration layer and of the ion cloud on the SAXS curve is required. The influence of the hydration layer on SAXS data has been investigated both using experiments [98], with a simple spherical model [99], and using MD simulations [57, 88–90, 100, 101]. Such studies established that the density of the hydration layer of proteins in aqueous solutions is, on average, often increased compared to the bulk, which manifests in an increased radius of gyration, R_g, as extracted from a Guinier fit to the SAXS curve. Notably, the increase of R_g due to the hydration layer is protein-dependent, suggesting that also the hydration layer is protein-dependent [90]. In contrast, the influence of the ion cloud of charged proteins on SAXS curves is less well understood. Zhang *et al.* investigated the influence of ionic strength on SAXS and small-angle neutron scattering (SANS) data on the charged model protein bovine serum albumin (BSA), with a focus on protein-protein interactions [271, 272]. In addition, Kim *et al.* disentangled contributions from water and ions on the hydration layer of supercharged proteins by combining SAXS with SANS [102]. However, to our knowledge, the effect of the ion cloud on R_g has not been systematically addressed.

Computationally efficient methods for SAXS curve prediction, such as CRYSOL, FoXS, or SASTBX account for the hydration layer with simplified descriptions, for instance by modelling a uniform excess density around the protein surface, or by scaling the

atomic form factors of solvent-exposed atoms [224, 225, 273, 274]. To improve the agreement with experiment, such method allow the fitting of a free hydration layer-related parameter against the data. In the case of charged proteins, which exhibit a counter ion cloud, the hydration layer-related parameter will presumably also absorb the contrast contributions from the ion cloud. However, such fitting parameters, just like any free adjustable model parameter, may be problematic. Fitting parameters may (i) increase the risk of overfitting [275], and (ii) they reduce the amount of structural information that can be harvested from the data. For instance, we observed in the context of a protein-detergent complex that the free parameter for hydration layer may be overfitted, thereby absorbing inaccuracies in the structural model [71].

SAXS curve predictions based on explicit-solvent MD simulations avoid such a hydration layer-related fitting parameter [87–90, 101]. Notably, such methods may, in addition, avoid fitting parameters related to the excluded volume – an advantage that is not further discussed in this article. When using these methods, the solvation shell included into the SAXS predictions should contain all solvent density modulations caused by the biomolecule; for uncharged or only weakly charged proteins, it was found that a solvation shell thickness of 7–8 Å is sufficient to account for the water density modulations in the hydration layer [88, 90]. For highly charged proteins, however, additional solvent density modulations appear owing to the counter ion cloud [102], which decay exponentially into the bulk with a characteristic length scale, termed Debye length, λ_D. For an approximately physiological 100 mM salt solution of monovalent ions, one obtains $\lambda_D \approx 9.7$ Å, suggesting that density modulations due to ions reach much farther into the bulk as compared to the water density modulations in the hydration layer. Hence, to account for effects from the ion cloud during explicit-solvent SAXS predictions, a large simulations system is required, as a solvation shell with a thickness of multiple Debye lengths must be included into the SAXS calculations. As we show here, such calculation are feasible and may predict the influence of the ion cloud on R_g; however, because such calculations are computationally quite expensive, they may not be suitable for routine applications. Therefore, we present computationally efficient alternatives for predicting the ion effects on R_g, based on non-linear PB calculations, and, even simpler, based on a spherical protein model in conjunction with linearized PB theory. The simplified calculations provide (i) computationally efficient predictions of the ion cloud effect on R_g; (ii) reveal whether explicit-solvent SAXS prediction with a finite solvation shell thickness are biased by missing contributions from far-distant ions (outside the included solvation shell); and (iii) offer quantitative corrections for such missing contributions.

In this work, we studied the influence of the counter ion cloud on SAXS curves of two charged proteins: (i) moderately charged ($Q = -16e$) bovine serum albumin (BSA) and

(ii) highly charged ($Q = -60e$) glucose isomerase (GI), in a series of alkali chloride solutions. BSA and GI carry considerable charge and have been frequently used as model proteins in SAXS studies, making them ideal test cases for the present study. We considered systems with a relatively low concentration (100 mM), implying a large Debye length of $\lambda_D \approx 9.7$ Å, and systems with increased salt concentration (500 mM), implying a short Debye length of $\lambda_D \approx 4.3$ Å. Remarkably, for such common buffer conditions, we found that the influence of the ion cloud on R_g is in a similar range as the influence of the hydration layer on R_g (up to ∼2 Å), demonstrating that an accurate interpretation of SAXS data requires accurate modelling of the ion cloud. We used four different methods to obtain the influence of the ions on R_g, namely (i) atomistic MD simulations in conjunction with explicit-solvent SAXS calculations, (ii) a spherical protein model in conjunction with linearized Poisson-Boltzmann (PB) calculations, (iii) an atomic protein model in conjunction with non-linear Poisson Boltzmann (PB) calculations, and (iv) SAXS experiments. We found reasonable agreement between the four methods, suggesting that the simplified methods may estimate ion cloud effect on R_g. Further, we analyze the trends of R_g as a function of ion type, ion concentration, protein charge, and protein size, providing a reference for the analysis of SAXS data of charged proteins in future studies.

A Python implementation of the spherical model, as well as a modified Gromacs version that implements the explicit-solvent SAXS calculations are provided on the authors' website at https://biophys.uni-saarland.de/software.html.

3.3 Methods

3.3.1 SAXS experiments of BSA

Bovine serum albumin (BSA) (product no. A7638) was purchased from Sigma-Aldrich and used without further purification. Solutions with BSA concentrations of 2.5 mg/ml and 5.0 mg/ml were prepared by dissolving the lyophilized powder in buffer solutions containing 50 mM KOH, 50 mM HEPES, pH = 7.5 and 100 mM (or 500 mM) of different alkali chlorides (i.e. LiCl, NaCl, RbCl, or CsCl). Sample and buffer solutions were syringe filtered with 0.22 μm pore size (Merck Millipore, Germany) and centrifuged at 13500 rpm for 10 min in a tabletop centrifuge (Eppendorf, Germany) prior to each measurement.

SAXS measurements were carried out at beamline BM29, ESRF, Grenoble [276] at an X-ray wavelength λ of 0.99 Å, using a sample-to-detector distance of 2.87 m and a Pilatus 1M detector, resulting in a usable q-range of 0.06 to 3.0 nm^{-1} ($q = 4\pi sin(\theta)\lambda^{-1}$,

where θ is the total scattering angle). For each sample concentration, 10 runs with an exposure time of 4 s in 'flow' mode were conducted at room temperature. Buffer samples were measured using identical procedures before and after each sample measurement. Sample and buffer data from each run were analyzed for radiation damage; no damage was observed in any of the measurements. Matching sample and buffer profiles were averaged and buffer profiles were subtracted for background correction.

We performed Guinier analyses to determine radii of gyration (R_g) for all measured samples by fitting the logarithm of the scattering intensity as a function of q^2 to a straight line for small values of q. Mean and the reported standard deviations of the R_g values for each sample were computed from 10 R_g calculations that systematically varied the maximum q of the fitting range between $1.0 < q_{\max} \cdot R_g < 1.3$. For comparison, we also computed the standard deviations via error propagation using a fixed fitting range, which provided very similar standard deviation estimates. We found that R_g estimates from 2.5 mg/ml samples were systematically larger than estimates from the 5.0 mg/ml samples, presumably due to protein-protein interactions; specifically, R_g was larger by $\sim$ 1 Å at 100 mM salt concentration, and by $\sim$ 0.5 Å at 500 mM salt concentration. However, the modulations of R_g upon varying the alkali cation was very similar for the 2.5 mg/ml and 5.0 mg/ml samples, suggesting that protein-protein interactions have only a small effect on the *change* of R_g due the ion cloud (Fig. 3.6). Hence, as we here focus purely on the increase of R_g due the ion cloud, we averaged R_g taken from the 2.5 mg/ml and 5.0 mg/ml samples.

3.3.2 Method 1: Spherical model in conjunction with linearized Poisson-Boltzmann calculations

At low electrostatic potentials, $e_c\Phi \ll k_BT$, the non-linear Poisson-Boltzmann (PB) equations can be linearlized, thereby allowing analytic solutions for simple geometries. Here, e_c denotes the unit charge, Φ the electrostatic potential, k_B the Boltzmann constant, and T the temperature. As a model that allows such analytic solution of the linearized PB equation, we modeled the charged protein as a uniform sphere defined by its volume V, electron density ρ_{sph}, and total charge Q. Values for V and ρ_{sph} were taken such that the ρ_{sph} and the forward scattering intensity $\tilde{I}(0)$ match with the values computed from the protein structure, while neglecting any ion cloud effects. Accordingly, ρ_{sph} and $\tilde{I}(0)$ of the protein structure were first computed with the CRYSOL [224], using the default CRYSOL parameters, and the volume of the sphere was calculated as

$$V = \tilde{I}(0)^{1/2}/(\rho_{\text{sph}} - \rho_{\text{water}}), \tag{3.1}$$

where ρ_{water} denotes electron density of the water. The radius of the sphere is

$$R_{\text{sph}} = (3V/4\pi)^{1/3}, \tag{3.2}$$

and its radius of gyration is

$$\tilde{R}_g = (3/5)^{1/2} R_{\text{sph}}. \tag{3.3}$$

Hence, the absolute radius of gyration of the spherical model may differ from the radius of gyration of the protein; however, we found that matching $\tilde{I}(0)$ (or the total contrast) between protein and spherical model is more relevant for predicting the change of the radius of gyration owing to the ion cloud (ΔR_g) than matching the absolute radius of gyration.

For a uniformly charged sphere, the linearized PB equation can be solved analytically, as described in many monographs [277, 278]. Accordingly, the space is divided into three regions (Fig. 3.9): (1) the protein, $r < R_{\text{sph}}$; (2) an ion-exclusion layer, $R_{\text{sph}} < r < a$, taking into account that the centers of finite-sized ions can not move arbitrarily close to the protein surface. Hence, the thickness of the ion-exclusion layer is typically given by the ion radius; and (3) the solvent, $r > a$. The electrostatic potential in the region (3) is [278]:

$$\Phi(r) = \frac{Qe^{\kappa a}}{\varepsilon(1+\kappa a)} \cdot \frac{e^{-\kappa r}}{r} \tag{3.4}$$

where a is the sum of R_{sph} and the exclusion layer thickness. The symbol ε denotes the permittivity, and $\kappa = 1/\lambda_D$ is Debye-Hückel parameter, given by:

$$\kappa = \left(\frac{8\pi M e_{\text{c}}^2}{\varepsilon k_B T} \right)^{1/2}. \tag{3.5}$$

Here, M is number density of the ions in bulk at a large distance from the protein. Given the potential, the number density of the ion species i with ionic charge q_i is given by

$$n_i(\mathbf{r}) = M \mathrm{e}^{q_i \Phi(|\mathbf{r}|)/k_B T}, \tag{3.6}$$

where the index i indicates either the anion or the cation. The number densities $n_i(\mathbf{r})$ were written as functions of a Cartesian vector $\mathbf{r}$ to keep the following equations applicable also for non-spherical geometries (Method 2, see next paragraph). From the number densities $n_i(\mathbf{r})$, the effect of the counter ion cloud on the electron density contrast may be computed via

$$\delta\rho_{\text{CI}}(\mathbf{r}) = \sum_i n_i(\mathbf{r}) \Delta N_i^e \tag{3.7}$$

where the ΔN_i^e denotes the contrast per ion, given in units "number of electrons" e. The contrast per ion ΔN_i^e is given by

$$\Delta N_i^e = N_i^e - V_i \, \rho_{\text{water}}, \tag{3.8}$$

N_i^e and V_i denote the number of electrons and the volume of ion species i, respectively, and ρ_{water} is the density of water. For chloride, for instance, N_i^e would take the value $18e$. The volumes V_i of the ions were taken from the ionic radii in aqueous solutions reported in Ref. 279. The radius of gyration of the spherical protein model, including the effect of the counter ion cloud, is defined via

$$R_g^2 = \left(\int \Delta\rho(\mathbf{r}) \, \mathrm{d}\mathbf{r} \right)^{-1} \int \Delta\rho(\mathbf{r}) (\mathbf{r} - \mathbf{r}_s)^2 \, \mathrm{d}\mathbf{r}, \tag{3.9}$$

where $\mathbf{r}_s$ denotes the center of mass of the protein, and we used:

$$\Delta\rho(\mathbf{r}) = \begin{cases} \rho_{\text{sph}} - \rho_{\text{bulk}} & \text{if } |\mathbf{r}| \leq R_{\text{sph}} \\ 0 & \text{if } R_{\text{sph}} < |\mathbf{r}| \leq a \\ \delta\rho_{\text{CI}}(|\mathbf{r}|) & \text{if } |\mathbf{r}| > a \end{cases} \tag{3.10}$$

where ρ_{bulk} denotes the bulk density of salt solution. The integrals in eq. 3.9 was evaluated numerically. Finally, the change in the radius of gyration due to the ion cloud was taken as $\Delta R_g = R_g - \tilde{R}_g$.

3.3.3 Method 2: Non-linear Poisson-Boltzmann calculations

In contrast to the linearized PB equations used for the spherical model (last paragraph), the non-linear PB equations remain valid at high electrostatic potentials where $e_c\Phi(\mathbf{r}) \gtrsim k_B T$. Therefore, as the second method for computing the ion distribution around the charged proteins, we used non-linear Poisson-Boltzmann calculations. The calculations were carried out with the Adaptive Poisson-Boltzmann Solver (APBS) [280]. The structures of BSA and GI were prepared with the PDB2PQR software [281, 282]. The total size of the grid was $50{\times}50{\times}50\,\text{nm}^3$, using a grid spacing 2.6 Å. The center of mass of the protein was placed at the center of the grid. All calculations were conducted using the non-linear Poisson-Boltzmann equation. The radius of gyration was calculated using the same approach as described in previous paragraph (eqs. 3.7 through 3.10). However, the ion densities $n_i(\mathbf{r})$ were not taken from Eq. 3.6, but instead from the APBS result, and the integrals of eq. 3.9 were evaluated by using a discrete sum over the APBS grid points. Here, the density of the protein was accounted for by assigning the electrons of the protein atoms to the nearest APBS grid point.

3.3.4 Method 3: SAXS and ion density calculations from explicit-solvent MD simulations

Among the three computational models applied in this study, MD simulations provide the highest level of physical detail. The MD simulations are based on all-atom models of the protein, water, and ions. In contrast to the PB calculations employed here, MD simulations account for fluctuations of water, ions, and protein atoms, and they naturally include ion-ion correlations. Given that the employed force field accurately models ion-protein contacts, the simulations may further account for transient ion-protein salt bridges, for instance between cations and acidic amino acids. Accurate parameters for ion-protein contacts are available for certain ions [283, 284], and remain a matter on ongoing force field development.

SAXS curves were computed from the MD simulations using explicit-solvent SAXS calculations [90, 226]. Accordingly, all explicit water molecules and ions within a predefined distance from the protein contributed to the SAXS calculations, as defined by a spatial envelope (Fig. 3.1A/B, blue surfaces). Here, the envelope should be chosen large enough such that the solvent at the envelope surface is bulk-like, or, more precisely, that density correlations between the inside and the outside of the envelope are due to bulk solvent [89]. Following previous work [88, 90], the buffer-subtracted SAXS curve was computed from the scattering of atoms inside the envelope volume, as taken from MD simulation frames of two systems: (i) containing the protein in solvent and (ii) containing purely solvent (Figure 3.5).

To find an appropriate protein-envelope distance d, we computed ΔR_g with the spherical model (Method 1, see above) using increasing solvation shell thicknesses. We found that ΔR_g was mainly determined by the ions within approx. three Debye lengths from the protein surface (Fig. 3.7C). Hence, for production calculations, we used envelopes at distances of 3.0 nm and 1.6 nm from the proteins in systems with 100 mM and 500 mM salt, respectively. Ions at larger distances from the protein had a smaller effect on ΔR_g, which could in principle be captured by using even larger envelopes. However, with very large envelopes, calculations revealed numerical instabilities because (i) the SAXS curves are computed as a difference between two increasingly large numbers (scattering of protein including the hydration layer minus pure buffer scattering); and (ii) presumably due to tiny density mismatches between the protein and the pure-buffer simulation systems, which cannot be fully corrected by our density correction scheme [90]. Therefore, we used the spherical model (Method 1, see above) to estimate the ΔR_g contribution from ions outside the envelope, leading to corrections of 33%, 11%, 19% and 7% for the

systems BSA/100mM, BSA/500mM, GI/100mM and GI/500mM, respectively. This estimate is justified by the fact that the ion densities at large protein distances follow the Debye-Hückel behavior both in MD simulations and in the spherical model (Fig. 3.7).

The radius of gyration R_g was computed by fitting the Guinier approximation to the computed SAXS curves at small q, following $\ln[I(q)/I(0)] = -(qR_g)^2/3$, where $I(0)$ is the forward intensity. Here, in contrast to Guinier fit to the experimental SAXS curves (see above), we used a fixed q range because the calculated $I(q)$ curves follow exactly the Guinier approximation at small q [90]. Scattering contributions from the excluded solvent were computed with simulation boxes which contained the same concentration of the same ion species as the simulation boxes with the proteins, as described previously [90, 226]; otherwise, artifacts due to buffer mismatch would arise, in particular in systems with electron-rich ions (rubidium and cesium) and high salt concentration (500 mM). Notably, such buffer mismatch artifacts would arise not purely due to a mismatch in the mean solvent density (which can be corrected [89, 90]), but also due a mismatch owing to ion-ion correlations (Fig. 3.8). The error bars for R_g were computed by block averaging, using blocks of 3 ns.

The number density of ions versus distance from the protein was computed by constructing a series of envelopes around the protein, where the distance of the envelope vertices from the protein atoms was increased in steps of 0.25 Å. Then, the number densities were taken from the average number of ions between adjacent envelopes, averaged over the MD trajectories.

3.3.5 MD setup and parameters

The initial structures of bovine serum albumin (BSA) and glucose isomerase (GI) were taken from the Protein Data Bank (PDB; codes 4F5S [285] and 1MNZ [286], respectively). Organic molecules from the crystallization buffer were removed. The structures were placed into a simulation box of a rhombic dodecahedron, keeping a distance of at least 6.0 nm and 3.5 nm to the box boundary in the 100 and 500 mM salt concentration, respectively. The simulation boxes were filled with explicit TIP3P water [287]. Alkali metal and chloride ions were added by replacing the appropriate number of water molecules with ions. To neutralize the system, the number of added cations was larger by 16 (BSA) or 60 (GI) than the number of added Cl^- ions. The energy of each system was minimized with the steepest-descent algorithm.

The simulations were carried out using the Gromacs simulation software, version 5.1.2 [288]. During all simulations, the positions of the heavy atoms of the protein were restrained at the crystallographic positions with a harmonic potential (force constant

$1000\,\mathrm{kJ\,mol^{-1}nm^{-2}}$). This procedure ensured that modulations of the SAXS curve are purely a consequence of the ion cloud and the buffer, but not a consequence of varying protein conformations. Electrostatic interactions were calculated with the particle-mesh Ewald method [154, 155], and dispersive interactions were described by a Lennard-Jones potential with a cutoff at 1 nm. Bond lengths and angles of water molecules were contained with the SETTLE algorithms [289], and all other bond lengths were constrained with P-LINCS [290]. Each system was equilibrated for 30 ns. Longer equilibration had no effect on the SAXS curves. Production simulations were run for another 70 to 400 ns. The temperature was controlled at 300 K through velocity rescaling [291] (τ = 1 ps) and using a stochastic dynamics integration scheme [292] during equilibration and production runs, respectively. The pressure was kept at 1 bar with a Berendsen [159] and the Parrinello-Rahman barostat [160] (τ = 0.4 ps) during equilibration and production runs, respectively, thereby allowing stable box relaxation during equilibration and guaranteeing the correct ensemble during production.

3.3.6 Force field parameters

Interactions of the protein and ions in KCl, RbCl and CsCl solutions were modeled with the Amber99SB-ILDN force field [135, 163] and using the Joung-Cheatham ion parameters [293]. Li^+ and Na^+ ions modeled with the Joung-Cheatham parameters were found to bind strongly to aspartate and glutamate residues, leading to effects in the SAXS curves that did not match experimental findings. Therefore, systems containing NaCl we instead modeled with the CHARMM36 force field [294, 295], version of November 2016, translated into Gromacs [296]. The CHARMM36 force field implements a corrected Lennard-Jones (LJ) diameter acting between Na^+ and the oxygen atoms of carboxyl groups, termed NBFIX, thereby avoiding overbinding of Na^+ [283].

However, we observed strong binding of lithium Li^+ ions to carboxyl groups, also if simulated with CHARMM36. Hence, we hypothesized that the original force fields overestimate the lithium-carboxylate salt bridges, as previously reported for Na^+ [283]. As a simple measure to avoid such overbinding, we increased the LJ diameter $\sigma_{\text{Li-OC}}$ acting between the Li^+ atom type and the OC atomtype of the carboxylate oxygen atoms, thereby overruling the Lorentz-Berthelot combination rule. We tested a series of $\sigma_{\text{Li-OC}}$ and found that an increase of $\sigma_{\text{Li-OC}}$ by 6% is sufficient to avoid excessive overbinding. Hence, we used $\sigma_{\text{Li-OC}}$ increased by 6% for production simulations, $\sigma_{\text{Li-OC}} = 0.283\,\mathrm{nm}$. This increase is in the same order of magnitude as the increase by 3.8% for the Na^+–OC LJ diameter implemented in the NBFIX [284]. For comparison, we also tried to avoid overbinding of Li^+ modeled with the Joung-Cheatham parameters to carboxylate groups modeled with Amber99SB-ILDN. We found that the respective LJ diameter must be

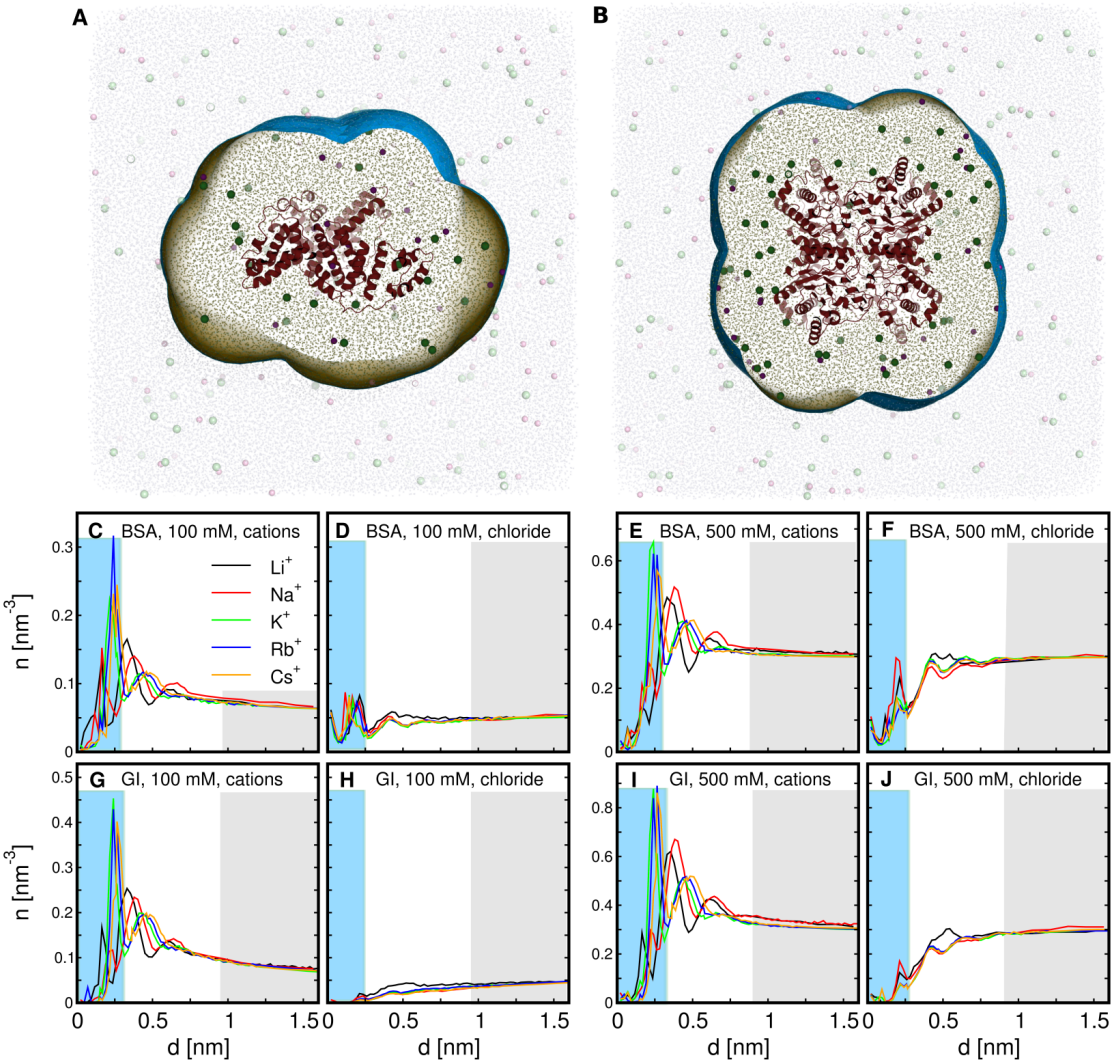

FIGURE 3.1: Simulation systems of (A) bovine serum albumine (BSA) and (B) glucose isomerase (GI), each with 100 mM NaCl buffer. Proteins are shown in red cartoon representation, water as sticks, and Na^+ and Cl^- ions as green and purple spheres. Envelopes (blue surface) at a distance of 30 Å from the protein atoms. Explicit water and ions inside the envelope were included into the SAXS predictions (opaque sticks and spheres), whereas water and ions outside the envelope are not included into SAXS predictions (transparent sticks and spheres). (C-J) Number density of alkali and chloride ions as a function of distance d from the protein, taken from atomistic MD simulations of (C-F) BSA and (G-J) GI. Bulk ion concentrations of alkali chloride were either 100 mM (C/D/G/H) or 500 mM (E/F/I/J). The color code indicates the alkali species present in the system (Li^+, Na^+, K^+, Rb^+, or Cs^+, see legend). Blue rectangles indicate regions of ion-protein salt bridges, while gray rectangles represent regions where distribution of ions obey Debye-Hückel theory.

increased by ∼50% to avoid overbinding, suggesting that the Amber99SB-ILDN/Joung-Cheatham force field combination is more prone to Li^+ overbinding. Further, we note that for cations that do not overbind (K^+, Rb^+, and Cs^+), the calculated SAXS curves computed form CHARMM36 or Amber99SB-ILDN simulations were nearly identical.

3.4 Results and discussion

3.4.1 Counter ion cloud and SAXS curves from MD simulations

To investigate the influence of the ion clouds on SAXS curves, we carried out MD simulations of two negatively charged proteins: bovine serum albumine (BSA, charge $Q = -16e$, Fig. 3.1A) and glucose isomerase (GI, $Q = -60e$, Fig. 3.1B). The simulations were conducted with ten different alkali chloride buffers composed of LiCl, NaCl, KCl, RbCl, or CsCl, either at 100 mM or 500 mM concentrations. Figure 3.1C-J presents the ion concentration of alkali cations and chloride anions in the ion cloud, as averaged over at least 50 ns of simulation, and plotted versus the distance from the nearest protein atom. Overall, the concentration curves show that cations are attracted whereas anions are repelled from the proteins, as expected for negatively charged proteins. More specifically, sharp peaks at small distances reflect transient salt bridges formed with protein residues ($d \approx 0.25$ nm, light blue background) [297]. Smaller maxima and minima at intermediate distances indicate ions of the second and third solvation layer (d approx. 0.4 nm to 0.7 nm) [241, 242]. At large distances, the concentrations gradually decay to the bulk level following the Debye-Hückel theory (gray background). This decay is slower at 100 mM compared to 500 mM (compare Fig. 3.1 C/D/G/H with E/F/I/J) rationalized by the fact that the Debye length is larger at 100 mM as compared to 500 mM. Hence, the MD simulations provide a detailed model of the ion cloud, including effects from (i) specific ion-protein salt bridges, (ii) solvation layer effects due to ion-ion correlations, captured only beyond a mean-field description of the ion cloud, as well as (iii) Debye-Hückel behavior at large distances.

Figure 3.2A-D presents SAXS curved of BSA in the ten different alkali chloride buffers, either computed from MD simulations (Fig. 3.2A/B) or obtained experimentally (Fig. 3.2 C/D). SAXS curves of GI were purely obtained from MD simulations (Fig. 3.2E/F). Critically, the SAXS curves were computed from MD simulations with position restraints on the heavy atoms, making sure that variations in the SAXS curve are purely a consequence of the ion cloud and not owing due to a conformational change in the protein. The general agreement between MD and experimental profile is reasonable. Exceptions are the systems with 500 mM RbCl or 500 mM CsCl case, presumably due to buffer mismatch between the protein solution and the buffer solution (Fig. 3.2D, blue and orange curves). In addition, Fig. 3.2 demonstrates that the overall SAXS intensity decreases at high salt concentrations of electron-rich ions, in particular Rb^+ and Cs^+, because the contrast of the protein with respect to the buffer decreases (Fig. 3.2B/F, blue and orange curves).

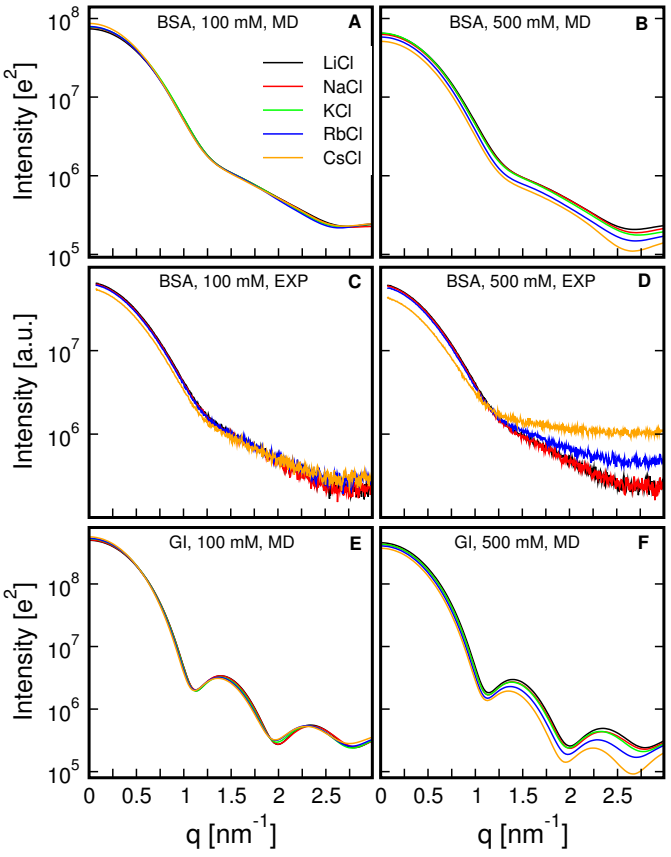

FIGURE 3.2: SAXS curves of (A-D) bovine serum albumin (BSA) and (E/F) glucose isomerase (GI) in different alkali chloride buffers (for color code, see legend) of 100 mM (A/C/E) or 500 mM (B/D/F) concentration. (A/B/E/F) SAXS curves computed from MD simulations, and (C/D) obtained experimentally.

3.4.2 Effect of the ion cloud on the radius of gyration

The influence of the ion cloud on the radius of gyration R_g is presented in Fig. 3.3, for BSA and GI systems with 100 mM or 500 mM salt buffers. Experimental data (red squares) are compared with results from our three theoretical models, namely (i) atomistic MD simulations (black dots), (ii) the spherical model (green triangle up), (iii) non-linear PB calculations (blue triangle down). The absolute R_g estimated from different theoretical methods may differ (Fig. 3.10), either owing to different modeling of the hydration layer, or because the absolute R_g of the spherical model may differ from the absolute R_g of the protein (see section on Method 1). In addition, the experimental absolute R_g were systematically larger as compared to the MD-based estimates, presumably due to a small fraction of aggregated BSA, in line with previous findings [271] (Fig. 3.10). Therefore, we purely compare the change of the radius of gyration, ΔR_g, relative to the system with NaCl salt. As evident from Fig. 3.3, we find reasonable agreement between experiment and calculations, suggesting that the trends in the calculated ΔR_g resemble the experimental conditions. A larger discrepancy between calculations

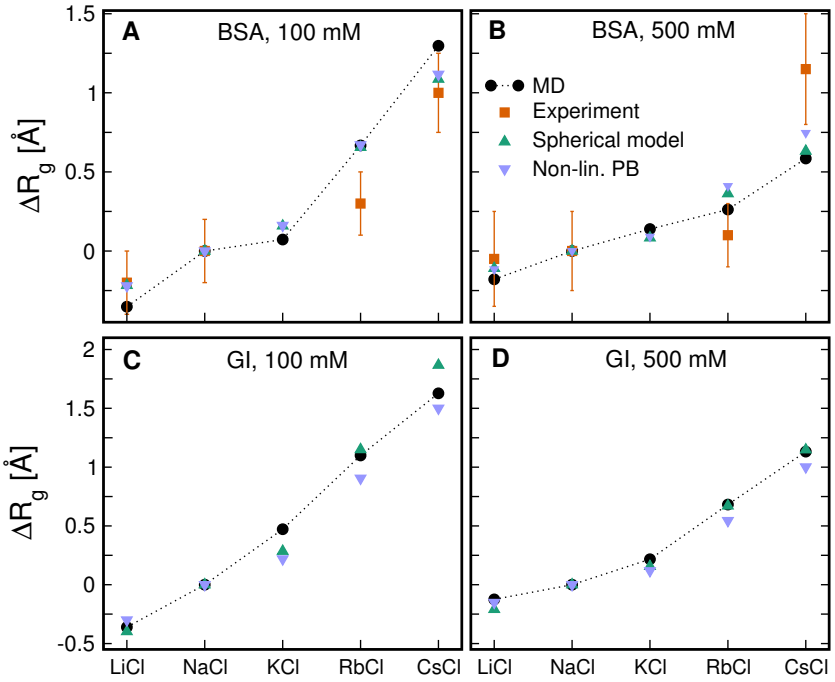

FIGURE 3.3: Change of a radius of gyration ΔR_g, shown relative to R_g in NaCl solution, as a function of salt type and concentration for (A/B) bovine serum albumin (BSA) and (C/D) glucose isomerase (GI). Results are shown as taken from all-atom MD simulations (black circles), non-linear Poisson-Boltzmann calculations (blue triangles down), the spherical model (green triangles up), and experiment (red squares). The absolute values of R_g are shown in Fig. 3.10.

and experiment were found purely with the 500 mM CsCl buffer, possibly due to buffer mismatch between the protein solution and the buffer solution as suggested from the experimental SAXS curve (Fig. 3.2D, orange curve).

Overall, we find that R_g increases upon varying the alkali cations from Li^+ to Cs^+, i.e., upon increasing the number of electrons of the cation. This trend is explained by the combination of two points: (a) the alkali cations are enriched in the counter ion cloud of the negatively charged proteins (Fig. 3.1 C/E/G/I), i.e. the number densities $n_i(\mathbf{r})$ of the cations are increased near the proteins as compared to bulk solvent. Notably, because only the ion charge enters the PB calculations, $n_i(\mathbf{r})$ does not depend on the type of alkali ions within PB theory; (b) larger alkali cations impose a larger electron density contrast *per* ion. In combination, upon varying the alkali cations from Li^+ to Cs^+, electron density contrast of the counter ion cloud increases, which eventually manifests in an increased R_g.

In addition, by comparing the 100 mM- with the 500 mM-systems, Fig. 3.3 demonstrates that R_g increases more strongly with smaller bulk salt concentration (compare Fig. 3.3A/C with B/D). This findings is rationalized by a longer Debye length at lower salt concentration: with longer Debye lengths, the ion densities decay more slowly to the bulk density, leading to a modified density contrast at larger distances from the protein. Since

the electron contrast enters the radius of gyration weighted by the squared distance from the protein center (Eq. 3.9), a longer Debye length eventually imposes a larger influence of salt on R_g.

3.4.3 Comparison of three theoretical methods for estimating ΔR_g

Next, we compare the ΔR_g estimates from the three theoretical methods used here, namely MD simulations, non-linear PB calculations with atomic protein models using APBS, and linearized PB calculations in conjunction with a spherical protein model. Despite the fact that the level of molecular detail captured by the three methods greatly differ, we find reasonable agreement between the ΔR_g estimates (Fig. 3.3, black, orange, blue and green symbols). The agreement between MD simulations and PB calculations suggests that molecular details of salt-protein interactions, such as transient salt bridges or the ionic solvation layers, have only a small effect on ΔR_g, and that, instead, ΔR_g is dominated by the long-range decay of the ion clouds towards bulk solvent, which is similar among the three models. This finding further implies that remaining uncertainties in ion-protein interactions during MD simulations, which might influence the ion density near the protein surface (Fig. 3.1C-J, blue background) have only a small influence on ΔR_g.

In addition, the agreement between non-linear PB calculations and the spherical model suggest that the non-spherical shape of BSA has a small effect on ΔR_g, and that linearized PB calculations as used by the spherical model are sufficiently accurate for estimating ΔR_g. These findings further imply that estimating the ion effect on R_g does not strictly require computationally expensive calculations such as MD simulations, but instead simplified calculations with reduced computational cost are sufficient. Hence, we expect that the spherical model provides a reasonable ΔR_g estimate for many common experimental conditions. Small discrepancies between the ΔR_g estimates of the spherical model and MD simulations may be explained by different ion distribution near the protein surface (Fig. 3.7).

Moreover, we tested influence of the exclusion layer on the ΔR_g estimates by the spherical model (Fig. 3.11), assuming exclusion layers between 0 and 2Å, i.e. in the range of typical ionic radii. We found that the exclusion layer has only small effect on the ΔR_g, suggesting that the exclusion layer may be omitted for predicting ΔR_g.

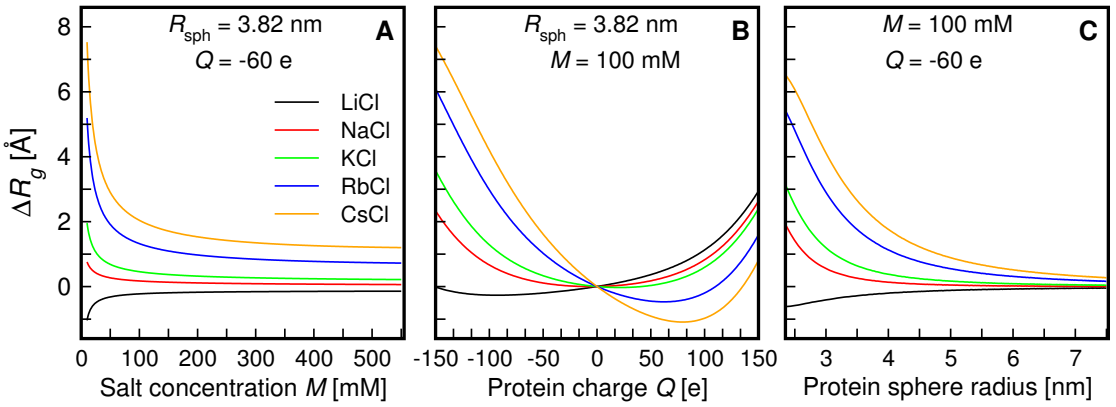

FIGURE 3.4: Systematic analysis of the effect of (A) salt concentration, (B) protein charge, and (C) protein size on the increase of the radius of gyration ΔR_g due to the counter ion cloud, as calculated with the spherical model. (A) ΔR_g versus salt concentration, at fixed protein charge $Q = -60e$ and fixed protein size ($R_{\mathrm{sph}} = 3.82\,\mathrm{nm}$) taken from glucose isomerase (GI). The color indicates the salt type, see legend. (B) ΔR_g versus protein charge, at fixed protein size taken from GI ($R_{\mathrm{sph}} = 3.82\,\mathrm{nm}$) and fixed salt concentration of 100 mM. (C) ΔR_g versus versus size of the protein, plotted as the radius of the sphere, at fixed protein charge $-60e$ and fixed salt concentration 100 mM. The trends for Li^+ are inverted with respect to the other alkali cations because Li^+ imposes a significantly smaller contrast per ion as compared to Cl^-.

3.4.4 A systematic analysis of ΔR_g: effect of salt concentration, salt type, protein charge, and protein size

We used the spherical model to systematically analyze how the following quantities influence ΔR_g: (i) bulk salt concentration M, (ii) protein charge Q, and (iii) protein size quantified by the radius R_{sph} of the sphere. To this end, we computed ΔR_g as a function of each of these three quantities, while keeping the other two quantities fixed, see Figure 3.4. Here, the respectively fixed quantities were taken from the spherical model for glucose isomerase, using $Q = -60e$, $R_{\mathrm{sph}} = 3.82\,\mathrm{nm}$ and $\rho = 431.6\,\mathrm{e\,nm^{-3}}$ as electron density of the sphere. In scans with fixed bulk salt concentration we used $M = 100\,\mathrm{mM}$. ΔR_g was computed for the five alkali chloride salts considered in this study (Fig. 3.4, color code). We note that at such high charge, the linearized PB solution is only an approximation; consequently, the spherical model may slightly overestimate the ion cloud effect on ΔR_g as compared to the non-linear PB solution, in line with ΔR_g computed for GI shown above (Fig. 3.3C/D, compare blue triangles down with green triangles up). The qualitative trends of ΔR_g, however, are expected to be correct.

Overall, in agreement with the results presented above, Fig. 3.4 shows that larger alkali ions have a larger effect on ΔR_g since they impose a larger electron density contrast per ion. An exceptional case is Li^+ that imposes a much smaller contrast per ion as compared to Cl^-. Here, in case of a negatively charged protein, the positive contrast owing to Li^+ enrichment is outweighed by the negative contrast owing to Cl^- depletion

(and vice versa in case of a positively charged protein); in consequence, the LiCl salt may lead to inverted trends in ΔR_g as compared to all other alkali chloride salts (Fig. 3.4, black curves).

Figure 3.4A presents the effect of the bulk salt concentration M, illustrating that ΔR_g increases with decreasing M (except for Li+, which inverts the trend). As pointed out above, this finding is mainly rationalized by the fact that, upon decreasing M, the Debye length increases, thereby leading to a modified density contrast at larger distances from the protein. The concentration M may, in addition, slightly influence the ΔR_g by shifting the relative effects from cation enrichment versus anion depletion from the surface, or vise versa in case of $Q > 0$ [258]. Notably, in the spherical model, ΔR_g diverges with $M \to 0$ as a consequence of a diverging Debye length, which deserves an additional comment: the diverging ΔR_g only applies at infinitely low protein concentration as assumed for the spherical model. Under experimental conditions of finite protein concentrations, in contrast, effects from protein-protein interactions would appear in the SAXS curves, which are not captured by this model. As such, the divergence at $M \to 0$ is primarily a consequence of the idealized model of a single protein in an infinite salt solution. The qualitative trends in Figure 3.4A, however, are relevant to experimental conditions.

Figure 3.4B presents the effect of the protein charge Q on ΔR_g. Only at small $|Q|$, the $\Delta R_g(Q)$ curves are approximately linear, where the slope strongly depends on the type of ion; the more negative slopes for larger alkali ions are again a consequence of the larger contrast per ion. At large $|Q|$, in contrast, the $\Delta R_g(Q)$ curves exhibit highly non-linear trends. Specifically, the curves pass through a minimum, whereby the position of the minimum depends on the type of alkali ion. These findings are rationalized as follows: The ion cloud of negatively charged proteins, such as GI and BSA, is characterized by an enrichment of cations (here: alkali cations) and a depletion of anions (here: chloride). Inversely, the ion cloud of positively charged proteins is characterized by a depletion of cations and an enrichment of anions. Both chloride and the alkali ions except Li^+ impose a positive electron density contrast, i.e., they carry more electrons than the water molecules that are displaced the by ion. Consequently, as Q increases from large negative values to large positive values, the concentration of the alkali ions in the counter ion cloud decrease, rationalizing the decreasing trends of $\Delta R_g(Q)$, as most apparent for CsCl and RbCl (Fig. 3.4B, blue and orange curves). However, since the concentration of the alkali cation in the counter cloud can not be depleted below zero concentration, the effects from anion enrichment may outweigh the effect from cation depletion at large positive Q, leading to an increasing ΔR_g at large positive Q. In other words, the non-linear relation between local concentration and potential (eq. 3.6) manifests in the highly non-linear $\Delta R_g(Q)$ relation visible in Fig. 3.4B.

Figure 3.4C shows the effect of the protein size on ΔR_g, as quantified by the sphere radius R_{sph}. Evidently, ΔR_g decreases with increasing size of the protein. This finding is explained by the fact that the contrast of the sphere increases proportional to the sphere volume, $\propto R_{\mathrm{sph}}^3$, whereas, owing to the fixed protein charge Q, the counter ion cloud adds approximately a constant contrast. Hence, with increasing R_{sph}, the additional contrast due to the counter ion cloud yields decreases relative to the total contrast.

3.5 Concluions

We have presented a systematic analysis on the influence of the ion cloud of charged proteins on SAXS curves, with a focus on the modulation of the radius of gyration ΔR_g as detected by a Guinier analysis. We found that ΔR_g strongly depends on the protein charge and size, on salt concentration, and on the ion type. For common test proteins such as BSA and GI, ΔR_g was found in the range of $-0.5\,\text{Å}$ to $2\,\text{Å}$ relative to a standard NaCl buffer, suggesting that such modulations are detectable with modern SAXS experiments, in particular with high-precision SEC-SAXS setups. Notably, these ΔR_g values are of similar magnitude as ΔR_g modulations imposed by water density modulations in the hydration layer [90], suggesting that understanding of both, hydration layer and the ion cloud effects on R_g are equally relevant. We expect our calculations to be useful for the interpretation of SAXS data of proteins, since they help to disentangle ΔR_g modulation owing to the ion cloud from ΔR_g modulation owing to other sources, such as functionally relevant conformation transitions or water density modulations in the hydration layer [102].

We found that MD simulations in conjunction with explicit-solvent SAXS calculations may account for ion effects on SAXS curves, in addition to the influence of the modified water density in the hydration layer of proteins studied previously [57, 88–90, 100, 101]. However, because the ion cloud may impose density modulations at relatively large distances from the protein, large, computationally expensive MD simulation systems are required in order to account for the entire ion cloud in such calculations. To estimate the ion cloud effect on SAXS data with reduced computational cost, we introduced two simplified methods either based on non-linear PB calculations with an atomic protein model, or based on a simple spherical model in conjunction with linearized PB calculations. The PB calculations provide accurate estimates for the ion cloud effects on the radius of gyration at greatly reduced computational cost.

In addition, the PB calculations can be used to test whether explicit-solvent SAXS calculations with a finite hydration layer thickness are biased by lacking contributions from ions at large distance from the protein, outside the layer that is explicitly included in

the calculations. If so, the PB calculations offer quantitative corrections to the explicit-solvent SAXS calculations. In combination, explicit-solvent calculations complemented by PB calculations provide SAXS predictions that fully account for the hydration layer, for specific ion-protein salt bridges, for ionic solvation layers, as well as for the long-range decay of the ion concentration following Debye-Hückel theory.

Acknowledgment We thank Dr. Martha Brennich for assistance at beamline BM29 at the ESRF (Grenoble, France). MTI, RS and JSH were supported by the Deutsche Forschungsgemeinschaft (DFG) through grant numbers HU 1971/1-1, HU 1971/1-3, and HU 1971/1-4. LKB and JL acknowledge support by the DFG through grant number SFB 863/A11.

3.6 Supplementary figures

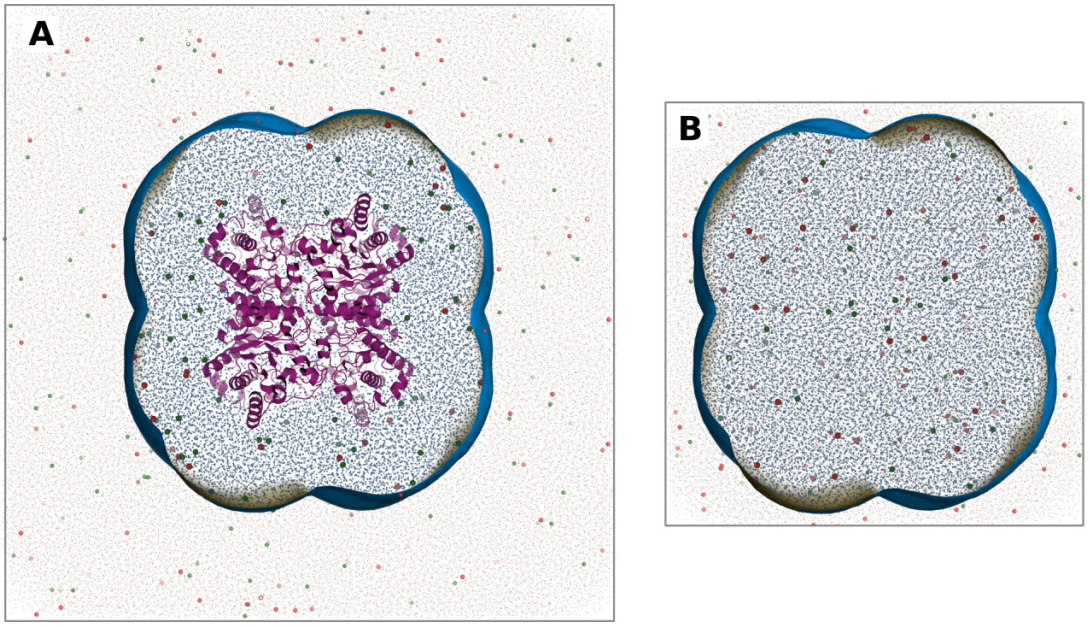

FIGURE 3.5: Simulation systems of (A) glucose isomerase (GI) and (B) pure solvent, each containing 100 mM NaCl. GI is shown in purple cartoon representation, Na^+ and Cl^- ions as green and red spheres, respectively, and water as blue spheres inside the envelope and as brown spheres outside of the envelope. The envelope is represented by a blue surface, constructed at a distance of 30 Å from the protein atoms in (A). The same envelope is used in the pure-solvent system (B). To compute the buffer-subtracted SAXS curve, scattering contributions from all atoms inside the envelopes of the protein/solvent (A) and the pure-solvent system (B) were taken into account, averaged over 3000 MD simulation frames. For more details, we refer to previous work (Chen and Hub, Biophys. J., 2014, 107, 435–447).

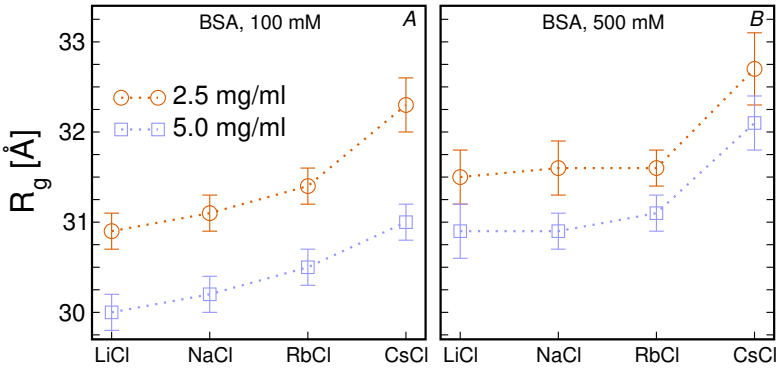

FIGURE 3.6: On systematic errors of the R_g estimates from experimental SAXS curves: absolute values of experimental R_g of BSA in different salt solutions from Guinier analysis, for protein concentrations of 2.5 mg/ml (orange circles) and 5.0 mg/ml (blue squares), at 100 mM salt (A) and 500 mM salt (B). R_g estimates from 5.0 mg/ml samples are systematically smaller compared to 2.5 mg/ml samples, presumably due to increased protein-protein interactions at higher protein concentrations. Such systematic errors are reduced at 500 mM salt concentration (B) compared to 100 mM salt concentration (A), reflecting reduced protein-protein interactions in consequence of increased screening of the protein charge. Critically, the *increase* of R_g, ΔR_g upon varying the alkali cation from Li^+ to Cs^+ is similar between the 2.5 mg/ml and 5.0 mg/ml samples, suggesting that ΔR_g is hardly affected by systematic errors due to protein-protein interactions. Hence, we averaged ΔR_g estimates from 2.5 mg/ml and 5.0 mg/ml samples before comparing with the calculated ΔR_g values.

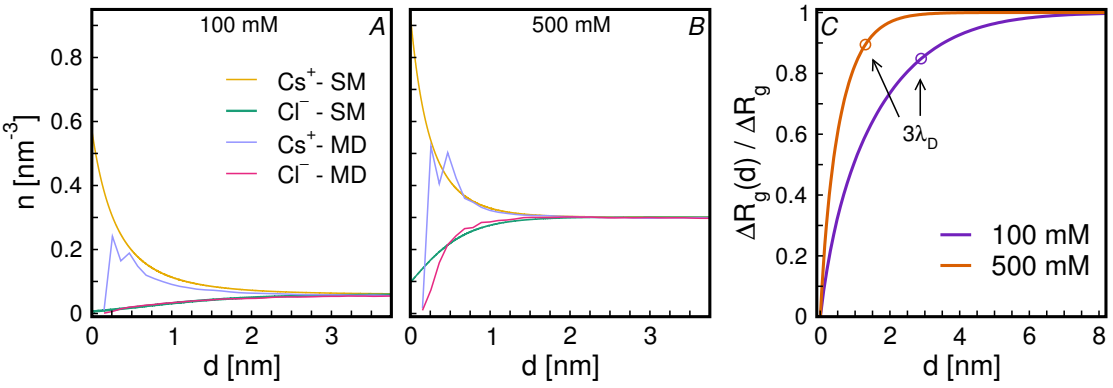

FIGURE 3.7: Number density of caesium (Cs) and chloride (Cl) ions as a function of distance d from the protein surface, taken from the spherical model (SM) and atomistic MD simulations (MD) of glucose isomerase at 100 mM CsCl (A) and 500 mM CsCl (B). The ion distributions from the SM and from MD agree at large distances from the protein, following the decay according to Debye-Hückel theory. At smaller distances, in contrast, large deviations between the SM and MD are found; here, specific salt bridges and solvation layers of ions revealed in MD simulations are not captured by the SM. (C) Cumulative contributions to ΔR_g from ions computed from the spherical model, plotted versus the distance d from the sphere surface. The circles indicate the points at $d = 3\lambda_D$, where λ_D is the Debye length.

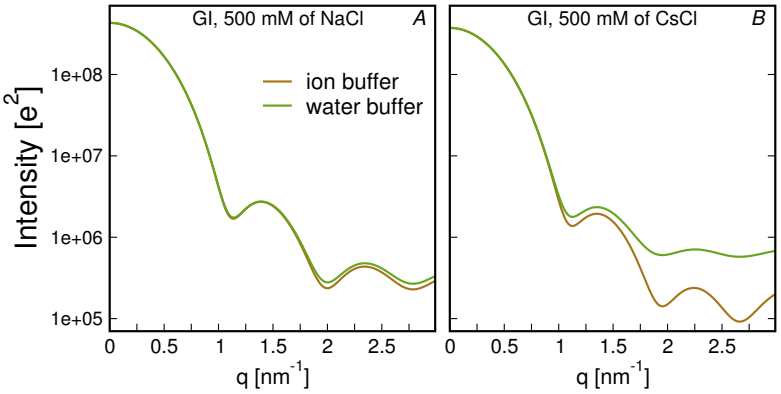

FIGURE 3.8: On the importance of buffer matching during explicit-solvent SAXS calculations with ionic buffers. (A) Calculated SAXS curves of GI in 500 mM of NaCl and (B) 500 mM CsCl. SAXS curves were computed for the pure-solvent simulations either with a matching salt buffer (brown lines) or using a non-matching pure-water buffer (green lines). Evidently, if a non-matching buffer is applied, large buffer mismatch artifacts appear with 500 mM CsCl (B), and small artifacts appear with 500 mM of NaCl (A). In these SAXS calculations, the mean buffer density was corrected as described previously (Chen and Hub, Biophys. J., 2014), suggesting that such artifacts do *not* appear owing to a mismatch of the mean density between solute and solvent systems, but instead owing to a mismatch in ion-ion correlations. This analysis highlights the importance of using identical buffers in solute and pure-solvent simulations, as done throughout in this study, in particular in systems with high salt concentration and large electron-rich ions, even when correcting the mean density.

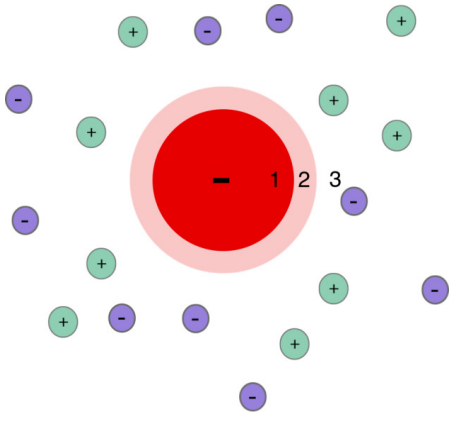

FIGURE 3.9: Spherical model of a charged protein in a salt solution, represented by 3 regions: (1) protein (red); (2) exclusion layer (pale red); (3) solvent with salt (green/purple spheres). The ion-exclusion layer accounts for the finite size of the ions, avoiding that the ions overlap with the protein.

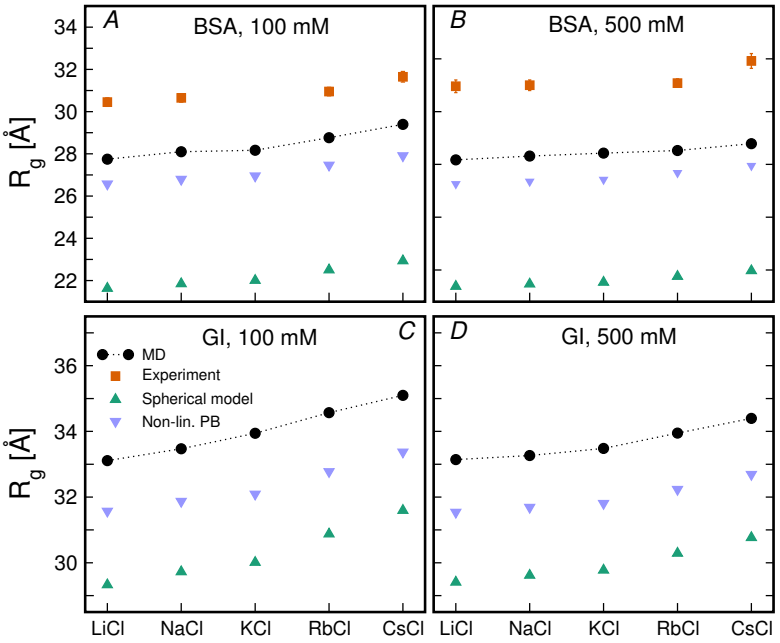

FIGURE 3.10: Absolute radii of gyration R_g taken from MD simulations (black circles), experiment (red squares), the spherical model (green triangles down), and non-linear Poisson-Boltzmann calculations (blue triangles right). R_g is shown for bovine serum albumin (BSA, A/B) and glucose isomerase (GI, C/D), at bulk salt concentrations of 100 mM (A/C) and at 500 mM (B/D). For a discussion on the differences, see main text.

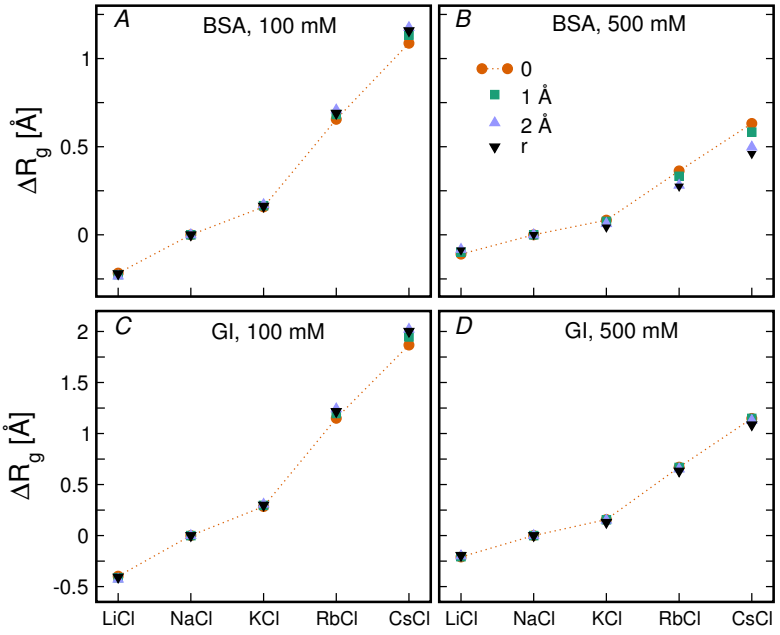

FIGURE 3.11: ΔR_g computed with the spherical model, using different values for exclusion layer thicks 0 Å, 1 Å, 2 Å, and ion radius (r), as indicated in the legend.

Chapter 4

Temperature-Dependent Atomic Models of Detergent Micelles Refined Against Small-Angle X-ray Scattering Data

THIS chapter, under its title, is published in Ref. 298. We acknowledge the permission from the *Angewandte Chemie International Edition* to reproduce the article.

Authors of the letter are Miloš T. Ivanović, Linda K. Bruetzel, Jan Lipfert and Jochen S. Hub, with the following contributions:
JSH and JL motivated the project;
MTI performed and analyzed MD simulations;
JL and LB performed SAXS experiment and density measurements;
all authors checked and discussed results;
MTI, JSH and JL wrote the article.

4.1 Abstract

Surfactants have found a wide range of industrial and scientific applications. In particular, detergent micelles are used as lipid membrane mimics to solubilize membrane proteins for functional and structural characterisation. However, an atomic-level understanding of surfactants remains limited because many experiments provide only low-resolution structural information on surfactant aggregates. Here, we combine small-angle X-ray scattering with molecular dynamics simulations to derive fully atomic models of

two maltoside micelles, at temperatures between 10°C and 70°C. We find that the micelles take the shape of general tri-axial ellipsoids and decrease in size and aggregation number with increasing temperature. Density profiles of hydrophobic groups and water along the three principal axes reveal that the minor micelle axis closely mimics lipid membranes. Our results suggest that coupling atomic simulations with low-resolution data allows for a structural characterisation of surfactant aggregates.

4.2 Introduction

As fundamental building blocks of soft matter systems, surfactants (surface-active agents) have found a wide range of industrial, scientific, and consumer applications [299, 300]. For instance, surfactants and their micellar aggregates may accelerate or inhibit chemical reactions as compared to an aqueous medium [301, 302]. Since surfactants may alter their structure in response to external stimuli [303], they have been used as a carriers for therapeutic molecules, to build confined reactions platforms for sustainable chemistry, and for modifying the characteristics of food products [304–306]. Surfactants are routinely applied in cosmetics, personal care, and cleaning products, hence the world market of surfactants was valued at ∼30 billion US dollars in 2015 [307]. Beside such applications, surfactants serve as model systems in soft matter research, as their self-organization into structures such a micelles or bilayers is still not fully understood [300]. A functional understanding of such soft matter systems remains limited, partly due to the lack of reliable atomic models.

Detergents are the most widely used type of surfactant. Above the critical micelle concentration *cmc*, detergent monomers in solution self-assemble and form micelles of various shapes such as spheres, ellipsoids or cylinders. The formation of micelles involves a delicate balance of free-energy contributions from enthalpy and both solute and solvent entropy [119, 120]. Consequently, many fundamental properties of micelles, such as the *cmc*, the aggregation number N, and their size and shape are temperature-dependent [308].

Since the cross section of certain detergent micelles resemble lipid membranes, micelles are frequently used as membrane mimics to solubilize membrane proteins for biochemical and structural characterization [309]. In the experiments, protein-detergent complexes are often exposed to varying temperatures, for instance in NMR studies to accelerate diffusion and hence to improve the quality of spectra, or to dissect thermodynamic equilibria involving conformational changes or ligand binding [310, 311]. Choosing a suitable detergent for membrane protein solubilization is often accomplished by trial and error; however, recent work suggests that matching of the hydrophobic thickness between

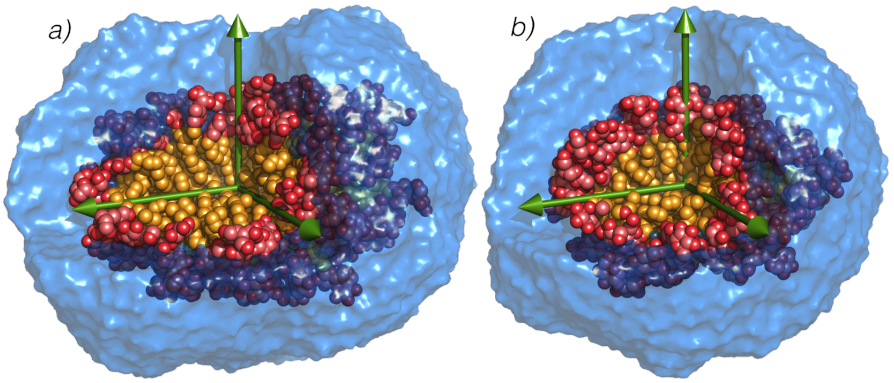

FIGURE 4.1: Atomic models of DDM (a) and DM (b) micelles at 25°C, refined against experimental SAXS data. Red spheres: head groups; orange spheres: tails; blue surface: explicit water included into the calculations of SAXS profiles. Green arrows indicated principal axes (length 4 nm).

the micelle and the membrane protein of interest can provide a route towards rational selection and design of detergent micelles [105, 106]. Hence, accurate information on the shape and size of detergent micelles, also as function of temperature, represents a starting point to improve the stability of protein-detergent complexes.

Detergent micelles have been studied using a range of methods, including small-angle X-ray and neutron scattering (SAXS/SANS) [107–109, 111], NMR self-diffusion [312], and several others [313–315]. The information content of such data is often insufficient to derive atomic models of micelles. Complementary, molecular dynamics (MD) simulations have been used to gain atomic insight into micelles [113, 115–117], however it remains unclear to which extent force field imperfections bias the structure and shape of the simulated micelle [114]. Hence, methods that integrate experimental data into MD simulations are needed to obtain reliable atomic models of micelles.

4.3 Results and discussion

Here, we derived atomic models of two maltoside micelles, n-dodecyl-β-D-maltoside (DDM) and n-decyl-β-D-maltoside (DM) at temperatures between 10°C and 70°C, by combining experimental SAXS data with all-atom MD simulations. The data were collected at beam line 12ID of the Advanced Photon Source [316] (SI Methods). We incorporated the SAXS data as a energetic restraint into MD simulations, allowing us to refine micellar models against scattering data. Because all SAXS predictions were based on explicit-solvent models, the calculations involve accurate physical models for the hydration layer and the excluded solvent, thereby avoiding any solvent-related fitting parameters and, in turn, enabling highly predictive structural modelling [90].

4.3.1 Micelle aggregation number

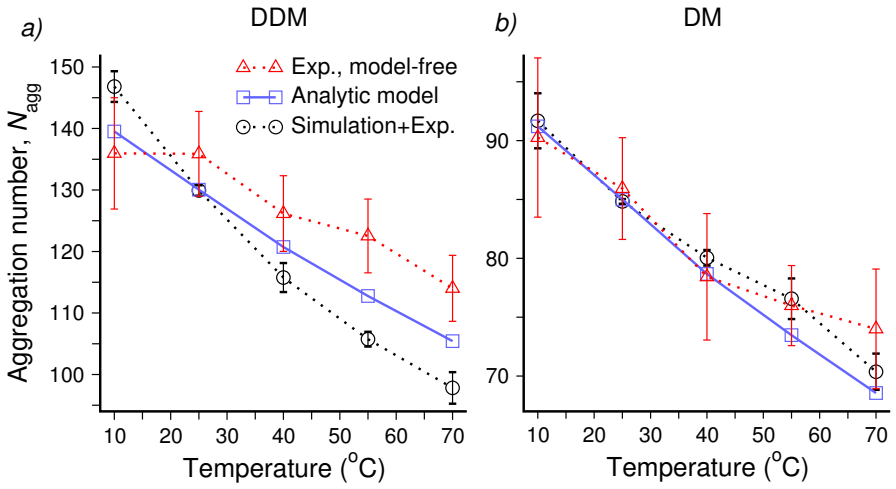

FIGURE 4.2: Aggregation number of DDM and DM micelles versus temperature, derived from SAXS data in a model-free procedure (red), from SAXS and MD simulations (black), and from an analytic model [119]. Error bars denote 1 SEM. Error bars on red symbols were obtained from repeated measurements and taking into account a 5% uncertainty in the intensity calibration. Error bars on black symbols were obtained from error propagation, see SI Methods.

We determined the aggregation number N_{agg} (i.e. the number of detergent monomers per micelle) as function of temperature from the experimental SAXS data using two independent methods. First, we determined N_{agg} from the extrapolated forward scattering intensity I_0, obtained by Guinier analysis of the data in the low q-region (Fig. 4.6), together with the expected scattering intensity from a detergent monomer. This approach is model-free in the sense that it does not rely on a particular representation of micelle structure, and it has been successfully applied to a range of different detergent micelles at room temperature [107, 108]. Here, we measured and explicitly took into account the temperature dependencies of the buffer and detergent densities (SI Methods and Figs. 4.7, 4.8, 4.9). The data show that N_{agg} decreases with increasing temperature (Fig. 4.2, red triangles), by ≈20% going from 10°C to 70°C.

Second, we estimated the experimental N_{agg} using a series of free MD simulations with increasing N_{agg} at each temperature. From each simulation, we computed the scattering profile and compared the position of the pronounced minimum at $q \approx 1\,\mathrm{nm}^{-1}$ to the experimental data to determine the best fitting simulations and thus the estimated experimental N_{agg} (Fig. 4.2, black circles; SI Methods and Figs. 4.10, 4.11).

The two approaches to determining N_{agg} give good agreement both for the absolute values and for the temperature dependencies; for DM, the agreement is even excellent and within experimental error. For DDM, some deviations in particular at the highest temperatures are apparent. These deviations might stem from the fact that the lower

density for DDM compared to DM makes the analysis more sensitive to uncertainties in the temperature dependence of the detergent density, both for estimates from I_0 and from MD simulations. Notably, our N_{agg} estimates for 25°C agree with previous reports [107–109].

Having determined the temperature dependence of N_{agg} enables us to test the analytic model for micellar aggregation proposed by Chandler and coworkers [119]. Using plausible values for the alkyl-chain length parameter (SI Methods), the Chandler model (Fig. 4.2, blue line and symbols) provides a good description of the experimental data. The remaining discrepancy can be explained by the fact that Chandler's theory assumes a spherical micelle, whereas DDM micelles and to a lower degree DM micelles take an ellipsoidal shape (see below).

4.3.2 Micelle shape

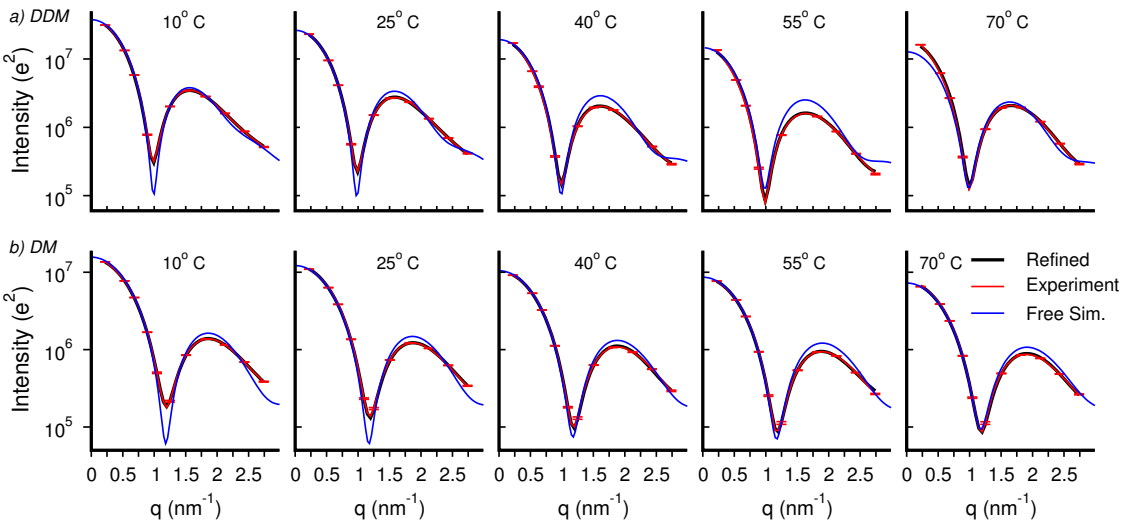

FIGURE 4.3: Experimental SAXS curves (red, representative errors obtained from repeated measurements), computed from free (blue) and refined simulations (black) of DDM (top row) and DM (bottom row) micelles at the different temperatures as indicated in the subplots.

The scattering profiles computed from 50 ns of free MD simulations of DM and DDM micelles at the temperatures between 10°C and 70°C conducted at the estimated N_{agg} (Fig. 4.2) give reasonable, but not perfect agreement with the experimental data (Fig. 4.3, blue and red curves, respectively). The experimental data suggest that the micelles under our conditions are monodisperse (see SI Text), making it unlikely that heterogeneous ensembles over different N_{agg}, i.e. polydispersity in micelle size, accounts for the remaining residuals. In addition, we found that modelling polydispersity in N_{agg} hardly improved the agreement (Fig. 4.14), indicating that not a distribution over N_{agg}, but

instead incorrect micellar shapes account for the residuals between simulation and experiment. Specifically, the too pronounced minima ($q \approx 1\,\mathrm{nm}^{-1}$) and maxima ($q \approx 2\,\mathrm{nm}^{-1}$) in the calculated profiles suggest that micelles in free simulations were slightly too spherical, which is likely a consequence of small imperfections of the CHARMM36 force field [317] used in our simulations.

To refine the micelles against the experimental SAXS data, we introduced the experimental curves as an energetic restraint into the simulations [91]. Ultraweak restraints of 2-3 kJ/mol for the entire system (Fig. 4.12) were sufficient to restrain the micelles to shapes in quantitative agreement with the data (Fig. 4.3, red and black curves). We found that the SAXS-derived restraints hardly influenced the distribution of tail length as compared to free simulations (Figs. 4.15, 4.16), but the restraints modified the shape of the micelle, as apparent from the distributions of radii of gyration around the three principal axes (Fig. 4.19). Representative snapshots of refined DDM and DM micelles at 25°C shown in Fig. 4.1 reveal slightly elongated ellipsoidal shapes.

To characterise the refined micellar shapes more quantitatively, we computed the density profiles along the three principal axes, decomposed into contributions from the hydrophobic tails, head groups, and water (Fig. 4.4b-g). The density profiles were computed from 300 ns of SAXS-restraint simulations at 25°C, and from 40–60 ns for all other temperatures, suggesting that conformational fluctuations of the micelles, to the extent allowed by the SAXS restraint, are included in the density profiles. As evident from the three distinct semi-axes, the micelles did not take the shape of a spheroid (an ellipsoid with two identical semi-axes), but instead took the shape of a general tri-axial ellipsoid. Figure 4.4h/i summarize the semi-axes for temperatures between 10°C and 70°C, taken from the density of tails and head groups. The micelles shrink with increasing temperatures along the major and middle axes, as expected from the decreasing N_{agg}. The overall micelle dimensions shrink by reducing the thickness of the hydrophobic region and not by reducing the thickness of the head group region (Figs. 4.17 and 4.18). In contrast, the extension along the minor axes is nearly temperature-invariant, likely because the shortest axes is constrained by the extensions of the detergent tails, which are almost independent of temperature (Fig. 4.17). Further, along all three axes, the water density decays gradually over a range of ∼1.5 nm between the bulk and hydrophobic tail regions (Fig. 4.4b-g, cyan), similar to the water density detected by neutron reflectometry for phosphatidylcholine membranes [318].

For both DDM and DM at most temperatures, the micelles are mainly characterised by one long and two approximately equal shorter semi-axes (Fig. 4.4h,i, 4.17, and 4.19), i.e., the micelles rather resemble prolate than oblate ellipsoids; DDM micelles at 55°C and 70°C even take on close to ideal prolate shapes. This finding seems to contrast

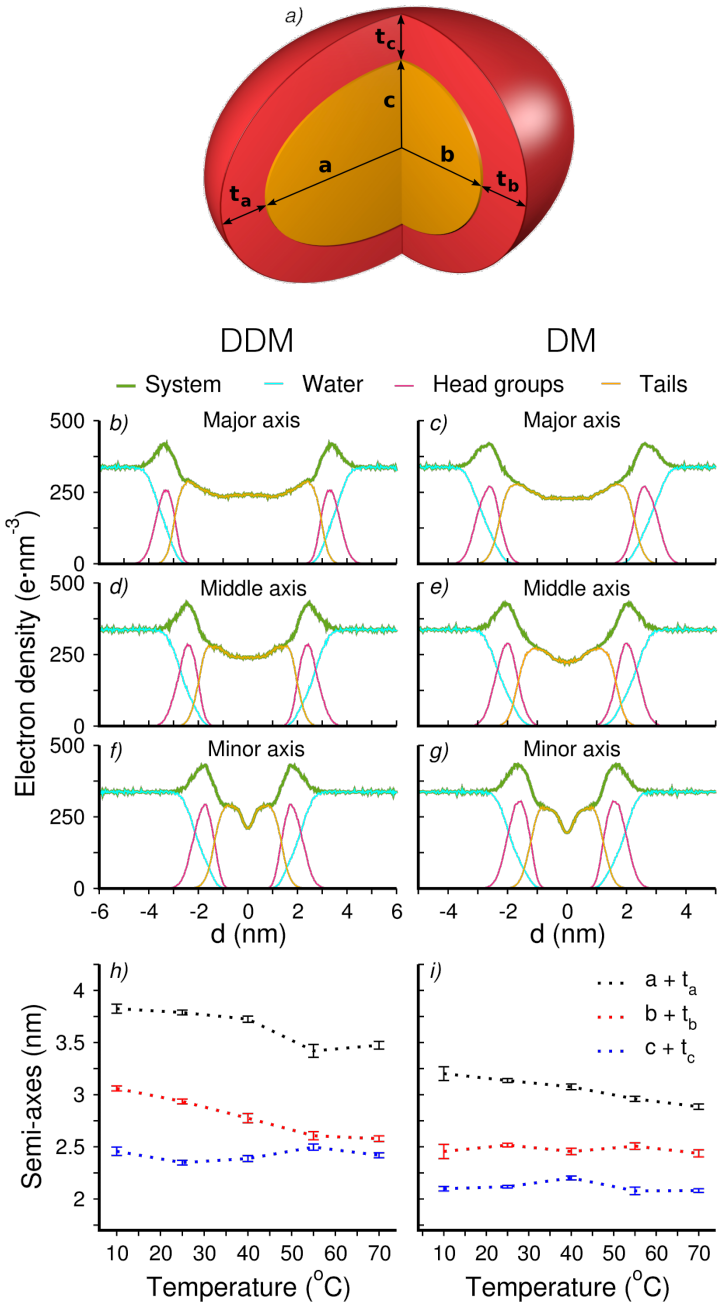

FIGURE 4.4: (a) Schematic representation of a micelle as a two-component tri-axial ellipsoid. Orange: detergent tails; red: head groups. (b-g) Density profiles along the major, middle, and minor principal axes of refined DDM (b, d, f) and DM (c, e, g) micelles at 25°C. (h, i) Semi-axes of DDM (g) and DM (h) micelles versus temperature, defined as FWHM of the densities (SI Methods). Hydrophobic semi-axes a, b, c (see also Fig. 4.17), and head group thicknesses t_a, t_b, t_c are illustrated in (a). Error bars denote 1 SEM obtained from block averaging.

previous fits of an implicit two-component micelle model to SAXS data, deserving further explanation. Previously, micellar shapes were extracted from SAXS data by fitting symmetrized geometric models, namely prolate or oblate ellipsoids with only two independent semi-axes (Fig. 4.4a with $a = b$ and $t_a = t_b = t_c$) [107, 108]. Such fits, however,

may become bistable, thereby yielding prolate and oblate models with similar agreement with the data. In case of bistable fits, the physically relevant solution was chosen such that the short semi-axis is shorter than the detergent chain, thereby avoiding a vacuum cavity in the micelle [120]. For DDM and DM, this procedure has led to the proposition of oblate micellar shapes [107]. By contrast, for the present study, we avoided any assumptions about the symmetry, but we let the simulation decide which micellar shape is most plausible in the light of the data and the force field. Thereby, we obtained qualitatively different, namely more elongated prolate-like micellar shapes.

4.3.3 Comparison with membrane models

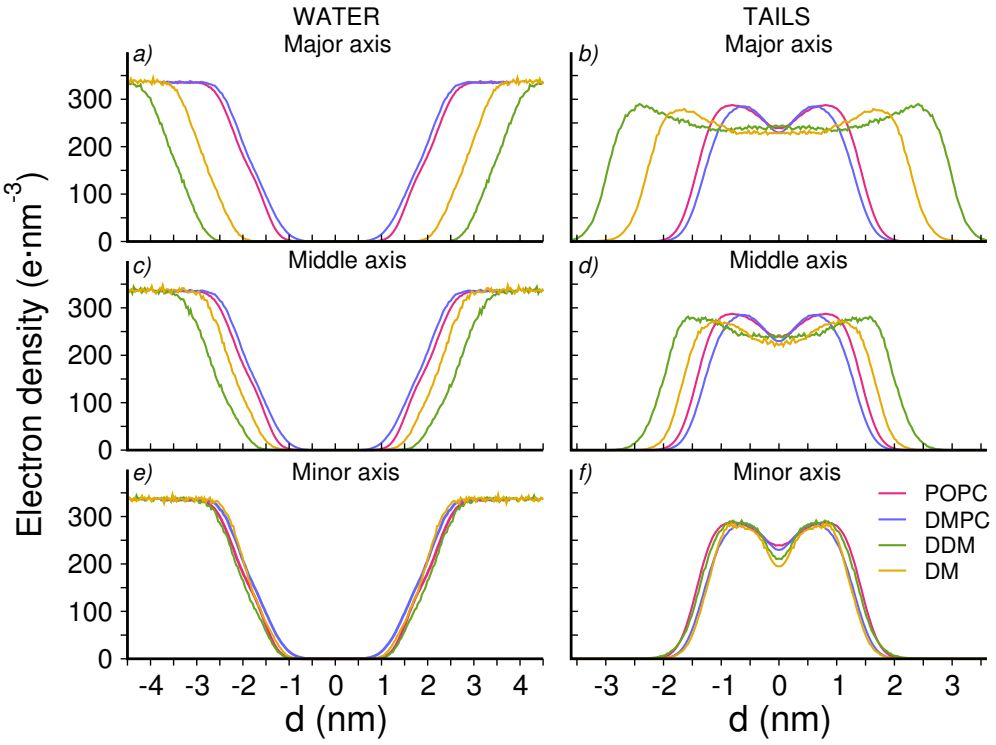

FIGURE 4.5: Density profiles at 25°C of water (left column) and hydrophobic detergent tails (right column) along the major axis (a,b), middle axis (c,d), and minor axis (e,f) of micelles of DDM (green lines) and DM (dark yellow lines). For comparison, density profiles of water and hydrophobic tails across lipid membranes of POPC (red lines) and DMPC (blue lines) are shown.

Recent work suggested that the stability of a protein-detergent complex improves if the hydrophobic thickness matches between micelle and protein [105, 106]. To test if the refined DDM and DM micelles provide an accurate mimic for lipid membranes, we compared the hydrophobic and the polar profiles along the three principal axes with the profiles of two typical model membranes, composed of either palmitoyl oleoyl phosphatidylcholine (POPC) or of dimyristoyl phosphatidylcholine (DMPC). For the density of water as well as for the hydrophobic tails, we found an excellent match between

the micelle's minor axis and the lipid membranes, in particular between DDM and POPC on the one hand, and between DM and DMPC on the other hand (Fig. 4.5e,f). In contrast, the profiles along the middle and major axes are too wide to match the lipid membranes (Fig. 4.5a-d). This suggests that membrane proteins are predominantly embedded into DDM and DM micelles such that the protein's membrane-normal axis is aligned along the minor micelle axis. Further, since maltosides have been frequently used to solubilize membrane proteins, our analysis supports the view that a match between the minor micelle axis and the lipid membrane is the key determinant for successful protein solubilization [105].

4.4 Conclusions

To conclude, we have derived fluctuating atomic models of two maltoside micelles by combining experimental SAXS data with MD simulations and explicit-solvent SAXS predictions. Free simulations revealed reasonable, but not quantitative agreement with experimental SAXS curves; hence a weak experiment-derived energetic bias was required to obtain simulations that accurately agree with experimental conditions. We found that DDM and DM micelles take the shape of a general tri-axial ellipsoid, where major and middle axes decreased with increasing temperature, whereas the minor axis was approximately constant between 10°C and 70°C. Density profiles along the principal axes showed that the cross section along the minor axis of the micelles closely mimics lipid membranes, with implications on the rational design of stable protein-detergent complexes. We found the aggregation number N_{agg} to decrease moderately with increasing temperature, predominantly by shrinking the major and middle axis of the micelle.

The study highlights that a direct coupling between experiment and simulation provides more spatially detailed and more reliable structures of soft matter systems, as compared to each of the methods alone. Specifically, SAXS provides information on the overall shape and size, but does not provide information at the atomic level. MD simulations provide atomic details with reasonably accurate potential energy functions (force fields), and they naturally account for thermal fluctuations; however, MD simulations have difficulties with obtaining large-scale features a priori. As such, SAXS and MD provide highly complementary physicochemical information. Our work may provide a starting point for a rational selection of detergent for solubilizing membrane proteins, and for further improvements of detergent force fields. The fact that the minor axes are approximately independent of temperature (Fig. 4.4) implies that a match between the length of the hydrophobic part of the protein and the micelle thickness would be maintained over a significant temperature range and, consequently, that the maltosides investigated

here can be used for solubilization over a range of temperatures. Future work will have to test to what extent these trends also hold for other detergents. While the refinement of DM and DDM micelles was simplified by their monodispersity in size, the refinement of highly polydisperse micelle solutions will likely require the explicit treatment of heterogeneous ensembles. In addition, we are currently extending the protocol for incorporating complementary SANS data, providing a framework for integrative structural modelling of soft matter systems.

Acknowledgment We thank Linda Columbus and Sebastian Doniach for discussions, Linda Columbus for providing detergent samples, Sönke Seifert for help with measurements at beam line 12ID at the Advanced Photon Source, and Annalena Salditt for performing the density measurements. We acknowledge funding from the Deutsche Forschungsgemeinschaft (HU 1971/1-1, HU 1971/1-3, HU 1971/1-4, Nanosystems Initiative Munich). This research used resources of 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.

4.5 Supplementary material

4.5.1 Methods

4.5.1.1 Analytic model for the aggregation number

Maibaum, Dinner, and Chandler have proposed an analytic model for the aggregation number of spherical detergent micelles [119],

$$N_{\mathrm{agg}}^{\mathrm{ana}} = (49\pi/48)\beta\gamma\delta^2, \tag{4.1}$$

where $\beta = (k_B T)^{-1}$ denotes the inverse temperature, and γ is the interfacial oil-water surface tension. Values for γ for DDM and DM at the different temperatures were taken from Ref. 319. δ is a length parameter of the alkyl chain of n_c carbon atoms. Because the exact value of δ for a given type of detergent is not obvious, we adjusted δ such that $N_{\mathrm{agg}}^{\mathrm{ana}}$ matched N_{agg} estimated by comparing the MD simulations with the experimental SAXS curves at 25°C (Fig. 4.2, black circles). This procedure yields $\delta = 17.8\,\text{Å}$ for DDM and $\delta = 14.5\,\text{Å}$ for DM, in reasonable agreement with the maximum length of the tails given by Tanford's formula [120], $\delta = (0.15 + 0.1265 n_c)\,\mathrm{nm}$, where n_c denotes the number of carbon atoms in the tail (10 for DM and 12 for DDM), suggesting that the parameter δ was adjusted to physically reasonable values (see also Fig. 4.15 below).

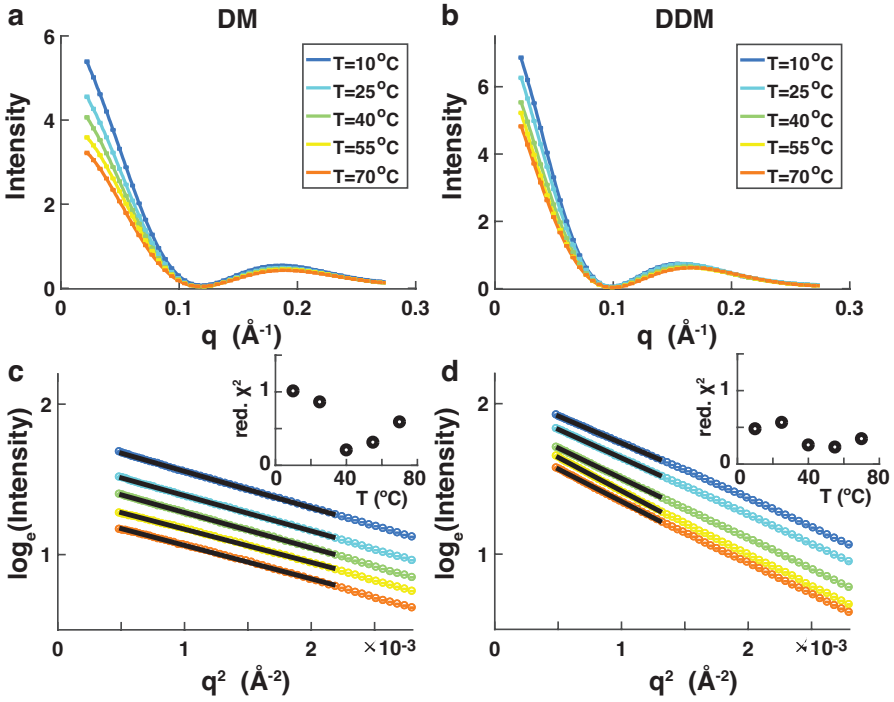

FIGURE 4.6: SAXS data and Guinier analysis for DM and DDM micelles as a function of temperature. a), b) Scattering profiles recorded for 45 mM DM (a) and 45 mM DDM (b) in the temperature range from 10°C to 70°C (see legends in panels a and b). Error bars are shown only for every 10$^{\text{th}}$ point for clarity. c), d) Guinier analysis of the same data as show in panels a and b, using the same color code. The black lines indicate typical fitting ranges, which where chosen such that $R_g \cdot q_{\max} \leq 1.3$ where $q_{\max}$ is the largest q-value included in the fit. Error bars are smaller than symbols. The insets show the reduced-χ^2 values (computed as squared difference between data and fit, divided by the variance, and normalized to the number of points included in the fit) for the Guinier fits presented in the main panel. Reduced-χ^2 values ≤ 1 indicate an excellent fit. The good linearity of the data on the Guinier range is apparent, suggesting monodisperse samples and the absence of bias in the data by micelle-micelle correlations.

4.5.1.2 SAXS experiments

Experimental data were collected at beam line 12ID of the Advanced Photon Source [320] (APS), essentially as described previously (Table 4.1) [107, 108, 321]. Measurements used custom-made sample cells and a cell holder that was temperature controlled in the range from 10°C to 70° with a circulating water bath [316]. Data were collected at an X-ray energy of 12 keV (corresponding to a wavelength of $\lambda = 1$ Å) using a sample-to-detector distance of 1.8 m, resulting in a useable q-range of 0.02 Å^{-1} to 0.275 Å^{-1}. Scattering angles were calibrated using a silver behenate standard sample. DM and DDM were purchased from Anatrace and measured in 20 mM phosphate buffer, pH 6.2, with 150 mM NaCl added at a detergent concentration of 45 mM. It has been previously shown that under these conditions, both DM and DDM micelles are monodisperse and interparticle interference (finite concentration) effects are negligible [107, 216]. For each

TABLE 4.1: Data-collection parameters. Data read out, normalization and circular averaging were performed using custom routines at beam line 12ID, APS, IL, USA. Buffer subtraction and Guinier analysis were performed using custom routines in Matlab (Mathworks) that are available from the authors upon request. Software for SAXS predictions and model refinement are described in detail in the SI text below.

Beam line	Beam line 12ID Advanced Photon Source, IL, USA
Wavelength (Å)	1.0
Useable q-range ($Å^{-1}$)	0.02 - 0.275
Exposure time (s)	3.0
Monomer concentration (mM)	45
Temperature range (K)	283 - 343

TABLE 4.2: Detergent parameters used in the calculation of the aggregation number from the forward scattering intensity.

Symbol	**Quantity**	**Value at 25°C**	
		DM	**DDM**
c	Detergent concentration	45 mM	45 mM
cmc	Critical micelle concentration	1.8 mM	0.17 mM
ρ_{det}	Electron density of the detergent	0.407 e $Å^{-3}$	0.398 e $Å^{-3}$
ρ_{sol}	Electron density of the solvent	0.340 e $Å^{-3}$	0.340 e $Å^{-3}$
V_{mon}	Volume of a detergent monomer	644.0 $Å^3$	697.8 $Å^3$

condition, three exposures of 1.0 s each were taken, image corrected, and circularly averaged. The three resulting profiles for each condition were compared to confirm the absence of radiation damage and averaged to improve signal. Matching buffer profiles were collected with identical procedures and subtracted for background correction. We employed 8 mg/ml horse heart cytochrome c (Sigma), in 100 mM acetate buffer, pH 4.6, with 0.5 M guanidinium hydrochloride as a scattering standard. All samples were centrifuged at $11000 \cdot g$ for 10 min prior to data collection.

4.5.1.3 Model-free determination of micelle aggregation number from forward scattering

The aggregation number N of micelles can be determined from the forward scattering intensity via the relationship [107]:

$$N = \frac{I(0)_{\mathrm{exp}}}{I(0)_{\mathrm{mon}}} = \frac{I(0)_{\mathrm{exp}}}{K\left[c(T) - cmc(T)\right]\left[\rho_{\mathrm{det}}(T) - \rho_{\mathrm{sol}}(T)\right]^2 V_{\mathrm{mon}}^2(T)} \tag{4.2}$$

$I(0)_{\mathrm{exp}}$ is the experimentally determined forward scattering intensity of a detergent solution above the critical micelle concentration cmc that is obtained from Guinier analysis of the SAXS data (Figure 4.6). The denominator is the expected forward scattering signal from a detergent monomer (which is too weak to be measured directly). The

proportionality constant K is setup specific and was determined from measurements of the scattering standard cytochrome *c*. The remaining parameters in the denominator are properties of the detergent and the solvent: ρ_{det} and ρ_{sol} are the electron densities of the solvent and detergent, respectively, c is the detergent concentration, *cmc* the critical micelle concentration, and V_{mon} the volume of a detergent monomer. An overview of the parameters and their known room temperature values from the literature are given in Table 4.2. The SAXS data at 25°C obtained in this study were analyzed using the parameter values from Table 4.2. The solvent and detergent parameters are, however, temperature dependent, as discussed in the next paragraph.

Size monodispersity In line with previous findings [322], our SAXS data suggests that DDM and DM micelles are reasonably monodisperse in size (or in aggregation number N_{agg}), based on two lines of arguments:

Firstly, if the micelle ensembles would exhibit a wide size distribution, the distribution would likely shift to larger sizes when increasing the detergent concentration, thereby leading to non-trivial changes in the scattering profiles. However, for concentrations low enough such that interparticle interference is negligible, the scattering profiles are superimposable after rescaling by concentration [107], suggesting that the micelles are reasonably monodisperse.

Secondly, polydisperse systems typically exhibit deviations from linearity in the Guinier region. However, we observe good linearity of the experimental data in the Guinier plots (Fig. 4.6c,d), providing additional indication that the micelles are reasonably monodisperse.

Moreover, we found that modelling the micelles as heterogeneous ensembles over various N_{agg} does not improve the agreement between experimental SAXS data and the SAXS data computed from free, unbiased simulations (Fig. 4.14). This suggests that (i) residuals between calculated and experimental SAXS curves are not caused by polydipersity in N_{agg} but instead reflect differences in micelle shape, and (ii) that explicit modelling of polydispersity in N_{agg} would mainly increase the risk of overfitting. Hence, we did not consider polydispersity in N_{agg} during SAXS-driven simulations.

Temperature dependence of the concentration The concentration depends on temperature since the volume of the solvent changes with temperature. Stock solutions were prepared at room temperature (∼25°C) and aliquots from the same stock solution were measured at different temperatures. We approximate the volume expansion of the solvent by the tabulated values for the density of water (Figure 4.7). Since the density of

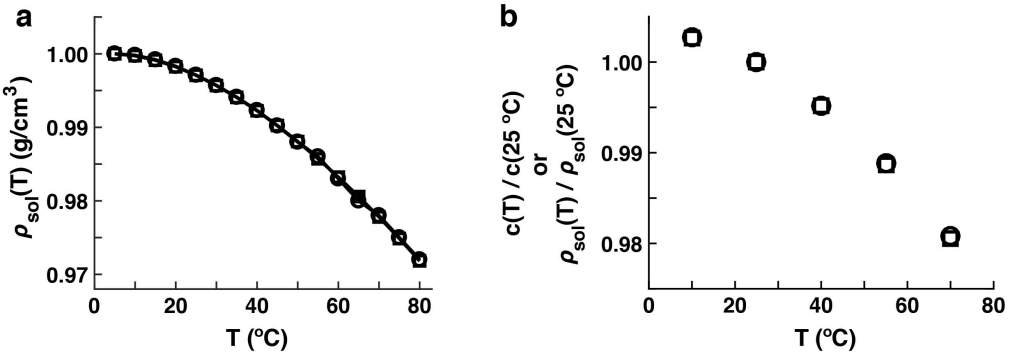

FIGURE 4.7: Temperature dependence of solvent density. a) Density of water as a function of temperature. Circles are data from http://www.engineeringtoolbox.com/water-thermal-properties-d_162.html; squares are data from Ref. 323. b) Relative temperature dependence of the density of the solvent and of the concentration, normalized to the values at 25°C. The values were computed from the water densities in panel a).

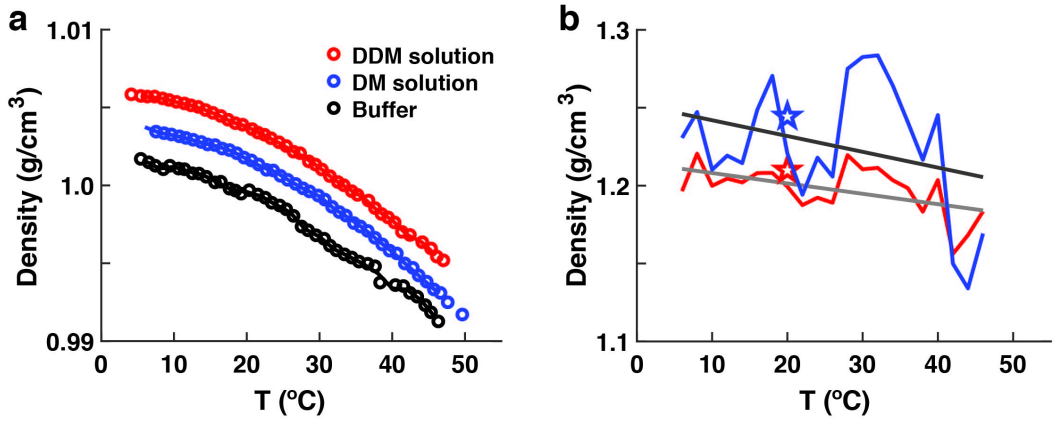

FIGURE 4.8: Temperature dependence of DM and DDM density. a) Density of buffer, DM, and DDM solutions as a function of temperature. Circles are from measurements with a Gay-Lussac pycnometer (see Methods). Lines are spline interpolated values. b) Density of DM and DDM as a function of temperature computed from the interpolated data in panel a) (colored lines; same color code as in panel a). The black and grey lines are linear fits to the DM and DDM data, respectively. The density values at room temperature obtained from the literature [107] are shown as stars for comparison.

water only changes by about 2 % in the temperature range investigated, the temperature dependent change in concentration is a small correction.

Temperature dependence of the electron density of the solvent The electron density of our aqueous buffer with 150 mM NaCl (Table 4.2) was used previously [107] and taken from Ref. 324. Its temperature dependence was again approximated using the temperature dependence of the density of water (Fig. 4.7). Even though the density of the solvent changes only by about 2 % over the temperature range studied, the fact that the two densities in the $(\rho_{\mathrm{det}} - \rho_{\mathrm{sol}})^2$ term in the denominator of Eq. 4.2 are similar means that even small changes are significant.

Temperature dependence of the electron densities and monomeric volumes of the detergents The monomeric volumes V_{mon} and corresponding electron densities ρ_{det} at room temperature were previously computed [107] from the specific densities [325], using the Tanford formula for alkyl chain volumes to adjust for different chain lengths [120]. formula [120] for the alkyl chain volume. To determine the temperature dependence of the density of DM and DDM, we used a Gay-Lussac pycnometer (Brand, Cat. No. 43305). The masses of the pycnometer filled with DM and DDM solutions in buffer and with buffer only were measured as a function of temperature in the range 10°C - 50°C. The buffer was 20 mM phosphate buffer, pH 6.2, with 150 mM NaCl added, i.e. the same buffer that was used for SAXS measurements. Temperature control was achieved through a water bath-thermostat (Biosan, WB-4MS). We limited the measurements to temperatures ≤ 50°C, as for higher temperatures the accuracy decreases according to vendor specifications and bubble formation in the pycnometer made accurate sample handling challenging. From the measured masses $m(T)$, the temperature-dependent densities of the buffer and the DM and DDM solutions were computed by taking into account the (temperature-independent) volume V and mass m_{pyc} of the pycnometer as $\rho(T) = (m(T) - m_{\mathrm{pyc}})V^{-1}$ (Figure 4.8a). From the temperature-dependent densities of the buffer and DM solutions, ρ_{buf} and $\rho_{\mathrm{DM,sol}}$, the temperature-dependent densities of DM ρ_{DM} was computed using the relationship:

$$\rho_{\mathrm{DM}}(T) = \frac{\rho_{\mathrm{buf}}(T)}{\frac{\rho_{\mathrm{buf}}(T)}{\rho_{\mathrm{DM,sol}}(T)}\left(1 + \frac{1}{\chi}\frac{FW_{H_2O}}{FW_{\mathrm{DM}}}\right) - \frac{1}{\chi}\frac{FW_{H_2O}}{FW_{\mathrm{DM}}}} \tag{4.3}$$

where FW_{DM} is the formula weight (i.e. the molecular mass) of DM, FW_{H_2O} the formula weight of water, and χ the molality of the DM solutions, in number of DM molecules per solvent molecule. An analogous expression was used for DDM. We used 18.02, 482.56, and 510.62 Da for the FW of water, DM and DDM, respectively. The resulting densities of DM and DDM are shown in Figure 4.8. The data are relatively noisy, due to the difficulty of measuring very small changes in density with the pycnometer. Nonetheless, the values are consistent, within experimental error, with the density values at room temperature from the literature [107, 325] (Figure 4.8, stars). To obtain a robust estimate of the temperature dependence, and to extrapolate the data beyond the measured temperature range, we applied a linear fit to the measured data (Figure 4.8, black and grey lines). The fitted linear temperature dependence was used to evaluate Eq. 4.2.

Temperature dependence of the critical micelle concentrations The critical micelle concentrations in Table 4.2 are the values used previously [107] and were taken

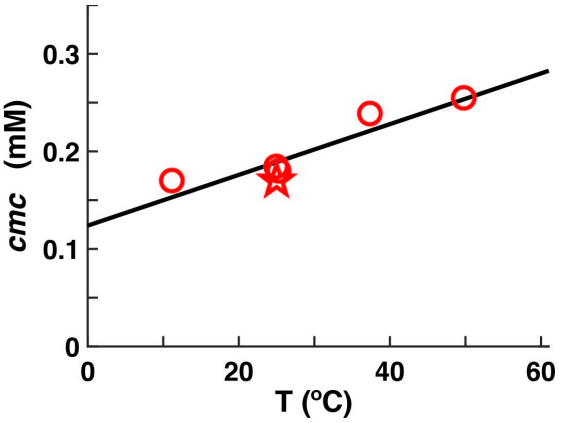

FIGURE 4.9: Temperature dependence of the critical micelle concentration. Critical micelle concentration of DDM as a function of temperature. The circles show the data from Ref. 326. The star is the room temperature value used previously [107]. The black solid line shows the linear fit to the data that was used to evaluate the temperature-dependence of the cmc for both DM and DDM.

from the Anatrace catalogue (https://www.anatrace.com/). The temperature dependence of the *cmc* for DDM was measured spectrofluorometrically by Aoudia *et al.* [326]. The temperature-dependent data by Aoudia *et al.* are well described by a simple linear relationship (Figure 4.9) and we used this linear dependence on temperature to compute the aggregation number by evaluating Equation 4.2 for both DM and DDM.

4.5.1.4 MD simulations and SAXS calculations

MD setup and simulation parameters Structures of single detergent molecules of the n-dodecyl-β-D-maltoside (DDM) and n-decyl-β-D-maltoside (DM) were taken from the CHARMM-GUI web site. citecheng2013charmm. To build the micelle, initial coordinates were generated by placing the detergent molecules in a spherical and uniformly distributed arrangement. The structures were placed into a simulation box of a dodecahedron, keeping a distance of at least 3 nm to the box boundary. The simulation boxes were filled by CHARMM-modified TIP3P water [294, 295]. Water molecules were removed from the hydrophobic cores of the micelles. The energy of each system was minimized with the steepest-decent algorithm. Micelles of different aggregation numbers ($N_{\mathrm{agg}}^{\mathrm{sim}}$) were setup and initially equilibrated at 30°C. For DDM, we set up micelles of 70 to 210 detergent molecules in steps of 5. For DM we we set up micelles of 60 to 110 detergent molecules in steps of 5. The final structures from these equilibration simulations were used as a starting structures for the free production simulations. These simulations were run for another 100 ns, if not stated otherwise, at the temperatures of 10°C, 25°C, 40°C, 55°C, and 70°C. SAXS curves were calculated from 1000 snapshots of the last 50 ns of the simulations, using the explicit-solvent SAXS calculations described previously (see below for details) [90].

Unbiased, free simulations were carried out using the Gromacs simulation software, version 5.0.4 [288]. SAXS curve predictions and SAXS-driven MD simulations (with experiment-derived energetic restraints) were conducted with an in-house modification of Gromacs 4.6. Detergent interactions were modeled with the CHARMM36 lipid force-field [317], version of March 2014, translated into Gromacs [296]. The temperature was controlled at the desired value using velocity rescaling [291] during free simulations ($\tau = 1\,\mathrm{ps}$), and using a stochastic dynamics integrator during SAXS-driven simulations ($\tau = 0.2\,\mathrm{ps}$). The pressure was kept at 1 bar using the Berendsen barostat [159] ($\tau = 5\,\mathrm{ps}$). Long-range electrostatic interactions were calculated using the particle-mesh Ewald method [154, 155]. Dispersive interactions and short-range repulsion were described together by a Lennard-Jones potential with a cutoff at 1.2 nm.

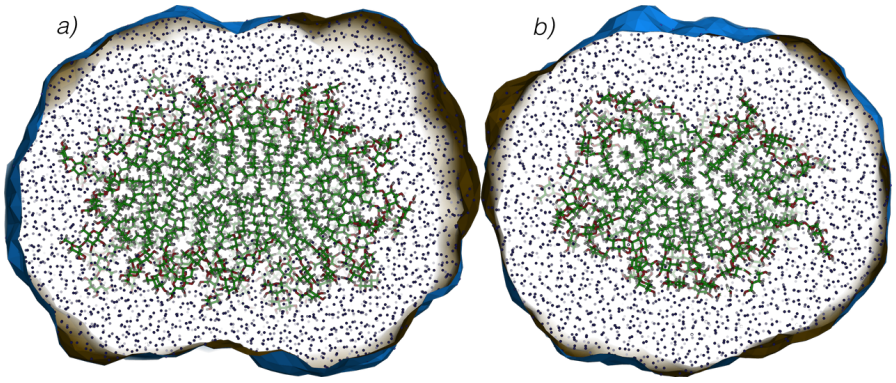

FIGURE 4.10: Spatial envelope around micelles of DDM (a) and DM (b), separating the micelles and the solvation layer from the bulk water. The envelope was constructed at the ∼1.5 nm distance from the micelles surface. Explicit water inside the envelope contributed to the SAXS calculations, thereby accurately accounting for scattering contributions from the hydration layer.

SAXS curve predictions and SAXS-driven MD simulations SAXS curves were computed using the explicit-solvent calculations described previously [90]. Explicit water molecules that contributed to the SAXS calculations were defined by a spatial envelope that enclosed the micelle and the hydration layer (Fig. 4.10, blue surface). The envelope was constructed such that the vertices of the envelope had a distance of at least 0.6 nm from all micelle atoms during the simulation. Due to substantial fluctuation of the micelle, this procedure led to a distance between micelle and envelope of ∼1.5 nm in most simulation frames. To carry out the orientational average (or spherical quadrature), scattering amplitudes were computed for 1000 **q**-vectors per absolute value of the momentum transfer q, which were distributed by the spiral method. Because the density of the applied TIP3P water slightly differs from the experimental density, we corrected the solvent density to $334\,\mathrm{e\,nm^{-3}}$ as described previously [90].

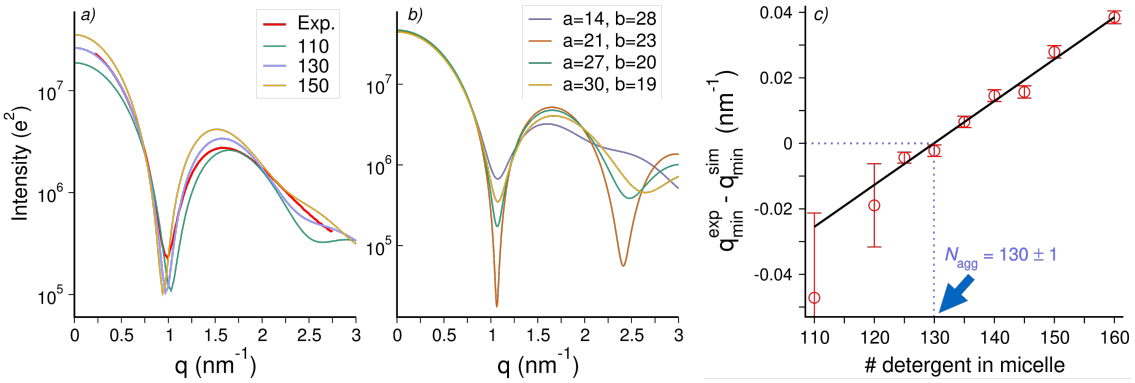

FIGURE 4.11: (a) SAXS curves computed from the simulations of micelles composed of 110, 130 and 150 DDM molecules (for color coding, see legend). Red curve: experimental SAXS curve for DDM micelles at 25°C. Evidently, the location of the minimum at $q \approx 1\,\mathrm{nm}^{-1}$ is shifted to lower q with increasing aggregation number. (b) SAXS curves of DDM micelle computed with the analytic model by Lipfert *et al.* [107] (c) Linear fit of the difference between the q-positions of the experimental and the calculated SAXS curve minima, here shown for DDM at 25°C. The blue arrow indicates the simulated detergent number that leads to the best match of the minimum between simulation and experiment.

4.5.1.5 Extracting the aggregation number from SAXS data using MD simulations

Figure 4.11b shows SAXS curves of DDM micelle computed with the analytic model by Lipfert *et al.* [107], which models a micelle as two-component density (for head groups and tails) in the shape of a spheroid, with one semi-axis a and two semi-axes b. SAXS curves were computed for various prolate ($a > b$) and oblate ($a < b$) micelles with different semi-axes a and b (see Fig. 4.11b, legend), but with constant overall volume (and hence modelling a constant aggregation number) as apparent from the constant forward intensity at $q = 0$. For the analytic model, the head group thicknesses were taken as $t_a = t_a = t_c = 6.06\,\text{Å}$, while the densities of the core, the head groups, and the solvent were taken as $0.227\,\mathrm{e\,Å}^{-3}$, $0.520\,\mathrm{e\,Å}^{-3}$, and $0.334\,\mathrm{e\,Å}^{-3}$, respectively [107]. Evidently, although a and b strongly influenced the SAXS curves, the position of the first minimum at $q \approx 1\,\mathrm{nm}^{-1}$ is well conserved, suggesting that the position of the minimum encodes mainly the volume and much less the shape of the micelle. Indeed, SAXS curves computed from free micelle simulations with increasing number of detergent molecules reveal a systematic left-shift of the minimum with increasing detergent number (Fig. 4.11a). Consequently, we estimated the aggregation number by comparing the position of the SAXS curve minimum between (i) the experimental SAXS curves, $q^{\mathrm{exp}}_{\mathrm{min}}$, and (ii) the SAXS curves computed from unbiased MD simulations, $q^{\mathrm{sim}}_{\mathrm{min}}$, with different number of detergent molecules in the simulated micelle.

The q-positions of the SAXS curve minima q_{min} were extracted by fitting a parabola

TABLE 4.3: Aggregation numbers (N_{agg}) obtained by comparing the position of the SAXS curve minimum between experimental curves and MD-computed curves, statistical error (δN_{agg}), and aggregation number used for production SAXS-driven simulations (Used) at different temperatures.

Temperature	**DDM**			**DM**		
(°C)	$N_{\mathbf{agg}}$	$\delta N_{\mathbf{agg}}$	**Used**	$N_{\mathbf{agg}}$	$\delta N_{\mathbf{agg}}$	**Used**
10°	146.8	2.5	145	91.7	2.4	90
25°	129.9	1	130	84.8	0.2	85
40°	115.8	2.4	115	80	0.7	80
55°	105.7	1.2	105	76.6	2	75
70°	97.8	2.6	100	70.4	1.5	70

to the minimia, $I_{\mathrm{fit}}(q) = a(q - q_{\mathrm{min}})^2 + c$, within a small q-range around the minima, using the Levenberg-Marquardt algorithm. The fitted q-range was $0.1\,\mathrm{nm}^{-1}$ for the experimental SAXS curves, and $0.3\,\mathrm{nm}^{-1}$ and $0.4\,\mathrm{nm}^{-1}$ for the calculated curves for DDM and DM, respectively. All fitted parabolas closely matched the data in the fitted range. This procedure was repeated for SAXS curves computed from simulations with various detergent numbers $N_{\mathrm{agg}}^{\mathrm{sim}}$, yielding a series of computed minima positions, $q_{\mathrm{min}}^{\mathrm{sim}}(N_{\mathrm{agg}}^{\mathrm{sim}})$. Finally, we fitted a straight line to $\Delta q(N_{\mathrm{agg}}^{\mathrm{sim}}) = q_{\mathrm{min}}^{\mathrm{exp}} - q_{\mathrm{min}}^{\mathrm{sim}}(N_{\mathrm{agg}}^{\mathrm{sim}})$ (Fig. 4.11c, black line), and we obtained the experimental aggregation number N_{agg} by extrapolating to $\Delta q = 0$ (Fig. 4.11c, blue arrow). Statistical errors of $\Delta q(N_{\mathrm{agg}}^{\mathrm{sim}})$ were obtained from the Levenberg-Marquardt algorithm and using error propagation. The error of N_{agg} was taken from the Levenberg-Marquardt algorithm. We stress that these errors represent purely statistical errors due to the extraction of the SAXS curve minimum. Putative systematic errors are not included, which could, for instance, appear in case of imperfections of the detergent densities in the simulations; however, because the estimated N_{agg} well agrees with N_{agg} obtained with the model-free approach (see above, and Fig. 4.2), such systematic errors are probably small. The estimated N_{agg} are summarized in Table 4.3.

SAXS-driven MD simulations Final structures obtained by free simulations were used as starting structures for the SAXS-driven simulations. These simulations were run for 300 ns at 25°C, and between 40 ns and 60 ns at all other temperatures. SAXS-derived forces applied to the detergent molecules were calculated from the SAXS-derived potential E_{SAXS} [91]:

$$E_{\mathrm{SAXS}}(\mathbf{R}, t) = \alpha(t) k_c \frac{k_B T}{n_q} \sum_{i=1}^{n_q} \frac{[\langle I_c(\mathbf{R}, q_i)\rangle_{t;\tau} - f I_e(q_i)]^2}{\sigma^2(q_i)} \tag{4.4}$$

where $I_{\mathrm{e}}(q_i)$ denotes the experimental SAXS intensity. f is a fitting constant for the absolute intensity scale, which was adjusted at every step such that E_{SAXS} is minimized. In contrast to previous work [91], no fitting constant for a constant intensity offset was

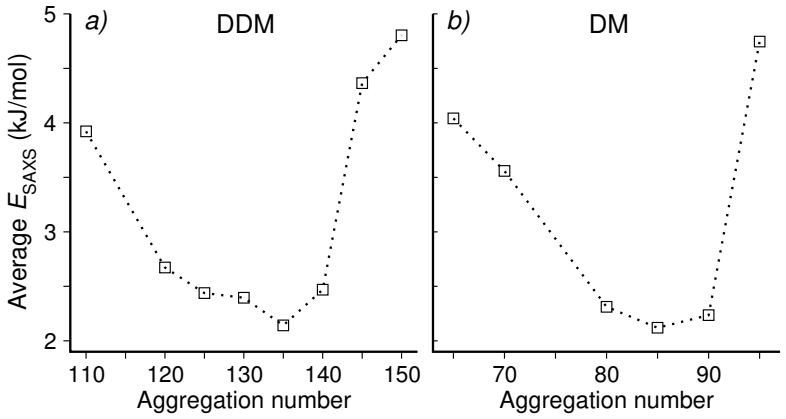

FIGURE 4.12: Average SAXS-derived potential E_{SAXS} (Eq. 4.4) during SAXS-driven simulations of DDM (a) and DM (b). E_{SAXS} micelles at different aggregation number at 25°C, calculated from the last 10 ns of the simulations.

applied because adjusting an offset was not required to obtain quantitative agreement between experiment and simulation. $\langle I_c(\mathbf{R}, q_i)\rangle_{t;\tau}$ is the SAXS intensity computed on-the-fly from the simulation coordinates $\mathbf{R}$. The symbol $\langle\cdot\rangle_{t;\tau}$ denotes the running average at time t, using weights that decay exponentially into the past with a memory time τ. In this work, we used $\tau = 300\,\mathrm{ps}$. As such, the time-averaged $\langle I_c(\mathbf{R}, q_i)\rangle_{t;\tau}$ represents an average over fluctuations that occur on a time scale of a few hundred picoseconds, implying that a fluctuation-averaged SAXS curve is compared with the experimental curve in Eq. 4.4. The overall uncertainty $\sigma(q_i)$ accounts for experimental and statistical calculated errors, as well as for a systematic error that originates from an uncertainty of the buffer density [91]. To estimate the latter, we assumed a relative uncertainty of the solvent density of 0.1%. The experimental errors were modelled as 1% of the experimental intensity. The symbol k_c is a force constant set to 1 in this study, n_{q} is the number of intensity q-points, k_BT the thermal energy, and $\alpha(t)$ is a time-dependent function that allows a gradual introduction of the SAXS-derived potential at the beginning of the simulation ($0 < \alpha(t) \leq 1$). The first 8 ns of SAXS-driven simulations were not used for analysis in order to account for equilibration.

Figure 4.12 presents average E_{SAXS} values obtained from SAXS-driven simulations of DDM and DM micelles with various aggregation numbers at 25°C. Evidently, E_{SAXS} takes small values in DDM and DM simulations if N_{agg} is close to 130 and 85 for DDM and DM, respectively, corresponding to the N_{agg} estimates that we obtained by matching the position of the SAXS curve minimum at $q \approx 1\,\mathrm{nm}^{-1}$ between experiment and unbiased simulation (see Fig. 4.11). In contrast, E_{SAXS} is increased in SAXS-driven simulation with N_{agg} values that strongly deviate from the optimal values, reflecting that micelles with an incorrect N_{agg} cannot be refined to shapes that accurately agree with the experimental data.

To exclude that the SAXS-driven simulations are biased by the initial conformation, we started SAXS-driven simulations for DDM at 25°C from multiple initial frames picked from a free simulation, such that the micelle exhibited different shapes in the initial frames (some more prolate-, some more oblate-like). These SAXS-driven simulations consistently led to rather prolate-like shapes, suggesting that (i) the SAXS-driven simulations were not biased by the initial frames, and (ii) that the simulations do not suffer from sampling problems owing to multiple energetic minima.

The modified Gromacs source code used for SAXS predictions and SAXS-driven MD simulations is available on the authors' website (http://cmb.bio.uni-goettingen.de/).

Electron density calculations Electron densities along the principal axes were computed as an average over SAXS-driven simulations, as follows (Fig. 4.4a-f). For each simulation frame, the mass-weighted principal axes were computed from all micelle atoms. Subsequently, a cylinder of 0.5 nm was aligned along each axis. The electron densities were computed from the atoms of the respective atom type (representing tails, head groups, water, or all atoms) within the cylinder along the respective axis (minor, middle, or major). Here, the hydrocarbon chain was defined as "tail", and all other atoms (including the oxygen bound to the hydrocarbon chain) as "head group".

The length of the three full semi-axes, $a + t_a$, $b + t_b$, and $c + t_c$ (Fig. 4.4g/h), were defined as the distance from the micelle center of mass, where the density of detergent dropped below $120\,\mathrm{e\,nm^{-3}}$, corresponding to approximately half the density of micelle core. Likewise, the semi-axes of the hydrophobic core (Fig. 4.17a/b) were defined by the distance where the density of the tails dropped below $120\,\mathrm{e\,nm^{-3}}$. The error bars were computed by block averaging, using blocks of 4 ns.

The electron densities of lipid membranes of DMPC and POPC were computed from 20 ns and 40 ns of equilibrium simulations, respectively. The membrane simulations contained 128 lipids plus 40 water molecules per lipid. Interactions were described by the CHARMM36 force field and the CHARMM-modified TIP3P model [287, 317]. The simulation of DMPC was taken from a recent study [327]. The POPC setup and MD parameters were identical to the DMPC simulation described previously [327]. The tail densities of the membranes were computed purely from the hydrocarbon lipid tails, starting with the first carbon atom below the ester groups, i.e., the ester and glycerol groups were not considered as part of the "tails".

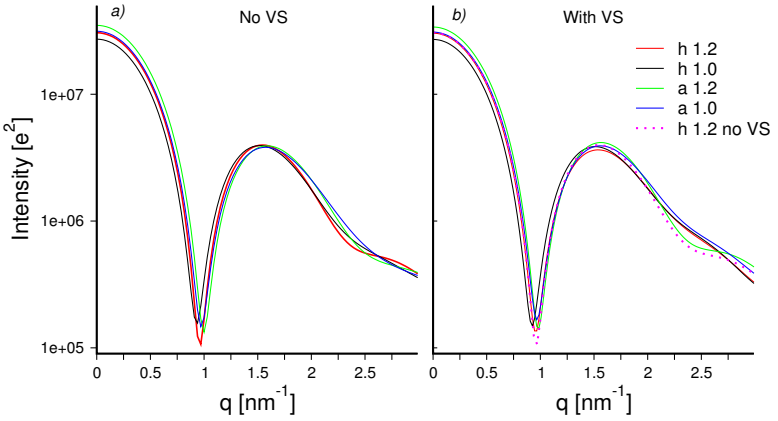

FIGURE 4.13: SAXS curves computed from free simulations of a DDM micelle composed of 140 detergent molecules at 25°C, used to test the influence of various MD parameters on the calculated SAXS curves. (a) Simulations modelling hydrogen atoms as normal atoms, and (b) modelling hydrogen atoms at virtual sites (VS). Numbers 1.2 and 1.0 represents the cutoff distance for Lennard-Jones and short-range Coulomb interactions. Letters "a" and "h" indicate bond constraints of all atoms or purely involving hydrogen atoms, respectively.

4.5.1.6 Computational tests

Influence of MD parameters and force field details on calculated SAXS curves Before running production simulations, we have carefully evaluated the influence of various MD parameters on free micelle simulations: (i) the influence of cutoff distances for non-bonded interactions; (ii) the effect of constraining all bonds instead of purely bonds of hydrogen atoms; (iii) the effect of modelling hydrogen atoms as virtual sites (v-sites), allowing one to increase the integration time step from 2 to 4 fs; (iv) influence of the water model.

All test simulations of a DDM micelle were conducted under the same conditions (aggregation number 140, temperature 25°C). SAXS curves calculated from these test simulations are shown in the Figure 4.13. We found that the cutoff distance as well as the constraints settings may influence the SAXS curves. In contrast, modelling hydrogen atoms as virtual sites had only a small effect on the SAXS curves. Likewise, using the standard TIP3P water model [287] instead of the CHARMM-modified TIP3P model (with Lennard-Jones interactions of hydrogen atoms) did not influence the SAXS curves. We decided to follow the settings that closely resemble the default settings for the CHARMM36 force field: cutoff at 1.2 nm, bond constraints applied purely to hydrogen atoms, and hydrogen atoms *not* modelled as v-sites.

Test of convergence To exclude that the calculated SAXS curves were biased by sampling problems, we conducted two independent simulations of the DDM micelle using increased temperatures and simulated annealing, as follows: (i) 50 ns of simulation at

either 370 K or 420 K; (ii) annealing down to 300 K within 20 ns; and (iii) 100 ns at 300 K. SAXS curves were calculated from the last 50 ns of the two simulations and compared with the previously calculated curves for the same system simulated purely at 300 K. All three calculated SAXS curves were nearly identical, suggesting that our simulations and SAXS calculations were not biased by sampling problems.

In addition, to exclude that the initial detergent conformation influences the computed SAXS curve, the system of 140 DDM detergent molecules was set up following three different procedures: (i) placing the detergent molecules in a spherical and uniformly distributed arrangement; (ii) by building a preassembled micelle with the CHARMM-GUI server [328]; and (iii) via a simulation of micelle aggregation, starting from a random distribution of 140 detergent molecules in a water box. Here, the micelle formed within 200 ns. SAXS curves calculated from the three different setup procedures were nearly identical, suggesting that our approach of generating the initial micelle coordinates does not bias the results.

4.5.2 Additional supplementary figures

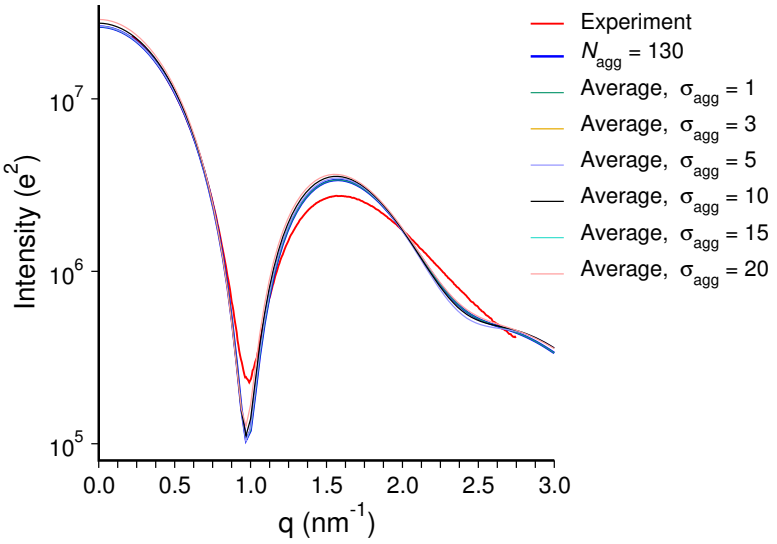

FIGURE 4.14: SAXS curves for DDM at 25°C, demonstrating that modelling a heterogeneous ensemble over various aggregation numbers (polydispersity in micelle size) does *not* improve the agreement between the experimental curve (red) and the curve calculated from free MD simulations. Blue: SAXS curve from free simulation with 130 DDM molecules. All other curves: Average SAXS curves $I_{\mathrm{av}}(q;\sigma_{\mathrm{agg}}) = W^{-1}\sum_n w_n(\sigma_{\mathrm{agg}})I(q;n)$ computed as a weighted average over curves $I(q;n)$ computed from free MD simulations with $n = 110, 120, 125, 130, 135, 140, 145, 150$ or 160 detergent molecules. Here, w_n is the weight taken from a Gaussian distribution with mean 130 and width σ_{agg}, and $W = \sum_n w_n(\sigma_{\mathrm{agg}})$ is the normalization constant.

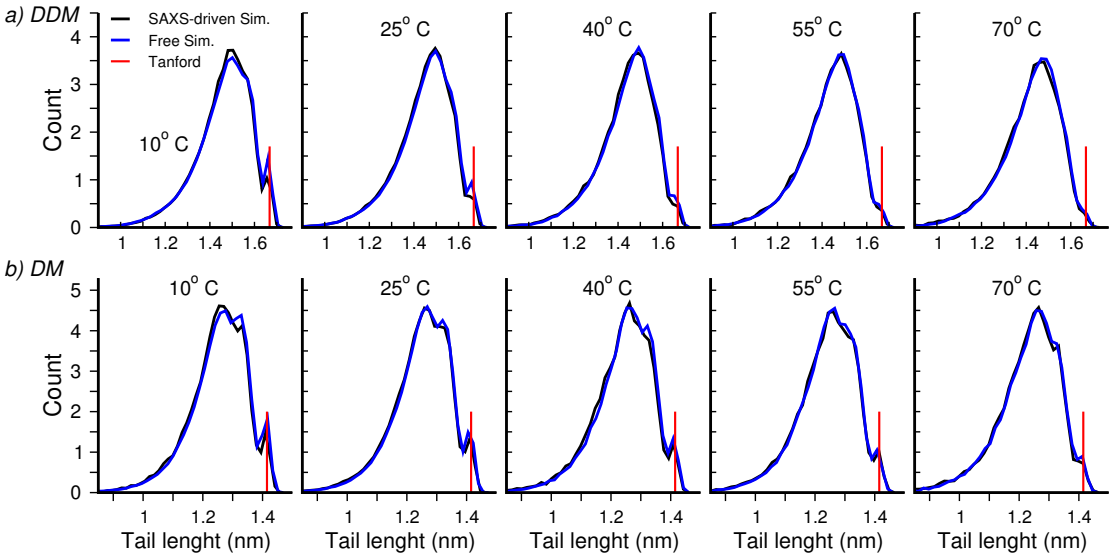

FIGURE 4.15: Distributions of the detergent tail length for DDM (a) and DM (b), computed as the distance between the the first carbon atom (C_1) of the tail (ii) the terminal carbon atom (C_t) of the tail, corrected by 0.21 nm due to (i) the Van-der-Waals radius of the terminal methyl group (0.15 nm) and (ii) half of the bond length between C_1 and the neighboring oxygen atom (0.06 nm), thus following Tanford's defintion [120]. Blue curves: free simulations; black curves: SAXS-driven simulations. Results are shown for simulations at temperatures between 10°C and 70°C, see labels. Red vertical marks indicate the maximum tail length estimated by Tanford's equation, 1.668 nm for DDM tail and 1.415 nm for DM tail. The curves demonstrate that the SAXS-derived restraints hardly influence tail length distribution and, hence, hardly influence the tail structure.

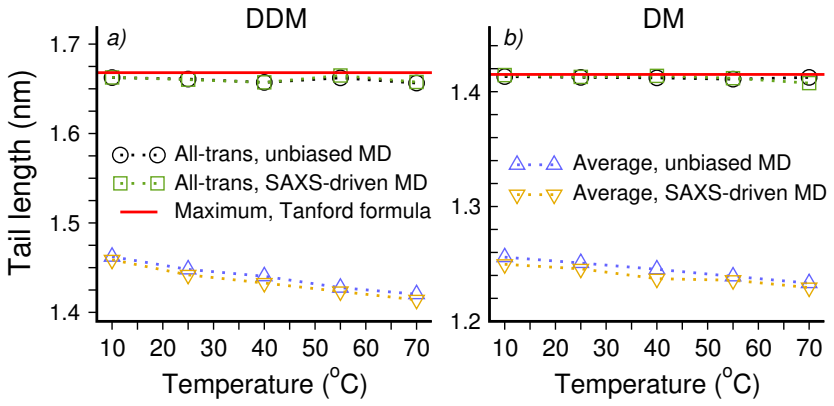

FIGURE 4.16: Average detergent tail length of DDM (a) and DM (b) at various temperatures, taken from the last 50 ns of free, unbiased simulations (blue) or last 10 ns of SAXS-driven simulations (orange). The tail length was computed as the distance between the the first carbon atom (C_1) of the tail (ii) the terminal carbon atom (C_t) of the tail, corrected by 0.21 nm due to (i) the Van-der-Waals radius of the terminal methyl group (0.15 nm) and (ii) half of the bond length between C_1 and the neighboring oxygen atom (0.06 nm), following Tanford's defintion [120]. The data demonstrate that the SAXS-derived restraints influence the average tail length only marginally. Further, the average tail length slightly decreases with increasing temperature, as expected since disordered tails are favoured by entropy. For comparison, black and green symbols show the tail length averaged purely over all-trans configurations of detergent molecules, revealing close agreement with the maximum extension estimated by Tanford's equation (red lines; DDM: 1.668 nm; DM: 1.415 nm).

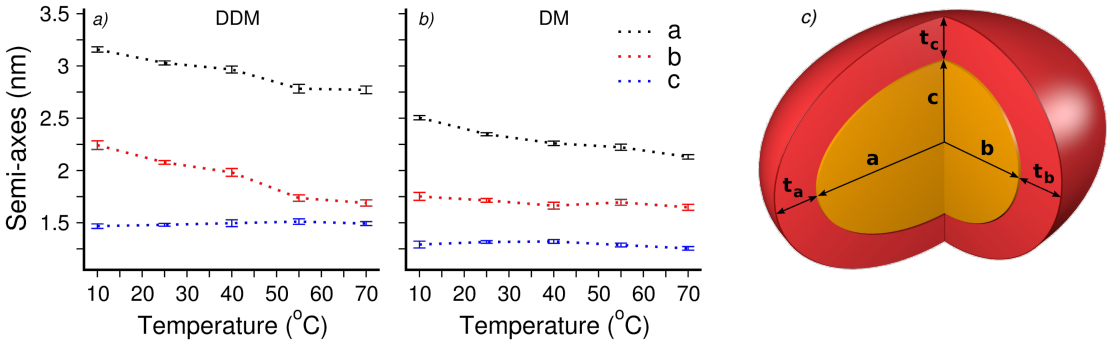

FIGURE 4.17: Length of the semi-axes a, b, and c of the hydrophobic cores along the three principal axes of refined micelles of DDM (a) and DM (b). a, b, and c are plotted versus temperature. Errors were computed by binning analysis. The major and middle semi-axes (b and c) shrink with increasing temperature, whereas the minor semi-axes c are approximately temperature-invariant in both DDM and DM. Error bars denote 1 SEM computed from block averaging. (c) Schematic model of a micelle, illustrating the hydrophobic core with semi-axes a, b, and c (orange), and the head groups with thicknesses t_a, t_b, and t_c (red). The lengths of the full semi-axes including the head groups, $a + t_a$, $b + t_b$, and $c + t_c$, are shown in main text Fig. 4.4.

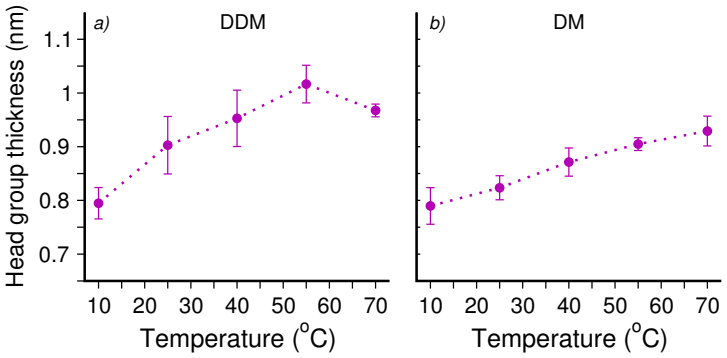

FIGURE 4.18: Thicknesses of head groups in a DDM and DM micelle plotted as a function of temperature. Because the head group thicknesses along the three principal axes were identical within statistical errors, we here averaged the thicknesses over the three principal axes. The thicknesses were computed as FWHM of the Gaussian-like head group electron density distributions (Fig. 4.4b-g, magenta lines). Statistical errors (1 SEM) are slightly increased owing to occasional long-living head group/head group contacts, leading to slower sampling of head group conformation as compared to tail conformations. Overall, the head group thickness slightly increases with temperature, rationalized by increased fluctuations and disorder at higher temperatures.

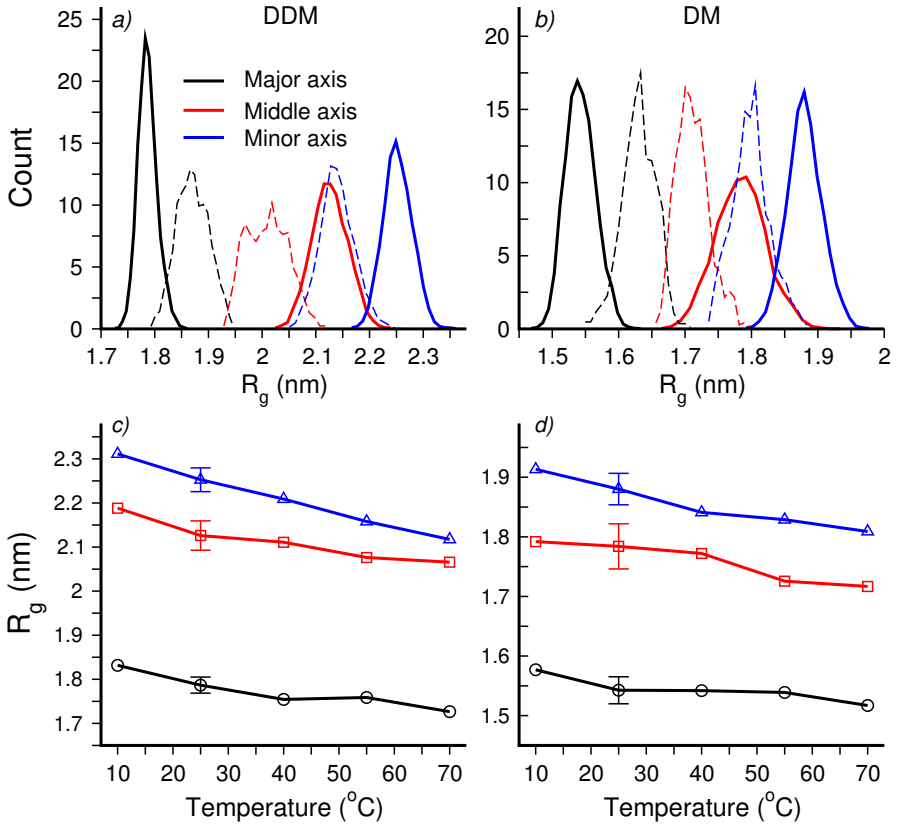

FIGURE 4.19: Distributions of radii of gyration R_g around three principal axes at the 25°C for DDM (a) and DM (b). Radii of gyration are related to the moments of inertia (MOI) $I^{(\mathrm{m})}$ via $R_{\mathrm{g}}^{(\mathrm{m})} = (I^{(\mathrm{m})}/M)^{1/2}$ where $m = 1, 2, 3$ indicates the major, middle or minor principal axis, and M is mass of the micelle. Solid lines represent the results from 300 ns of SAXS-driven simulations. For comparison, dashed lines show distributions from free simulations, demonstrating that micelles in free simulations, which agree with the SAXS data only approximately, were too spherical. (c,d) Temperature dependence of average radii of gyration around the three principal axes, $R_{\mathrm{g}}^{(\mathrm{m})}$, for refined DDM (c) and DM (d) micelles. The error bars at the 25°C indicate the standard deviations of the R_g distributions. Two large and one small $R_{\mathrm{g}}^{(\mathrm{m})}$ (see blue and red versus black line) indicate rather prolate-like than oblate-like micellar shapes for all temperatures.

Chapter 5

SAXS Curves of Detergent Micelles: Effects of Asymmetry, Shape Fluctuations, Disorder, and Atomic Details

THIS chapter, under its title, is deposited in bioRxiv (doi: https://doi.org/10.1101/815266). At the date of thesis submission article is under revision at *The Journal of Physical Chemistry Letters*. Corresponding editor is informed about the article publication in this thesis. Authors of the article are Miloš T. Ivanović, Markus R. Hermann, Maciej Wójcik, Javier Pérez and Jochen S. Hub with the following contributions:
JP, JSH and MTI motivated the project;
MTI performed and analyzed MD simulations and analytic calculations;
JP provided experimental SAXS data;
MRH developed multi-replica code used for the simulations;
MW implemented multi-replica code in the new Gromacs version;
all authors checked and discussed results;
MTI and JSH wrote the article.

5.1 Abstract

Small-angle X-ray scattering (SAXS) is a widely used experimental technique, providing structural and dynamic insight into soft-matter complexes and biomolecules under near-native conditions. However, interpreting the one-dimensional scattering profiles in terms of three-dimensional structures and ensembles remains challenging, partly because it is

poorly understood how structural information is encoded along the measured scattering angle. We combined all-atom SAXS-restrained ensemble simulations, simplified continuum models, and SAXS experiments of a n-dodecyl-β-D-maltoside (DDM) micelle to decipher the effects of model asymmetry, shape fluctuations, atomic disorder, and atomic details on SAXS curves. Upon interpreting the small-angle regime, we find remarkable agreement between (i) a two-component tri-axial ellipsoid model fitted against the data with (ii) a SAXS-refined all-atom ensemble. However, continuum models fail at wider angles, even if they account for shape fluctuations, disorder, and asymmetry of the micelle. We conclude that modelling atomic details is mandatory for explaining the SAXS curves at wider angles.

5.2 Introduction

Detergent micelles are utilized in a wide spectrum of industrial, consumer, and scientific applications [299, 300]. For instance, because the cross-section of detergent micelles resembles lipid membranes [270], micelles are frequently used as lipid membrane mimics for solubilizing membrane proteins, thereby enabling further biophysical and structural studies [309]. For a rational design of such protein-detergent complexes, and for modelling biophysical experiments, understanding of micellar shapes would be highly desirable [105, 106], with respect to both the overall shape and atomic details. However, owing to their intrinsic disorder and pronounced shape fluctuations, obtaining reliable models of micelles remains a major challenge.

Small-angle scattering, either with X-rays (SAXS) or neutrons, is a popular technique providing structural insight into soft-matter systems and biomolecules under near-native conditions [61, 67, 68, 107, 108, 112, 329–333]. However, the interpretation of the one-dimensional scattering profiles in terms of structural models is challenging for several reasons [15]: (i) the information content of the SAXS profile is low and by far insufficient for defining all degrees of freedom of the solute, leading to a significant risk of overfitting the data; (ii) because the SAXS profile reports on the overall electron density contrast of the biomolecule, the data reflects the modulation of the solvent density in the hydration layer, suggesting that the hydration layer must be modelled upon interpreting the data. These challenges prompted the development of methods for the interpretation of SAXS data based on explicit-solvent molecular dynamics (MD) simulations because the simulations (i) add physicochemical information to the low-information SAXS data, thereby reducing the risk of overfitting the data [91], and (ii) MD simulations may naturally account for the hydration layer of the solute [88–90, 101].

However, additional challenges emerge from a lack of understanding on how structural and dynamic information is encoded along the measured scattering angle. An exception is the very low-angle Guinier regime, which provides the radius of gyration of the solute [58]. Further, for SAXS data of detergent micelles, the position of a broad maximum of the intensity curve $I(q)$ was shown to correlate with the headgroup-headgroup distance across the shortest micelle diameter (see Fig. 5.1A/B) [107], while the position of the first minimum was shown to be sensitive to the overall micelle volume [298]. However, the information in the magnitude of the $I(q)$ features, and in particular the information at wider scattering angles is poorly understood, which complicates the interpretation of the data. For instance, given that an experimental SAXS curve differs from a curve computed from a structural model, it is often unclear if such discrepancy originates from experimental problems or from a simplification in the model, such as an assumed model symmetry, neglect of shape fluctuations, or neglect of atomic details.

To investigate the structural information in SAXS curves of soft-matter complexes, we measured the SAXS curve of a n-dodecyl-β-D-maltoside (DDM) detergent micelle [123] up to $q = 6\,\mathrm{nm}^{-1}$, where $q = 4\pi \sin(\theta)/\lambda$ with the X-ray wavelength λ and the scattering angle 2θ. Using a recently developed method for coupling parallel-replica MD simulations to experimental SAXS data [93], we refined a heterogeneous atomic ensemble against the data with commitment to the principle of maximum entropy [207, 208]. Having the atomistic ensemble in agreement with the data as a reference, we deciphered step-by-step the influence of model symmetry, shape fluctuations, disorder, and atomic detail on SAXS curve, by comparing the results from MD simulations with simplified micelle models. In addition, to shed more light on the complementarity of SAXS and MD, we investigated which q-range of the SAXS curve are most critical for improving the agreement of MD simulations with experimental conditions.

5.3 Results and discussion

5.3.1 Deriving ensemble of structures in agreement with experimental data

To obtain the atomic ensemble of the DDM micelle under experimental conditions, we used all-atom MD simulations. First, we used a series of free, unbiased MD simulations to determine the most likely aggregation number N_{agg}, i.e. the number of detergent monomers per micelle. By comparing the position of the pronounced minimum at $q \approx 1\,\mathrm{nm}^{-1}$ between the experimental curve and calculated curves [298], we found that the most likely N_{agg} under the given experimental conditions is 135 (Supporting Material).

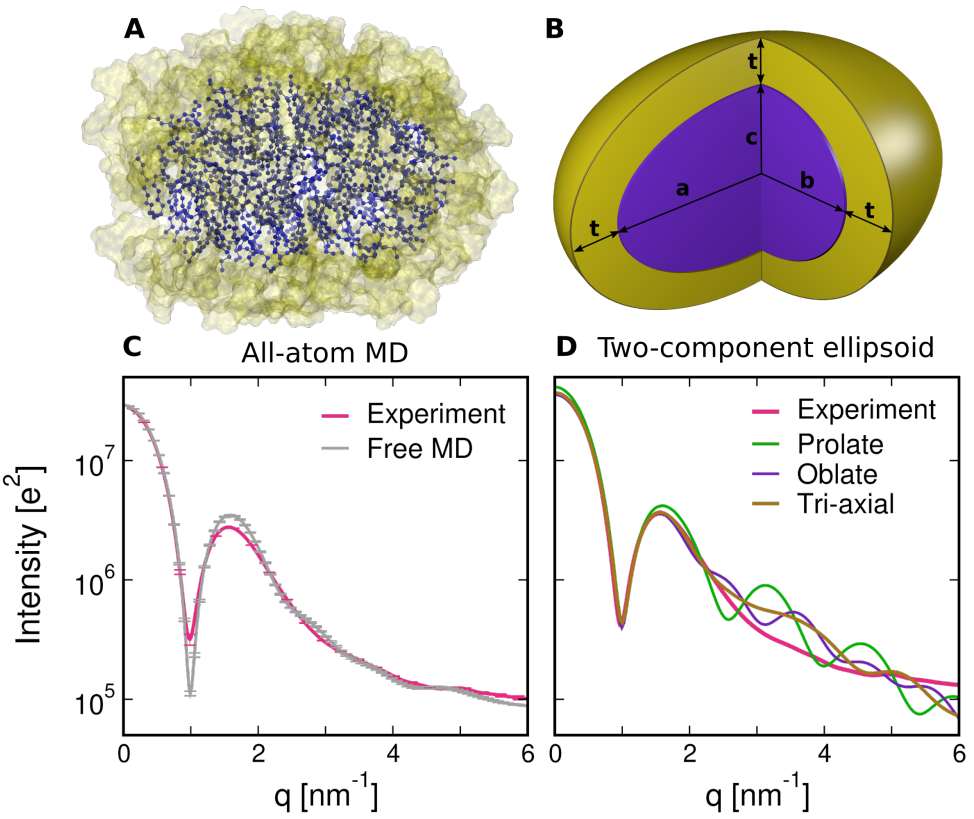

FIGURE 5.1: (A) Snapshot of a DDM micelle from all-atom MD simulation. The hydrophobic core is represented by violet spheres and sticks, and the hydrophilic headgroups are shown as yellow surface. Solvent is omitted for clarity. (B) Two-component ellipsoid model of a micelle. The same color scheme was used as in panel A. a, b and c denote the lengths of the semi-axes of the hydrophobic core. For $b = c > a$, the ellipsoid would be oblate; for $b = c < a$ the ellipsoid would be prolate. Throughout this study, the thickness of the headgroup region was set to $t = 0.55\,\text{nm}$ (SI Methods). (C) Comparison of the experimental curve (red) with curves calculated from free, unbiased MD simulation (gray). For clarity, every 5th error bar is shown. (D) Comparison of the experimental SAXS curve (red) with the best-fitting curves computed from a two-component model: prolate model (green), oblate model (purple), and general tri-axial ellipsoid ($a \neq b \neq c$, brown).

This value is in agreement with previously determined values at similar temperatures [107, 109, 298]. In line with previous findings [298], the SAXS curve calculated from a free simulation of DDM micelle with N_{agg} monomers yield reasonable but not perfect agreement with the experiment, presumably as a consequence of minor imperfections of the applied CHARMM36 force-field [317] (Fig. 5.1C).

Next, to overcome force-field imperfection, we refined the MD ensembles with an energetic restraint against the experimental curve. To apply only a minimal bias we ran several parallel replicas and coupled the replica-averaged SAXS curve to the experiment [93]. This procedure follows Jaynes' maximum entropy principle in the limit of a larger number of replicas [207, 208], and hence enforces that only the ensemble-averaged SAXS curve matches the experiment, but not necessarily the SAXS curve of each simulation

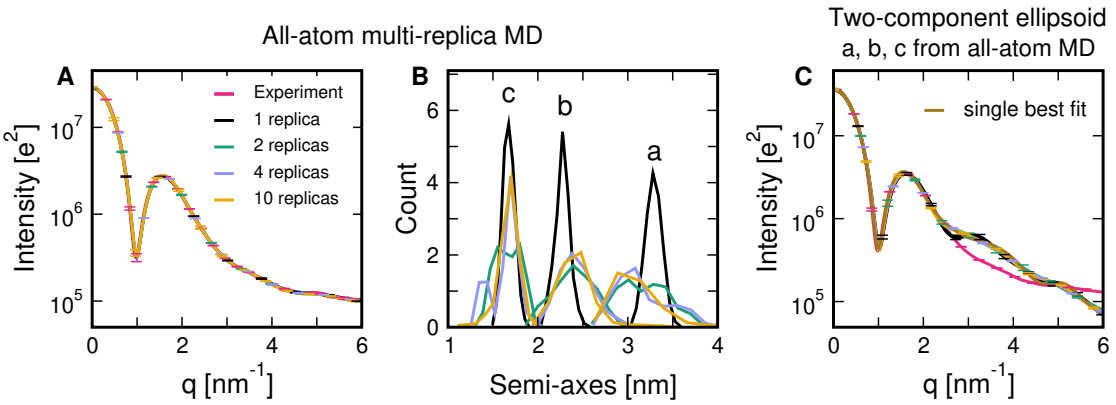

FIGURE 5.2: (A) Favorable agreement between the experimental SAXS curve (red) with curves from multi-replica SAXS-driven simulations refined against the experimental curve, shown for different numbers of parallel replicas (see legend). A few representative errors are shown as 1 SEM computed from independent runs. (B) Histograms of semi-axes a, b, c calculated from multi-replica SAXS-driven simulations obtained from the refined atomic ensembles. (C) Averages of curves calculated from the two-component tri-axial ellipsoid using semi-axes distributions from the refined atomic ensembles in panel (B). Representative errors show 1 SEM, computed via semi-axes distributions from independent MD simulation runs.

TABLE 5.1: Average semi-axes calculated from multi-replica SAXS-driven MD simulations (top rows) and from fitting a single two-component ellipsoid (bottom rows). Errors of MD simulations are given as 1 SEM of averages between independent runs. The prolate solution is unphysical because all semi-axes are significantly larger than the maximum extension of ∼1.67 nm of the hydrophobic tail [120].

All-atom MD # of replicas	**a [nm]**	**b [nm]**	**c [nm]**	**t [nm]**
1	3.29 ± 0.02	2.27 ± 0.02	1.68 ± 0.02	
2	3.17 ± 0.03	2.40 ± 0.04	1.67 ± 0.01	
4	3.16 ± 0.01	2.43 ± 0.02	1.63 ± 0.01	
10	3.17 ± 0.03	2.42 ± 0.02	1.67 ± 0.01	
Two-component model				
prolate (unphysical)	3.39	1.97	1.97	0.57
oblate	2.84	2.84	1.60	0.55
tri-axial	3.20	2.47	1.65	0.55

frame. To test the effect of the number of parallel replicas, we refined ensembles using an increasing number of 1, 2, 4 (Movie S1 - https://vimeo.com/364987188), or 10 replicas, and we computed the SAXS curves and the micelle shape distributions from the refined heterogeneous ensembles (Fig. 5.2A/B, Table 5.1). In this work, the micelle shape was quantified via the semi-axes a, b, c of the hydrophobic core (see Fig. 5.1B), which were obtained from simulation frames every 10 ps via the instantaneous moments of inertia (SI Methods).

As expected, coupling only a single replica yields relatively narrow distributions of a, b,

c, indicating an overly restrained ensemble (Fig. 5.2B, black) and a violation of the maximum entropy principle. Using multi-replica refinement, in contrast, the distributions become wider, in accordance to the maximum entropy principle (Fig. 5.2B, colored). The average semi-axes agree among multiple-replica simulations, and they differ by only $\approx 0.1\,\mathrm{nm}$ from the values obtained from the single-replica simulations. However, irrespective of the number of replicas, all refined ensembles reveal quantitative agreement with the experimental curve (Fig. 5.2A), suggesting that the SAXS curve encodes mainly the information about the mean micelle shape, and much less the information about the heterogeneity of the ensemble. Instead, MD simulations coupled to SAXS data with a minimal bias, as done here, are required to derive both the mean shape and the shape fluctuations.

Notably, in simulations with 10 or 20 parallel replicas, we reproducibly observed an unexpected horseshoe-shaped micelle in one or two replicas, respectively (Fig. 5.6). Although we cannot exclude that DDM micelles occasionally adopt elongated shapes, as reported for other detergent micelles [112, 334, 335], these shapes may indicate a force field limitation and hence may provide a starting point for further refinements of the CHARMM36 parameters (see SI Text).

The atomic ensembles of micelles in agreement with experimental SAXS data derived above provide a reference to study the influence of model symmetry, shape fluctuations, and atomic details on SAXS curves of detergent micelles. To this end, we investigated which parts of the SAXS curve may be explained with a greatly simplified two-component ellipsoidal micelle model, composed of uniform densities for head group and tail regions, as illustrated in Fig. 5.1B. Such models, constrained to oblate ($b = c > a$) or prolate ($b = c < a$) shapes, have well explained experimental curves of DDM micelles up to $\sim 2.7\,\mathrm{nm}^{-1}$ [107, 108, 121, 122], and they were successfully applied to derive the aggregation number of micelles [107, 108, 120]. Critically, fitting such models often leads to two disparate solutions, one prolate and one oblate, that match the data equally well [336], Since the existence of the water droplet or vacuum void in the micelle hydrophobic core would be energetically unfavourable [119], the physically relevant solution was chosen by requesting that at least one semi-axis is shorter than the tail length, thereby avoiding a void at the micelle core. Following this procedure, and in line with previous findings [107], we confirmed that both the oblate and prolate solutions fit the data well at small angles, where only the oblate solution avoids a vacuum void at the micelle core. At wide angles, however, where the experimental curve continuously decays along q, both the oblate and prolate solutions reveal several minima and maxima in sharp contrast to the data (Fig. 5.1D). Moreover, a, b, c determined with the oblate/prolate fits disagree with the values determined using SAXS-driven MD (Table 5.1).

The disagreement at wide angles may potentially be consequence of several simplifications: (i) the two-component oblate/prolate model allows for only two independent semi-axes, while the micelle under the experimental conditions most likely adopts the shape of a less symmetric, general tri-axial ellipsoid, as suggested by our previous study [298]; (ii) by fitting a single model, shape fluctuations of the micelle in solution are ignored; (iii) two-component ellipsoid model assumes sharp core/headgroup and headgroup/water boundaries, while in reality these boundaries are more disordered and smeared out over a range of $\sim$1 nm [298]; (iv) atomic details of both the micelle and the solvent may have a significant effect on the SAXS curve at $q > 2.5\,\mathrm{nm}^{-1}$. In the following, we disentangle the contribution of these potential sources of disagreement between model and experiment, with the aim to obtain an intuitive interpretation of the structural information of the wide-angle data.

5.3.2 Asymmetry

First, to test the influence of the model asymmetry, we dropped the constraint to prolate/oblate shapes and instead fitted a two-component model of a general tri-axial ellipsoid to the experimental curve. The SAXS curve of the two-component tri-axial ellipsoid was computed following Ref. 337 (SI Methods), and the fits carried out by Powell optimization rapidly converged to a well-defined single optimum. The tri-axial ellipsoid fits the data only slightly better as compared to the prolate/oblate model (Fig. 5.1D brown). Specifically, the minima and maxima exhibited by the prolate and oblate models at $q > 2.5\,\mathrm{nm}^{-1}$ are less pronounced in the case of a tri-axial ellipsoid reflecting the reduced symmetry. Nonetheless, the overall agreement to experiment at $q > 2.5\,\mathrm{nm}^{-1}$ remains poor, suggesting that asymmetry is not the key to rationalize the wide-angle data. It is interesting to note, however, that the semi-axes a, b, c of the fitted tri-axial ellipsoid (i) were quite robust, irrespective of the fitted q-range, and (ii) favourably agree with the ensemble-refined MD simulations within 0.5 Å (Table 5.1). This finding suggests that the overall DDM micelle shape is well encoded in the $q < 2.5\,\mathrm{nm}^{-1}$ range of the SAXS curve and may be extracted by fitting a tri-axial two-component model.

5.3.3 Shape fluctuations

Second, to investigate the influence of the shape fluctuations on the SAXS curve, we generalized the two-component tri-axial ellipsoid model to a fluctuating model by averaging over a distribution of semi-axes a, b, c. Here, samples of the semi-axes were taken from snapshots of the multi-replica SAXS-driven MD simulations with 1, 2, 4, or 10 replicas; as such, the samples of semi-axes are compatible with the experimental conditions and,

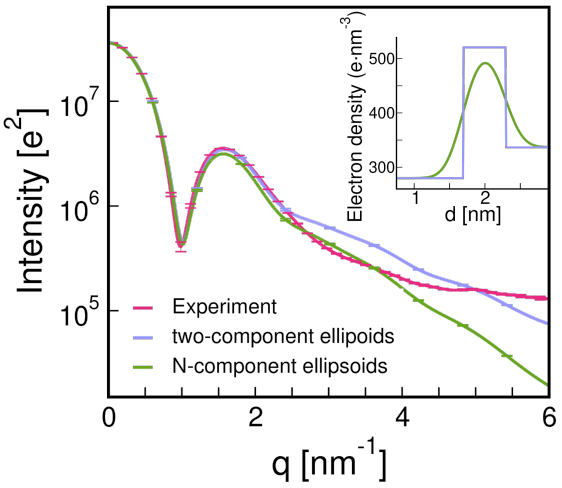

FIGURE 5.3: SAXS curves from experiment (red), the from a fluctuating two-component tri-axial ellipsoid with a piecewise constant electron density profile (blue) and after smoothing the density profile with Gaussian filter (green). The inset shows an example for a piecewise constant and the smoothed electron density profiles along the minor micelle axis. The same samples of semi-axes were used as for the calculations for Fig. 5.2C. For clarity, only curves calculated from sets of semi-axes taken from the four-replica simulations are shown. Curves calculated with semi-axes from ten-replica simulations are nearly identical and shown in Fig. 5.8.

given a sufficient number of parallel replicas, reflect physically realistic shape fluctuations. However, including shape fluctuations into the two-component tri-axial model only marginally improves the agreement to experiment, as compared to a single best-fitting model (Fig. 5.2C). Namely, although the spurious bumps $q > 2.5\,\mathrm{nm}^{-1}$ are partly smeared out, the calculated curves decay too rapidly with q as compared to experiment. This finding further confirms that the SAXS curve of DDM micelle is mainly given by the average micelle shape, whereas shape fluctuations have only the minor impact on the SAXS curve. Notably, this finding is not trivial because SAXS curves of other disordered ensembles, such as ensembles of intrinsically disordered proteins (IDPs), could not be explained by a single average structure [24, 338]. This difference is likely a consequence of the moderate magnitude the micelle fluctuations as compared to the large fluctuations carried out by many IDPs. Further, the fact that greatly different distributions of a, b, c (from a single set up to heterogeneous distributions) lead to nearly identical SAXS curves up to $6\,\mathrm{nm}^{-1}$ (Fig. 5.2C) implies that micelle fluctuations can not be derived from a SAXS experiment alone.

5.3.4 Disorder

Third, to investigate the effect of disorder at the core-headgroup and headgroup–water interfaces, we smeared out the density contrast along the radial direction with a simple Gaussian filter, providing a more realistic density profile (Fig. 5.3, inset). The SAXS

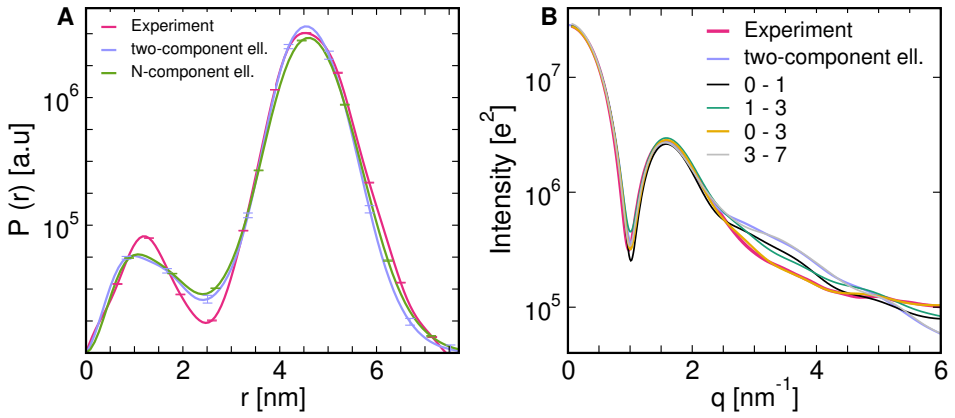

FIGURE 5.4: (A) $P(r)$ functions calculated from the SAXS curves shown in (A) using the GNOM software [339]. Every 10th error bar is shown, for clarity. $P(r)$ curves calculated from the SAXS curves of oblate, prolate, or tri-axial fits and from the MD simulations are shown in Fig. 5.9 (B) To test the reason for discrepancy of $P(r)$ functions calculated from experimental curve and shape-fluctuating two-component ellipsoids, we replaced parts of $P(r)$ function of the shape-fluctuating two-component ellipsoids with the experimental $P(r)$. Replaced r-ranges are shown in legend (units are all in nm).

curve were computed analytically by generalizing the two-component model to a N-component model, following Ref. 337 (SI Methods). In line with the previous paragraph, shape fluctuations were included by averaging over semi-axes sampled from frames of the four-replica SAXS-driven MD simulation. However, upon smearing out the electron density profile, the agreement with the experimental SAXS curve becomes even worse, as apparent from an even more rapid decay of the calculated SAXS curve at $q > 2.5\,\mathrm{nm}^{-1}$ (Fig. 5.3, green and red). The rapid decay of the SAXS curve of the N-component model (Fig. 5.3, green) may be rationalized by the loss of density–density correlations as a consequence a smeared out density.

Taken together, the analysis demonstrates that accounting for shape asymmetry, fluctuations, and disorder may improve the agreement with experiment at wide angles only in a qualitative manner. Specifically, they largely remove the marked maxima and minima exhibited by the SAXS curve of the two-component oblate/prolate models, which indicated a spuriously high degree of symmetry and order in the oblate/prolate models (compare Fig. 5.1D with Fig. 5.3). To explain the wide-angle data quantitatively, we propose that, in addition, account for atomic details of both the micelle and the solvent are required, as captured by the MD simulations.

5.3.5 Atomic details

To decipher the role of atomic details on the SAXS curve at wider larger angles, we calculated the pair-distance distribution function $P(r)$ from the calculated and from the experimental SAXS curves using the GNOM software [339] (Fig. 5.4A and Fig.

5.9). The $P(r)$ function is sensitive to density distribution of the micelle and provides more intuitive, real-space structural information. Evidently, the $P(r)$ function from the ensemble-refined MD simulations favourably agree with the experiment (Fig. 5.9), as expected from the agreement of the SAXS curve. In contrast, the $P(r)$ function obtained from the two-component models differ from the experiment, most prominently at small distances r between 0.9 nm and 3 nm. Namely, the features in the experimental or MD-based $P(r)$ are smeared out by the two-component models, indicating a lack of structure at short-range, molecular distances (Fig. 5.4A). In addition, the $P(r)$ from the prolate/oblate fits strongly differ at larger distances from the experimental $P(r)$, reflecting a too high degree of symmetry as compared to the experiment (Fig. 5.9). To test how the lack of short-range structure propagates into the SAXS curve, we overwrote the $P(r)$ from the shape-fluctuating two-component model with the experimental $P(r)$ in different r-intervals, and subsequently back-calculated the SAXS curve via [340] $I(q) = \int P(r)\frac{\sin(qr)}{qr}\mathrm{d}r$, using the pddffit module of the ATSAS software (Fig. 5.4B) [220]. Although there is no simple one-by-one relation between specific r-regions of $P(r)$ with q-regions of $I(q)$, this analysis confirms that the short-range order ($0\,\mathrm{nm} < r < 3\,\mathrm{nm}$) has a strong effect on the SAXS curve at wider angles ($q > 2.5\,\mathrm{nm}^{-1}$). Remarkebly, by replacing the region $r = 0\,\mathrm{nm}$ to $r = 3\,\mathrm{nm}$ of the the shape-fluctuating two-component model with the experimental $P(r)$, we obtained very good agreement between experimental and calculated curve, suggesting that the discrepancy between experimental and SAXS curve calculated from fluctuating two-component ellipsoids is mainly recorded in the $r = 0\,\mathrm{nm}$ to $r = 3\,\mathrm{nm}$ region of the $P(r)$.

5.3.6 Testing the force field accuracy at different q-ranges

To further investigate the structural information in different q-regions, and to test which part of the q-region of the SAXS curve plays the most important role in overcoming force-field imperfections during MD simulations, we performed the series of multi-replica SAXS-driven MD simulations using only specific q-intervals of the experimental curve as a target (Table 5.2 and Fig. 5.10). Table 5.2 lists the semi-axes of the refined micelles. The difference between the SAXS curves from experiment and from the refined MD ensembles were quantified with a non-weighted χ^2 measure on a log scale, denoted χ^2_{ln}. Computed SAXS curves (Fig. 5.10) as well as the calculated semi-axes show that: (i) applying the $0\,\mathrm{nm}^{-1} \leq q \leq 3\,\mathrm{nm}^{-1}$ region leads to results that are very close to results obtained with using the whole experimental curve. If instead even smaller-angle regions are applied ($\leq 2\,\mathrm{nm}^{-1}$ or $\leq 1\,\mathrm{nm}^{-1}$), the agreement with the whole experimental curve still greatly improve. This finding demonstrates that the micelle shape is mainly encoded

TABLE 5.2: Semi-axes calculated from four- and ten-replica SAXS-driven simulations, using different q-intervals of the experimental curve as target curve. The deviation between calculated and experimental SAXS curves was quantified by a non-weighted χ^2 on a logarithmic scale defined via $\chi^2_{\ln} = N^{-1} \sum_{i=1}^{N} [\ln(I_{c,i}) - \ln(I_{exp,i})]^2$, where $I_{c,i}$ and $I_{exp,i}$ denote the calculated and experimental SAXS intensities at the data point i, and N is the number of data points. Each four-replica (ten-replica) simulation was carried out for at least 100 ns (70 ns) per replica. Errors of a, b, c were computed using block averaging with 4 ns blocks, and errors were typically smaller than 0.02 nm. For reference, values from a free MD simulation are: $a = 2.80$ nm, $b = 2.44$ nm, $c = 1.77$ nm and $\chi^2_{\ln} \cdot 10^3 = 28.1$

	4 replicas				**10 replicas**			
q-range	**a [nm]**	**b [nm]**	**c [nm]**	$\chi^2_{\ln} \cdot 10^3$	**a [nm]**	**b [nm]**	**c [nm]**	$\chi^2_{\ln} \cdot 10^3$
0 - 6	3.15	2.44	1.63	0.5	3.17	2.42	1.67	0.6
0 - 4	3.17	2.43	1.65	0.8	3.19	2.39	1.68	0.8
0 - 3	3.18	2.40	1.64	1.0	3.22	2.36	1.67	1.1
0 - 2	3.13	2.48	1.61	5.3	3.15	2.43	1.62	4.7
0 - 1	3.17	2.26	1.70	6.3	3.19	2.25	1.72	7.1
0 - 0.5	2.85	2.48	1.73	23.2	2.81	2.47	1.76	28.9
1 - 2	3.09	2.45	1.62	5.3	3.13	2.44	1.62	3.9
2 - 4	3.13	2.40	1.69	1.8	3.05	2.41	1.68	3.0
2 - 3	3.08	2.42	1.68	2.5	3.05	2.43	1.67	4.3
3 - 4	2.96	2.38	1.75	18.5	2.89	2.44	1.73	21.5
3 - 6	2.94	2.41	1.73	24.0	2.93	2.42	1.73	18.9
4 - 6	2.86	2.41	1.78	26.9	2.87	2.42	1.76	27.8

in the $0\,\mathrm{nm}^{-1} \leq q \leq 3\,\mathrm{nm}^{-1}$ region, as already indicated by fitting two-component triaxial ellipsoid (see above). However, the Guinier region alone ($\leq 0.5\,\mathrm{nm}^{-1}$) is insufficient for obtaining good agreement with the experiment; (ii) applying various intervals of the $q > 3\,\mathrm{nm}^{-1}$ range leads only to a minor improvement compared to the free MD simulation. Taken together, these findings suggests that the MD force field already provides accurate description of the short-range order mainly encoded by the $q > 3\,\mathrm{nm}^{-1}$ range, hence adding experimental data hardly improves the simulation. However, the force field alone has problems with defining the overall shape, as encoded by the $q < 3\,\mathrm{nm}^{-1}$; hence experimental data in this range greatly improves the simulation.

5.3.7 Conclusions

To conclude, we obtained an heterogeneous atomic ensemble of a DDM detergent micelle by coupling a set of parallel-replica MD simulations to an experimental SAXS curve. Because the multi-replica ensemble refinement method applies only a minimal bias, as requested by Jaynes' maximum entropy principle, the shape fluctuations of the free simulations were maintained. We found that scattering data at small angles ($q < 3\,\mathrm{nm}^{-1}$) may guide the simulation into quantitative agreement with experiment, whereas scattering data at wider angles is matched already by free simulations with reasonable accuracy.

This suggests that the force field is capable of reproducing the short-range structure of the micelle at atomic and molecular scales, but experimental data is needed to obtain the correct overall shape. According to the refined ensemble, the DDM micelle at 15°C adopts on average the shape of a general tri-axial ellipsoid. The major and middle semi-axes fluctuate by ∼20% and the minor semi-axis by 5–10%.

The refined atomic ensemble provided a reference to test whether the fitting of simplified analytic models to the data may provide physically correct micellar shapes. Remarkably, by fitting a two-component general tri-axial ellipsoid to the data, we obtained a micellar shape in quantitative agreement with the multi-replica ensemble refinement, suggesting that (i) the two-component tri-axial model was not overfitted, and (ii) that the SAXS curve up to $q \approx 2.5\,\mathrm{nm}^{-1}$, in the case of DDM micelles, contains sufficient information for defining three independent semi-axis as well as the headgroup thickness. Upon restricting the fit to prolate or oblate shapes, however, we obtained different semi-axes, and the long-range structure quantified by the $P(r)$ function disagreed with the experiment. Further, by increasing the complexity of the analytic micelle model step by step, we analyzed the role of model asymmetry, shape fluctuations, and disorder on the SAXS curve of the micelle. We found that these features partly improve the agreement with the experiment at wider angles, but, even when combined, they are insufficient for obtaining quantitative agreement. Taken together, atomic and molecular details, as naturally included in the MD simulation, are required to quantitatively explain the SAXS curve over the entire q-range.

Acknowledgment M.T.I., M.R.H. and J.S.H. were supported by the Deutsche Forschungsgemeinschaft (DFG) (HU 1971/3-1 and HU 1971/4-1). M.T.I. was additionally supported by the DFG via SFB 803/A12.

5.4 Supplementary material

5.4.1 Methods

SAXS experiment The SAXS curve of the DDM micelle was taken from a recent study [123]. Details of sample preparation and data collection are provided in Ref. 123. For the purpose of this work, the experimental SAXS curve was smoothed using the default options of GNOM software [339].

SAXS curves for tri-axial ellipsoid models Detergent micelles have previously been modeled as two-component ellipsoids of revolution, i.e. ellipsoids with two density contrasts for tails and headgroups, and with only two independent semi-axes (prolate/oblate ellipsoids) [107]. These calculations can be generalised to a general tri-axial ellipsoid, and/or to models composed of N instead of two concentric shells. Here, the N shells are constructed by superimposing N ellipsoids. Each of the N ellipsoids is defined by three semi-axes a_i, b_i and c_i, the volume $V_i = \frac{4}{3}\pi a_i b_i c_i$, and by the solvent-subtracted electron density ρ_i, where $i = 1, \ldots, N$. The semi-axes are denoted in decreasing order, for instance $a_1 > a_2 > \ldots > a_N$. Following the Eqs. 55, 57, and 62 of Ref. 337, the scattering intensity of the N-component tri-axial ellipsoid is given by:

$$I(q, a_1, \ldots, a_N, b_1, \ldots, b_N, c_1, \ldots, c_N) = \frac{2}{\pi} \int_0^{\pi/2} \int_0^{\pi/2} F_3^2(q, r_1, \ldots, r_N) \sin\alpha \, \mathrm{d}\alpha \, \mathrm{d}\beta \qquad (5.1)$$

where

$$r_i(a_i, b_i, c_i, \alpha, \beta) = \left[(a_i^2 \sin^2\beta + b_i^2 \cos^2\beta) \sin^2\alpha + c_i^2 \cos^2\alpha \right]^{1/2}. \qquad (5.2)$$

Here,

$$F_3(q, r_1, \ldots, r_N) = \rho_1 V_1 F_1(q, r_1) + \sum_{i=2}^{i=N} (\rho_i - \rho_{i-1}) V_i F_1(q, r_i) \qquad (5.3)$$

is the scattering amplitude of an N-shell ellipsoid, and

$$F_1(q, r_i) = \frac{3\left[\sin(qr_i) - qr_i \cos(qr_i)\right]}{(qr_i)^3}. \qquad (5.4)$$

For the two-component tri-axial ellipsoid (with only two shells), Eq. 5.3 simplifies to

$$\tilde{F_3} = \rho_1 V_1 F_1[q, r(a+t, b+t, c+t, \alpha, \beta)] + (\rho_2 - \rho_1) V_2 F_1[q, r(a, b, c, \alpha, \beta)], \qquad (5.5)$$

Here, the hydrophobic core is described as a tri-axial ellipsoid with semi-axes a, b, c (Fig. 5.1B, purple region) with electron density ρ_{core}. The headgroup region has a constant thickness t along the three axes (Fig. 5.1B, yellow region) and electron density

ρ_{hg}. Accordingly, the parameters in Eq. 5.5 are given by $\rho_1 = \rho_{\mathrm{hg}} - \rho_{\mathrm{sol}}$, $\rho_2 = \rho_{\mathrm{core}} - \rho_{\mathrm{sol}}$, $V_1 = \frac{4}{3}\pi(a+t)(b+t)(c+t)$, and $V_2 = \frac{4}{3}\pi abc$. We validated numerically and analytically that the mathematical expressions shown above reduce to the expression for a two-component ellipsoid of revolution (oblate/prolate) used by Lipfert *et al* [107].

The computed scattering intensity is sensitive to the electron densities, suggesting that reasonably accurate electron density estimates are required. For the density of the hydrophobic core and the headgroup region at the temperature of 15°C we used $\rho_{\mathrm{core}} = 279.8\,\mathrm{e\,nm^{-3}}$ and $\rho_{\mathrm{hg}} = 520.5\,\mathrm{e\,nm^{-3}}$. These values were determined taking into the account that: (i) electron densities of the core and headgroup region at 25°C are $277\,\mathrm{e\,nm^{-3}}$ and $520\,\mathrm{e\,nm^{-3}}$, respectively [107]; (ii) the density of the DDM detergent at 15°C is increased by 0.56% compared to 25°C [298]; (iii) the density of the alkyl tails at 15°C is increased by 1.02% compared to 25°C, as estimated by the temperature dependence of the density of alkanes with similar chain length [341]. The solvent density was set to $\rho_{\mathrm{sol}} = 336.7\,\mathrm{e\,nm^{-3}}$ to match the electron density of the 150 mM NaCl aqueous solution at the temperature of 15°C.

The thickness of the headgroup region, t, was determined once to 0.55 nm by fitting the general tri-axial ellipsoid to the data. This value is close to the value of $t = 0.6$ nm, as previously determined by fitting a ellipsoid of revolution (prolate/oblate) at the temperature of 25°C [107]. It was shown previously that modulating t in the range of 0.6 nm to 0.63 nm or using different headgroup thicknesses hardly influence the fits to the experimental data in the case of two-axial ellipsoids [107]. Likewise, we found here that small modulations of t along the three axes hardly influences the fits in the case of a general tri-axial ellipsoid and hardly influences the results from modeling shape fluctuation. Therefore, in all follow-up calculations with the ellipsoid model, the value of $t = 0.55$ nm was used.

The smeared out electron density was modeled using an N-component tri-axial ellipsoid, where $N = 200$ was used. The density profiles along the three axes were obtained by convoluting the piecewise constant density (corresponding to a two-component ellipsoid) with a Gaussian filter with $\sigma = 0.2$ nm (inset in Fig. 5.3 A). Using this filter, the smoothed density profiles were qualitatively similar to the density profiles determined from MD simulations [298]. The largest semi-axes of the N-component tri-axial ellipsoid was chosen such that Gaussian tails up to 3σ of the smoothed density were taken into account.

5.4.1.1 MD simulations and SAXS calculations

MD setup and simulation parameters Unbiased, free simulations were carried out similar to previous work [298]. In short, detergent, water and ion interactions were modeled using CHARMM36 lipid force-field [317] and CHARMM-modified TIP3P water [294]. Free simulations were carried out with Gromacs 2018.3 [288]. If not stated otherwise, the micelle was solvated in a 150 mM NaCl aqueous solution [236]. Likewise, a 150 mM NaCl solution was simulated as pure-solvent system. The temperature was controlled at 15°C to match the experimental conditions using velocity rescaling [291] ($\tau = 1$ ps). The pressure was controlled at 1 bar using the Berendsen barostat [159] ($\tau = 5$ ps). Electrostatic interactions were calculated using the particle-mesh Ewald method [154, 155]. Dispersive interactions and short-range repulsion were described together by a Lennard-Jones potential with a cutoff at 1.2 nm.. Length of the free simulations, with and without added salt were 500 ns. The convergence of the free simulation ensemble was validated by comparing SAXS curves computed from 50-nanosecond blocks of the trajectory. The first 50 ns of all free simulations were removed for equilibration.

Explicit-solvent SAXS curve predictions SAXS curves were calculated using explicit-solvent SAXS predictions described previously [90]. Explicit solvent atoms contributing to the SAXS curve were defined by a spatial envelope. Here, the envelope was constructed at a distance of at least 1 nm from all detergent atoms in all frames of an equilibrium simulation. Because the micelle heavily fluctuates, this procedure led to the distance of ∼2 nm between micelle and envelope in most frames. The same envelope was used for all SAXS-driven simulations. The SAXS curve was calculated using the positions of atoms inside the envelope each 10 ps. Scattering amplitudes were computed using 1200 **q**-vectors per q-point, which were distributed by the spiral method. Because the TIP3P solvent density differs from the experimental value, we corrected the solvent density to 335.7 e nm^{-3}, corresponding to a 150mM NaCl solution, following the procedure described previously [90].

SAXS-driven MD simulations The initial configurations of SAXS-driven simulations was taken from free simulations. In the case of parallel-replica simulations, frames for the replicas were taken from free MD snapshots at 5-nanosecond intervals. Apart from using 150 mM NaCl aqueous solution instead of pure-water solution, single-replica SAXS-driven simulations [91] were performed as described previously [298]. Details of the recently developed multi-replica SAXS-driven simulations, following the principle of maximum entropy, are described in Ref. 93. In all single-replica and parallel-replica simulations, the temperature was controlled at 15°C using a stochastic dynamics integrator

($\tau = 0.2$ ps). To validate that the refined ensembles are reproducible, we ran multiple independent simulations for each setup of 1, 2, 4, 10 or 20 replicas. The number of independent runs, the applied force constant used to couple the simulation to the data, and total simulation times are listed in Table 5.3. Further, we excluded that the memory time τ used for on-the-fly averaging of the SAXS curve [91] has a significant effect on the refined ensembles. We found that the choice for τ between 100 ps and 500 ps did not influence the calculated SAXS curve or semi-axes. Here, for all the production runs, we used $\tau = 200$ ps. During the first 5 ns of SAXS-driven simulations, the SAXS-derived forces were gradually switched on, and the first 8 ns of all SAXS-driven simulations were omitted from the analysis for equilibration. In contrast to previous work [298], the overall scale f and a constant offset c of the experimental curve were marginalized out on-the-fly during the SAXS-driven simulations. As shown previously [92], marginalizing out f and c (in a Bayesian sense) is equivalent to adjusting f and c at each step to the value that leads to the smallest biasing energy. Without adjusting the constant c, we did not achieve good agreement between calculated and experimental curve at wider angles, possibly owning to a small buffer subtraction uncertainty. Critically, the adjusted value of f and c values were nearly identical among all SAXS-driven MD simulations, suggesting that f and c were not overfitted. After the SAXS-driven simulations had finished, we computed the SAXS curve from the entire refined ensemble. To compare this ensemble-averaged calculated curve with the experimental curve, a constant set of f and c was applied throughout this study ($f = 7.62433 \times 10^7$, $c = -22883e^2$), motivated from the fact that the adjusted f and c were highly similar in all SAXS-driven simulations.

TABLE 5.3: Number of independent simulations N_{runs}, force constant k_c and simulation time per replica T_{sim} (after removing 8 ns for equlibration) for a runs with 1, 2, 4 10 and 20 replicas

# of replicas	k_c	N_{runs}	T_{sim}
1	12	20	450
2	5	8	460
4	3	4	290
10	1 and 0.5	6	340
20	0.5	4	190

Calculations of semi-axes and shape averages. In our previous work, semi-axes of the hydrophobic core (a, b, c) were calculated from the density profiles of the hydrophobic core [298]. This procedure requires averaging over a few nanoseconds for obtaining reasonable estimates of a, b, c. However, for obtaining distributions along a, b, c as derived here, instantaneous values for a, b, c are required. Hence, we estimated the instantaneous a, b, c from the instantaneous moments of inertia (MOI) of the micelle, while assuming an ellipsoidal shape. We found that a, b, c calculated from the MOI are

systematically larger by $\sim 0.2\,\mathrm{nm}$ as compared to the values calculated from the density profiles reported previously [298].

The MOI of a tri-axial ellipsoid of mass m and semi-axes a, b and c are given as:

$$\begin{aligned} I_a &= \frac{m}{5}(b^2 + c^2) \\ I_b &= \frac{m}{5}(a^2 + c^2) \\ I_c &= \frac{m}{5}(a^2 + b^2) \end{aligned} \tag{5.6}$$

From the MOI, semi-axes of the tri-axial ellipsoid can be calculated as:

$$\begin{aligned} a &= \left(\frac{5}{2m}(I_b + I_c - I_a)\right)^{1/2} \\ b &= \left(\frac{5}{2m}(I_a + I_c - I_b)\right)^{1/2} \\ c &= \left(\frac{5}{2m}(I_a + I_b - I_c)\right)^{1/2} \end{aligned} \tag{5.7}$$

To compute semi-axes from MD simulations, we first computed the three MOI of the micelle core every 10 ps using the Gromacs tool `gmx principal`. Assuming that the density of the core can be approximated by the ellipsoid of uniform density, the semi-axes were computed using Eqs. 5.7, and the distributions were computed from the refined ensemble.

Modeling shape fluctuations. For modeling shape fluctuations, samples of a, b, c, were drawn from the respective distributions. Here, we drew the samples from the one-dimensional distributions of a, b, c, thereby neglecting correlations between the semi-axes. To ensure that all semi-axes samples model a micelle with a constant volume, we normalized each drawn set (a, b, c) via

$$\begin{aligned} \tilde{a_i} &= a_i \left(\frac{\langle V^{\mathrm{core}} \rangle}{V_i^{\mathrm{core}}}\right)^{1/3} \\ \tilde{b_i} &= b_i \left(\frac{\langle V^{\mathrm{core}} \rangle}{V_i^{\mathrm{core}}}\right)^{1/3} \\ \tilde{c_i} &= c_i \left(\frac{\langle V^{\mathrm{core}} \rangle}{V_i^{\mathrm{core}}}\right)^{1/3} \end{aligned} \tag{5.8}$$

where $V_i^{\mathrm{core}} = 4\pi abc/3$ is the volume of the ellipsoid with semi-axes a, b, c. The mean volume was taken as $\langle V^{\mathrm{core}} \rangle = 4\pi \langle a \rangle \langle b \rangle \langle c \rangle /3$, where $\langle \cdot \rangle$ denotes the average over the respective refined ensemble. $\langle V^{\mathrm{core}} \rangle$ differed only marginally from $4\pi \langle abc \rangle /3$. Subsequently, the SAXS curve was computed using the analytic model of the tri-axial ellipsoid

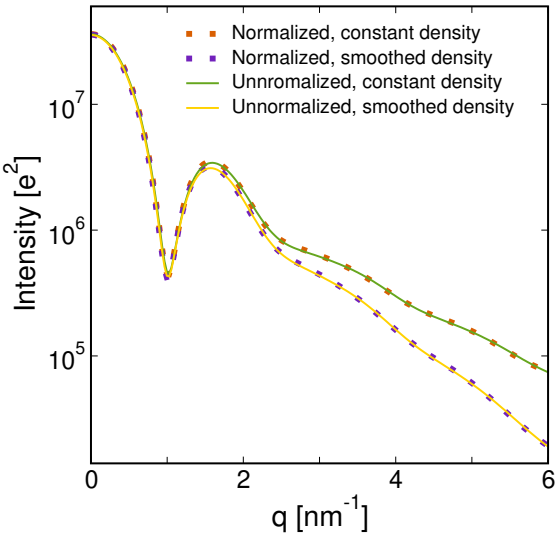

FIGURE 5.5: SAXS curves computed from the two-component tri-axial model. The density profiles were taken as piecewise constant (green line) or smoothed with a Gaussian filter (yellow curve). Samples of normalized and not normalised sets of semi-axes are taken from a 100 ns 4-replica simulation.

using the normalized set of semi-axes $(\tilde{a}, \tilde{b}, \tilde{c})$, and the SAXS curves were averaged. Notably, the correction of the semi-axes via Eqs. 5.8 had only a marginal effect on the SAXS curves (Fig. 5.5). Errors were computed as 1SEM using only independent multi-replica simulations as independent data points, providing a conservative error estimate.

Notes on force field imperfections During all ten-replica simulations, the micelle in one of the replicas adopted an unexpected horseshoe shape, leading to slightly larger value for the large semi-axes a semi-axis (Fig. 5.6). Consistent with this observation, the micelle in two of 20 replicas adopted a horseshoe shape during twenty-replica simulations. With fewer replicas, no such horseshoe shapes were observed. Although we can not strictly exclude that such shapes occasionally exist under experimental conditions, we speculated that either (i) unknown systematic errors in the data, (ii) overly restrained ensembles, or (iii) force field imperfections may be responsible for such occasional shapes.

To shed more light on these observation, additional test simulations were carried out. First, motivated by the fact that the experimental curve exhibits some uncertainty around the pronounced minimum ($q \approx 1\,\mathrm{nm}^{-1}$), we performed two addition sets of refinement simulations using target SAXS curves, whose error was increased by factors of 3 or 5 in the q-region between $0.85\,\mathrm{nm}^{-1}$ and $1.25\,\mathrm{nm}^{-1}$, leading to strongly reduced weights in this q-region. However, also in these additional simulations, the horseshoe shapes were reproduced in one out of ten replicas, suggesting that the relatively high uncertainty at the $q \approx 1\,\mathrm{nm}^{-1}$ region does not cause the horseshoe shapes. Second, we performed series of test simulations with weaker coupling to the experimental curve,

by reducing the force constant (k_c). Only with very low $k_c = 0.1$, the horseshoe shape vanished, but now the agreement to the experimental data was significantly reduced.

Taken together, it seems unlikely that systematic experimental errors induced the horseshoe shapes, or that the simulations were overly restrained to the data. Instead, we hypothesise that, with increasing number of replicas, and hence an increasing number of degrees of the freedom, ensemble refinement becomes more sensitive to force-field imperfections. Hence, ensemble refinement, as conducted here, is also a starting point for future developments of soft matter force fields.

To exclude that the occasional horseshoe shape influences the conclusions of this manuscript, we report results from four-replica simulation, in which no horseshoe shapes were observed, along with results from ten-replica simulations. The distributions of a, b, c were similar in four- and ten-replica simulations (except for contributions from the horseshoe-shaped micelles, Fig. 5.6), and the mean values of a, b, c are nearly identical. This suggests that the key conclusions were not affected by the force field imperfections.

TABLE 5.4: Average semi-axes calculated from multi-replica SAXS-driven MD simulations (top rows) and from fitting two-component ellipsoid models (bottom rows), here assuming pure-water solvent. Errors of a, b, c were computed using block averaging with 4 ns blocks or estimated as 1 SEM of averages between independent runs, and errors were typically smaller than 0.03 nm.

All-atom MD # of replicas	**a [nm]**	**b [nm]**	**c [nm]**	**t [nm]**
1	3.30	2.30	1.66	
4	3.19	2.38	1.68	
8	3.19	2.43	1.66	
24	3.19	2.43	1.69	
Two-component model				
prolate (unphysical)	3.40	1.99	1.99	0.55
oblate	2.85	2.85	1.60	0.53
tri-axial	3.22	2.48	1.66	0.53

Note on solvent simulations All SAXS-driven simulations were conducted in 150 mM NaCl solution, while the SAXS experiment was performed in pure water solvent. To exclude that the details of the solvent influence the conclusions of this study, we repeated the key simulations with the DDM micelle in pure-water solvent. In addition, to exclude that the solvent density affects the calculations with the two-component model, we repeated the fits with the two-component model assuming a solvent density of 334.7 e nm^{-3} to match the electron density of water at 15°C [123]. The fitted parameters, as shown in Table 5.4, are nearly identical to the results found with the 150 mM NaCl solution (Table 5.1).

5.4.2 Supplementary figures

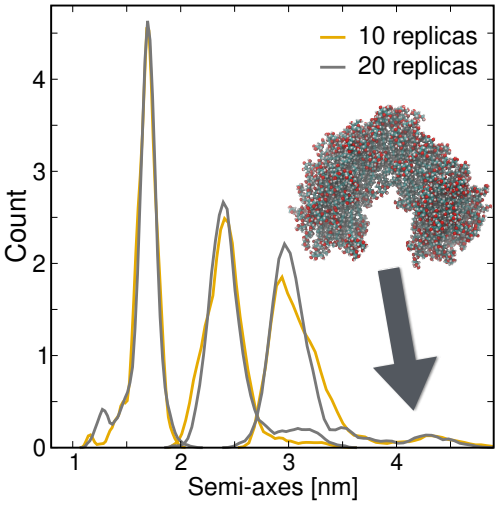

FIGURE 5.6: Distributions of semi-axes in simulations with 10 or 20 replicas, see legend. During these simulations, the micelle adopted in one of ten replicas a horseshoe shape. See SI Text for further discussion. Further, the agreement among the ten- and twenty-replica simulations suggests that 10 replica are sufficient to achieve a minimally biased ensemble.

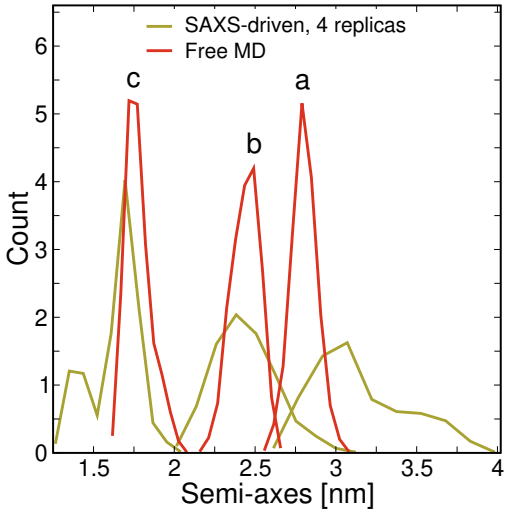

FIGURE 5.7: Distributions of semi-axes, calculated from a four-replica SAXS driven simulation (yellow) or free a MD simulation (red). Evidently, the three axes are slightly more similar in free as compared to SAXS-driven simulations, corresponding to a slightly too spherical shape [298].

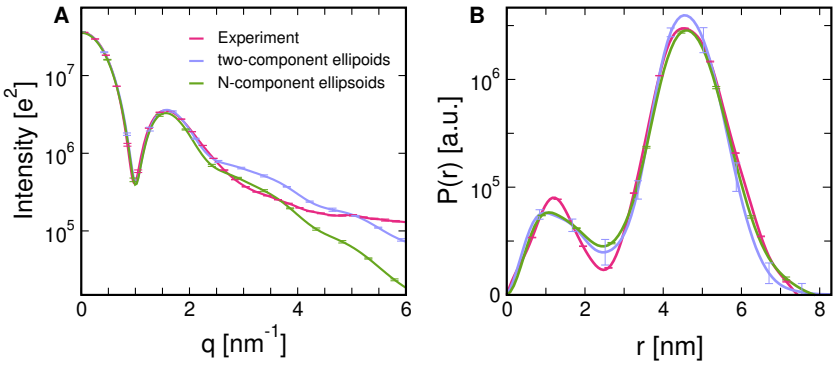

FIGURE 5.8: (A) SAXS curve computed as an average of sets of two-component tri-axial ellipsoids (blue), or of N-component tri-axial ellipsoids with smoothed electron densities around the headgroup region (green). Both models exhibit poor agreement with experimental data (red) at wider angles, suggesting that modeling of atomic details is mandatory at wide angles. Here, the sets of semi-axes were taken from ten-replica MD simulations. Curves calculated with semi-axes from four-replica simulations are nearly identical and shown in Fig. 5.3.

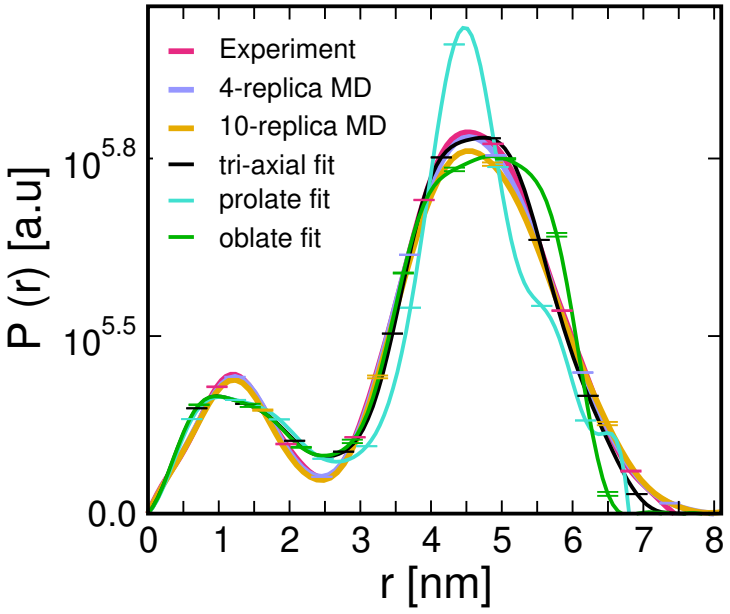

FIGURE 5.9: Pair-distance distribution functions $P(r)$ with the representative errors, obtained from the experimental SAXS curve, from MD-derived SAXS curves, or from curves of fitted oblate, prolate and tri-axial models (see legend for color code). The $P(r)$ curves were computed using GNOM [339] with default settings. The parameter for the maximum diameter ($D_{\max}$) of the particle was taken from the MD frames or from the fitted structural models. $P(r)$ from experiment and from four-replica MD agree favorably. The small deviations between experiment and ten-replica MD at $r \sim 4.5$ nm is a consequence of the horseshoe-shaped micelle in one of ten replicas, see SI Text for discussion and Fig. 5.6. $P(r)$ from the fitted tri-axial ellipsoid give a reasonable agreement to experiment, while $P(r)$ from prolate/oblate fits reveals major discrepancies to the experiment.

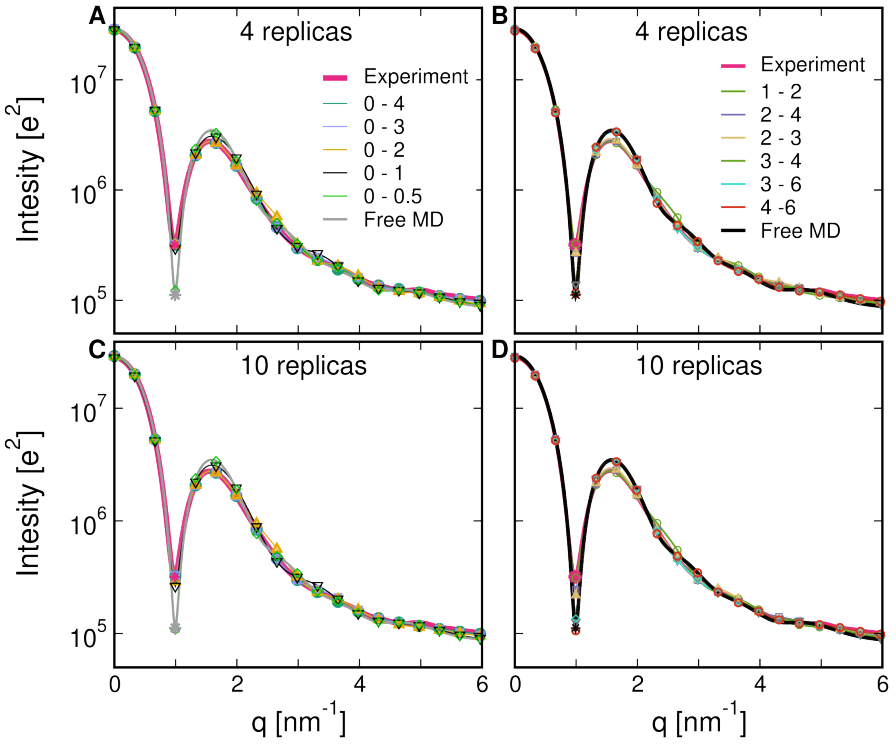

FIGURE 5.10: SAXS curves computed from four-replica SAXS-driven simulations (A and B) and ten-replica SAXS-driven simulations (C and D), using only q-intervals of the experimental curve as a target. The applied q-ranges are indicated in the legends. Representative symbols with the same colors codes are shown to guide the eye.

Chapter 6

Merging In-Solution X-ray and Neutron Scattering Data Allows Fine Structural Analysis of Membrane-Protein Detergent Complexes

THIS chapter, under its title, is published in Ref. 342. We acknowledge the permission from the *The Journal of Physical Chemistry Letters* to reproduce the article.

Authors of the letter are Gatan Dias Mirandela, Giulia Tamburrino, Miloš T. Ivanović, Felix M. Strnad, Olwyn Byron, Tim Rasmussen, Paul A. Hoskisson, Jochen S. Hub, Ulrich Zachariae, Frank Gabel and Arnaud Javelle, with the following contributions: GDM performed the expression, purifcation, SAXS/SANS experiment on the different biological samples and the experimental modelling. FG supervised and trained GDM to perform the SAXS/SANS experiments and checked the data analysis and experimental modelling. MTI and JSH performed the theoretical SANS/SAXS calculations and a part of the MD simulations. GT and UZ performed part of the MD simulations. AJ designed the project. GDM and AJ made the figures for the manuscript. All the authors checked and discussed the results presented in the manuscript. All the authors contributed to the writing of the manuscript.

6.1 Abstract

In-solution small-angle X-ray and neutron scattering (SAXS/SANS) have become popular methods to characterize the structure of membrane proteins, solubilized by either detergents or nanodiscs. SANS studies of protein-detergent complexes usually require deuterium-labeled proteins or detergents, which in turn often lead to problems in their expression or purification. Here, we report an approach whose novelty is the combined analysis of SAXS and SANS data from an unlabeled membrane protein complex in solution in two complementary ways. First, an explicit atomic analysis, including both protein and detergent molecules, using the program WAXSiS, which has been adapted to predict SANS data. Second, the use of MONSA which allows one to discriminate between detergent head- and tail-groups in an ab initio approach. Our approach is readily applicable to any detergent-solubilized protein and provides more detailed structural information on proteindetergent complexes from unlabeled samples than SAXS or SANS alone.

6.2 Introduction

Integral membrane proteins form the entry and exit routes for nutrients, metabolic waste and drugs in biological cells, and they are involved in key steps of signaling and energy transduction. They thus play a central role in a variety of biological processes with exceptional medical relevance [343]. Structural information on membrane proteins has traditionally been obtained by X-ray crystallography aided by detergent molecules that replace the lipids during the purification and crystallization processes. Detergents stabilize membrane proteins by shielding the hydrophobic domains from the aqueous environment [344]. However, the translocation cycle underpinning membrane transporter activity requires substantial conformational variability and, in many cases, the static structural insight achieved by X-ray crystallography has proven insufficient to capture the essential functional information on these systems [345]. For this reason, there is considerable interest in the application of small angle scattering (SAS) methods to structurally characterize membrane proteins. Recently, efforts have been dedicated to develop combined in-solution small-angle X-ray/neutron scattering (SAXS/SANS) approaches to investigate membrane proteins stabilized by detergents or nanodiscs [346–348]. Further developments in these areas have faced important obstacles. Crucially, the electron density of the detergent shell encompassing the hydrophobic domains of membrane proteins differs from the electron density of the protein. Hence, it is difficult to obtain a model of a protein-detergent complex using ab initio SAXS-based methods, which typically assume a uniform electron density across the entire complex. To circumvent this problem,

SANS experiments making use of contrast variation either by using deuterium-labeled proteins and/or detergent molecules have been employed. However, difficulties are often encountered in the expression and purification of deuterated proteins, as well as the limited availability of deuterated detergents [346]. To overcome these issues, we report a new methodology that combines SAXS and SANS from unlabeled (i.e., nondeuterated) proteins and/or detergent samples to obtain detailed structural information on protein-detergent complexes. This approach is readily applicable to any detergent-solubilized protein.

6.3 Results and discussion

We used the ammonium transporter AmtB from Escherichia coli, a structurally well-studied member of the ubiquitous and medically important Amt/rhesus family of proteins, to develop and validate our methodology [349]. To stabilize AmtB, the detergent n-dodecyl-β-D-maltosid (DDM) was used throughout the purification process (Supporting Information). Size exclusion chromatography in-line with multiangle light scattering (SEC-MALS) analysis showed that the AmtB-detergent complex comprises 285 ± 12 DDM molecules (Figure 6.1 and Table 6.1). Independently conducted analytical ultracentrifugation (AUC) experiments revealed a detergent shell of 321 ± 1 DDM molecules (Figure 6.2 and Table 6.1). Taken together, these independent findings indicate that the detergent corona around AmtB is likely to include between 260 and 320 DDM molecules.

We next exploited atomistic molecular dynamics (MD) simulations of the AmtB-DDM complex and scored the models against SAXS data to resolve the experimental uncertainty regarding the size of the detergent corona. AmtB in the physiologically functional trimeric form (PDB ID: 1U7G) [350]. was simulated surrounded by DDM coronas of 260, 280, 300, 320, 340, and 360 molecules. A representative model obtained for a detergent corona containing 320 molecules of DDM is shown in Figure 6.1. During the equilibration phase, the DDM molecules adopted the typical toroidal shape reported for other protein-detergent complexes [71, 270], with their hydrophilic heads facing the aqueous solution and their hydrophobic tails oriented toward the inside of the complex (Figure 6.1). As previously shown, the detergent corona further adapted to the shape of the transmembrane surface of the protein [270]. Our simulations indicate that the proteindetergent complexes are stable, and although some reorientation of DDM was observed, in particular during the first stages of the simulations, no dissociation of detergent molecules from the protein was detected after 20 ns of simulation time. We next computed SAS curves for the simulated complexes and compared them with experimental SAS measurements (Figures 6.2 and 6.3). It has previously been shown that

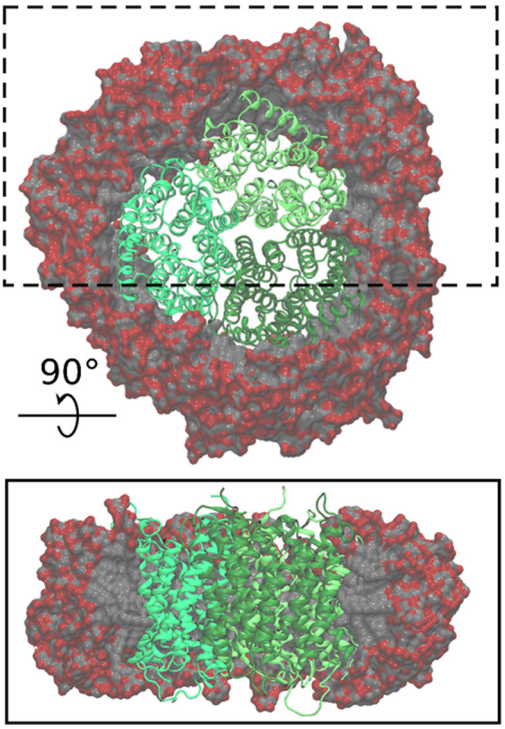

FIGURE 6.1: Atomistic model of the AmtB-DDM complex containing 320 DDM molecules. The model displays an equilibrated complex. In the trimer, each AmtB monomer is shown in a different shade of green, and the DDM carbon and oxygen atoms are shown in gray and red, respectively. The upper panel shows the complex seen from the top; the lower panel is a side-view of the complex where the DDM molecules outside of the box highlighted in the top panel are omitted, to illustrate the interior of the micelle.

single structures extracted from MD trajectories do not fully capture the characteristics of the solution ensemble [71]. We therefore calculated the predicted SAXS curves from conformational ensembles comprising 9000 individual configurations as observed in $70 - 160$ ns simulations of each differently sized complex. The SAXS curves were obtained using explicit-solvent calculations as implemented in the WAXSiS method, thereby taking into account accurate atomic models for both the hydration layer and the excluded solvent, and consequently avoiding any solvent-related fitting parameters (Figure 6.2) [90, 226].

SAS experiments are very demanding in terms of requirements of sample quality [73, 351], therefore, before recording SAS data, we ascertained that our samples were monodisperse and that AmtB was pure, stable, and critically active in detergent (Supporting Information, Figures 6.2 - 6.3). We subsequently collected experimental SAXS data following size-exclusion chromatography of the AmtBDDM complex. The radius of gyration (R_g) was found to be constant across the elution peak (Figure 6.1), indicating the monodispersity of the complex and good data quality. Importantly, the scattering curves predicted for the models containing 260, 280, 300, 340, and 360 DDM molecules

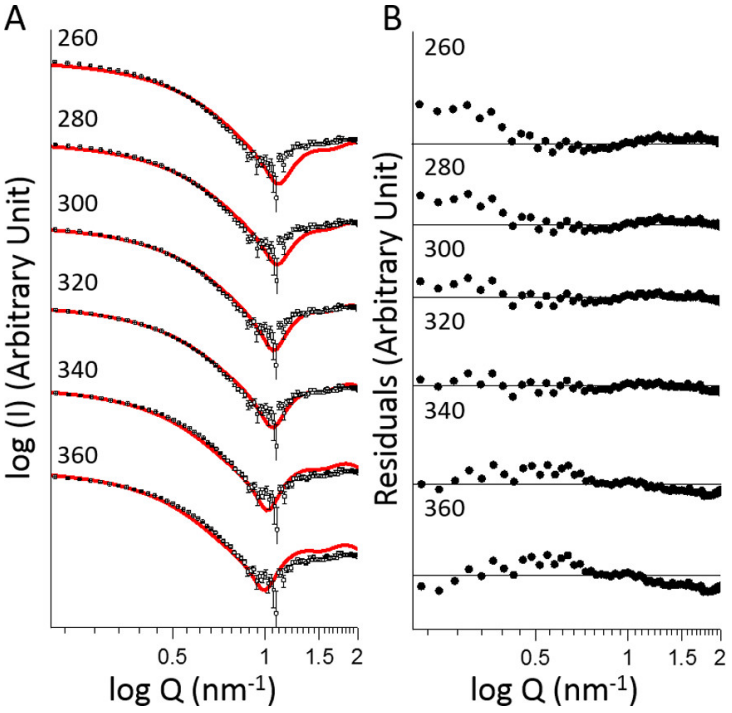

FIGURE 6.2: (A) Comparison of the experimental (symbols) and computed (red line) SAXS curves for the AmtB-DDM complex containing between 260 and 360 DDM molecules. For all plots, the maximum and minimum values for the y-axis are 1011 and 105. (B) Residual error plot expressed as the experimental minus computed scattering intensity. For all plots, the maximum and minimum values for the y-axis are 40 and 40. $Q = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle.

deviate slightly from the experimental data (Figures 6.2 and 6.4). By contrast, the curve computed for the MD model containing 320 DDM molecules was nearly indistinguishable from the experimental SAXS data (Figure 6.2 and 6.4). Furthermore, the values for R_g obtained by the Guinier approximation from the experimental data and from for the MD model containing 320 DDM molecules were in quantitative agreement (Table 6.3 and Figure 6.9). This suggests that the overall dimension of the simulated proteindetergent complex containing 320 molecules of DDM is identical to that in solution. It is important to note that the overall information content of SAXS is relatively low, and thus agreement between experimental and back-calculated curves may be insufficient to serve as unambiguous evidence for a structural model [352]. Specifically, in the context of a proteindetergent complex, SAXS data reports on the overall shape of the complex, whereas they do not provide independent information on the individual contributions from the protein and the detergent corona. Therefore, we employed SANS together with contrast variation to more firmly validate our computational model.

We collected SANS data at four contrast points (0%, 22%, 42% and 60% (v/v) D_2O) to differentiate between the individual components of the proteindetergent complex. To ensure that the samples were stable over the course of the SANS experiment, the

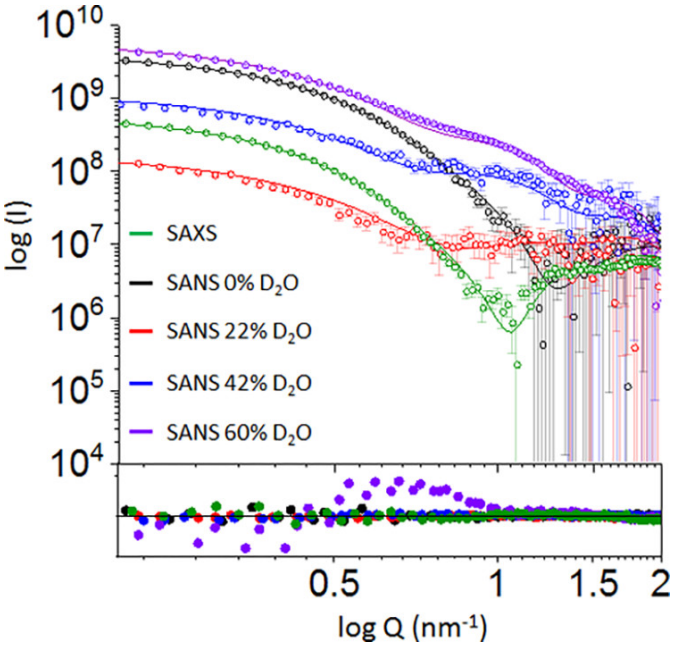

FIGURE 6.3: Comparison of the experimental (symbols) and computed (red line) SAXS/SANS curves for the model containing 320 DDM molecules. Residual error plot expressed as the experimental minus computed scattering intensity. The maximum and minimum values for the y-axis are 40 and 40, respectively.

hydrodynamic behavior of the proteins were analyzed before and after the SANS measurements by analytical size exclusion chromatography. No differences were observed in the elution profile, confirming the stability of the protein during the SANS experiment (Figure 6.10). To ascertain the reproducibility and the quality of our measurements, two independent sets of SANS data were acquired, using two batches of AmtB purified independently. The two data sets were found to be identical within the limits of the observed experimental noise (Figure 6.11). It has previously been shown that in the absence of D_2O in the buffer, neutron scattering from DDM micelles originates primarily from the hydrophilic head groups [353]. We calculated (Supporting Information) the overall contrast match point of DDM to be at 22% D_2O, while the contrast match point for typical proteins is around 42% D_2O [346, 354]. Consequently, the scattering contribution is dominated by the protein and the DDM hydrophilic headgroup in a buffer containing 0% D_2O, by the protein at 22% D_2O and by the complete detergent corona at 42% D_2O. To compare the experimental neutron scattering data with the MD-generated models, SANS curves were calculated using WAXSiS for 9000 individual configurations observed during $70 - 160$ ns MD trajectories of each of the complexes. To this end, we extended the WAXSiS method, originally developed for SAXS predictions, to also allow SANS predictions with explicit-solvent models at various D_2O concentrations (Supporting Information). The experimental curves were fitted to the calculated curves following $I_{fit} = f \cdot I_{exp} + c$, thereby accounting for scattering contributions from the incoherent background with the fitting parameter c. However, neither the hydration layer nor

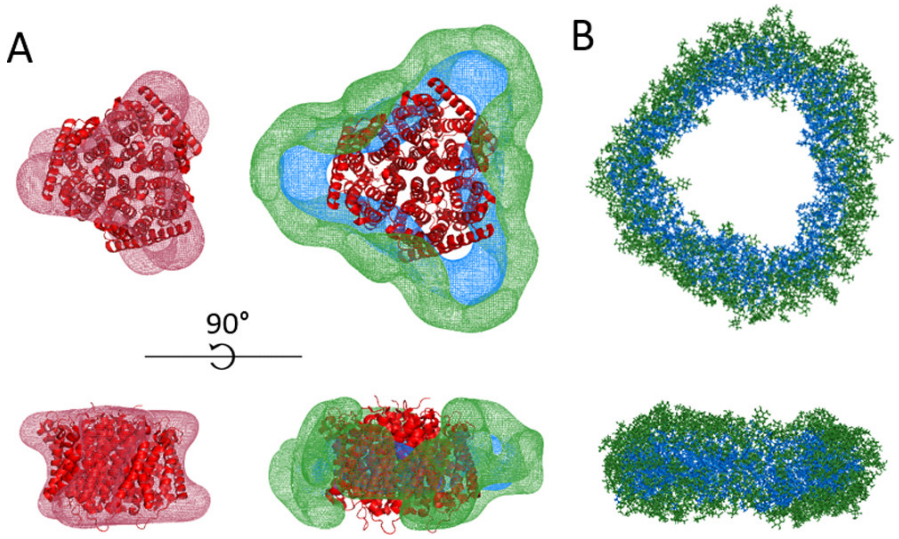

FIGURE 6.4: (A) MONSA multiphase modeling using experimental SAXS and SANS data. The phase corresponding to the protein is represented in red mesh, while the hydrophilic and hydrophobic detergent densities are represented in green and blue, respectively. (B) Molecular-dynamics generated model of the detergent corona (320 molecules) surrounding AmtB.

the excluded volume were adjusted. Congruent with the analysis of the SAXS data, all SANS data sets were best fitted by the curves calculated for the model incorporating 320 molecules of DDM (Figures 6.3 and 6.12). Hence, the SANS and SAXS data consistently validate our MD model with 320 DDM molecules. Second, the excellent agreement we observe between the experimental and calculated SAXS curves shows that the overall organization of the complex is accurately reflected by the atomistic model. Finally, the good agreement between experimental and computed SANS curves indicates that the MD model describes accurately the hydrophobic and hydrophilic phase of the detergent ring as well as the position of AmtB inside the corona.

Importantly, the crystal structure of AmtB was used to produce our MD trajectories, which precludes the possibility of applying this combined MD/SAXS/SANS approach to membrane proteins of unknown structure. We therefore applied, in the final step, an independent "MD-free" approach to obtain a full *ab initio* model that captures detailed structural information on the complex without using the crystal structure of AmtB. To achieve this, we merged our complete SAXS and SANS data and conducted a multiphase volumetric analysis of the complex using MONSA [355, 356] (Figure 6.4). Importantly, we introduced two separate phases to describe the head and tail groups of the DDM detergent corona. Assuming the volume of a DDM molecule to be 690 $Å^3$ (350 $Å^3$ and 340 $Å^3$ for the head and the tail, respectively),(4) we imposed a volume of 112000 $Å^3$ and 108800 $Å^3$ for the hydrophilic and hydrophobic phases of the 320 DDM molecules. The volume of AmtB (166864,$Å^3$) was calculated based on its amino acid sequence alone (Supporting Information). Moreover, since the trimeric nature of AmtB in solution was

confirmed by our SEC-MALS and AUC data (Figures 6.1 and 6.2 and Table 6.2), we imposed a P3 symmetry on the complex. Ten MONSA runs (Figure 6.13) were performed yielding similar *ab initio* envelopes for AmtB. A representative MONSA model is shown in Figure 6.4, which faithfully reflects both the size and shape of the MD-generated model. The protein envelope is a good representation of the crystallographic structure of AmtB and is, furthermore, confined inside the detergent corona. Importantly, the joint use of both SAXS and multiple SANS data sets allowed us to distinguish the head- and tail-groups of the detergent corona and place them correctly with respect to the protein surface and solvent. Such detailed insight is usually not achieved with ab initio models unless additional contact restraints are applied [357]: the detergent ring fits the contours of the protein and the positions of the two detergent phases (head- and tail-groups) are particularly clear. The hydrophobic phase is strictly contained between AmtB and the hydrophilic ring, with only the tails of DDM being in contact with the hydrophobic surface of the transmembrane domain. Hence, without using deuterated protein or detergent, and without information about the 3D structure of AmtB, the combination of SAXS and SANS data capture the essential structural details contained in membraneprotein detergent complexes in solution.

6.4 Conclusions

In summary, there is considerable interest in developing SAS methodology further to allow routine investigation of membrane proteins. We have adapted WAXSiS to account for SANS data and therefore open up this software package for future projects including both types of scattering data. Using our methodology, based upon a combination of SAXS/SANS measurements and MD simulations, we have been able to propose an atomic model of a protein-detergent complex. Our integrative approach demonstrates that combining SAXS, SANS, and iterative simulations provides much more detailed structural information than each of the methods alone.

It is widely recognized that cryo-electron microscopy (cryo-EM) will revolutionize the structural analysis of membrane proteins in the near future [358, 359]. It is our belief that a hybrid approach, combining in solution SAS techniques, *in silico* modeling, and cryo-EM will allow for better tracking and description of conformational changes of membrane proteins in solution, induced by ligand or cofactor binding. In this context, it was important to account accurately for the bound detergent molecules, which is greatly improved by combining SAXS and SANS data at various contrasts. Second, our multiphase analysis, which merges SAXS and SANS data, without using deuterated protein or detergent, allowed us to obtain unprecedented structural information

on the phase density of the detergent, in particular to distinguish head- and tail-groups in the assembled membrane proteindetergent complexes. This is particularly relevant as deuterated media/detergents are often expensive and/or toxic for bacteria, leading to decreased protein yields [360]. Crucially, the multiphase analysis does not require information on the 3D structure of the protein, which opens up the possibility of applying this methodology to a wide range of important membrane proteins that have so far remained inaccessible to high resolution structural analysis. While SAS has become a popular technique among structural biologists, combinations of SANS, SAXS and MD simulations have remained underexploited by the community. In this context, our work represents a significant advancement in data acquisition, model validation, development of new software, and multiphase volumetric analysis to firmly establish SAS technology as a standard method for membrane protein structural biology.

Acknowledgments G.D.M. and A.J. were supported by a Ph.D. and a Chancellors Fellowship from Strathclyde University, respectively, G.T. and U.Z. acknowledge funding from the Scottish Universities' Physics Alliance (SUPA). P.A.H. acknowledges the support of the Natural Environment Research Council (NE/M001415/1) and A.J. the support of Tenovus Scotland (Project S17-07). M.T.I., F.M.S., and J.S.H. acknowledge support by the Deutsche Forschungsgemeinschaft (HU 1971-1/1, HU 1971-3/1, HU 1971-4/1). We thank the ILL Block Allocation Group (BAG) system for SANS beamtime at D22 and Dr A. Martel for help with the setup of the instrument. We acknowledge Diamond Light Source for time on Beamline B21. We thanks Dr. P. Soule (NanoTemper Technologies GmbH) and Dr. M. Tully (DIAMOND, U.K.) for help with the microscale thermophoresis experiments and SEC-SAXS data acquisition, respectively.

6.5 Supplementary material

Protein purification and stability AmtB(His_6), cloned into the pET22b vector [361], was overproduced and purified as described previously [361] except 0.03% (0.58 mM) of n-dodecyl-β-D-maltoside (DDM) was use instead of 6 mM *N*, *N*-dimethyldodecylamine-N-oxide (LDAO) in the final Size Exclusion Chromatography (SEC) buffer (Tris/HCl 50mM, pH 7.8, NaCl 100 mM, 0.58 mM DDM). AmtB was kept in SEC buffer at 4°C for subsequent characterisation. AmtB stability was assessed before and after each SAS experiment by SEC using a Superdex 200 10/300 (Ge Healthcare) gel filtration column (Figure 6.10).

Multiangle light scattering (SEC-MALS) analysis SEC-MALS analysis of the AmtB-DDM complex was carried out using Superdex 200 10/300 column (Ge Healthcare) attached on an Agilent 1100 HPLC system. 70 μl of AmtB at 75 μM in SEC buffer was injected at a flow rate of 0.5 ml/min. Light scattering, refraction index and absorbance at 280 nm were measured using a multi-angle light scattering mini DAWN TREOS detector (Wyatt Technology), a refractometer Optilab T-rEX detector (Wyatt Technology) and a Jasco UV-2077 Plus UV/vis spectrophotometer respectively. We used the ASTRA software package version 5.3.2.10 (Wyatt Technologies) to import the signals from the three detectors and analysed the data according to Slotboom *et al.* (2008) [362].

Analytical ultracentrifugation (AUC) AmtB at 10, 22 and 87 μM was submitted to sedimentation velocity using a Beckman Coulter Optima XL-I analytical ultracentrifuge mounted with an An-50 Ti 4-hole rotor (49000 rpm at 4°C). The reference buffer used was the SEC buffer without detergents (50 mM Tris pH 7.8, 100 mM NaCl). Data were acquired every 6 min for 12 hrs, with interference and absorbance optics and were subsequently analysed using SEDFIT [362] with the continuous c(s) distribution model. SEDNTERP was used to determine the molar mass (46647 g/mol) and the partial specific volume (0.749 ml/g) of AmtB. The partial specific volume of DDM used was 0.82 ml/g. The viscosity (1.567 cP) and the density (1.00557 g/ml) of the SEC buffer were determined using SEDNTERP. The ratio of detergent bound to the protein and the molecular weight of the complex (Table 6.1) were calculated using a method described previously [363].

Micro-Scale Thermophoresis (MST) AmtB(His_6) was labelled using the kit Monolith His-Tag according to manufacturer instructions (NanoTemper Technologies). Labelled samples of AmtB in the concentration range [6 μM − 200 mM] were loaded into 16 hydrophobic coating grade capillaries and analyzed using the Monolith NT.115 (NanoTemper Technologies) analyser. The data were processed using the MO.Affinity Analysis software v2.2.4 (NanoTemper Technologies) as previously described [364].

Characterisation of the sample SAS experiments are very demanding in terms of sample quality [73, 351], therefore, before recording SAS data, we assessed the purity and monodispersity of the samples as follows: the mass of the complex calculated from our SEC-MALS analysis are constant across the elution peak (Figure 6.5). Secondly, SDS page and SEC analyses of our sample before and after the SAS experiments show that the protein is pure and stable for weeks at 4°C (Figure 6.10). Finally, to ensure that the protein was purified in an active form, we measured AmtB NH_4^+ binding activity

by microscale thermophoresis. Clear NH_4^+ dependent binding activity (Kd 0.6 mM) was measured, which indicates that AmtB is correctly folded and active (Figure 6.7). Taken together, these results show that our sample is pure, monodisperse and that the protein is active in detergent. Hence our sample is highly suitable for SAS analysis.

MD simulations The AmtB crystal structure at 1.35 Å obtained by Kademi *et al.* (PDB ID: 1U7G) [350] was used for the molecular dynamics simulations. The protein was processed using the CHARMM-GUI web interface [294, 365]. The mutations F68S, S126P, and K255L inserted in the crystallographic construct were reverted to the wild-type, all the selenomethionine modifications were changed back to methionine, and the N- and C-terminal residues were capped using acetyl and N-methyl amide groups, respectively. The protein was initially embedded in a DDM bilayer using CHARMM-GUI Membrane Builder plugin [366, 367]. The system was subsequently solvated in water and K^+ and Cl^- ions were used to neutralise the system and reach a salt concentration of 50 mM. In-house code was then used to remove excess DDM molecules and to form a detergent torus around the protein of 260, 280, 300, 320, 340 and 360 DDM molecules, respectively, in order to reflect the experimental conditions.
All molecular dynamics simulations where performed with the Gromacs 5.1.4 software package [368, 369]. The CHARMM36 force field was used for the protein, the ions, and DDM [295, 370]. The water molecules were modelled with the TIP3P model [371]. Water bonds and distances were constrained by the Settle method [289], and all other bonds by the LINCS method [149]. After a steepest descent minimization, the system was equilibrated by six consecutive equilibration steps with position restraints on heavy atoms of 1000 $kJ(mol)^{-1}nm^{-2}$. The first three equilibration steps were carried under a NVT ensemble using a Berendsen thermostat to keep the temperature at 310 K. The subsequent steps were conducted under a NPT ensemble, switching on a Berendsen barostat [159] with isotropic coupling, to keep the pressure at 1 bar. Production MD simulations were carried using a v-rescale thermostat [291] with a time constant of 0.2 ps, and a Berendsen barostat with isotropic coupling. A Verlet pair-list scheme was used for describing non-bonded interactions, and two different cut-off values, of 1.2 Å and 1.5 Å were tested; no significant difference was observed between these two. A timestep of 2 fs was used throughout the simulations. The first 70 ns of production simulations were discarded from the analysis to allow rearrangement of the DDM molecules around the protein.

SEC-SAXS analysis Synchrotron SAXS data were collected on the B21 bioSAXS beamline at the DIAMOND Synchrotron. The exact same conditions were used than for the SEC-MALS analysis in terms of sample, column and running conditions. 50 μl of

AmtB at 75 μM were injected into the SEC-system. The running buffer used was 50 mM Tris pH 7.8, 100 mM NaCl and 0.03% DDM. 15 frames of the elution peak corresponding to the membrane protein were averaged and subtracted to the running buffer using ScÅtter software. 57 frames corresponding to the buffer curve were averaged prior to the subtraction. SAXS data were collected in 255 time frames with 3 s per frame (13 min in total). The scattering images were averaged and the buffer scattering intensities subtracted using the program ScAÅtter and the same program was used to evaluate the radius of gyration (R_g). The data-collection parameters are presented in Table 6.2.

SAXS curve predictions SAXS curves were computed using the explicit-solvent calculations described previously [90], as implemented in the WAXSiS method [226]. Accordingly, a spatial envelope was constructed around the AmtB-DDM complex, such that the distance of the envelopes vertices have a distance of 6 Å from all atoms in all simulation frames. Because the detergent exhibited substantial fluctuations, this procedure yielded an envelope that had a larger distance from the complex in most of the MD frames, suggesting that solvent density modulations due to the hydration layer were captured by the envelope volume. The spherical average was conducted using 1200 q-vectors per absolute value of the scattering vector q. Vector q is given as $q = 4\pi \sin(\theta)\lambda^{-1}$, where 2θ is the scattering angle. The bulk solvent density was corrected to 334 e nm^{-3} to correct for the slightly incorrect density of the CHARMM-modified tip3p model, as described previously [90].

SANS data measurement and analysis To ascertain the reproducibility and the quality of our measurements, two independent set of SANS data were measured (September 2016 and March 2018) using two batches of AmtB purified independently. SANS experiments were conducted at 6°C using the large dynamic range diffractometer D22 at the Institut Laue-Langevin (Grenoble, France) in Hellmar quartz cuvettes 100QS with 1 mm optical pathlength. 300 μl of samples at a concentration of 110 μM were extensively dialysed (3 times 12 hrs) against the size-exclusion chromatography buffer (50 μM Tris pH 7.8, 100 mM NaCl, 0.03% DDM and D_2O as required) and used for the SANS experiment. The final dialysis buffer was used in the SANS experiment as the reference and subtracted to the protein signal. The samples were recorded at a 4 m/4 m detector/collimator distance, using a neutron wavelength of $\lambda = 6$ Å. For each condition, H_2O/D_2O buffers, the empty beam, an empty quartz cuvette as well as a boron sample (electronic background) were measured. Exposure times varied between 20 min (empty cell, boron) and 3 hours for the protein samples and buffers. Transmissions were measured for 1 min. The raw data were reduced (detector efficiency, electronic background and angular averaging) using a standard ILL software package. Finally, the corrected

scattered intensities $I(q)$, where $q = 4\pi \sin(\theta)\lambda^{-1}$, and 2θ is the scattering angle, from the different q-ranges and the respective buffer signals were subtracted using the program PRIMUS from the ATSAS suite [220]. The data-collection parameters are presented in Table 6.2.

Contrast Variation Series The contrast match point of DDM was experimentally determined by measuring SANS contrast series of DDM (5 mg/ml) at 0, 20, 40, 60, 80 and 100 % D2O and used to plot $(I_0/T_sC)^{0.5}$ as a function of percentage of D_2O in the solvent (T_s is the measured sample transmission). The DDM contrast match point (22.2%) was determined by the intersection of a linear fit through all points with the abscissa as previously described [372].

Guinier plot Guinier plots were used to calculate the R_g based on the following equation: $\ln I(q) = (1/3) \ln I(0) - (qR_g)^2$ with $qR_g < 1.3$. AmtB molecular weight, was determined from the I_0 intensity at 22% D_2O using absolute calibration against H_2O under the assumption that the detergent (free micelles and bound) had a negligible contribution at the contrast match point according to Compton *et al.*, (2011) [372] (Figure 6.9 and Table 6.1).

MONSA multiphase modelling The multiphase volumetric analysis using MONSA [356, 373] (extended version of DAMMIN) was used to obtain a three phases dummy atom models of the AmtB-DDM complex reporting the protein, DDM head and DDM tail phases respectively. The analysis was done using all SAS (SAXS and SANS at 0, 22, 42 and 60% D2O) data. The parameters used for the analysis were (i) the volume of the AmtB trimer (calculated from the amino acid sequence using the Biomolecular Scattering Length Density Calculator available online (http://psldc.isis.rl.ac.uk/Psldc). The volume obtained was 166864 $Å^3$; (ii) the volume of the 320 molecules of DDM head and tail (112000 $Å^3$ and 108800 $Å^3$ respectively) [346]. The SEC-MALS, AUC and SAXS analysis shows that AmtB is trimeric (Table 6.1), hence a P3 symmetry constraint was applied. The MONSA analysis (200 annealing steps) were done using DAMESV28-29 models. 10 model were generated, superimposed and checked for consistency (Figure 6.13). All models were very similar in size and shape for all three phases.

6.5.1 Supplementary tables and figures

TABLE 6.1: Determination of the number of DDM molecules in the AmtB-DDM complex. Molecular masses are given in kDa. OS stands for oligomeric state.

	M_w AmtB-DDM	**M_w AmtB/OS**	**M_w DDM**	**DDM molecules**
SEC-MALS	287.2 ± 16.7	144.4 ± 11.1 / trimer	142.8 ± 5.9	285 ± 2
AUC	312.6 ± 14.5	148.8 ± 6.8 / trimer	163.7 ± 0.7	321 ± 1
SANS 123	n.a.	146.0 ± 29.2 trimer	n.a.	n.a.

TABLE 6.2: SEC-SAXS and SANS data-collection parameters.

	from SEC-SAXS	**from SANS batch**
Instrument/data processing	B21 BioSAXS Diamond Light Source synchrotron beamtime (UK)	D22 beamline Institut Laue-Langevin Neutron source with a multidetector (3He) 16K resolution elements
Beam geometry (mm)	1 x 5	55 x 40
Camera length (m)	4.014	4/4 offset
Flow (ml/min)	0.05	n.a.
Wavelength (Å)	0.99	6.0 ($\delta\lambda/\lambda = 0.1$)
s range (Å^{-1})	0.0022 to 0.42	0.045 to 0.4
Exposure time (s)	3 (every 5 seconds)	∼ 3600 to 10800 / sample
Sample configuration		cell path-lenght of 1 mm
Concentration range (mg/mL)	See supporting info.	See supporting info.
Temperature (K)	298	279.15

TABLE 6.3: R_g calculated from Guinier approximation of the experimental and computed SAXS curves of the AmtB-DDM complex containing between 260 and 360 DDM molecules. All values are given in angstroms.

260	280	300	320	340	360	Experimental SAXS data
43.1 ± 0.24	44.0 ± 0.30	45.6 ± 0.10	**45.6 ± 0.10**	47.8 ± 0.16	48.6 ± 0.13	**45.5 ± 0.09**

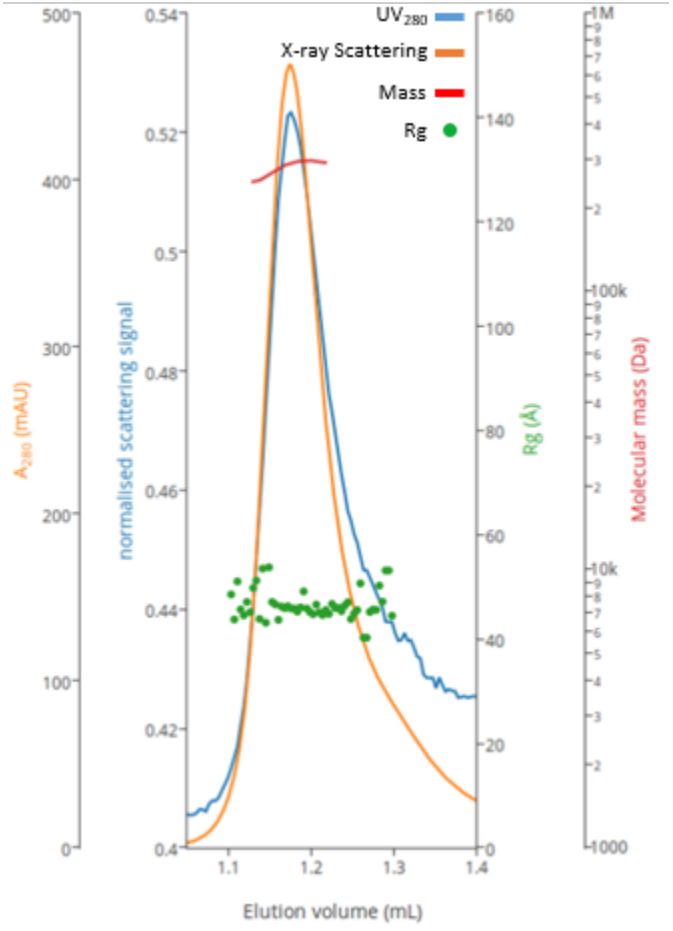

FIGURE 6.5: Elution profile of AmtB purified in 0.03% DDM from a superdex 200 10 times 300 column measured by OD at 280 nm and scattering.

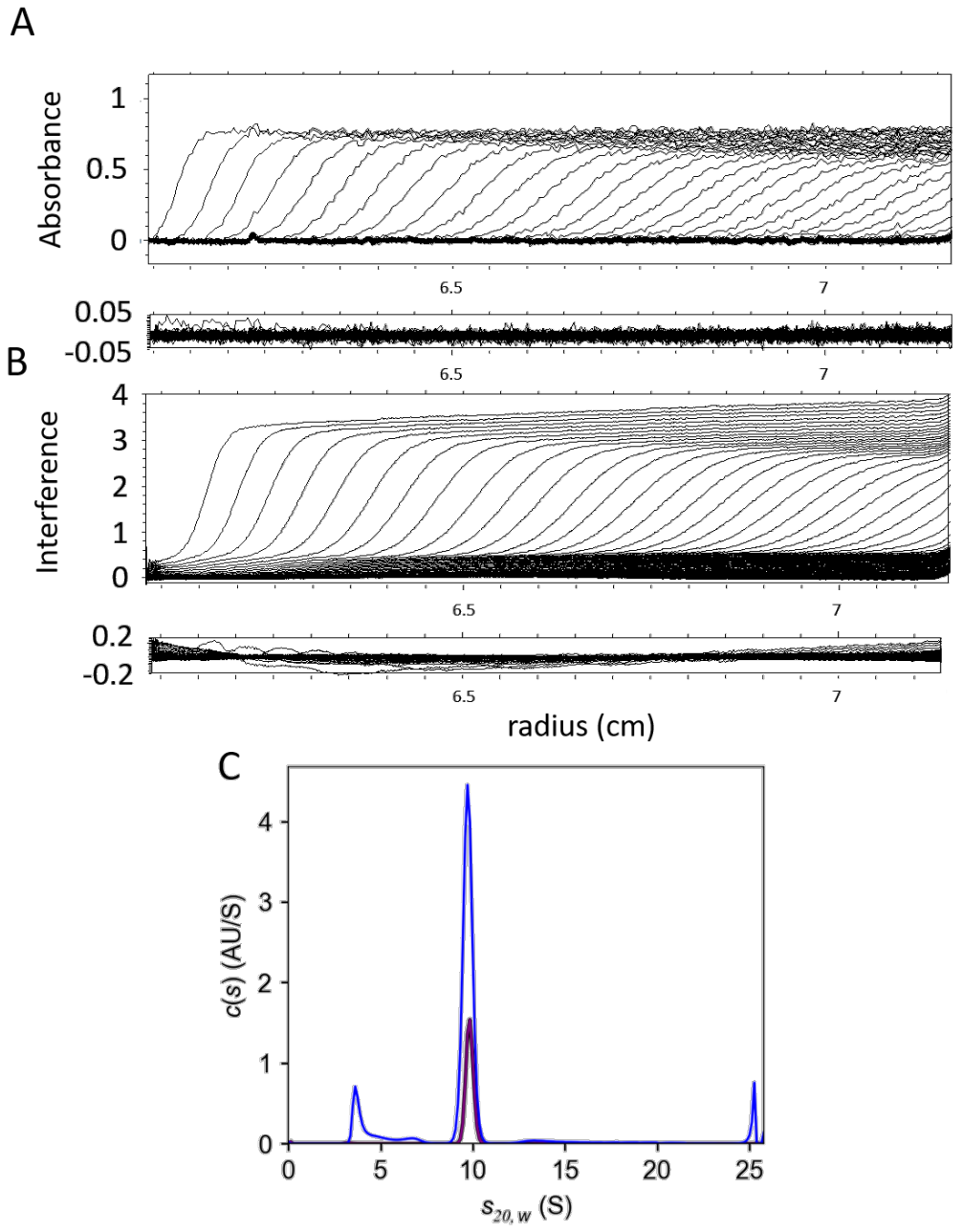

FIGURE 6.6: (A) Absorbance at 280nm and (B) interference signals of analytical ultracentrifugation sedimentation profile of AmtB solubilised in 0.03% DDM. (C) Superposition of the c(s) distributions expressed at 20°C in water (blue: $A_{280\text{nm}}$, purple: Interference).

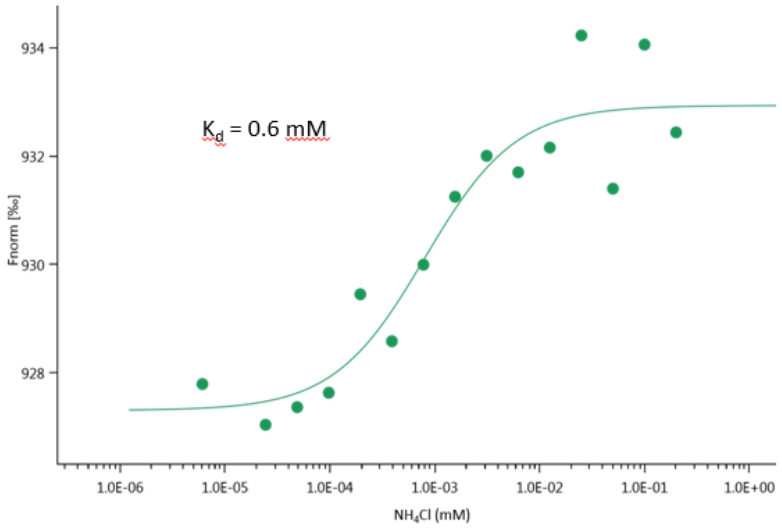

FIGURE 6.7: Measurement of the ammonium binding on AmtB solubilised in 0.03% DDM by microscale thermophoresis.

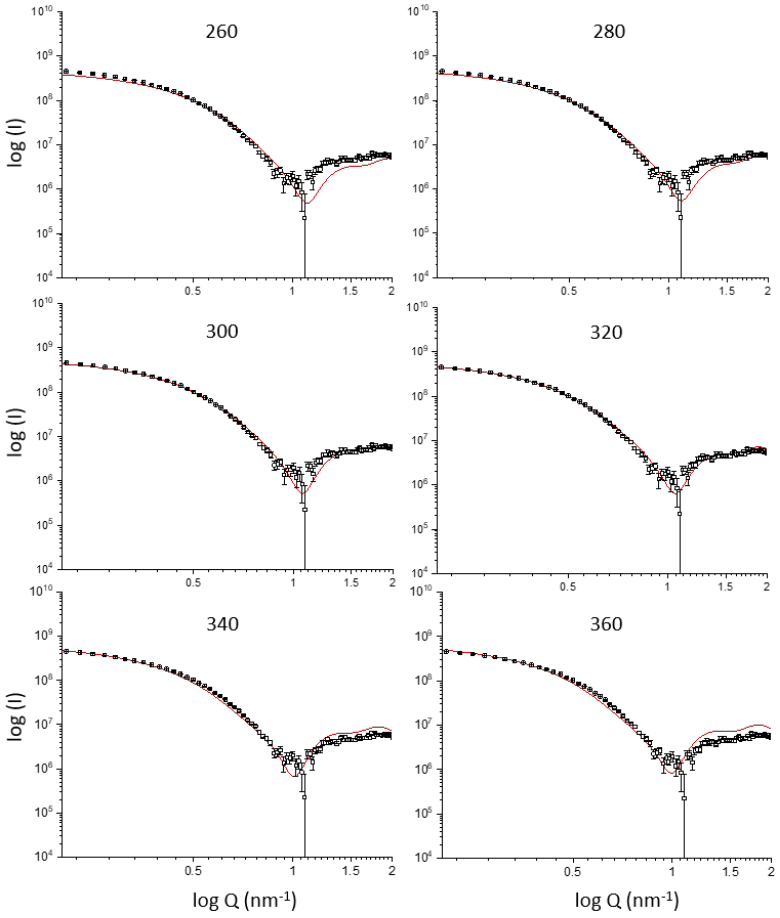

FIGURE 6.8: Comparison of the SAXS curves from experiment (symbols) and computed from MD simulations of the AmtB-DDM complex containing between 260 and 360 DDM molecules (red lines).

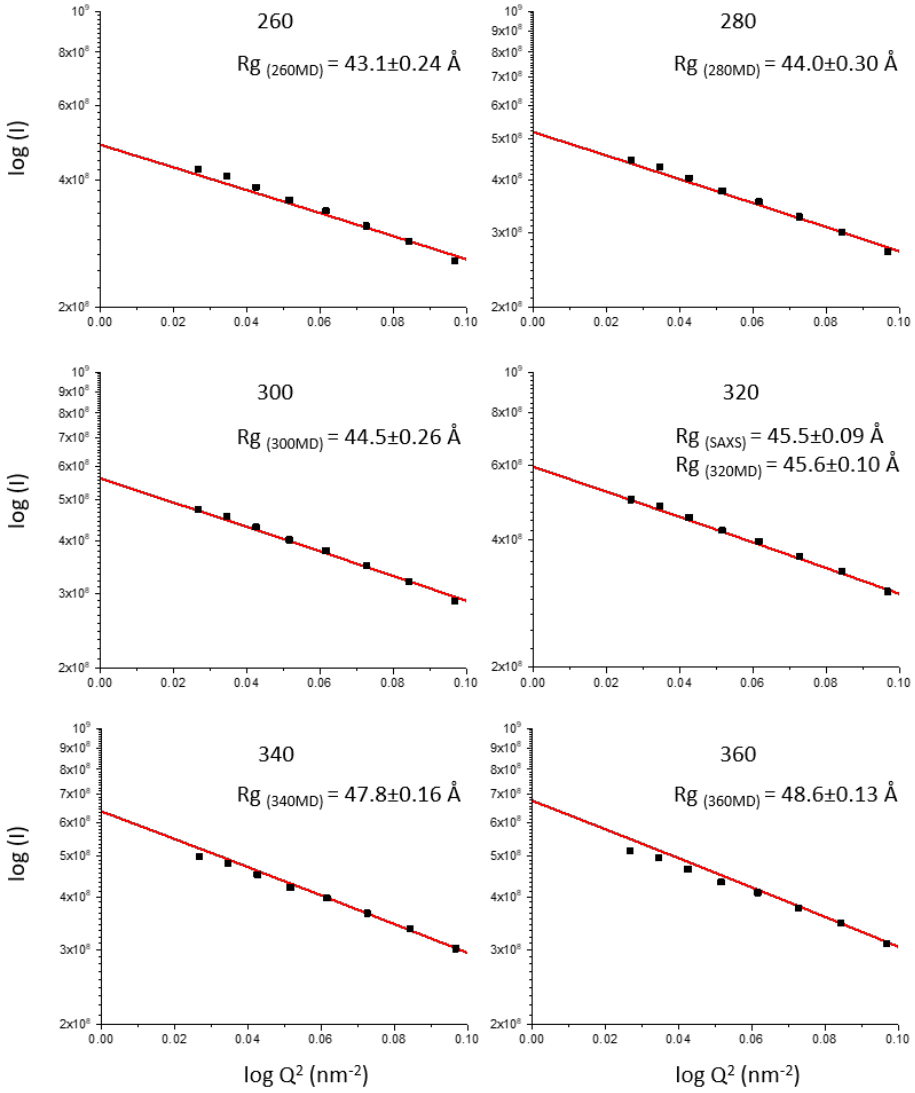

FIGURE 6.9: Guinier plot comparison of the experimental (symbols) and computed (red line) SAXS curves of the AmtB-DDM complex containing between 260 and 360 DDM molecules.

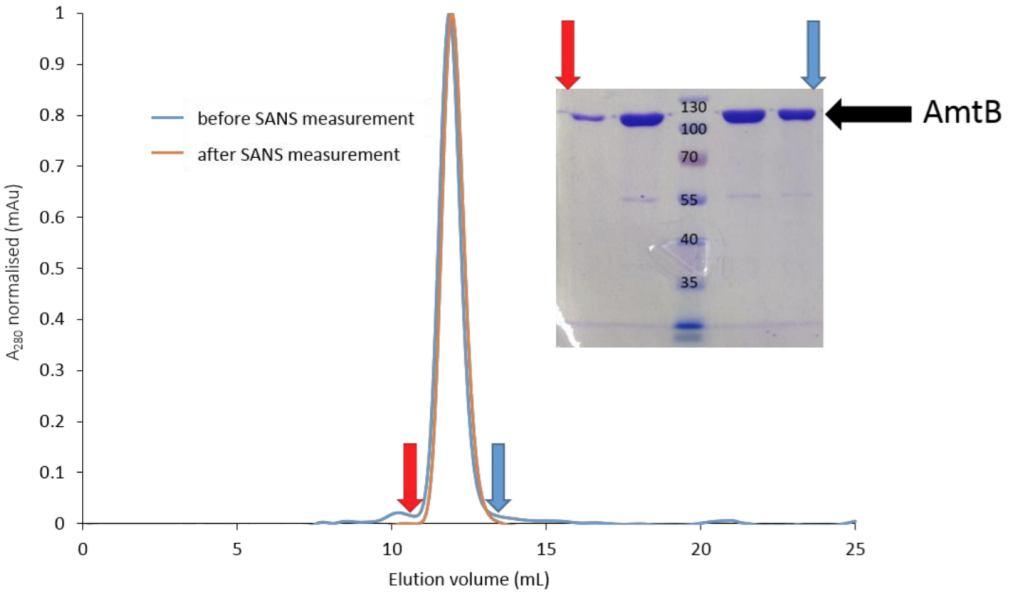

FIGURE 6.10: Size exclusion chromatography profile of AmtB solubilised in 0.03% DDM, 0% D_2O, before and after SANS experiments. Before SANS measurement, peak elution 11.87 ml and after SANS measurement, peak elution 11.96 ml. (insert) 12.5% SDS-PAGE fraction analysis based of the elution peak.

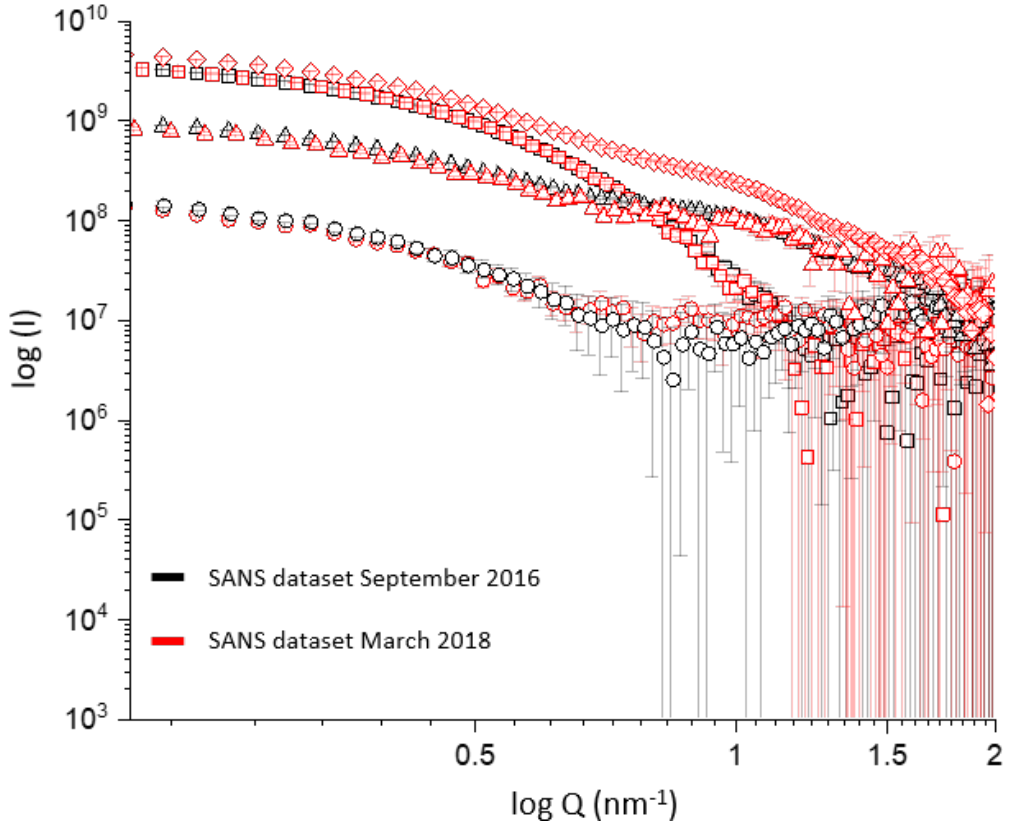

FIGURE 6.11: Comparison of the two SANS dataset. (square symbols) 0% D_2O, (curcle symbols) 22% D_2O, (triangle symbols) 42% D_2O. The dataset at (diamond symbols) 60% D_2O have only been measured in March 2018.

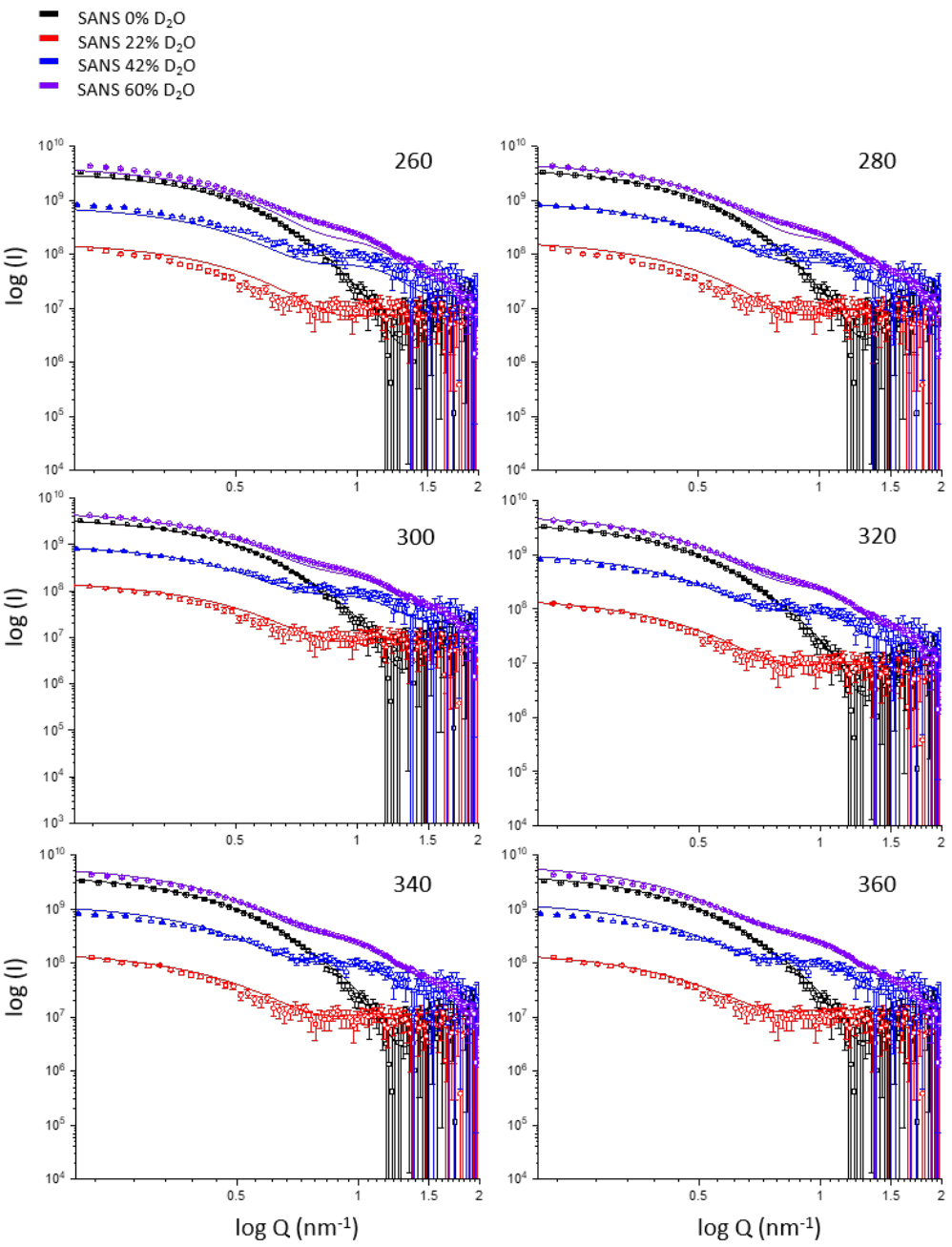

FIGURE 6.12: Comparison of experimental (symbols) and computed (lines) SANS curves from MD simulations containing between 260 and 360 DDM molecules

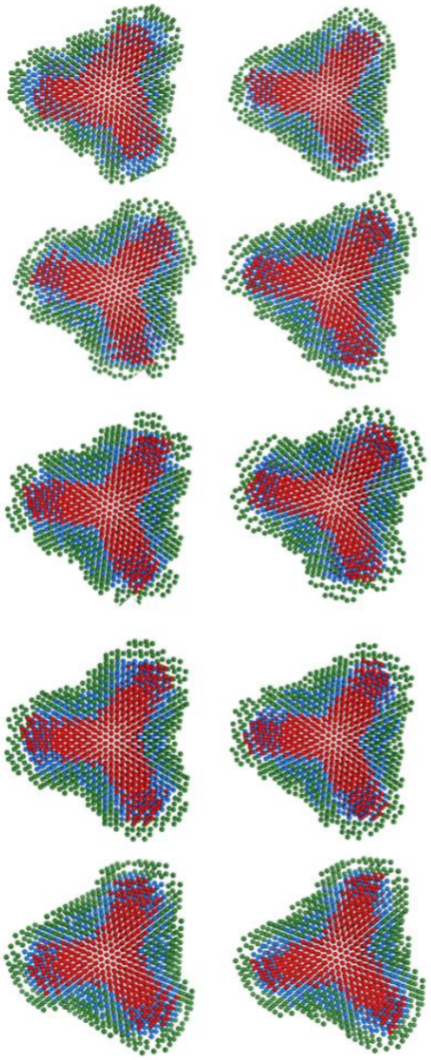

FIGURE 6.13: Top view of 10 different MONSA multiphase modelling using the SAXS and SANS data. The phase corresponding to the protein is represented by red beads, the hydrophilic and hydrophobic detergent density are represented by green and blue beads, respectively.

Chapter 7

Conclusions and Outlook

Detailed conclusions regarding specific projects presented in this thesis are given in the conclusion parts of Chapters 3-6 and will not be repeated here. Instead, we wish to note that the common thread connecting all projects presented in this thesis is the fact that the results obtained by combining SAS and MD simulations would be out of reach for SAS or MD alone. Therefore, the work presented here contributes to the emerging view that the direct coupling of various experiments with simulations represents a particularly promising direction in answering many important biophysical questions [16, 80, 374–377]. Of course, the integration of data from different sources (often referred to as integrative modeling or integrative structural biology) should be done with special attention to advantages, disadvantages and sources of uncertainties of each technique. Accordingly, the Bayesian formalism as well as the maximum entropy principle offer mathematically and statistically rigorous framework for further method developments for data integration.

To conclude this thesis, we present the outlook of the specific projects described in Chapters 3-6.

Ion cloud of charged proteins This project demonstrates how the influence of the ion cloud on R_g of the charged proteins can be accurately estimated using a greatly simplified spherical model of two proteins (BSA and GI) in conjunction with analytic calculations based on the Debye-Hückel theory. Good agreement between the results obtained using this simplified approach with the experimental and MD results, even for the nonspherical BSA, strongly suggests that the simplified approach could be successfully applied to any charged macromolecule in a salt solution. However, further simulations and SAXS experiments of other charged proteins, along with DNA and RNA molecules in various salt solutions would be necessary to draw final conclusions.

Atomic models of detergent micelles The size and the shape of detergent micelles described in Chapter 4 are obtained by directly coupling the experimental SAXS curve to the MD simulations. The fact that a very weak experimentally derived potential was sufficient to correct for the minor force field imperfection suggests that the derived models are not overfitted. However, a direct comparison with additional, orthogonal experimental data would be desirable. Unfortunately, to this end, additional experimental data for DDM and DM micelles are not available.
Furthermore, detecting a discrepancy between unbiased MD simulations and experimental data, as done here, is the first step towards improving force fields. Notably, in response to our work, we were contacted by the CHARMM developers. To support further CHARMM force field developments, we shared the output of our simulations.
Finally, while the *average* shape of detergent micelles obtained here was expected to be correct, we did not expect the magnitude of shape fluctuations to be correct. This is because shape fluctuations may be suppressed upon restraining the single-replica simulation with a harmonic restraint on the data, thereby violating the principle of maximum entropy, as already demonstrated in Ref. 206, for different type of experimental data. Indeed, in the follow-up study (Chapter 5), we demonstrated that the shape of the DDM micelle can be accurately determined with single-replica simulations, but multi-replica refinement with commitment to the principle of maximum entropy is necessary to obtain a more reliable *ensemble* of micelle structures.

Effects of asymmetry, shape fluctuations, disorder, and atomic details on the SAXS curve of detergent micelles Size monodispersity of DDM micelles along with the statistically rigorous approach to integrate SAXS data into MD simulations enabled us to study contributions of different effects on the SAXS curve of detergent micelles. However, not all micelles are monodisperse. Some detergent micelles adopt a wide distribution of sizes in solutions. Different studies of micelle size polydispersity were conducted in the past, but it remained unclear to which extent the assumption about the underlying distribution biased the interpretation of low-resolution experimental data [112, 301, 331, 378–382]. In principle, integration of SAXS data into multi-replica simulations should enable us to study the size polydispersity of micelles that adopt a wide distribution of sizes in solutions. Currently, in collaboration with Lise Arleth Group (University of Copenhagen), we are trying to understand the size polydispersity and shape of sodium dodecyl sulfate (SDS) micelles. Specifically, in addition to hydrophobic and polar effects that determine the aggregation number (N_{agg}) of the electrostatically neutral micelles (Chapters 4 and 5), N_{agg} of SDS micelles is influenced by the effects originating from the negative charge of detergent headgroups. Namely, electrostatic repulsion of the negatively charged SDS headgroups at a low concentration of the salt in a

solution would favor low N_{agg}. By increasing the salt concentration, N_{agg} is increasing, as the screening of headgroup-headgroup repulsion is increasing with the concentration of positive ions in a solution. Furthermore, the increase of detergent concentration causes a non-linear increase of N_{agg} [112, 331, 378–381]. A previous study suggested that varying detergent and salt concentrations may cause tremendous changes of N_{agg}. For example, at low detergent and moderate salt concentration, N_{agg} was estimated to be around 50. In contrast, the increase of detergent and salt concentrations led to the creation of extremely large, worm-like SDS micelles, whose average N_{agg} was estimated to be close to 9000 [112]. To validate and quantify the influence of detergent and salt concentrations on average N_{agg} of SDS micelles, but also to understand how the size polydispersity depends on salt and detergent concentrations, we are trying to interpret the SAXS curves of SDS micelles measured at different detergent and salt concentrations. We believe that the direct coupling of experimental SAXS data and MD simulations with a large number of replicas provides a reasonable framework for our investigation.
Finally, we wish to note that the two-component tri-axial ellipsoid models of micelles presented in this thesis may be easily applied to different systems. For example, even the two-component spherical model enabled us to predict the SAXS curves of golden and silica nanoparticles in different organic solvents with the reasonable accuracy. Golden and silica nanoparticles are studied because they represent a promising tool for the drug delivery [383]. Therefore, fast calculations of SAXS curves using simplified models, as described in Chapter 5, could be useful in choosing the optimal particle size and the type of solvent for further SAXS experiments.

Combining SAXS and SANS data to study protein-detergent complexes In the work presented in Chapter 6, we demonstrated how SAXS and SANS data can be combined to enable a fine structural analysis of protein-detergent complexes. However, we want to note that the single matching point (22% D_2O) was used to render the detergent "invisible" for the SANS experiment. While 22% D_2O represents the matching point of the DDM detergent *on average*, DDM headgroups and tails have distinct matching points. Therefore, conducting SANS experiments with two matching points vis-a-vis the detergent - one that renders detergent headgroups "invisible" and one that makes detergent tails "invisible" - would represent a more rigorous procedure for further studies. Additionally, we wish to point out that the structure of the protein-detergent complex was assumed to be nearly identical at each D_2O concentration. To quantify the influence of varying D_2O concentration on the protein-detergent complex structure, conducting SAXS experiment at different D_2O concentrations would be valuable.

List of Figures

List of Tables

Acronyms

AUC	Analytical ultracentrifugation
BSA	Bovine serum albumin
CG	Coarse-grained
cmc	critical micelle concentration
cryo-EM	Cryogenic electron microscopy
CV	Collective variable
DDM	n-dodecyl-β-D-maltoside
DM	n-decyl-β-D-maltoside
EM	Electron microscopy
FRET	Förster resonance energy transfer
GI	Glucose isomerase
IDPs	Intrinsically disordered proteins
LJ	Lennard-Jones
MD	Molecular dynamics
NMR	Nuclear magnetic resonance
PB	Poisson-Boltzmann
PDB	Protein data bank
R_g	Radius of gyration
SAS	Small-angle scattering
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size Exclusion Chromatography
SEC-MALS	Size exclusion chromatography in-line with multiangle light scattering
SLD	Scattering length density
TR-SAXS	Time-resolved SAXS
WAXS	Widel-angle X-ray scattering

Bibliography

[1] Jack, A.; Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. *Acta Crystallogr. A* **1978**, *34*, 931–935.

[2] Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **2007**, *40*, 191–285.

[3] Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole Ii, F. L.; Tsutakawa, S. E.; Jenney Jr, F. E.; Tainer, J. A. Robust, high-throughput solution structural analyses by small angle X-ray scattering SAXS. *Nat. Methods* **2009**, *6*, 606.

[4] Heller, W. T. Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures. *Acta Cryst. D* **2010**, *66*, 1213–1217.

[5] Schneidman-Duhovny, D.; Kim, S. J.; Sali, A. Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct. Biol.* **2012**, *12*, 17.

[6] Rambo, R. P.; Tainer, J. A. Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu. Rev. Biophys.* **2013**, *42*, 415–441.

[7] Schneidman-Duhovny, D.; Pellarin, R.; Sali, A. Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* **2014**, *28*, 96–104.

[8] RCSB PDB. `http://www.rcsb.org/`.

[9] Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol.* **2014**, *10*, e1003406.

[10] Schröder, G. F. Hybrid methods for macromolecular structure determination: experiment with expectations. *Curr. Opin. Struct. Biol.* **2015**, *31*, 20–27.

[11] Fernandez-Leiro, R.; Scheres, S. H. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **2016**, *537*, 339–346.

[12] Ravera, E.; Sgheri, L.; Parigi, G.; Luchinat, C. A critical assessment of methods to recover information from averaged data. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5686–5701.

[13] Allison, J. R. Using simulation to interpret experimental data in terms of protein conformational ensembles. *Curr. Opin. Struct. Biol.* **2017**, *43*, 79–87.

[14] Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116.

[15] Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 18–26.

[16] Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355–360.

[17] Brosey, C. A.; Tainer, J. A. Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Curr. Opin. Struct. Biol.* **2019**,

[18] Biochem., A. R. Biophysical Techniques in Structural Biology. *Annu. Rev. Biochem.* **2019**, *88*.

[19] Burley, S. K.; Berman, H. M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J. M.; Dutta, S. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* **2018**, *47*, D464–D474.

[20] Ganten, D.; Ruckpaul, K. *Encyclopedic reference of genomics and proteomics in molecular medicine*; Springer, 2006.

[21] Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R.; Wyckoff, H.; Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **1958**, *181*, 662–666.

[22] Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature* **1960**, *185*, 416.

[23] Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst. D* **2019**, *75*, 861–877.

[24] Kikhney, A. G.; Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577.

[25] Frank, J. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu. Rev. Biophys.* **2002**, *31*, 303–319.

[26] James, T. L. Fundamentals of NMR. *Online Textbook: Department of Pharmaceutical Chemistry, University of California, San Francisco* **1998**, 1–31.

[27] Wüthrich, K. The way to NMR structures of proteins. *Nat. Struct. Mol. Biol.* **2001**, *8*, 923.

[28] Takashima, H. High-resolution protein structure determination by NMR. *Annu. Rep. NMR Spectros.* **2006**, *59*, 235–273.

[29] Doniach, S.; Lipfert, J. *Small and Wide Angle X-ray Scattering from Biological Macromolecules and their Complexes in Solution*; Elsevier BV, 2012.

[30] Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. V.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; Wang, Y.; Gumbart, J. C. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *J. Chem. Inf. Model* **2016**, *56*, 721–733.

[31] Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *PNAS* **1997**, *94*, 12366–12371.

[32] Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. NMR analysis of a 900K GroEL–GroES complex. *Nature* **2002**, *418*, 207.

[33] Milne, J. L.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S. Cryo-electron microscopy–a primer for the non-microscopist. *FEBS J.* **2013**, *280*, 28–45.

[34] Frank, J.; Penczek, P.; Grassucci, R.; Srivastava, S. Three-dimensional reconstruction of the 70S Escherichia coli ribosome in ice: the distribution of ribosomal RNA. *J. Cell Biol.* **1991**, *115*, 597–605.

[35] Gabashvili, I. S.; Agrawal, R. K.; Spahn, C. M.; Grassucci, R. A.; Svergun, D. I.; Frank, J.; Penczek, P. Solution structure of the E. coli 70S ribosome at 11.5 Å resolution. *Cell* **2000**, *100*, 537–549.

[36] Bai, X.-C.; McMullan, G.; Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015**, *40*, 49–57.

[37] Kühlbrandt, W. The resolution revolution. *Science* **2014**, *343*, 1443–1444.

[38] Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **2015**, *13*, 24.

[39] Vinothkumar, K. R.; Henderson, R. Single particle electron cryomicroscopy: trends, issues and future perspective. *Q. Rev. Biophys.* **2016**, *49*.

[40] Frank, J. Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nat. Protoc.* **2017**, *12*, 209.

[41] Murata, K.; Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2018**, *1862*, 324–334.

[42] Kim, J. S.; Afsari, B.; Chirikjian, G. S. Cross-validation of data compatibility between small angle X-ray scattering and cryo-electron microscopy. *Journal of Computational Biology* **2017**, *24*, 13–30.

[43] Torres, J.; Stevens, T. J.; Samsó, M. Membrane proteins: the 'Wild West' of structural biology. *Trends Biochem. Sci.* **2003**, *28*, 137–144.

[44] Cantor, C. R.; Schimmel, P. R. *Biophysical chemistry: Part II: Techniques for the study of biological structure and function*; Macmillan, 1980.

[45] Dewey, T. G. Biophysical and biochemical aspects of fluorescence spectroscopy. **1991**,

[46] Truong, K.; Ikura, M. The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Curr. Opin. Struct. Biol.* **2001**, *11*, 573–578.

[47] Chan, F. K.-M.; Siegel, R. M.; Zacharias, D.; Swofford, R.; Holmes, K. L.; Tsien, R. Y.; Lenardo, M. J. Fluorescence resonance energy transfer analysis of cell surface receptor interactions and signaling using spectral variants of the green fluorescent protein. *Cytometry* **2001**, *44*, 361–368.

[48] Heyduk, T. Measuring protein conformational changes by FRET/LRET. *Curr. Opin. Biotech.* **2002**, *13*, 292–296.

[49] Aznauryan, M.; Delgado, L.; Soranno, A.; Nettels, D.; Huang, J.-r.; Labhardt, A. M.; Grzesiek, S.; Schuler, B. Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *PNAS* **2016**, *113*, E5389–E5398.

[50] Pollok, B. A.; Heim, R. Using GFP in FRET-based applications. *Trends Cell Biol.* **1999**, *9*, 57–60.

[51] Maurel, D.; Comps-Agrar, L.; Brock, C.; Rives, M.-L.; Bourrier, E.; Ayoub, M. A.; Bazin, H.; Tinel, N.; Durroux, T.; Prézeau, L. Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat. Methods* **2008**, *5*, 561.

[52] Silvius, J. R.; Nabi, I. R. Fluorescence-quenching and resonance energy transfer studies of lipid microdomains in model and biological membranes. *Mol. Membr. Biol.* **2006**, *23*, 5–16.

[53] Leavesley, S. J.; Rich, T. C. Overcoming limitations of FRET measurements. *Cytometry A* **2016**, *89*, 325.

[54] Haris, P. I. Infrared spectroscopy of protein structure. *Encyclopedia of Biophysics* **2013**, 1095–1106.

[55] Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Mol. Biol* **2000**, *7*, 724.

[56] Bunaciu, A. A.; Aboul-Enein, H. Y.; Hoang, V. D. Raman Spectroscopy for Protein Analysis. *Appl. Spectrosc. Rev.* **2015**, *50*, 377–386.

[57] 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.

[58] Guinier, A.; Fournet, G.; Yudowitch, K. L. Small-angle scattering of X-rays. **1955**,

[59] Lipfert, J.; Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 307–327.

[60] Koch, M. H.; Vachette, P.; Svergun, D. I. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quart. Rev. Biophys.* **2003**, *36*, 147–227.

[61] Graewert, M. A.; Svergun, D. I. Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* **2013**, *23*, 748–754.

[62] Hammel, M.; Amlanjyoti, D.; Reyes, F. E.; Chen, J.-H.; Parpana, R.; Tang, H. Y.; Larabell, C. A.; Tainer, J. A.; Adhya, S. HU multimerization shift controls nucleoid compaction. *Sci. Adv.* **2016**, *2*, e1600650.

[63] Rutsdottir, G.; Härmark, J.; Weide, Y.; Hebert, H.; Rasmussen, M. I.; Wernersson, S.; Respondek, M.; Akke, M.; Højrup, P.; Koeck, P. J.; Emanuelsson, C. Structural model of dodecameric heat-shock protein Hsp21: Flexible N-terminal arms interact with client proteins while C-terminal tails maintain the dodecamer and chaperone activity. *J. Biol. Chem.* **2017**, *292*, 8103–8121.

[64] De Nardis, C.; Lössl, P.; Van Den Biggelaar, M.; Madoori, P. K.; Leloup, N.; Mertens, K.; Heck, A. J.; Gros, P. Recombinant expression of the full-length ectodomain of LDL receptor-related protein 1 (LRP1) unravels pH-dependent conformational changes and the stoichiometry of binding with receptor-associated protein (RAP). *J. Biol. Chem.* **2017**, *292*, 912–924.

[65] Zhang, Z.; Liang, W. G.; Bailey, L. J.; Tan, Y. Z.; Wei, H.; Wang, A.; Farcasanu, M.; Woods, V. A.; McCord, L. A.; Lee, D.; Tang, W.-J. Ensemble cryoEM elucidates the mechanism of insulin capture and degradation by human insulin degrading enzyme. *eLife* **2018**, *7*, e33572.

[66] Boldon, L.; Laliberte, F.; Liu, L. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Rev.* **2015**, *6*, 25661.

[67] Bizien, T.; Durand, D.; Roblina, P.; Thureau, A.; Vachette, P.; Pérez, J. A brief Survey of State-of-the-Art BioSAXS. *Protein Pept. Lett.* **2016**, *23*, 217–231.

[68] Trewhella, J. Small-angle scattering and 3D structure interpretation. *Curr. Opin. Struct. Biol.* **2016**, *40*, 1–7.

[69] Hura, G. L.; Budworth, H.; Dyer, K. N.; Rambo, R. P.; Hammel, M.; McMurray, C. T.; Tainer, J. A. Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* **2013**, *10*, 453.

[70] Tang, H. Y.; Tainer, J. A.; Hura, G. L. *Biological Small Angle Scattering: Techniques, Strategies and Tips*; Springer, 2017; pp 167–181.

[71] Chen, P.-c.; Hub, J. S. Structural Properties of Protein-Detergent Complexes from SAXS and MD Simulations. *J. Phys. Chem. Lett.* **2015**, *6*, 5116–5121.

[72] Hammel, M. Validation of macromolecular flexibility in solution by small-angle X-ray scattering SAXS. *Eur. Biophys. J.* **2012**, *41*, 789–799.

[73] Jeffries, C. M.; Graewert, M. A.; Blanchet, C. E.; Langley, D. B.; Whitten, A. E.; Svergun, D. I. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat. Protoc.* **2016**, *11*, 2122–2153.

[74] Konarev, P. V.; Petoukhov, M. V.; Volkov, V. V.; Svergun, D. I. ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2006**, *39*, 277–286.

[75] Pérez, J.; Vachette, P. *Biological Small Angle Scattering: Techniques, Strategies and Tips*; Springer, 2017; pp 183–199.

[76] Malaby, A. W.; Chakravarthy, S.; Irving, T. C.; Kathuria, S. V.; Bilsel, O.; Lambright, D. G. Methods for analysis of size-exclusion chromatography–small-angle X-ray scattering and reconstruction of protein scattering. *J. Appl. Crystallogr.* **2015**, *48*, 1102–1113.

[77] Cho, H. S.; Dashdorj, N.; Schotte, F.; Graber, T.; Henning, R.; Anfinrud, P. Protein structural dynamics in solution unveiled via 100-ps time-resolved X-ray scattering. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 7281–7286.

[78] Arnlund, D.; Johansson, L. C.; Wickstrand, C.; Barty, A.; Williams, G. J.; Malmerberg, E.; Davidsson, J.; Milathianaki, D.; DePonte, D. P.; Shoeman, R. L. Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat. Methods* **2014**, *11*, 923–926.

[79] Levantino, M.; Schiro, G.; Lemke, H. T.; Cottone, G.; Glownia, J. M.; Zhu, D.; Chollet, M.; Ihee, H.; Cupane, A.; Cammarata, M. Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nature Commun.* **2015**, *6*, 6772.

[80] Russel, D.; Lasker, K.; Webb, B.; Velázquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS biol.* **2012**, *10*, e1001244.

[81] Round, A.; Brown, E.; Marcellin, R.; Kapp, U.; Westfall, C. S.; Jez, J. M.; Zubieta, C. Determination of the GH3. 12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography. *Acta Crystallogr. D* **2013**, *69*, 2072–2080.

[82] Chen, B.; Zuo, X.; Wang, Y.-X.; Dayie, T. K. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.* **2011**, *40*, 3117–3130.

[83] Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Girona, G. E.; Milles, S.; Mishra, A. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *PNAS* **2017**, *114*, E6342–E6351.

[84] Meisburger, S. P.; Sutton, J. L.; Chen, H.; Pabit, S. A.; Kirmizialtin, S.; Elber, R.; Pollack, L. Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: Beyond the wormlike chain model. *Biopolymers* **2013**, *99*, 1032–1045.

[85] Best, R. B.; Zheng, W.; Borgia, A.; Buholzer, K.; Borgia, M. B.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A. Comment on Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **2018**, *361*, eaar7101.

[86] Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schuler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A. Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J. Am. Chem. Soc.* **2016**, *138*, 11714–11726.

[87] Merzel, F.; Smith, J. C. 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 Cryst. D* **2002**, *58*, 242–249.

[88] 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*, 134114.

[89] Köfinger, J.; Hummer, G. Atomic-resolution structural information from scattering experiments on macromolecules in solution. *Phys. Rev. E.* **2013**, *87*, 052712.

[90] Chen, P.; Hub, J. S. Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophys. J.* **2014**, *107*, 435–447.

[91] Chen, P.; Hub, J. S. Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics. *Biophys. J.* **2015**, *108*, 2573–2584.

[92] Shevchuk, R.; Hub, J. S. Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005800.

[93] 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.

[94] 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.

[95] Feigin, L. A.; Svergun, D. I. *Structure analysis by small-angle X-ray and neutron scattering*; Springer, 1987.

[96] Guo, X.; Zhao, N.; Chen, S.; Teixeira, J. Small-angle neutron scattering study of the structure of protein/detergent complexes. *Biopolymers* **1990**, *29*, 335–346.

[97] Gabel, F.; Bicout, D.; Lehnert, U.; Tehei, M.; Weik, M.; Zaccai, G. Protein dynamics studied by neutron scattering. *Q. Rev. Biophys.* **2002**, *35*, 327–367.

[98] Svergun, D.; Richard, S.; Koch, M.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: experimental observation by X-ray and neutron scattering. *PNAS* **1998**, *95*, 2267–2272.

[99] 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* **2015**, *71*, 57–66.

[100] Virtanen, J. J.; Makowski, L.; Sosnick, T. R.; Freed, K. F. Modeling the hydration layer around proteins: applications to small-and wide-angle X-ray scattering. *Biophys. J.* **2011**, *101*, 2061–2069.

[101] Oroguchi, T.; Hashimoto, H.; Shimizu, T.; Sato, M.; Ikeguchi, M. Intrinsic Dynamics of Restriction Endonuclease *Eco*O109I Studied by Molecular Dynamics Simulations and X-Ray Scattering Data Analysis. *Biophys. J.* **2009**, *96*, 2808–2822.

[102] 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.

[103] Makarov, V.; Pettitt, B. M.; Feig, M. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment.

[104] Persson, F.; Söderhjelm, P.; Halle, B. The geometry of protein hydration. *J. Chem. Phys.* **2018**, *148*, 215101.

[105] Columbus, L.; Lipfert, J.; Jambunathan, K.; Fox, D. A.; Sim, A. Y.; Doniach, S.; Lesley, S. A. Mixing and matching detergents for membrane protein NMR structure determination. *J. Am. Chem. Soc.* **2009**, *131*, 7320.

[106] O'Malley, M. A.; Helgeson, M. E.; Wagner, N. J.; Robinson, A. S. Toward rational design of protein detergent complexes: determinants of mixed micelles that are critical for the in vitro stabilization of a G-protein coupled receptor. *Biophys. J.* **2011**, *101*, 1938–1948.

[107] Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S. Size and shape of detergent micelles determined by small-angle X-ray scattering. *J. Phys. Chem. B* **2007**, *111*, 12427–12438.

[108] Oliver, R. C.; Lipfert, J.; Fox, D. A.; Lo, R. H.; Doniach, S.; Columbus, L. Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLOS ONE* **2013**, *8*, e62488.

[109] Dupuy, C.; Auvray, X.; Petipas, C.; Rico-Lattes, I.; Lattes, A. Anomeric effects on the structure of micelles of alkyl maltosides in water. *Langmuir* **1997**, *13*, 3965–3967.

[110] Tse-Ve-Koon, K.; Tremblay, N.; Constantin, D.; Freyssingeas, É. Structure, thermodynamics and dynamics of the isotropic phase of spherical non-ionic surfactant micelles. *Colloid Interface Sci.* **2013**, *393*, 161–173.

[111] Hayter, J. B.; Zemb, T. Concentration-dependent structure of sodium octanoate micelles. *Chem. Phys. Lett.* **1982**, *93*, 91–94.

[112] Arleth, L.; Bergström, M.; Pedersen, J. S. Small-angle neutron scattering study of the growth behavior, flexibility, and intermicellar interactions of wormlike SDS micelles in NaBr aqueous solutions. *Langmuir* **2002**, *18*, 5343–5353.

[113] Watanabe, K.; Ferrario, M.; Klein, M. L. Molecular dynamics study of a sodium octanoate micelle in aqueous solution. *J. Phys. Chem.* **1988**, *92*, 819–821.

[114] Abel, S.; Dupradeau, F.-Y.; Raman, E. P.; MacKerell Jr, A. D.; Marchi, M. Molecular simulations of dodecyl-β-maltoside micelles in water: influence of the headgroup conformation and force field parameters. *J. Phys. Chem. B* **2010**, *115*, 487–499.

[115] Tieleman, D.; Van der Spoel, D.; Berendsen, H. Molecular dynamics simulations of dodecylphosphocholine micelles at three different aggregate sizes: micellar structure and chain relaxation. *J. Phys. Chem. B* **2000**, *104*, 6380–6388.

[116] Marrink, S.; Tieleman, D.; Mark, A. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *J. Phys. Chem. B* **2000**, *104*, 12165–12173.

[117] Bruce, C. D.; Berkowitz, M. L.; Perera, L.; Forbes, M. D. Molecular dynamics simulation of sodium dodecyl sulfate micelle in water: micellar structural characteristics and counterion distribution. *J. Phys. Chem. B* **2002**, *106*, 3788–3793.

[118] Hoffmaster, A.; Bonnett, B.; Faramarzi, S.; Grodi, D.; Mertz, B.; Harvey, E. Molecular dynamics investigation of dodecylmaltoside micelle properties. *Proc. West Virg. Acad. Sci.* **2016**, *88*.

[119] Maibaum, L.; Dinner, A. R.; Chandler, D. Micelle formation and the hydrophobic effect. *J. Phys. Chem. B* **2004**, *108*, 6778–6781.

[120] Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes 2d Ed*; J. Wiley., 1980.

[121] Iijima, H.; Kato, T.; Yoshida, H.; Imai, M. Small-angle X-ray and neutron scattering from dilute solutions of cesium perfluorooctanoate. Micellar growth along two dimensions. *J. Phys. Chem. B* **1998**, *102*, 990–995.

[122] Pedersen, M. C.; Arleth, L.; Mortensen, K. WillItFit: a framework for fitting of constrained models to small-angle scattering data. *J. Appl. Crystallogr.* **2013**, *46*, 1894–1898.

[123] Gabel, F.; Engilberge, S.; Pérez, J.; Girard, E. Medical contrast media as possible tools for SAXS contrast variation. *IUCrJ* **2019**, *6*.

[124] Khusenov, M. A.; Dushanov, E. B.; Kholmurodov, K. T. Molecular dynamics simulations of the DNA-CNT interaction process: Hybrid quantum chemistry potential and classical trajectory approach. *J. Mod. Phys.* **2014**, *5*, 137.

[125] Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.

[126] Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646.

[127] Karplus, M.; Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **1990**, *347*, 631.

[128] Berendsen, H. J. C. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*; Cambridge University Press, 2007.

[129] Born, M.; Oppenheimer, R. Zur quantentheorie der molekeln. *Annalen der physik* **1927**, *389*, 457–484.

[130] Arndt, M.; Juffmann, T.; Vedral, V. Quantum physics meets biology. *HFSP J.* **2009**, *3*, 386–400.

[131] Kiefer, C. Emergence of a classical Universe from quantum gravity and cosmology. *Philos. Trans. R. Soc. A* **2012**, *370*, 4566–4575.

[132] Joos, E.; Zeh, H. D.; Kiefer, C.; Giulini, D. J.; Kupsch, J.; Stamatescu, I.-O. *Decoherence and the appearance of a classical world in quantum theory*; Springer Science & Business Media, 2013.

[133] MacKerell Jr, A. D. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

[134] Ponder, J. W.; Case, D. A. *Advances in protein chemistry*; Elsevier, 2003; Vol. 66; pp 27–85.

[135] 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: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

[136] MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[137] MacKerell Jr, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105–120.

[138] Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

[139] Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.

[140] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

[141] Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

[142] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **2004**, *25*, 1157–1174.

[143] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

[144] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.* **1995**, *91*, 43–56.

[145] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

[146] 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*, 19–25.

[147] Hockney, R.; Goel, S.; Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974**, *14*, 148–158.

[148] 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.

[149] Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[150] Feenstra, K. A.; Hess, B.; Berendsen, H. J. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **1999**, *20*, 786–798.

[151] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. a.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

[152] Saito, M. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. *J. Chem. Phys.* **1994**, *101*, 4055–4061.

[153] Saito, M. Molecular Dynamic/Free Energy Study of a Protein in Solution with All Degrees of Freedom and Long-Range Coulomb Interactions. *J. Phys. Chem.* **1995**, *99*, 17043–17048.

[154] 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.

[155] 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.

[156] Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *64*, 253–287.

[157] Berendsen, H. *Computer Simulation in Materials Science*; Springer, 1991; pp 139–155.

[158] Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **2009**, *130*, 074101.

[159] Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[160] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

[161] Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

[162] Lindorff-Larsen, K.; Piana, S.; Dror, R.; Shaw, D. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.

[163] Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PLOS ONE* **2012**, *7*, e32131.

[164] Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129–138.

[165] Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71.

[166] Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *PNAS* **2018**, *115*, E4758–E4766.

[167] Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409–1414.

[168] Leonard, A. N.; Wang, E.; Monje-Galvan, V.; Klauda, J. B. Developing and testing of lipid force fields with applications to modeling cellular membranes. *Chem. Rev.* **2019**, *119*, 6227–6269.

[169] McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernández, F.; Hargus, C.; Law, K.-H.; Klepeis, J. L.; Shaw, D. E. Improving the accuracy of Møller-Plesset perturbation theory with neural networks. *J. Chem. Phys.* **2017**, *147*, 161725.

[170] Bereau, T.; DiStasio Jr, R. A.; Tkatchenko, A.; Von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning. *J. Chem. Phys.* **2018**, *148*, 241706.

[171] Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.

[172] Chen, J.; Chen, J.; Pinamonti, G.; Clementi, C. Learning effective molecular models from experimental observables. *J. Chem. Theory Comput.* **2018**, *14*, 3849–3858.

[173] Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable force fields: history, test cases, and prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

[174] Lopes, P. E.; Roux, B.; MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11–28.

[175] Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys. Condens. Matte* **2009**, *21*, 333102.

[176] Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettitt, B. M.; Phillips Jr., G. N. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3288–3292.

[177] Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *PNAS* **2005**, *102*, 13885–13890.

[178] Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J. Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the conference on high performance computing networking, storage and analysis. 2009; p 39.

[179] Kalyaanamoorthy, S.; Chen, Y.-P. P. Modelling and enhanced molecular dynamics to steer structure-based drug discovery. *Prog. Biophys. Mol. Biol.* **2014**, *114*, 123–136.

[180] Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H. More bang for your buck: improved use of GPU nodes for GROMACS 2018. *J. Comp. Chem.* in press.

[181] Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers1, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341–346.

[182] Orsi, M.; Sanderson, W. E.; Essex, J. W. Permeability of small molecules through a lipid bilayer: a multiscale simulation study. *J Phys Chem B* **2009**, *113*, 12019–12029.

[183] Bennett, W. D.; Tieleman, D. P. Water defect and pore formation in atomistic and coarse-grained lipid membranes: pushing the limits of coarse graining. *J. Chem. Theory Comput.* **2011**, *7*, 2981–2988.

[184] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

[185] Waheed, Q.; Tjörnhammar, R.; Edholm, O. Phase transitions in coarse-grained lipid bilayers containing cholesterol by molecular dynamics simulations. *Biophys. J.* **2012**, *103*, 2125–2133.

[186] Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423–12430.

[187] Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750–760.

[188] Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144.

[189] Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noe, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542.

[190] Noé, F.; Rosta, E. Special Topic: Markov Models of Molecular Kinetics. *arXiv preprint arXiv:1911.00774* **2019**,

[191] Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *J. Chem. Phys.* **2013**, *139*, 11B609_1.

[192] Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. Rapid equilibrium sampling initiated from nonequilibrium data. *PNAS* **2009**, *106*, 19765–19769.

[193] Hinrichs, N. S.; Pande, V. S. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.* **2007**, *126*, 244101.

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

[195] Valsson, O.; Parrinello, M. Variational approach to enhanced sampling and free energy calculations. *Phys. Rev. Lett.* **2014**, *113*, 090601.

[196] Laio, A.; Parrinello, M. Escaping free-energy minima. *PNAS* **2002**, *99*, 12562–12566.

[197] Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

[198] Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta,* **2015**, *1850*, 872–877.

[199] Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a practical method for adaptive QM/MM simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.

[200] Seabra, G. d. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package. *J. Phys. Chem. A* **2007**, *111*, 5655–5664.

[201] Senn, H. M.; Thiel, W. *Atomistic approaches in modern biology*; Springer, 2006; pp 173–290.

[202] Kamerlin, S. C.; Haranczyk, M.; Warshel, A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J. Phys. Chem. B* **2008**, *113*, 1253–1272.

[203] Schaeffer, R. D.; Fersht, A.; Daggett, V. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr. Opin. Struct. Biol.* **2008**, *18*, 4–9.

[204] Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian inference method for heterogeneous systems. *Sci. Adv.* **2016**, *2*, e1501177.

[205] Rangan, R.; Bonomi, M.; Heller, G. T.; Cesari, A.; Bussi, G.; Vendruscolo, M. Determination of structural ensembles of proteins: restraining vs reweighting. *J. Chem. Theory Comput.* **2018**, *14*, 6632–6641.

[206] Hummer, G.; Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **2015**, *143*, 12B634_1.

[207] Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138*, 084107.

[208] Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **2013**, *138*, 03B603.

[209] Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLOS Comput. Biol.* **2014**, *10*, e1003406.

[210] Putnam, D. K.; Lowe Jr, E. W.; Meiler, J. Reconstruction of SAXS Profiles from Protein Structures. *Comput. Struct. Biotechnol. J.* **2013**, *8*, e201308006.

[211] Cromer, D. T.; Mann, J. B. X-ray scattering factors computed from numerical Hartree-Fock wave functions. *Acta Cryst. A* **1968**, *24*, 321–324.

[212] Jacques, D. A.; Trewhella, J. Small-angle scattering for structural biologyExpanding the frontier while avoiding the pitfalls. *Protein Sci.* **2010**, *19*, 642–657.

[213] Skou, S.; Gillilan, R. E.; Ando, N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protoc.* **2014**, *9*, 1727–1739.

[214] Smolsky, I. L.; Liu, P.; Niebuhr, M.; Ito, K.; Weiss, T. M.; Tsuruta, H. Biological small-angle X-ray scattering facility at the Stanford Synchrotron Radiation Laboratory. *Appl. Crystallogr.* **2007**, *40*, s453–s458.

[215] Pernot, P.; Theveneau, P.; Giraud, T.; Fernandes, R. N.; Nurizzo, D.; Spruce, D.; Surr, J.; McSweeney, S.; Round, A.; Felisaz, F. New beamline dedicated to solution scattering from biological macromolecules at the ESRF. Journal of Physics: Conference Series. 2010; p 012009.

[216] Bruetzel, L. K.; Fischer, S.; Salditt, A.; Sedlak, S. M.; Nickel, B.; Lipfert, J. A Mo-anode-based in-house source for small-angle X-ray scattering measurements of biological macromolecules. *Rev. Sci. Instrum.* **2016**, *87*, 025103.

[217] Gulbis, J. M.; Kelman, Z.; Hurwitz, J.; O'Donnell, M.; Kuriyan, J. Structure of the C-terminal region of p21WAF1/CIP1 complexed with human PCNA. *Cell* **1996**, *87*, 297–306.

[218] Wong, E.; Vaaje-Kolstad, G.; Ghosh, A.; Hurtado-Guerrero, R.; Konarev, P. V.; Ibrahim, A. F.; Svergun, D. I.; Eijsink, V. G.; Chatterjee, N. S.; van Aalten, D. M. The Vibrio cholerae colonization factor GbpA possesses a modular structure that governs binding to different host surfaces. *PLOS Pathog.* **2012**, *8*, e1002373.

[219] Erzberger, J. P.; Stengel, F.; Pellarin, R.; Zhang, S.; Schaefer, T.; Aylett, C. H.; Cimermančič, P.; Boehringer, D.; Sali, A.; Aebersold, R.; Ban, N. Molecular architecture of the 40S eIF1 eIF3 translation initiation complex. *Cell* **2014**, *158*, 1123–1135.

[220] Franke, D.; Petoukhov, M.; Konarev, P.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.; Kikhney, A.; Hajizadeh, N.; Franklin, J.; Jeffries, C.; Svergun, D. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225.

[221] Mylonas, E.; Svergun, D. I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Applied Crystallography* **2007**, *40*, s245–s249.

[222] Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H.; Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2003**, *36*, 1277–1282.

[223] 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.* **2000**, *113*, 9149–9161.

[224] 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.

[225] Schneidman-Duhovny, D.; Hammel, M.; Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucl. Acids Res.* **2010**, *38*, W540–W544.

[226] 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.

[227] Clifton, L. A.; Hall, S. C.; Mahmoudi, N.; Knowles, T. J.; Heinrich, F.; Lakey, J. H. *Lipid-Protein Interactions*; Springer, 2019; pp 201–251.

[228] Whitten, A. E.; Trewhella, J. *Micro and nano technologies in bioanalysis*; Springer, 2009; pp 307–323.

[229] Johansen, D.; Jeffries, C. M.; Hammouda, B.; Trewhella, J.; Goldenberg, D. P. Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophys. J.* **2011**, *100*, 1120–1128.

[230] Stothart, P. H. Subunit structure of casein micelles from small-angle neutron-scattering. *J. Mol. Biol.* **1989**, *208*, 635–638.

[231] Sears, V. F. Neutron scattering lengths and cross sections. *Neutron News* **1992**, *3*, 26–37.

[232] Kimanius, D.; Pettersson, I.; Schluckebier, G.; Lindahl, E.; Andersson, M. SAXS-guided metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 3491–3498.

[233] Björling, A.; Niebling, S.; Marcellini, M.; van der Spoel, D.; Westenhoff, S. Deciphering solution scattering data with experimentally guided molecular dynamics simulations. *J. Chem. Theory Comput.* **2015**, *11*, 780–787.

[234] Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of Chemical Physics* **2013**, *138*, 084107.

[235] Rózycki, B.; Kim, Y. C.; Hummer, G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **2011**, *19*, 109–116.

[236] 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.

[237] Draper, D. E.; Grilley, D.; Soto, A. M. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 221–243.

[238] Wong, G. C.; Pollack, L. Electrostatics of strongly charged biological polymers: ion-mediated interactions and self-organization in nucleic acids and proteins. *Ann. Rev. Phys. Chem.* **2010**, *61*, 171–189.

[239] Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic acid–ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813–841.

[240] Wensel, T. G.; Meares, C. F.; Vlachy, V.; Matthew, J. B. Distribution of ions around DNA, probed by energy transfer. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3267–3271.

[241] Giambaşu, G. M.; Luchko, T.; Herschlag, D.; York, D. M.; Case, D. A. Ion counting from explicit-solvent simulations and 3D-RISM. *Biophys. J.* **2014**, *106*, 883–894.

[242] Kirmizialtin, S.; Pabit, S. A.; Meisburger, S. P.; Pollack, L.; Elber, R. RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations. *Biophys. J.* **2012**, *102*, 819–828.

[243] Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: theory and applications. *Annu Rev Biophys Biophys Chem* **1990**, *19*, 301–332.

[244] Pollack, L. SAXS studies of ion-nucleic acid interactions. *Annu. Rev. Biophys.* **2011**, *40*, 225–242.

[245] Kirmizialtin, S.; Silalahi, A. R.; Elber, R.; Fenley, M. O. The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophys. J.* **2012**, *102*, 829–838.

[246] Cheatham, T. E.; Young, M. A. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* **2000**, *56*, 232–256.

[247] Kirmizialtin, S.; Elber, R. Computational exploration of mobile ion distributions around RNA duplex. *J. Phys. Chem. B* **2010**, *114*, 8207–8220.

[248] Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. Exploring the counterion atmosphere around DNA: what can be learned from molecular dynamics simulations? *Biophys. J.* **2004**, *87*, 800–811.

[249] Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. Ion motions in molecular dynamics simulations on DNA. *Proc. Nat. Acad. Sci. U.S.A.* **2004**, *101*, 14771–14775.

[250] van Dam, L.; Lyubartsev, A. P.; Laaksonen, A.; Nordenskiöld, L. Self-diffusion and association of Li+, Cs+, and H2O in oriented DNA fibers. An NMR and MD simulation study. *J. Phys. Chem. B* **1998**, *102*, 10636–10642.

[251] Deserno, M.; Holm, C.; May, S. Fraction of condensed counterions around a charged rod: Comparison of Poisson-Boltzmann theory and computer simulations. *Macromolecules* **2000**, *33*, 199–206.

[252] Jayaram, B.; Beveridge, D. Modeling DNA in aqueous solutions: theoretical and computer simulation studies on the ion atmosphere of DNA. *Annu. Rev. Biophys.* **1996**, *25*, 367–394.

[253] Cheatham III, T. E.; Kollman, P. A. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435–471.

[254] Beveridge, D. L.; McConnell, K. J. Nucleic acids: theory and computer simulation, Y2K. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182–196.

[255] Nguyen, H. T.; Pabit, S. A.; Pollack, L.; Case, D. A. Extracting water and ion distributions from solution X-ray scattering experiments. *J. Chem. Phys.* **2016**, *144*, 214105.

[256] Das, R.; Mills, T.; Kwok, L.; Maskel, G.; Millett, I.; Doniach, S.; Finkelstein, K.; Herschlag, D.; Pollack, L. Counterion distribution around DNA probed by solution X-ray scattering. *Phys. Rev. Lett.* **2003**, *90*, 188103.

[257] Bai, Y.; Das, R.; Millett, I. S.; Herschlag, D.; Doniach, S. Probing counterion modulated repulsion and attraction between nucleic acid duplexes in solution. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1035–1040.

[258] Bai, Y.; Greenfeld, M.; Travers, K. J.; Chu, V. B.; Lipfert, J.; Doniach, S.; Herschlag, D. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* **2007**, *129*, 14981–14988.

[259] Pabit, S. A.; Meisburger, S. P.; Li, L.; Blose, J. M.; Jones, C. D.; Pollack, L. Counting ions around DNA with anomalous small-angle X-ray scattering. *J. Am. Chem. Soc.* **2010**, *132*, 16334–16336.

[260] Horkay, F.; Hecht, A. M.; Rochas, C.; Basser, P. J.; Geissler, E. Anomalous small angle X-ray scattering determination of ion distribution around a polyelectrolyte biopolymer in salt solution. *J. Chem. Phys* **2006**, *125*, 234904.

[261] Meisburger, S. P.; Pabit, S. A.; Pollack, L. Determining the locations of ions and water around DNA from X-ray scattering measurements. *Biophys. J.* **2015**, *108*, 2886–2895.

[262] Chu, V. B.; Bai, Y.; Lipfert, J.; Herschlag, D.; Doniach, S. Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory. *Biophys. J.* **2007**, *93*, 3202–3209.

[263] Eigen, M.; Wicke, E. The Thermodynamics of Electrolytes at Higher Concentration. *J. Phys. Chem.* **1954**, *58*, 702–714.

[264] Netz, R.; Orland, H. Beyond Poisson-Boltzmann: Fluctuation effects and correlation functions. *The European Physical Journal E* **2000**, *1*, 203–214.

[265] Borukhov, I.; Andelman, D.; Orland, H. Steric Effects in Electrolytes: A Modified Poisson-Boltzmann Equation. *Phys. Rev. Lett.* **1997**, *79*, 435–438.

[266] Luo, Y.; Roux, B. Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Letters* **2009**, *1*, 183–189.

[267] Noy, A.; Soteras, I.; Luque, F. J.; Orozco, M. The impact of monovalent ion force field model in nucleic acids simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596–10607.

[268] 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.

[269] Yoo, J.; Aksimentiev, A. Improved parametrization of Li+, Na+, K+, and Mg2+ ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.* **2011**, *3*, 45–50.

[270] Berthaud, A.; Manzi, J.; Pérez, J.; Mangenot, S. Modeling detergent organization around aquaporin-0 using small-angle X-ray scattering. *J. Am. Chem. Soc.* **2012**, *134*, 10080–10088.

[271] Zhang, F.; Skoda, M. W.; Jacobs, R. M.; Martin, R. A.; Martin, C. M.; Schreiber, F. Protein interactions studied by SAXS: effect of ionic strength and protein concentration for BSA in aqueous solutions. *J. Phys. Chem. B* **2007**, *111*, 251–259.

[272] Zhang, F.; Roosen-Runge, F.; Skoda, M. W.; Jacobs, R. M.; Wolf, M.; Callow, P.; Frielinghaus, H.; Pipich, V.; Prevost, S.; Schreiber, F. Hydration and interactions in protein solutions containing concentrated electrolytes studied by small-angle scattering. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2483–2493.

[273] Yang, S.; Park, S.; Makowski, L.; Roux, B. A rapid coarse residue-based computational method for X-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys. J.* **2009**, *96*, 4449–4463.

[274] Liu, H.; Hexemer, A.; Zwart, P. H. The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *J. Appl. Cryst.* **2012**, *45*, 587–593.

[275] 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.

[276] Pernot, P.; Round, A.; Barrett, R.; De Maria Antolinos, A.; Gobbo, A.; Gordon, E.; Huet, J.; Kieffer, J.; Lentini, M.; Mattenet, M. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **2013**, *20*, 660–664.

[277] Dill, K. A.; Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*, 2nd ed.; Garland Science, 2010.

[278] Holst, M. J. The Poisson-Boltzmann equation: Analysis and multilevel numerical solution. Ph.D. thesis, California institute of Technology, 1994.

[279] Marcus, Y. Ionic radii in aqueous solutions. *Chem. Rev.* **1988**, *88*, 1475–1498.

[280] Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *PNAS* **2001**, *98*, 10037–10041.

[281] Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

[282] Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525.

[283] 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.

[284] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi1$ and $\chi2$ dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

[285] Bujacz, A. Structures of bovine, equine and leporine serum albumin. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2012**, *68*, 1278–1289.

[286] Nowak, E.; Panjikar, S.; Tucker, P. A. Atomic structure of Glucose isomerase.

[287] 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.

[288] 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*, 19–25.

[289] Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J. Comp. Chem.* **1992**, *13*, 952–962.

[290] Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Mol Simul. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

[291] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

[292] Van Gunsteren, W.; Berendsen, H. A leap-frog algorithm for stochastic dynamics. *Mol Simul* **1988**, *1*, 173–185.

[293] Joung, I. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

[294] Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

[295] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. 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. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

[296] Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.

[297] Vrbka, L.; Vondrášek, J.; Jagoda-Cwiklik, B.; Vácha, R.; Jungwirth, P. Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15440–15444.

[298] 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.

[299] Jain, S.; Bates, F. S. On the origins of morphological complexity in block copolymer surfactants. *Science* **2003**, *300*, 460–464.

[300] Rosen, M. J.; Kunjappu, J. T. *Surfactants and interfacial phenomena*; John Wiley & Sons, 2012.

[301] Dwars, T.; Paetzold, E.; Oehme, G. Reactions in micellar systems. *Angew. Chem. Int. Ed.* **2005**, *44*, 7174–7199.

[302] Dwars, T.; Paetzold, E.; Oehme, G. Reaktionen in micellaren Systemen. *Angew. Chem.* **2005**, *117*, 7338–7364.

[303] Patist, A.; Kanicky, J. R.; Shukla, P. K.; Shah, D. O. Importance of micellar kinetics in relation to technological processes. *Colloid Interface Sci.* **2002**, *245*, 1–15.

[304] Geng, Y.; Dalhaimer, P.; Cai, S.; Tsai, R.; Tewari, M.; Minko, T.; Discher, D. E. Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat. Nanotechnol.* **2007**, *2*, 249–255.

[305] Diodati, S.; Dolcet, P.; Casarin, M.; Gross, S. Pursuing the Crystallization of Mono- and Polymetallic Nanosized Crystalline Inorganic Compounds by Low-Temperature Wet-Chemistry and Colloidal Routes. *Chem. Rev.* **2015**, *115*, 11449–11502.

[306] Balzer, D.; Luders, H. *Nonionic surfactants: Alkyl polyglucosides*; CRC Press, 2000; Vol. 91.

[307] *Global Surfactant Market*, 4th ed.; Acmite Market Intelligence, 2016.

[308] D'Andrea, M. G.; Domingues, C. C.; Malheiros, S. V.; Neto, F. G.; Barbosa, L. R.; Itri, R.; Almeida, F. C.; de Paula, E.; Bianconi, M. L. Thermodynamic and structural characterization of zwitterionic micelles of the membrane protein solubilizing amidosulfobetaine surfactants ASB-14 and ASB-16. *Langmuir* **2011**, *27*, 8248–8256.

[309] Seddon, A. M.; Curnow, P.; Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *BBA-Biomembranes* **2004**, *1666*, 105–117.

[310] Horst, R.; Horwich, A. L.; Wüthrich, K. Translational diffusion of macromolecular assemblies measured using transverse-relaxation-optimized pulsed field gradient NMR. *J. Am. Chem. Soc.* **2011**, *133*, 16354–16357.

[311] Kim, T. H.; Chung, K. Y.; Manglik, A.; Hansen, A. L.; Dror, R. O.; Mildorf, T. J.; Shaw, D. E.; Kobilka, B. K.; Prosser, R. S. The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *J. Am. Chem. Soc.* **2013**, *135*, 9465.

[312] Nilsson, F.; Söderman, O.; Hansson, P.; Johansson, I. Physical- chemical properties of C9G1 and C10G1 β-alkylglucosides. Phase diagrams and aggregate size/structure. *Langmuir* **1998**, *14*, 4050–4058.

[313] Hargreaves, R.; Bowron, D. T.; Edler, K. Atomistic structure of a micelle in solution determined by wide Q-range neutron diffraction. *J. Am. Chem. Soc.* **2011**, *133*, 16524–16536.

[314] Kameyama, K.; Takagi, T. Micellar properties of octylglucoside in aqueous solutions. *Colloid Interface Sci.* **1990**, *137*, 1–10.

[315] Mesa, C.; Bonincontro, A.; Sesta, B. Solution properties of octyl β-D glucoside. Part 1: Aggregate size. shape and hydration. *Colloid Polym. Sci.* **1993**, *271*, 1165–1171.

[316] Lipfert, J.; Millett, I. S.; Seifert, S.; Doniach, S. Sample holder for small-angle X-ray scattering static and flow cell measurements. *Rev. Sci. Instrum.* **2006**, *77*, 046108.

[317] Pastor, R.; MacKerell Jr, A. Development of the CHARMM force field for lipids. *J. Phys. Chem. Lett.* **2011**, *2*, 1526–1532.

[318] Soranzo, T.; Martin, D. K.; Lenormand, J.-L.; Watkins, E. B. Coupling neutron reflectivity with cell-free protein synthesis to probe membrane protein structure in supported bilayers. *Sci. Rep.* **2017**, *7*.

[319] Zeppieri, S.; Rodríguez, J.; López de Ramos, A. Interfacial tension of alkane+ water systems. *J. Chem. Eng. Data* **2001**, *46*, 1086–1088.

[320] Seifert, S.; Winans, R.; Tiede, D.; Thiyagarajan, P. Design and performance of a ASAXS instrument at the Advanced Photon Source. *J. Appl. Crystallogr.* **2000**, *33*, 782–784.

[321] Oliver, R. C.; Lipfert, J.; Fox, D. A.; Lo, R. H.; Kim, J. J.; Doniach, S.; Columbus, L. Tuning micelle dimensions and properties with binary surfactant mixtures. *Langmuir* **2014**, *30*, 13353–13361.

[322] Dupuy, C.; Auvray, X.; Petipas, C.; Rico-Lattes, I.; Lattes, A. Anomeric effects on the structure of micelles of alkyl maltosides in water. *Langmuir* **1997**, *13*, 3965–3967.

[323] 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*, 97–105.

[324] Schiel, J. E.; Hage, D. S. Density measurements of potassium phosphate buffer from 4 to 45 C. *Talanta* **2005**, *65*, 495–500.

[325] le Maire, M.; Champeil, P.; Møller, J. V. Interaction of membrane proteins and lipids with solubilizing detergents. *BBA-Biomembranes* **2000**, *1508*, 86–111.

[326] Aoudia, M.; Zana, R. Aggregation behavior of sugar surfactants in aqueous solutions: effects of temperature and the addition of nonionic polymers. *Colloid Interface Sci.* **1998**, *206*, 158–167.

[327] Hub, J. S.; Awasthi, N. Probing a continuous polar defect: A reaction coordinate for pore formation in lipid membranes. *J. Chem. Theory Comput.* **2017**, *13*, 2352–2366.

[328] Cheng, X.; Jo, S.; Lee, H. S.; Klauda, J. B.; Im, W. CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems. 2013.

[329] Reiss-Husson, F.; Luzzati, V. Small-angle X-ray scattering study of the structure of soap and detergent micelles. *J. Colloid Interface Sci.* **1966**, *21*, 534–546.

[330] Hayter, J.; Penfold, J. Determination of micelle structure and charge by neutron small-angle scattering. *Colloid Polym. Sci.* **1983**, *261*, 1022–1030.

[331] Quina, F. H.; Nassar, P. M.; Bonilha, J. B.; Bales, B. L. Growth of sodium dodecyl sulfate micelles with detergent concentration. *J. Phys. Chem.* **1995**, *99*, 17028–17031.

[332] Dreiss, C. A. Wormlike micelles: where do we stand? Recent developments, linear rheology and scattering techniques. *Soft Matter* **2007**, *3*, 956–970.

[333] Bengtsen, T.; Holm, V. L.; Midtgaard, S. R.; Johansen, N. T.; Bottaro, S.; Arleth, L.; Lindorff-Larsen, K. Structure and dynamics of a lipid nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *bioRxiv* **2019**,

[334] Angelov, B.; Angelova, A.; Papahadjopoulos-Sternberg, B.; Lesieur, S.; Sadoc, J.-F.; Ollivon, M.; Couvreur, P. Detailed structure of diamond-type lipid cubic nanoparticles. *J. Am. Chem. Soc.* **2006**, *128*, 5813–5817.

[335] Porte, G.; Poggi, Y.; Appell, J.; Maret, G. Large micelles in concentrated solutions. The second critical micellar concentration. *J. Phys. Chem. A* **1984**, *88*, 5713–5720.

[336] Vass, S.; Pedersen, J. S.; Pleštil, J.; Laggner, P.; Rétfalvi, E.; Varga, I.; Gilányi, T. Ambiguity in determining the shape of alkali alkyl sulfate micelles from small-angle scattering data. *Langmuir* **2008**, *24*, 408–417.

[337] Pedersen, J. S. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv. Colloid Interface Sci.* **1997**, *70*, 171–210.

[338] Potrzebowski, W.; Trewhella, J.; Andre, I. Bayesian inference of protein conformational ensembles from limited structural data. *PLoS Comput. Biol.* **2018**, *14*, e1006641.

[339] Semenyuk, A.; Svergun, D. GNOM–a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **1991**, *24*, 537–540.

[340] Debye, P. Zerstreuung von röntgenstrahlen. *Annalen der Physik* **1915**, *351*, 809–823.

[341] Linstrom, P. J.; Mallard, W. G. The NIST Chemistry WebBook: A chemical data resource on the internet. *J. Chem. Eng. Data* **2001**, *46*, 1059–1063.

[342] Dias Mirandela, G.; Tamburrino, G.; Ivanović, M.; Strnad, F. M.; Byron, O.; Rasmussen, T.; Hoskisson, P. A.; Hub, J. S.; Zachariae, U.; Gabel, F. Merging In-Solution X-ray and Neutron Scattering Data Allows Fine Structural Analysis of Membrane–Protein Detergent Complexes. *J. Phys. Chem. Lett.* **2018**, *9*, 3910–3914.

[343] Dubyak, G. R. Ion homeostasis, channels, and transporters: an update on cellular mechanisms. *Adv. Physiol. Educ.* **2004**, *28*, 143–154.

[344] Werten, P. J. L.; Rémigy, H. W.; de Groot, B. L.; Fotiadis, D.; Philippsen, A.; Stahlberg, H.; Grubmüller, H.; Engel, A. Progress in the analysis of membrane protein structure and function. *FEBS Lett.* **2002**, *529*, 65–72.

[345] Stahlberg, H.; Engel, A.; Philippsen, A. Assessing the structure of membrane proteins: combining different methods gives the full picture. *Biochem. Cell Biol.* **2002**, *80*, 563–568.

[346] Breyton, C.; Gabel, F.; Lethier, M.; Flayhan, A.; Durand, G.; Jault, J.-M.; Juillan-Binard, C.; Imbert, L.; Moulin, M.; Ravaud, S. Small angle neutron scattering for the study of solubilised membrane proteins. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2013**, *36*, 71.

[347] Kynde, S. A.; Skar-Gislinge, N.; Pedersen, M. C.; Midtgaard, S. R.; Simonsen, J. B.; Schweins, R.; Mortensen, K.; Arleth, L. Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 371–383.

[348] Skar-Gislinge, N.; Kynde, S. A.; Denisov, I. G.; Ye, X.; Lenov, I.; Sligar, S. G.; Arleth, L. Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment. *cta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 2412–2421.

[349] Merrick, M.; Javelle, A.; Durand, A.; Severi, E.; Thornton, J.; Avent, N.; Conroy, M.; Bullough, P. The Escherichia coli AmtB protein as a model system for understanding ammonium transport by Amt and Rh proteins. *Transfus. Clin. Biol.* **2006**, *13*, 97–102.

[350] Khademi, S.; O'Connell, J.; Remis, J.; Robles-Colmenares, Y.; Miercke, L. J. W.; Stroud, R. M. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Science* **2004**, *305*, 1587–1594.

[351] Trewhella, J.; Duff, A. P.; Durand, D.; Gabel, F.; Guss, J. M.; Hendrickson, W. A.; Hura, G. L.; Jacques, D. A.; Kirby, N. M.; Kwan, A. H. 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta. Crystallogr. D. Struct. Biol.* **2017**, *73*, 710–728.

[352] Petoukhov, M. V.; Svergun, D. I. Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 1051–1058.

[353] Oliver, R. C.; Pingali, S. V.; Urban, V. S. Designing mixed detergent micelles for uniform neutron contrast. *J. Phys. Chem. Lett.* **2017**, *8*, 5041–5046.

[354] Zaccai, N. R.; Sandlin, C. W.; Hoopes, J. T.; Curtis, J. E.; Fleming, P. J.; Fleming, K. G.; Krueger, S. *Methods Enzymol.*; Elsevier, 2016; Vol. 566; pp 159–210.

[355] Rambo, R. P. *Methods for Using New Conceptual Tools and Parameters to Assess RNA Structure by Small-Angle X-Ray Scattering*; Academic Press, 2015; Vol. 549; p 235.

[356] Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **1999**, *76*, 2879–2886.

[357] Koutsioubas, A. Low-resolution structure of detergent-solubilized membrane proteins from small-angle scattering data. *Biophys. J.* **2017**, *113*, 2373–2382.

[358] Vinothkumar, K. R. Membrane protein structures without crystals, by single particle electron cryomicroscopy. *Curr. Opin. Struct. Biol.* **2015**, *33*, 103–114.

[359] Rawson, S.; Davies, S.; Lippiat, J. D.; Muench, S. P. The changing landscape of membrane protein structural biology through developments in electron microscopy. *Mol. Membr. Biol.* **2016**, *33*, 12–22.

[360] Xie, X.; Zubarev, R. A. Effects of low-level deuterium enrichment on bacterial growth. *PLOS ONE* **2014**, *9*, e102071.

[361] Zheng, L.; Kostrewa, D.; Bernèche, S.; Winkler, F. K.; Li, X.-D. The mechanism of ammonia transport based on the crystal structure of AmtB of Escherichia coli. *PNAS* **2004**, *101*, 17090–17095.

[362] Slotboom, D. J.; Duurkens, R. H.; Olieman, K.; Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **2008**, *46*, 73–82.

[363] Ebel, C. Sedimentation velocity to characterize surfactants and solubilized membrane proteins. *Methods* **2011**, *54*, 56–66.

[364] Fisher, E.; Zhao, Y.; Richardson, R.; Janik, M.; Buell, A. K.; Aigbirhio, F. I.; Toth, G. Detection and characterization of small molecule interactions with fibrillar protein aggregates using microscale thermophoresis. *ACS Chem. Neurosci.* **2017**, *8*, 2088–2095.

[365] Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **2015**, *12*, 405–413.

[366] Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **2009**, *97*, 50–58.

[367] Jo, S.; Kim, T.; Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLOS ONE* **2007**, *2*, e880.

[368] Abraham, M. J. Performance enhancements for GROMACS nonbonded interactions on BlueGene. *J. Comput. Chem.* **2011**, *32*, 2041–2046.

[369] Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

[370] Klauda, J. B.; Venable, R. M.; Freites, J. A.; OConnor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr, A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

[371] MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[372] Compton, E. L.; Karinou, E.; Naismith, J. H.; Gabel, F.; Javelle, A. Low resolution structure of a bacterial SLC26 transporter reveals dimeric stoichiometry and mobile intracellular domains. *J. Biol. Chem.* **2011**, *286*, 27058–27067.

[373] Reyes, F. E.; Schwartz, C. R.; Tainer, J. A.; Rambo, R. P. *Methods Enzymol.*; Elsevier, 2014; Vol. 549; pp 235–263.

[374] Van Den Bedem, H.; Fraser, J. S. Integrative, dynamic structural biology at atomic resolutionit's about time. *Nat. Methods* **2015**, *12*, 307.

[375] Ozdemir, E. S.; Nussinov, R.; Gursoy, A.; Keskin, O. Developments in integrative modeling with dynamical interfaces. *Curr. Opin. Struct. Biol.* **2019**, *56*, 11–17.

[376] Brosey, C. A.; Ahmed, Z.; Lees-Miller, S. P.; Tainer, J. A. *Methods in enzymology*; Elsevier, 2017; Vol. 592; pp 417–455.

[377] Ward, A. B.; Sali, A.; Wilson, I. A. Integrative structural biology. *Science* **2013**, *339*, 913–915.

[378] Hayashi, S.; Ikeda, S. Micelle size and shape of sodium dodecyl sulfate in concentrated sodium chloride solutions. *J. Phys. Chem.* **1980**, *84*, 744–751.

[379] Chen, J. M.; Su, T. M.; Mou, C. Y. Size of sodium dodecyl sulfate micelle in concentrated salt solutions. *J. Phys. Chem.* **1986**, *90*, 2418–2421.

[380] Hassan, P.; Fritz, G.; Kaler, E. W. Small angle neutron scattering study of sodium dodecyl sulfate micellar growth driven by addition of a hydrotropic salt. *Colloid Interface Sci.* **2003**, *257*, 154–162.

[381] Sammalkorpi, M.; Karttunen, M.; Haataja, M. Ionic surfactant aggregates in saline solutions: sodium dodecyl sulfate (SDS) in the presence of excess sodium chloride (NaCl) or calcium chloride (CaCl2). *J. Phys. Chem. B* **2009**, *113*, 5863–5870.

[382] Almgren, M.; Löfroth, J.-E. Effects of polydispersity on fluorescence quenching in micelles. *J. Chem. Phys.* **1982**, *76*, 2734–2743.

[383] Doblas, D.; Kister, T.; Cano-Bonilla, M.; González-García, L.; Kraus, T. Colloidal solubility and agglomeration of apolar nanoparticles in different solvents. *Nano Lett.* **2019**, *19*, 5246–5252.