PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 26351

Quantifying the influence of the ion cloud on SAXS profiles of charged proteins†

Miloš T. Ivanović, ‡[a] Linda K. Bruetzel,[b] Roman Shevchuk,[a] Jan Lipfert [b] and Jochen S. Hub ‡*[a]

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 compared 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 among 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.

Received 15th May 2018,
Accepted 2nd October 2018

DOI: 10.1039/c8cp03080d

rsc.li/pccp

1 Introduction

The ion cloud is an integral part of charged biomolecules, since the ions may influence the biomolecules' stability, structure, aggregation, and function.[1–3] 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,[1,3–18] whereas less work has focused on the ion cloud of charged proteins.[19,20]

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).[19,21–26] 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 in solution with charged groups on the biomolecule's surface. Hence, several extensions to the native PB theory have been proposed.[27–30] 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.[10–14,16–18] Results from MD simulations are compatible with other theoretical predictions and experiments,[5,6,9,15,19] suggesting that MD simulations predict a reasonably accurate ion distribution, despite some well-known limitations of certain ion force fields.[31–34] Complementary, the 3D-RISM method has been suggested as computationally efficient method for modeling the solvation shell of biomolecules.[5]

SAXS is an increasingly popular method for obtaining low-resolution structural information of biomolecules in solution,

[a] *Georg-August-Universität Göttingen, Institute for Microbiology and Genetics, Justus-von-Liebig Weg 11, 37077 Göttingen, Germany*
[b] *Ludwig-Maximilian-Universität München, Department of Physics, Amalienstraße 54, 80799 München, Germany*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cp03080d

‡ Present address: Saarland University, Theoretical Physics, Campus E2.6, 66123 Saarbrücken, Germany. E-mail: jochen.hub@uni-saarland.de; Fax: +49 (0)681 302-6604; Tel: +49 (0)681 302-2740

described in a number of excellent reviews.[35–44] 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),[45] and established standards for sample preparation and validation.[46,47] 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.[38] 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,[48] with a simple spherical model,[49] and using MD simulations.[50–55] 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.[51] 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.[56,57] In addition, Kim *et al.* disentangled contributions from water and ions on the hydration layer of supercharged proteins by combining SAXS with SANS.[20] 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.[58–61] 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 freely adjustable model parameter, may be problematic. Fitting parameters may (i) increase the risk of overfitting,[62] 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.[63]

SAXS curve predictions based on explicit-solvent MD simulations avoid such a hydration layer-related fitting parameter.[51,53–55,64] 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 predicting SAXS curves with explicit-solvent MD simulations, 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.[51,53] For highly charged proteins, however, additional solvent density modulations appear owing to the counter ion cloud,[20] 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 $\sim$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.

2 Methods

2.1 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^{-1} and 5.0 mg ml^{-1} 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 13 500 rpm for 10 min in a tabletop centrifuge (Eppendorf, Germany) prior to each measurement.

SAXS measurements were carried out at beamline BM29, ESRF, Grenoble[65] at an X-ray wavelength λ of 0.99 Å, using a sample-to-detector distance of 2.87 m and a Pilatus 1 M 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^{-1} samples were systematically larger than estimates from the 5.0 mg ml^{-1} 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^{-1} and 5.0 mg ml^{-1} samples, suggesting that protein–protein interactions have only a small effect on the change of R_g due the ion cloud (Fig. S1, ESI†). 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^{-1} and 5.0 mg ml^{-1} samples.

2.2 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,[58] using the default CRYSOL parameters, and the volume of the sphere was calculated as

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

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

$$R_{sph} = (3V/4\pi)^{1/3}, \quad (2)$$

and its radius of gyration is

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

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.[66,67] Accordingly, the space is divided into three regions (Fig. S5, ESI†): (1) the protein, $r < R_{sph}$; (2) an ion-exclusion layer, $R_{sph} < r < a$, taking into account that the centers of finite-sized ions cannot 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:[67]

$$\Phi(r) = \frac{Qe^{\kappa a}}{\varepsilon(1+\kappa a)} \cdot \frac{e^{-\kappa r}}{r} \quad (4)$$

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 the Debye–Hückel parameter, given by:

$$\kappa = \left(\frac{8\pi M e_c^2}{\varepsilon k_B T}\right)^{1/2}. \quad (5)$$

Here, M is the 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

$$n_i(\mathbf{r}) = M e^{q_i \Phi(|\mathbf{r}|)/k_B T}, \quad (6)$$

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*

$$\delta\rho_{\mathrm{CI}}(\mathbf{r}) = \sum_i n_i(\mathbf{r})\Delta N_i^e \tag{7}$$

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

$$\Delta N_i^e = N_i^e - V_i\rho_{\mathrm{water}}, \tag{8}$$

where 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. 68. The radius of gyration of the spherical protein model, including the effect of the counter ion cloud, is defined *via*

$$R_{\mathrm{g}}^2 = \left(\int\Delta\rho(\mathbf{r})\mathrm{d}\mathbf{r}\right)^{-1}\int\Delta\rho(\mathbf{r})(\mathbf{r}-\mathbf{r}_{\mathrm{s}})^2\mathrm{d}\mathbf{r}, \tag{9}$$

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

$$\Delta\rho(\mathbf{r}) = \begin{cases} \rho_{\mathrm{sph}} - \rho_{\mathrm{bulk}} & \text{if } |\mathbf{r}| \le R_{\mathrm{sph}} \\ 0 & \text{if } R_{\mathrm{sph}} < |\mathbf{r}| \le a \\ \delta\rho_{\mathrm{CI}}(|\mathbf{r}|) & \text{if } |\mathbf{r}| > a \end{cases} \tag{10}$$

where ρ_{bulk} denotes the bulk density of the salt solution. The integrals in eqn (9) was evaluated numerically. Finally, the change in the radius of gyration due to the ion cloud was taken as $\Delta R_g = R_g - \tilde{R}_g$.

2.3 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 $e_c\Phi(\mathbf{r}) \gtrsim k_BT$. 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).[69] The structures of BSA and GI were prepared with the PDB2PQR software.[70,71] The total size of the grid was 50 × 50 × 50 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 the previous paragraph (eqn (7) through (10)). However, the ion densities $n_i(\mathbf{r})$ were not taken from eqn (6), but instead from the APBS result, and the integrals of eqn (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.

2.4 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,[72,73] and remain a matter on ongoing force field development.

SAXS curves were computed from the MD simulations using explicit-solvent SAXS calculations.[51,74] 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. 1A and 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.[55] Following previous work,[51,53] 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 (Fig. S2, ESI†).

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. S3C, ESI†). 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.[51] 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/100 mM, BSA/500 mM, GI/100 mM and GI/500 mM, 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. S3, ESI†).

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 the Guinier fit to the experimental

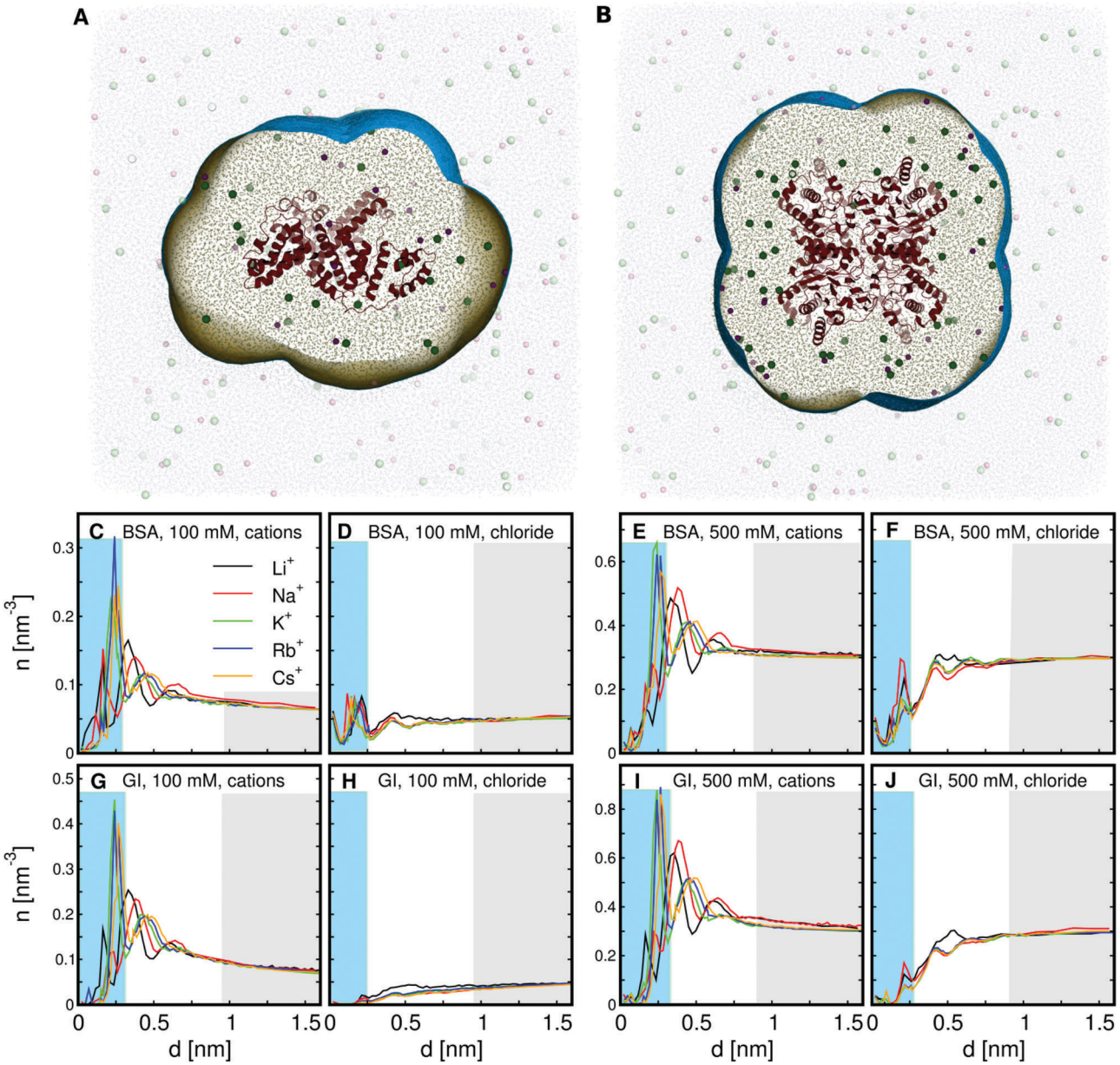

Fig. 1 Simulation systems of (A) bovine serum albumine (BSA) and (B) glucose isomerase (GI), each with 100 mM NaCl buffer. Proteins are shown in red cartoon representation, water as sticks, and Na^+ and Cl^- ions as green and purple spheres. Envelopes (blue surface) at a distance of 30 Å from the protein 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 and H) or 500 mM (E, F, I and 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 follow Debye–Hückel theory.

SAXS curves (see above), we used a fixed q range because the calculated $I(q)$ curves follow exactly the Guinier approximation at small q.[51] 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;[51,74] otherwise, artifacts due to buffer mismatch would arise, in particular in systems with electron-rich ions (rubidium and caesium) 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[51,55]), but also due a mismatch owing to ion–ion correlations (Fig. S4, ESI†). 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.

2.5 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[75] and 1MNZ,[76] 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.[77] 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.[78] 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 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,[79,80] 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,[81] and all other bond lengths were constrained with P-LINCS.[82] 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[83] (τ = 1 ps) and using a stochastic dynamics integration scheme[84] during equilibration and production runs, respectively. The pressure was kept at 1 bar with a Berendsen[85] and the Parrinello–Rahman barostat[86] (τ = 0.4 ps) during equilibration and production runs, respectively, thereby allowing stable box relaxation during equilibration and guaranteeing the correct ensemble during production.

2.6 Force field parameters

Interactions of the protein and ions in KCl, RbCl and CsCl solutions were modeled with the Amber99SB-ILDN force field[87,88] and using the Joung–Cheatham ion parameters.[89] 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 were instead modeled with the CHARMM36 force field,[90,91] version of November 2016, translated into GROMACS.[92] 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^+.[72]

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^+.[72] As a simple measure to avoid such overbinding, we increased the LJ diameter σ_{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 σ_{Li-OC} and found that an increase of σ_{Li-OC} by 6% is sufficient to avoid excessive overbinding. Hence, we used σ_{Li-OC} increased by 6% for production simulations, σ_{Li-OC} = 0.283 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.[73] 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 increased by $\sim$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 Results and discussion

3.1 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. 1A) and glucose isomerase (GI, $Q = -60e$, Fig. 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. Fig. 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).[93] 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).[5,6] 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. 1C, 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.

Fig. 2(A–D) presents SAXS curved of BSA in the ten different alkali chloride buffers, either computed from MD simulations (Fig. 2A and B) or obtained experimentally (Fig. 2C and D). SAXS curves of GI were purely obtained from MD simulations

(Fig. 2E and 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 due to a conformational change in the protein. The overall agreement between the SAXS curves from MD simulation and from experiment is reasonable. Exceptions are the systems with 500 mM RbCl or 500 mM CsCl, presumably due to a small buffer mismatch between the protein solution and the buffer solution (Fig. 2D, blue and orange curves). In addition, Fig. 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. 2B and F, blue and orange curves).

3.2 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, 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 triangles up), (iii) non-linear PB calculations (blue triangles down). The absolute R_g estimated from different theoretical methods may differ (Fig. S6, ESI†), 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[56] (Fig. S6, ESI†). 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, 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 and experiment was 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. 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. 1C, E, G and 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 systems with the 500 mM systems, Fig. 3 demonstrates that R_g increases more

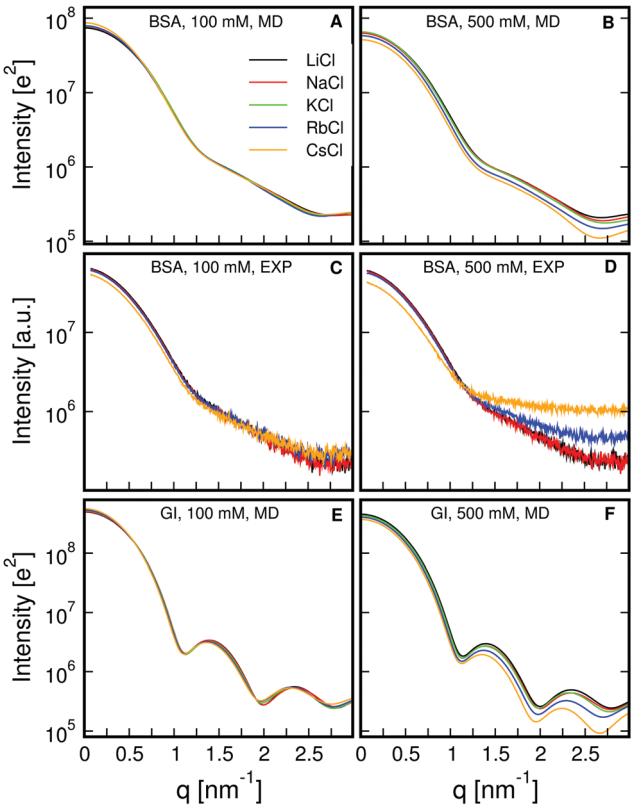

Fig. 2 SAXS curves of (A–D) bovine serum albumin (BSA) and (E and F) glucose isomerase (GI) in different alkali chloride buffers (for color code, see legend) of 100 mM (A, C and E) or 500 mM (B, D and F) concentration. (A, B, E and F) SAXS curves computed from MD simulations, and (C and D) obtained experimentally.

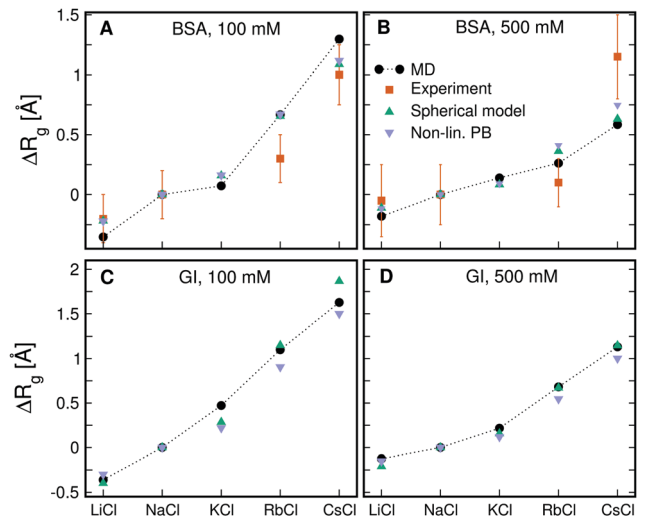

Fig. 3 Change of a radius of gyration ΔR_g, shown relative to R_g in NaCl solution, as a function of salt type and concentration for (A and B) bovine serum albumin (BSA) and (C and D) glucose isomerase (GI). Results are shown as taken from all-atom MD simulations (black circles), non-linear Poisson–Boltzmann calculations (blue triangles down), the spherical model (green triangles up), and experiment (red squares). The absolute values of R_g are shown in Fig. S6 (ESI†).

strongly with smaller bulk salt concentration (compare Fig. 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 (eqn (9)), a longer Debye length eventually imposes a larger influence of salt on R_g.

3.3 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, 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. 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. S3, ESI†).

Moreover, we tested influence of the exclusion layer on the ΔR_g estimates by the spherical model (Fig. S7, ESI†), assuming exclusion layers between 0 and 2 Å, *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.

3.4 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 Fig. 4. Here, the respectively fixed quantities were taken from the spherical model for glucose isomerase, using $Q = -60e$, $R_{sph} = 3.82$ nm and $\rho = 431.6$ e nm^{-3} as electron density of the sphere. In scans with fixed bulk salt concentration we used $M = 100$ mM. ΔR_g was computed for the five alkali chloride salts considered in this study (Fig. 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. 3C and 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. 4 shows that larger alkali ions have a larger effect on ΔR_g

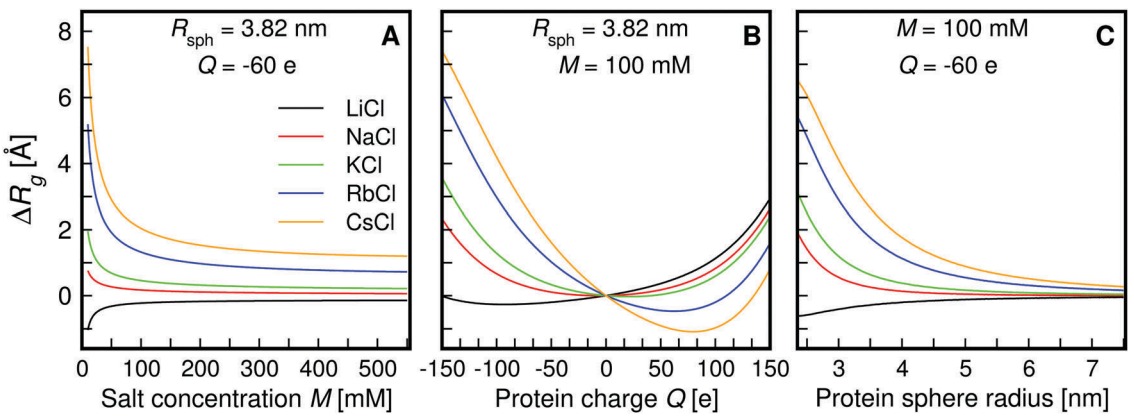

Fig. 4 Systematic analysis of the effect of (A) salt concentration, (B) protein charge, and (C) protein size on the increase of the radius of gyration ΔR_g due to the counter ion cloud, as calculated with the spherical model. (A) ΔR_g *versus* salt concentration, at fixed protein charge $Q = -60e$ and fixed protein size (R_{sph} = 3.82 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_{sph} = 3.82 nm) and fixed salt concentration of 100 mM. (C) ΔR_g *versus* size of the protein, plotted as the radius of the sphere, at fixed protein charge − 60*e* 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^-.

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. 4, black curves).

Fig. 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 *vice versa* in case of $Q > 0$.[23] 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 Fig. 4A, however, are relevant to experimental conditions.

Fig. 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 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. 4B, blue and orange curves). However, since the concentration of the alkali cation in the ion cloud cannot 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 (eqn (6)) manifests in the highly non-linear $\Delta R_g(Q)$ relation visible in Fig. 4B.

Fig. 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, $\propto {R_{sph}}^3$, 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.

4 Conclusions

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 Å to 2 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,[51] 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.[20]

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.[50–55] 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 protein 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.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

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 1971/3-1, and HU 1971/4-1. LKB and JL acknowledge support by the DFG through grant number SFB 863/A11.

References

1 D. E. Draper, D. Grilley and A. M. Soto, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 221–243.
2 G. C. Wong and L. Pollack, *Annu. Rev. Phys. Chem.*, 2010, **61**, 171–189.
3 J. Lipfert, S. Doniach, R. Das and D. Herschlag, *Annu. Rev. Biochem.*, 2014, **83**, 813–841.
4 T. G. Wensel, C. F. Meares, V. Vlachy and J. B. Matthew, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 3267–3271.
5 G. M. Giambasu, T. Luchko, D. Herschlag, D. M. York and D. A. Case, *Biophys. J.*, 2014, **106**, 883–894.
6 S. Kirmizialtin, S. A. Pabit, S. P. Meisburger, L. Pollack and R. Elber, *Biophys. J.*, 2012, **102**, 819–828.
7 K. A. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.*, 1990, **19**, 301–332.
8 L. Pollack, *Annu. Rev. Biophys.*, 2011, **40**, 225–242.
9 S. Kirmizialtin, A. R. Silalahi, R. Elber and M. O. Fenley, *Biophys. J.*, 2012, **102**, 829–838.
10 T. E. Cheatham and M. A. Young, *Biopolymers*, 2000, **56**, 232–256.
11 S. Kirmizialtin and R. Elber, *J. Phys. Chem. B*, 2010, **114**, 8207–8220.
12 M. Rueda, E. Cubero, C. A. Laughton and M. Orozco, *Biophys. J.*, 2004, **87**, 800–811.
13 S. Y. Ponomarev, K. M. Thayer and D. L. Beveridge, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14771–14775.
14 L. van Dam, A. P. Lyubartsev, A. Laaksonen and L. Nordenskiöld, *J. Phys. Chem. B*, 1998, **102**, 10636–10642.
15 M. Deserno, C. Holm and S. May, *Macromolecules*, 2000, **33**, 199–206.
16 B. Jayaram and D. Beveridge, *Annu. Rev. Biophys.*, 1996, **25**, 367–394.
17 T. E. Cheatham III and P. A. Kollman, *Annu. Rev. Phys. Chem.*, 2000, **51**, 435–471.
18 D. L. Beveridge and K. J. McConnell, *Curr. Opin. Struct. Biol.*, 2000, **10**, 182–196.
19 H. T. Nguyen, S. A. Pabit, L. Pollack and D. A. Case, *J. Chem. Phys.*, 2016, **144**, 214105.
20 H. S. Kim, A. Martel, E. Girard, M. Moulin, M. Härtlein, D. Madern, M. Blackledge, B. Franzetti and F. Gabel, *Biophys. J.*, 2016, **110**, 2185–2194.
21 R. Das, T. Mills, L. Kwok, G. Maskel, I. Millett, S. Doniach, K. Finkelstein, D. Herschlag and L. Pollack, *Phys. Rev. Lett.*, 2003, **90**, 188103.
22 Y. Bai, R. Das, I. S. Millett, D. Herschlag and S. Doniach, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 1035–1040.
23 Y. Bai, M. Greenfeld, K. J. Travers, V. B. Chu, J. Lipfert, S. Doniach and D. Herschlag, *J. Am. Chem. Soc.*, 2007, **129**, 14981–14988.
24 S. A. Pabit, S. P. Meisburger, L. Li, J. M. Blose, C. D. Jones and L. Pollack, *J. Am. Chem. Soc.*, 2010, **132**, 16334–16336.
25 F. Horkay, A. M. Hecht, C. Rochas, P. J. Basser and E. Geissler, *J. Chem. Phys.*, 2006, **125**, 234904.
26 S. P. Meisburger, S. A. Pabit and L. Pollack, *Biophys. J.*, 2015, **108**, 2886–2895.
27 V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag and S. Doniach, *Biophys. J.*, 2007, **93**, 3202–3209.
28 M. Eigen and E. Wicke, *J. Phys. Chem.*, 1954, **58**, 702–714.
29 R. Netz and H. Orland, *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 2000, **1**, 203–214.
30 I. Borukhov, D. Andelman and H. Orland, *Phys. Rev. Lett.*, 1997, **79**, 435–438.
31 Y. Luo and B. Roux, *J. Phys. Chem. Lett.*, 2009, **1**, 183–189.
32 A. Noy, I. Soteras, F. J. Luque and M. Orozco, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10596–10607.
33 R. M. Venable, Y. Luo, K. Gawrisch, B. Roux and R. W. Pastor, *J. Phys. Chem. B*, 2013, **117**, 10183–10192.
34 J. Yoo and A. Aksimentiev, *J. Phys. Chem. Lett.*, 2011, **3**, 45–50.
35 J. Lipfert and S. Doniach, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 307–327.
36 L. Feigin and D. Svergun, *Structure Analysis by Small-Angle X-Ray and Neutron Scatterin*, Plenum Press, New York, 1987.
37 M. H. Koch, P. Vachette and D. I. Svergun, *Q. Rev. Biophys.*, 2003, **36**, 147–227.
38 C. D. Putnam, M. Hammel, G. L. Hura and J. A. Tainer, *Q. Rev. Biophys.*, 2007, **40**, 191–285.
39 S. Doniach and J. Lipfert, *Small and Wide Angle X-ray Scattering from Biological Macromolecules and their Complexes in Solution*, Elsevier BV, 2012.
40 M. A. Graewert and D. I. Svergun, *Curr. Opin. Struct. Biol.*, 2013, **23**, 748–754.
41 R. P. Rambo and J. A. Tainer, *Annu. Rev. Biophys.*, 2013, **42**, 415–441.
42 L. Boldon, F. Laliberte and L. Liu, *Nano Rev.*, 2015, **6**, 25661.
43 T. Bizien, D. Durand, P. Roblina, A. Thureau, P. Vachette and J. Pérez, *Protein Pept. Lett.*, 2016, **23**, 217–231.
44 J. Trewhella, *Curr. Opin. Struct. Biol.*, 2016, **40**, 1–7.
45 A. Berthaud, J. Manzi, J. Pérez and S. Mangenot, *J. Am. Chem. Soc.*, 2012, **134**, 10080–10088.
46 S. Skou, R. E. Gillilan and N. Ando, *Nat. Protoc.*, 2014, **9**, 1727–1739.
47 C. M. Jeffries, M. A. Graewert, C. E. Blanchet, D. B. Langley, A. E. Whitten and D. I. Svergun, *Nat. Protoc.*, 2016, **11**, 2122–2153.

48 D. Svergun, S. Richard, M. Koch, Z. Sayers, S. Kuprin and G. Zaccai, *PNAS*, 1998, **95**, 2267–2272.
49 H. S. Kim and F. Gabel, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2015, **71**, 57–66.
50 F. Merzel and J. C. Smith, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5378–5383.
51 P. Chen and J. S. Hub, *Biophys. J.*, 2014, **107**, 435–447.
52 J. J. Virtanen, L. Makowski, T. R. Sosnick and K. F. Freed, *Biophys. J.*, 2011, **101**, 2061–2069.
53 S. Park, J. P. Bardhan, B. Roux and L. Makowski, *J. Chem. Phys.*, 2009, **130**, 134114.
54 T. Oroguchi, H. Hashimoto, T. Shimizu, M. Sato and M. Ikeguchi, *Biophys. J.*, 2009, **96**, 2808–2822.
55 J. Köfinger and G. Hummer, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2013, **87**, 052712.
56 F. Zhang, M. W. Skoda, R. M. Jacobs, R. A. Martin, C. M. Martin and F. Schreiber, *J. Phys. Chem. B*, 2007, **111**, 251–259.
57 F. Zhang, F. Roosen-Runge, M. W. Skoda, R. M. Jacobs, M. Wolf, P. Callow, H. Frielinghaus, V. Pipich, S. Prevost and F. Schreiber, *Phys. Chem. Chem. Phys.*, 2012, **14**, 2483–2493.
58 D. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.
59 S. Yang, S. Park, L. Makowski and B. Roux, *Biophys. J.*, 2009, **96**, 4449–4463.
60 D. Schneidman-Duhovny, M. Hammel and A. Sali, *Nucleic Acids Res.*, 2010, **38**, W540–W544.
61 H. Liu, A. Hexemer and P. H. Zwart, *J. Appl. Crystallogr.*, 2012, **45**, 587–593.
62 A. Grishaev, L. Guo, T. Irving and A. Bax, *J. Am. Chem. Soc.*, 2010, **132**, 15484–15486.
63 P.-c. Chen and J. S. Hub, *J. Phys. Chem. Lett.*, 2015, **6**, 5116–5121.
64 F. Merzel and J. C. Smith, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 242–249.
65 P. Pernot, A. Round, R. Barrett, A. De Maria Antolinos, A. Gobbo, E. Gordon, J. Huet, J. Kieffer, M. Lentini and M. Mattenet, *et al.*, *J. Synchrotron Radiat.*, 2013, **20**, 660–664.
66 K. A. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience, Garland Science*, 2nd edn, 2010.
67 M. J. Holst, PhD thesis, California institute of Technology, 1994.
68 Y. Marcus, *Chem. Rev.*, 1988, **88**, 1475–1498.
69 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *PNAS*, 2001, **98**, 10037–10041.
70 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res.*, 2004, **32**, W665–W667.
71 T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, *Nucleic Acids Res.*, 2007, **35**, W522–W525.
72 R. M. Venable, Y. Luo, K. Gawrisch, B. Roux and R. W. Pastor, *J. Phys. Chem. B*, 2013, **117**, 10183–10192.
73 R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. MacKerell Jr, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.
74 C. J. Knight and J. S. Hub, *Nucleic Acids Res.*, 2015, **43**, W225–W230.
75 A. Bujacz, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1278–1289.
76 E. Nowak, S. Panjikar and P. A. Tucker, *Atomic structure of Glucose isomerase*, DOI: 10.2210/pdb1mnz/pdb.
77 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
78 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.
79 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
80 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
81 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.
82 B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116–122.
83 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
84 W. Van Gunsteren and H. Berendsen, *Mol. Simul.*, 1988, **1**, 173–185.
85 H. J. C. Berendsen, J. P. M. Postma, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
86 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
87 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950–1958.
88 K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLoS One*, 2012, **7**, e32131.
89 I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.
90 B. R. Brooks, C. L. Brooks, A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels and S. Boresch, *J. Comput. Chem.*, 2009, **30**, 1545–1614.
91 R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. MacKerell Jr, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.
92 P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess and E. Lindahl, *J. Chem. Theory Comput.*, 2010, **6**, 459–466.
93 L. Vrbka, J. Vondrášek, B. Jagoda-Cwiklik, R. Vácha and P. Jungwirth, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15440–15444.