Accurate small and wide angle x-ray scattering profiles from atomic models of proteins and nucleic acids

Hung T. Nguyen,[1] Suzette A. Pabit,[2] Steve P. Meisburger,[2] Lois Pollack,[2] and David A. Case[1,3,a)]

[1]*BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey 08854, USA*
[2]*School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA*
[3]*Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA*

(Received 29 July 2014; accepted 10 September 2014; published online 29 September 2014)

A new method is introduced to compute X-ray solution scattering profiles from atomic models of macromolecules. The three-dimensional version of the Reference Interaction Site Model (RISM) from liquid-state statistical mechanics is employed to compute the solvent distribution around the solute, including both water and ions. X-ray scattering profiles are computed from this distribution together with the solute geometry. We describe an efficient procedure for performing this calculation employing a Lebedev grid for the angular averaging. The intensity profiles (which involve no adjustable parameters) match experiment and molecular dynamics simulations up to wide angle for two proteins (lysozyme and myoglobin) in water, as well as the small-angle profiles for a dozen biomolecules taken from the *BioIsis.net* database. The RISM model is especially well-suited for studies of nucleic acids in salt solution. Use of fiber-diffraction models for the structure of duplex DNA in solution yields close agreement with the observed scattering profiles in both the small and wide angle scattering (SAXS and WAXS) regimes. In addition, computed profiles of anomalous SAXS signals (for Rb^+ and Sr^{2+}) emphasize the ionic contribution to scattering and are in reasonable agreement with experiment. In cases where an absolute calibration of the experimental data at $q = 0$ is available, one can extract a count of the excess number of waters and ions; computed values depend on the closure that is assumed in the solution of the Ornstein–Zernike equations, with results from the Kovalenko–Hirata closure being closest to experiment for the cases studied here. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4896220]

I. INTRODUCTION

X-ray scattering of biomacromolecules in solution, albeit a low-resolution method, shows much promise as a complementary tool to crystallography and NMR.[1–5] The computation of scattering profiles from atomic models can be a difficult task, even in the simplest case where the solute molecule adopts a known, single, and relatively rigid conformation in solution. Because both solute and solvent contribute to the scattering, the perturbation of the solvent (usually water and ions) by the biomolecule must be understood and properly modeled in order to make comparisons to experiment.

Several methods have been developed to include the contribution of water to the overall scattering profiles.[6–15] Most rely on the simplified models of water to account for the scattering of excluded volume and hydration shell. Crysol,[9] for example, assumes a layer of uniform excess hydration density around the surface of the protein. However, the surface topology, electrostatics and hydrophobicity patterns surely play a role as well. The water shell, additionally, is composed of successive layers of excess and deficient density relative to the bulk. Different approaches have been considered to describe the hydration shell more realistically, from treating solvent as an assembly of freely orienting and interacting dipoles,[12] to reconstructing the three-dimensional hydration shell by combining a set of proximal radial distribution functions for different atom types extracted from molecular dynamics (MD) simulations.[14] In principle, molecular dynamics simulations can also provide such information,[15,16] but these are difficult to converge (especially for ions in the vicinity of charged biomolecules), and are computationally tedious and expensive.

The promising intermediate approach explored here uses integral equation theory to estimate thermally averaged water and ion distributions on a three-dimensional grid surrounding the biomolecule at a fraction of the cost of MD simulations.[17–19] These estimates are certainly imperfect ones, as they are based on simple force field models for the relevant atomic interactions, and use approximate closures and averaging procedures to treat molecular solvents like water. We show here that they are nevertheless accurate enough to provide good estimates of X-ray scattering out to angles corresponding to $q < 1.5$ Å^{-1} ($q = 4\pi \sin\theta/\lambda$ where 2θ is the scattering angle and λ is the wavelength). Once the force field and closure are determined, there are no adjustable parameters in this model. The results can be of particular use in treating salt solutions and for comparisons to experiments where an absolute calibration near $q = 0$ is available.

In this paper, we present our computational model, and make comparisons to both experiment and MD simulations for relatively rigid proteins and a DNA duplex. Our computational models have been added to the *AmberTools* suite of

a)Author to whom correspondence should be addressed. Electronic mail: case@biomaps.rutgers.edu

simulation programs, available at `http://ambermd.org`. We do not consider here the "inverse" problem of interpreting experimental data arising from an unknown structure, or from samples where an ensemble of structures is contributing to the scattering. Nevertheless, the computations described here are efficient enough to be applied to large numbers of proposed structures (or to structural ensembles), and are based on a physically-motivated model for solvent effects that appears to be more accurate than any other currently-available procedure. It is likely that these ideas could form the basis for model discovery and selection in a wide range of problems.

II. THEORY AND BACKGROUND

A. Reference interaction site model theory

Small angle X-ray scattering (SAXS) is a solution based technique. Therefore, it is important to have an accurate description not only of the macromolecule under investigation but also the water molecules and ions around it. The strength of the approach described here lies in our ability to describe the solvent environment in terms of the reference interaction site model (RISM). RISM is a microscopic approach to calculate the equilibrium distribution of the solvent and its associated thermodynamic properties.[19–22] The procedure begins with the Ornstein–Zernike (OZ) equation

$$h(\mathbf{r}_{12}, \Omega_1, \Omega_2) = c(\mathbf{r}_{12}, \Omega_1, \Omega_2) + \rho \int d\mathbf{r_3} d\Omega_3 c(\mathbf{r}_{13}, \Omega_1, \Omega_3) h(\mathbf{r}_{32}, \Omega_3, \Omega_2), \tag{1}$$

where $\mathbf{r}_{ij}$ is the vector connecting particles i and j, Ω_i and Ω_j are the orientation of particles i and j, respectively, relative to $\mathbf{r}_{ij}$, c is the direct correlation function, h is the total correlation function

$$h_{ij}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) \equiv g_{ij}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) - 1, \tag{2}$$

where g_{ij} is the pair distribution function.

The OZ equation effectively defines the direct correlation function, but to solve it, it is necessary to have a second, so-called closure, equation that relates h and c, which is conventionally written as

$$g(\mathbf{r}_{12}, \Omega_1, \Omega_2) = \exp[-\beta u(\mathbf{r}_{12}, \Omega_1, \Omega_2) + h(\mathbf{r}_{12}, \Omega_1, \Omega_2) - c(\mathbf{r}_{12}, \Omega_1, \Omega_2) + b(\mathbf{r}_{12}, \Omega_1, \Omega_2)] \tag{3}$$

or in a shorter form

$$g = \exp[-\beta u + h - c + b]. \tag{4}$$

Here u is the potential energy function and b is an unknown "bridge function." In the hypernetted-chain approximation (HNC), b is set to zero, giving

$$g_{HNC} = \exp[-\beta u + h - c]. \tag{5}$$

The HNC closure gives good results for ionic and polar systems, but poorer results for neutral systems, and it can be difficult to find converged solutions.[22–24] To address this, Kovalenko–Hirata introduced the KH closure as follows:[17]

$$g_{KH} = \begin{cases} \exp[-\beta u + h - c] & \text{if } g \le 1, \\ 1 - \beta u + h - c & \text{if } g > 1. \end{cases} \tag{6}$$

The partial series expansion of order-n (PSE-n) offers a way to interpolate between KH and HNC, and thus improves the results of KH closure while circumventing the convergence difficulty met in HNC closure[25]

$$g_{PSE-n} = \begin{cases} \exp[-\beta u + h - c] & \text{if } g \le 1, \\ \sum_{i=0}^{n} \frac{[-\beta u + h - c]^i}{i!} & \text{if } g > 1. \end{cases} \tag{7}$$

Hence, KH is the special case of PSE closure when $n = 1$; and when $n \to \infty$ HNC is obtained.

For the pairwise non-bonded potential u, we use the standard Coulomb and Lennard–Jones terms, with parameters taken from a molecular mechanics force field

$$u_{12}(r) = \frac{q_1 q_2}{r} + \varepsilon_{12} \left[\left(\frac{R_{min,12}}{r} \right)^{12} - 2 \left(\frac{R_{min,12}}{r} \right)^{6} \right]. \tag{8}$$

First a 1D-RISM calculation is carried out with only the solvent (water + ions) to obtain the solvent susceptibility $\chi_{\alpha\beta}^{VV}$ which contains all the information about the bulk solvent. This solution will be subsequently used for 3D-RISM to compute the solvent structure around a solute of choice.[17,18,26] Thus, at each condition (salt concentration, temperature, pressure ...), one needs to perform only one 1D-RISM step. The resulting χ^{VV} can be used for all subsequent 3D-RISM calculations. These latter calculations yield distribution functions that reflect the excess or deficit of each solvent site relative to bulk concentration on a grid around the solute in real space. From these density distributions, one can compute the excess electron density and SAXS profile for the solute.

B. Computing SAXS profiles

X-ray scattering experiments on biomolecules compare the scattering intensity from the sample of interest to a "blank" with just solvent present, and report the difference, or "excess" intensity

$$I(\mathbf{q}) = \langle |A(\mathbf{q})|^2 \rangle_t - \langle |B(\mathbf{q})|^2 \rangle_t, \tag{9}$$

where $A(\mathbf{q})$ and $B(\mathbf{q})$ are Fourier transforms of the scattering amplitudes for the sample and blank, respectively. The $\langle\rangle_t$ bracket indicates the intensities are averaged over the measurement time and volume. Following Park *et al.*,[16] we show in Sec. VII B that this expression can be approximately, but usefully, replaced by

$$I(\mathbf{q}) \simeq [\langle A_1(\mathbf{q}) \rangle - \langle B_1(\mathbf{q}) \rangle]^2, \tag{10}$$

where $A_1(\mathbf{q})$ and $B_1(\mathbf{q})$ are Fourier transforms for the sample and blank, respectively, but here only considering regions where there is excess/deficit electron density relative to the bulk. The $\langle\rangle$ bracket is now the ensemble average. This approximation is valid up to angles corresponding to $q \simeq 1.5$ Å^{-1}.

This approach has several considerable computational advantages for grid-based representations of the solvent. First,

we do not need to include the bulk into the calculation as in Eq. (9), so that the three-dimensional grid need only cover regions where the solvent is perturbed by the solute. Although we only consider local regions around the solute, the fact that we compute the difference between the two amplitudes as in Eq. (10) effectively treats an infinitely large system (in the bulk region, $A = B$) and therefore does not introduce any artificial boundary whose shape could influence the result. Second, the amplitude computation (Eq. (13), below) is linear in the number of grid points, whereas an intensity calculation (e.g., from a Debye sum) is quadratic in the number of grid points. (A recent study by Berlin *et al.*[27] describes an alternative approach to the Debye sum which scales with $O(N\log N)$, which might also be adapted to the grid representation used here.) The excess intensity calculation consists of two major steps, which are outlined in Secs. II B 1 and II B 2.

1. Computing the excess amplitude

We first compute the excess amplitude of the system (equivalent to $\langle A_1(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle$ in the context of Eq. (10)),

$$A_1(\mathbf{q}) - B_1(\mathbf{q}) = F(\mathbf{q}) = F_{solu}(\mathbf{q}) + F_{grid}(\mathbf{q}), \tag{11}$$

where

$$F_{solu}(\mathbf{q}) = \sum_j f_j(q) \exp\left(-\frac{B_j q^2}{16\pi^2}\right) \exp(-i\mathbf{q}.\mathbf{r_j}) \tag{12}$$

is the form factor of the solute. The Debye–Waller factor B_j roughly accounts for thermal motion, discussed in more detail in Sec. VII B 3.

The contribution from the solvent $F_{grid}(\mathbf{q})$ is computed by performing a 3D Fourier transformation of the excess electron density, using the so-called CUBE method[7,14]

$$F_{grid}(\mathbf{q}) = \sum_j^{N_{grid}} f_j(\mathbf{q}) e^{-i\mathbf{q}.\mathbf{r_j}}, \tag{13}$$

$$f_j(\mathbf{q}) = 8\frac{\sin\left(\frac{q_x a}{2}\right)\sin\left(\frac{q_y b}{2}\right)\sin\left(\frac{q_z c}{2}\right)}{q_x q_y q_z}\rho_{xe}^{(j)}, \tag{14}$$

where $\rho_{xe}^{(j)}$ is the excess electron density in the jth cell of the rectangular grid, with length, width, and height a, b, c. $\rho_{xe}^{(j)}$, in turn, is calculated by summing up all excess densities from individual atom types in the solution (for instance, H_{wat}, O_{wat}, Na^+, and Cl^- in NaCl solution)

$$\rho_{xe}^{(j)} = \sum_k Z_k \rho_k [g_k(\mathbf{r_j}) - 1] \tag{15}$$

with ρ_k and Z_k as the bulk density and the atomic number of the kth atom or ion, respectively. At each grid point the excess density is accounted for by the term $[g(\mathbf{r}) - 1]$ from the solution of the 3D-RISM equations. In Eq. (15), all the electrons of an atom/ion from the grid are assumed to reside within the cell where the nucleus is, in our case, a cube of length 0.5 Å. (We show in Sec. VII B 4 below that a more realistic assignment of electron density to the grid has a negligible effect on the computed profiles.)

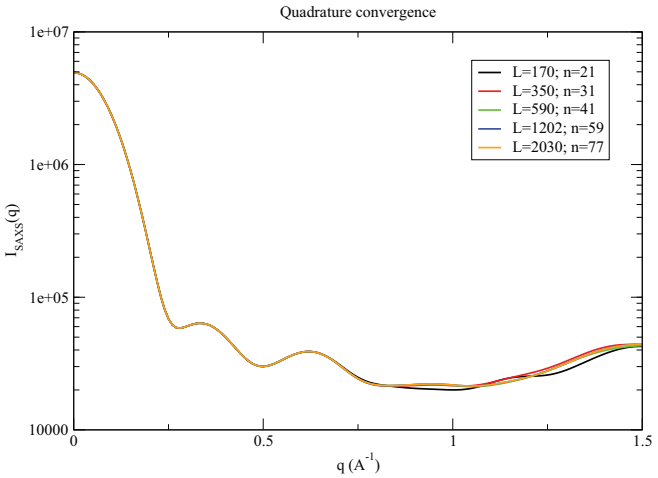

FIG. 1. Test of convergence of different quadrature orders. L is the number of points on the sphere surface, and n is the precision order (all spherical harmonics of order less than or equal to n will yield exact results). The two most accurate orders ($n = 59$ and 71) give relative errors within 0.2% for $q = 1.5$ Å^{-1}.

2. Computing the excess intensity

We next compute the excess intensity by performing spherical averaging

$$I(q) = \frac{1}{4\pi}\int |F(\mathbf{q})|^2 d\Omega. \tag{16}$$

One of the fastest and most accurate ways to perform spherical integration is to use Lebedev quadrature, which is analogous to Gaussian quadrature in a linear dimension[28,29]

$$\int |F(\mathbf{q})|^2 d\Omega \approx \sum_i^{N_p} w_i |F(\mathbf{q_i})|^2, \tag{17}$$

where the points are at pre-defined directions in a unit sphere (forming a two-dimensional grid on the sphere surface) with the weights w_i. Since $I(\mathbf{q}) = I(-\mathbf{q})$, we gain additional speed-up by evaluating the scattering vectors in only one hemisphere. As q increases, more points are needed to estimate the integral with high accuracy. For example, we use 38 grid points at $q = 0.01$ Å^{-1} and 1202 grid points at $q = 1.00$ Å^{-1} which is sufficient to keep the relative errors within 10^{-3} (Fig. 1).

III. RESULTS FOR PROTEINS

A. Myoglobin and lysozyme

It has long been recognized that the solvent shell around a protein significantly impacts the shape of the measured SAXS profile. As a first test of the RISM–SAXS method which efficiently generates the solvent distribution around a specified solute, Figure 2 compares the calculated profiles of lysozyme and myoglobin with experiment and with MD simulation results reported earlier.[16] The results are shown in both logarithmic and linear scale to exploit the benefits of both – the log scale can show the overall shape of the curve, whereas a linear scale can show more clearly the details at intermediate angles.

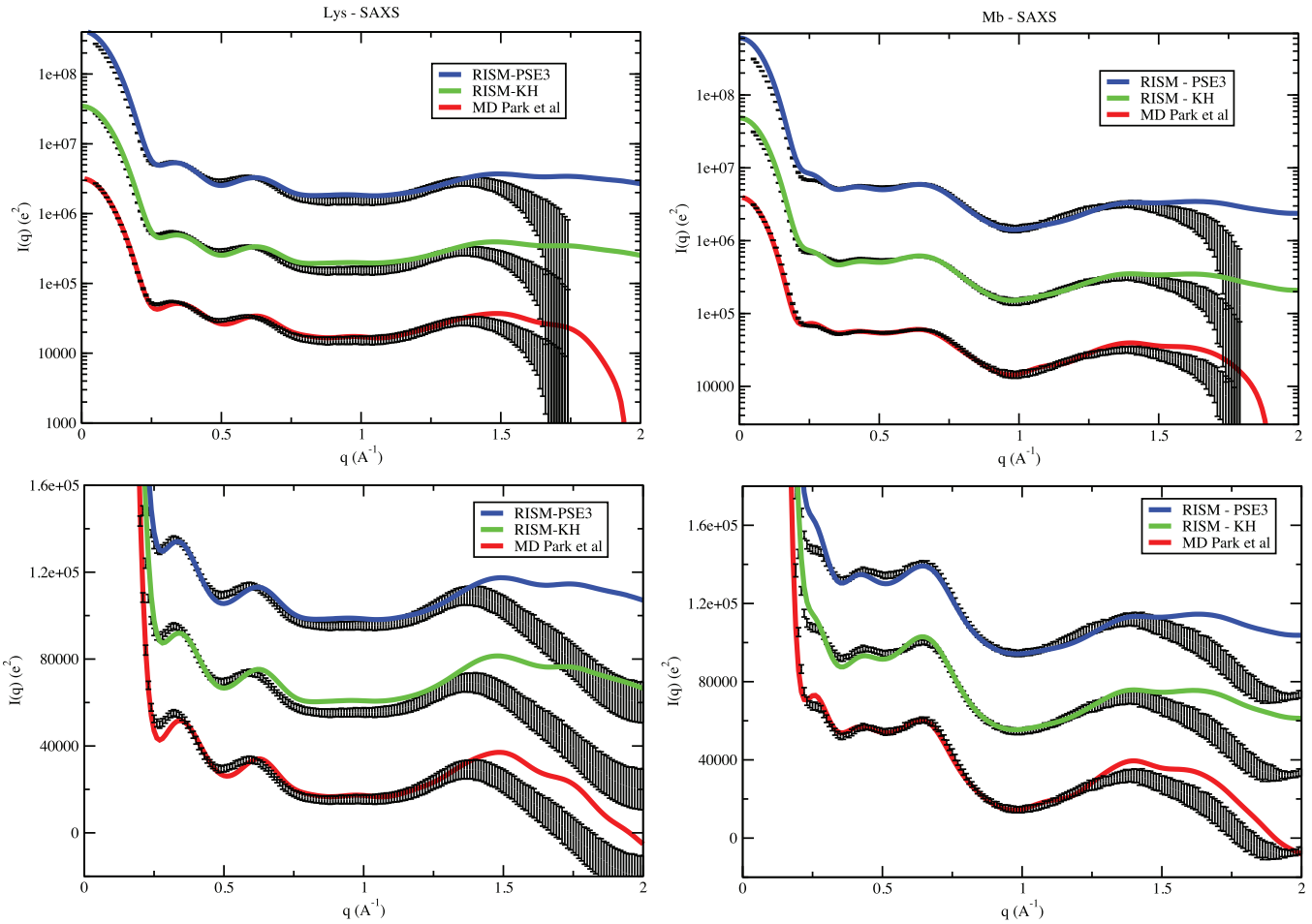

FIG. 2. SAXS profiles of lysozyme (*left*) and myoglobin (*right*) calculated from RISM with KH and PSE3 closures, plotted in log scale (*top*) and linear scale (*bottom*). Experimental data (error bars) and MD curve are taken from Park *et al.*[16] The data are offset for visual comparison with experiment (for the logarithmic plot, the scaling factor is 10 while for the linear plot, the offset factor is 4×10^4).

It can be seen in both cases that RISM reproduces the peaks and troughs of the SAXS profiles and is on par with MD simulations up to $q \simeq 1.5$ Å^{-1}. Obtaining good results beyond that threshold with RISM is difficult as fluctuation effects emerge that depress the excess intensities (see Figure 11 in Sec. VII B below). The computed results are promising if one considers that RISM is an "implicit" solvent theory; however it differs from other implicit solvent programs (for example, Crysol[9]) because RISM directly computes the solvent distribution around the solute considering only interactions between the solute and solvent, without further assumptions or fitting parameters. The agreement between RISM–SAXS with explicit solvent models (MD) and experiment indicates that we can capture the hydration shells and SAXS curves by using RISM theory, at a fraction of the computational time associated with MD.

B. Comparison to other methods

Figures 3 and 4 show SAXS profiles calculated by some widely used tools (Crysol,[9] FoXS,[10] AXES,[11] AquaSAXS,[12] and Hypred[14]). For lysozyme, all of them do relatively well at small angle. Crysol, despite its simplicity, is able to provide an excellent fit with the experiment up to 1.5 Å^{-1}, but overestimates scattering near $q = 0$ region. For myoglobin, no tools could predict the scattering curve satisfactorily, even at small angle region, except RISM–SAXS and, less satisfactorily, Hypred and AXES. It also should be noted that, a scaling factor is needed to plot the predicted profiles from other tools in order to match with the experiment, whereas nothing similar is needed in RISM–SAXS.

We have also performed calculations with structures that have experimental SAXS curves in the *BioIsis.net* database. A measure of the discrepancy between the experimental and predicted profiles is computed as

$$\chi^2 = \sum_i \left[\frac{I_{exp}(q_i) - a I_{cal}(q_i)}{\sigma_{exp}(q_i)} \right]^2 \qquad (18)$$

with a is the scaling constant and σ is the experimental uncertainty. As discussed below in Sec. IV B, we do not need any adjustable parameter to fit to experiments that have an absolute calibration (say against pure water); however, the experimental curves in BioIsis database are all relative, and a scaling factor is needed.

Table I reports χ values between the predicted and experimental SAXS curves for several widely used tools for predicting scattering profiles with comparison with RISM. The default parameters in all these tools are used (no fitting attempt has been made). The Hypred server is not able to

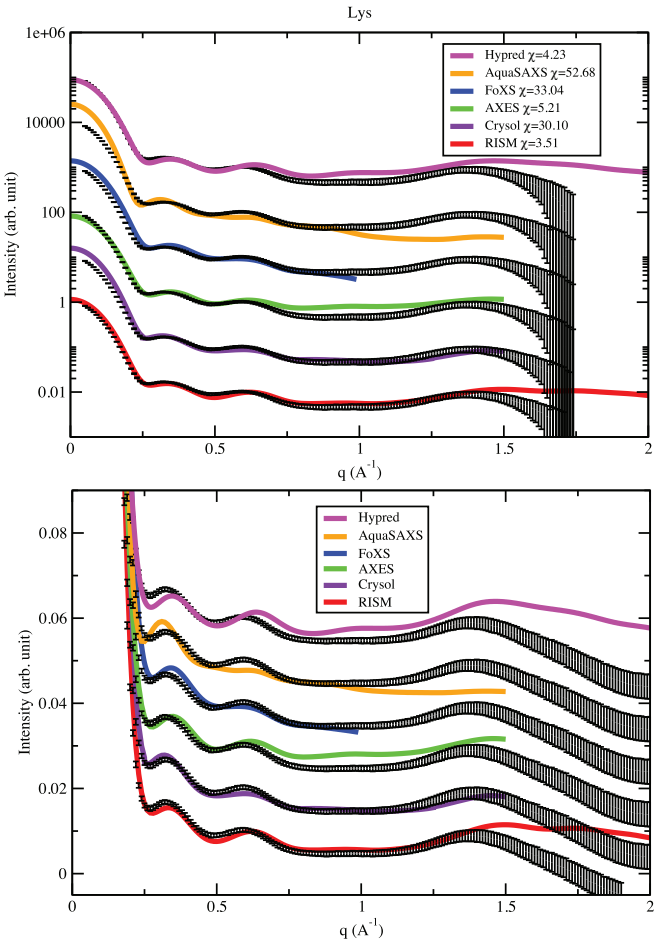

FIG. 3. Comparison between other methods for calculation SAXS of lysozyme: Crysol, FoXS, AXES, AquaSAXS, and Hypred, plotted in log scale (a) and linear scale (b).

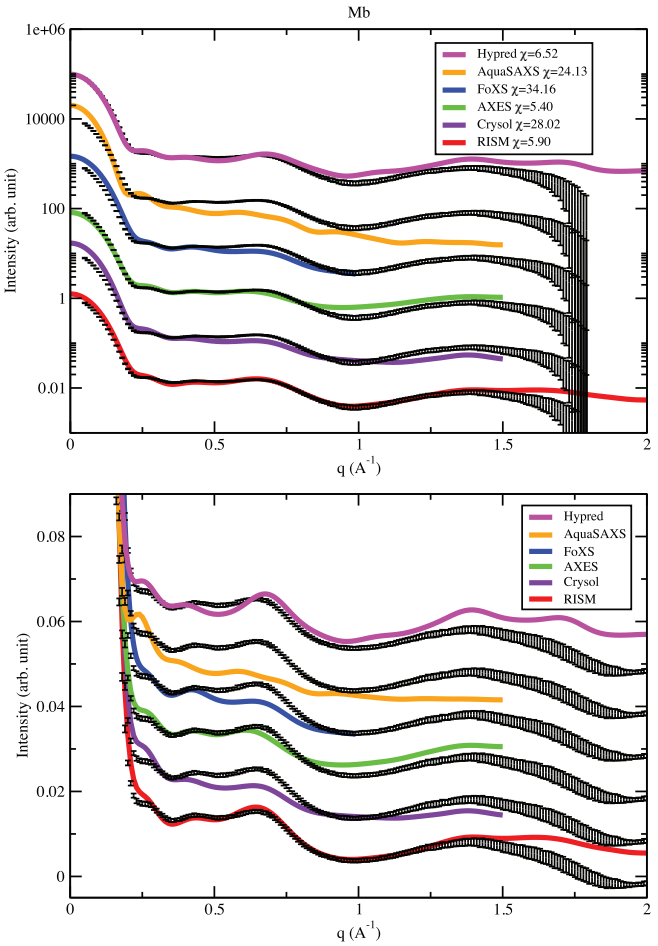

FIG. 4. Same in Fig. 3, but for myoglobin.

process nucleic acid pdb files (28BPDD and 2SAMRR), so those are not included in the statistics.

The RISM performance is encouraging, showing the best results for all tools tested with the average $\chi = 5.92$. However, it does encounter some difficulties, for instance in glucose isomerase ($\chi = 13.82$) and superoxide dismutase ($\chi = 7.69$). There are also structures that have highly flexible loops extending away from the protein core (Human regulator of chromosome condensation and glycosyl hydrolase+C-terminus), and thus are impossible to fit to experiment using only a single conformation. Whenever RISM fails, other tools do also. The model imposes a computational cost, discussed below; computation usually takes several minutes for small molecules to half an hour for biomolecules on a conventional desktop. This is faster than MD, but slower than most competing methods, and requires a force field representation before the integral equations are solved. Further study is needed to optimize this approach, especially in the presence of conformational heterogeneity.

IV. RESULTS FOR DUPLEX DNA IN SALT SOLUTION

A. The ion atmosphere around duplex DNA

Few experimental techniques can directly probe the spatial distribution of ions in the positively charged cloud around DNA.[30,31] Popular theoretical models include counterion condensation[32,33] or Poisson–Boltzmann (PB) theories,[34,35] although PB results for duplex DNA are in poor agreement with results from dialysis-type experiments.[26] MD simulation can also provide atomic details and in principle can describe the ionic atmosphere with great accuracy.[36–38] However, the computational cost is large, especially for the sampling of ions. Recently, RISM has been employed as a relatively cheap method to obtain a picture of the ion atmosphere around DNA, comparable to that of MD simulations.[26,39]

Figure 5 shows an experimentally acquired SAXS profile of a 25-bp duplex DNA in 100 mM NaCl solution. Experimental details are provided in Sec. VI. This mixed sequence duplex is expected to assume the B-form. Also shown are different scattering profiles computed by RISM, including the two helical forms of DNA that best resemble the data: B and B' forms. All models are built with the w3DNA web server[40] (see Fig. 6). The B'-form has a slightly wider major groove and narrower minor groove (the all heavy atom, i.e., without hydrogens, RMSD between these two structures is only 0.71 Å). This figure demonstrates the sensitivity of SAXS to the helical topology of the DNA, as mentioned above. The scattering from the B-form is in better agreement with experiment at the lowest and highest ($q > 0.6$ Å^{-1}) angles, but varies around $q = 0.4$ Å^{-1}. To further emphasize the difference in scattering profiles between these two forms at high

TABLE I. Performance comparison (χ value) for RISM–SAXS and other tools. Structures and experimental SAXS are all taken from the BioIsis.net database.

Molecule	PDB	BioIsis–ID	q_{max}(Å)	Crysol	AXES	FoXS	AquaSAXS	Hypred	RISM
28 bp DNA	...	28BPDD	0.33	1.11	2.19	1.66	1.42	...	0.73
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	3PXJ	LAR12P	0.33	4.52	8.05	3.82	24.30	3.72	3.53
S-adenosylmethionine riboswitch mRNA	2GIS	2SAMRR	0.30	2.52	2.53	2.10	9.22	...	2.12
Superoxide dismutase	1HL5	APSODP	0.62	16.21	7.43	30.71	28.34	14.27	7.69
Abscisic acid receptor PYR1	3K3K	1PYR1P	0.33	7.89	8.95	3.24	28.54	16.37	4.48
Glycosyl hydrolase + C-terminus	1EDG	AT5GHP	0.60	21.12	19.70	20.01	30.25	31.06	19.41
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	3T7E	ATG7CP	0.33	3.04	7.88	5.56	7.02	3.29	2.64
DNA double-strand break repair protein MRE11 + ATP	3AV0	MRERAP	0.33	2.00	20.44	5.10	9.34	26.77	4.26
Glucose isomerase	2G4J	GISRUP	0.56	16.95	46.43	36.75	26.35	78.30	13.82
Complement C3b + Efb-C	...	C3BEFP	0.33	4.02	21.70	5.74	...	2.47	3.44
Xylanase	1REF	1XYNTP	0.31	3.51	3.42	4.13	6.44	1.73	1.20
Pyrococcus furiosus decameric product	2E2G	1AHRHP	0.31	5.29	7.51	5.50	6.81	6.14	5.32
Ketoreductase-enoylreductase didomain	...	ZGDWKP	0.31	2.76	4.43	4.41	5.93	3.92	2.82
Human Regulator of Chromosome Condensation	...	HRCC1P	0.33	9.92	15.68	16.05	9.97	2.72	11.37
Average				7.20	12.60	10.34	14.92	15.90	5.92
Standard deviation				6.42	11.83	11.24	10.69	22.04	5.39
Median				4.27	7.97	5.30	9.34	5.03	3.90

angle, a Kratky plot of $I(q)q^2$ vs q is shown in Fig. 7. This presentation of the data emphasizes the shape of the scattering profiles at larger values of q, and suggests that the B-form is generally a better representation of the duplex DNA in solution. One possible explanation for this discrepancy is that the real structure is somewhere between B and B'-form. Another possible explanation for the deviation is that uncertainties in the computed distribution of ions and water in the solution affect the result.

The important contributions of both ion and water atmospheres to the overall scattering profiles of nucleic acids are even more pronounced than solvent effects in protein systems. To demonstrate the need to properly treat the solvent in computing SAXS profiles of nucleic acids, Figure 8 shows the important differences in the SAXS curves that arise from the interaction of DNA with solvent. The black curve shows a SAXS profile computed from DNA atoms in vacuo. The dashed blue curve represents the scattering of the DNA in water, accounting only for the displacement of water by the DNA duplex, and assuming that the water molecules around it do not feel its presence and behave like bulk water. The red curve includes all contributions from the hydration shell and ion layer to the DNA scattering, and should be the most realistic calculation. Water consistently perturbs the total curve up to very high angle. The most significant changes in the scattering profiles at mid to high angle reflect both the solute topology and the behavior of the hydration layer, and underscore the sensitivity of SAXS to these different aspects of nucleic acid structure. Thus, as suggested above, discrepancies between computed and measured profiles may be useful guides for improving the accuracy of calculations.

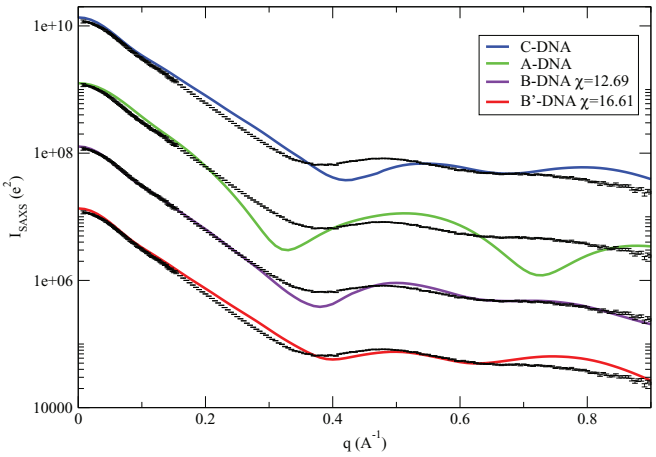

FIG. 5. SAXS profiles of different DNA structures (built with w3DNA) computed by RISM–SAXS in 0.1 M NaCl. All the curves are offset with the scaling factor of 10 for easy comparison with the experiment (shown in error bars).

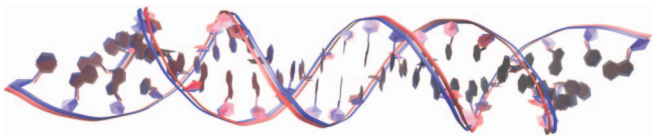

FIG. 6. B (blue) and B'-form (red) of the 25bp duplex DNA in the solution. The differences between these two structures are trivial, with the B'-form having a slightly larger major groove and smaller minor groove. RMSD for all heavy atom = 0.71 Å, backbone only = 0.86 Å.

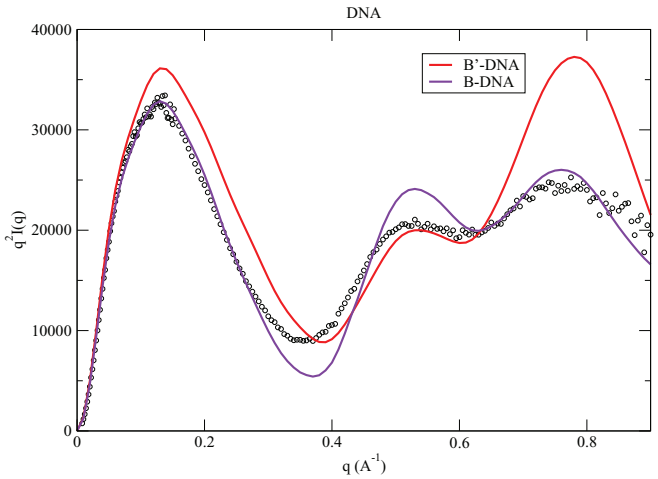

FIG. 7. Kratky plot comparison between B and B'-form of DNA.

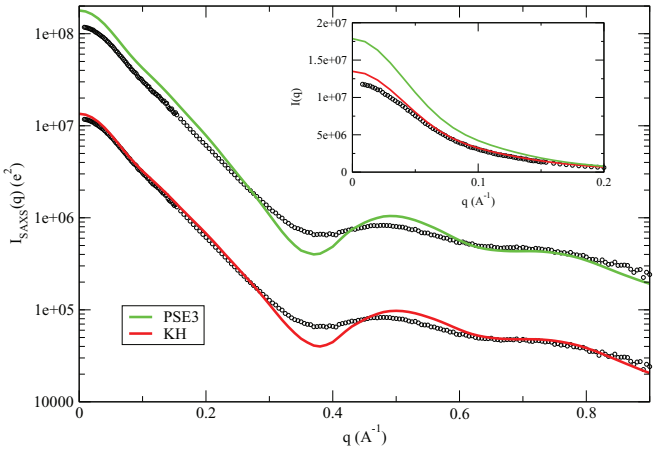

FIG. 9. SAXS profiles of B-DNA computed by RISM coupled with KH (red) and PSE3 (green) closures. The two profiles are offset by a factor of 10 for easy comparison. The inset zooms out the low angle region near $q = 0$, and is plotted without the offset factor.

B. Scattering behavior near $q = 0$

The excess form factor (as in Eq. (11)) is the 3D Fourier transform of the excess electron distribution. At $q = 0$ it is nothing but the number of excess electron in the system. Equation (11) becomes

$$A(0) = F_{solu}(0) + F_{grid}(0) = \sum_i f_i(0) + \sum_k Z_k \rho_k G_k,$$

where G_k is the Kirkwood–Buff integral[41]

$$G_k = \int [g_k(\mathbf{r}) - 1] d\mathbf{r} = \int h_k(\mathbf{r}) d\mathbf{r} \tag{19}$$

and $N_k = \rho_k G_k$ is the excess (or deficit) of the kth atom or ion around the solute. Therefore, the intensity at $q = 0$ is

$$I(0) = (Z_{solu} + Z_{wat} N_{wat} + Z_{cation} N_{cation} + Z_{anion} N_{anion})^2 \tag{20}$$

with Z_{solu} is the number of electrons in the solute, Z_{ion} is the number in the ion, and $Z_{wat} = 10$. (Note that at $q = 0$ the form factors become real numbers.) Previous work shows that the number of excess ions extracted from RISM calculations is in good agreement with "ion counting" dialysis experiments.[26] Therefore, when absolutely calibrated SAXS data are available (see Methods), the $q = 0$ value provides an absolute comparison between data and simulation. This comparison can be used to establish the effectiveness of RISM subject to different closures. Figure 9 compares measured and calibrated SAXS profiles of a 25-bp duplex DNA in 100 mM NaCl with profiles computed with two closures, KH and PSE3. The KH curve agrees better with the experiment near $q = 0$, implying that the total number of excess electrons in the system should be closer to those from KH as opposed to PSE3 closure.

Table II reports the number of excess waters and ions around the 25-bp DNA computed by RISM coupled with different closures. The neutral atomic form factors are used in the SAXS calculations, requiring a modification of the electron number N_e to account for the overall charge of the DNA: we correct the computed N_e to include the extra electrons accounting for the DNA charge. The number of excess water is approximately partitioned into contributions from the excluded volume of the DNA, and the remainder, which is termed the hydration shell. (Whether a cell belongs to the excluded volume or hydration shell depends on the distance d between it and its nearest atom of the solute j. If $d < r_j + r_{wat}$, where r_j is the atomic radii of atom j and $r_{wat} = 1.4$ Å is approximately the radius of water molecule then the cell is within the excluded volume. This is a somewhat arbitrary division, and the results in Table II provide only a general account of excluded-volume versus hydration shell effects).

From Eq. (20), an experimental estimate of N_{wat} can be extracted if all other terms are known. We assume here that the number of excess Na^+ is the same as that measured for Rb^+, which is 39 ± 2 (see below), which in turn implies that N_{Cl} is -9 ± 2, to achieve electroneutrality. The total number of electrons in the DNA is 7940 (assuming a net charge of -48), and $I(0) = 1.098 \pm 0.070 \times 10^7$, extrapolated using GNOM,[42] then the number of excess waters can be computed

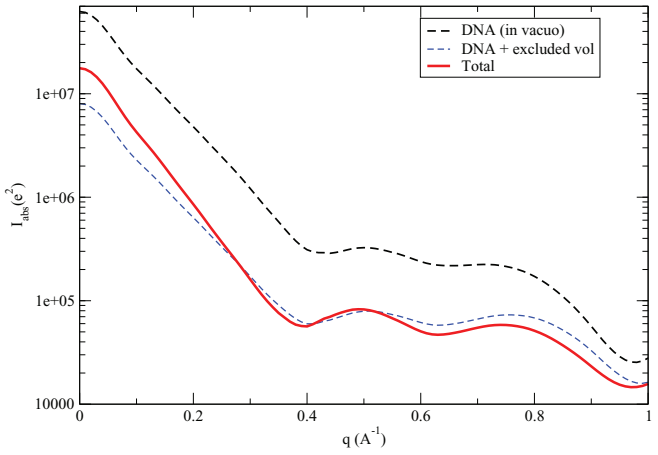

FIG. 8. Decomposition of the total SAXS curve into contributions of the DNA+excluded volume and solvent.

TABLE II. Number of excess water and ions around the 25-bp duplex DNA from SAXS experiment and RISM calculations with various closures. N_{wat} is partitioned into contributions from the excluded volume of the DNA N_{excl} and hydration shell N_{shell}, as described in the text.

	N_{wat}					
	N_{excl}	N_{shell}	Total	N_{Na}	N_{Cl}	$N_e = \sqrt{I(0)}$
KH		172	−420	30.5	−17.5	3722
PSE2		196	−396	35.6	−12.4	4104
PSE3	−592	207	−385	37.7	−10.3	4275
PSE4		211	−381	39.1	−8.9	4356
SAXS		107	−485 ± 16	39 ± 2	−9 ± 2	3, 300 ± 100

from Eq. (21),

$$1.098 \times 10^7 = (Z_{DNA} + 10N_{wat} + 10N_{Na} + 18N_{Cl})^2 = (7940 + 10 \times N_{wat} + 10 \times 39 + 18 \times [-9])^2. \quad (21)$$

This gives $N_{wat} = -485 \pm 16$. As shown in Table II, this total can be approximately viewed as the sum of a deficit of −592 waters (arising from the excluded volume of the DNA duplex), and an excess of 107 waters in the hydration shell. All of the RISM closures appear to overestimate the number of excess waters in the hydration shell. The KH results are closest to experiment, with an overestimate of about 2 water molecules per base-pair.

C. Anomalous SAXS (ASAXS)

The use of ASAXS (see Methods) provides another important degree of comparison between RISM and measurement. The ASAXS profile is the difference between the SAXS curves of the same sample but probed at two different beam energies. One of these energies is close to the absorption edge of a particular element, in this case Rb^+ or Sr^{2+}. The energy change only influences the scattering of the selected ions, but the ASAXS signal contains contributions from all terms involving ion scattering, including ion-ion, ion-solute, and ion-water cross terms. ASAXS experiments are restricted to elements whose K-edges are readily X-ray accessible. Past work focused on Rb, Sr, and Co. Lighter and more biologically relevant elements (O, C, N, Na, Mg ...) which have low energy K-edges are currently inaccessible for ASAXS.

Figure 10 shows a comparison of experimental ASAXS profiles of Rb^+ and Sr^{2+} around duplex DNA with profiles computed from RISM–SAXS, shown in the absolute scale. ASAXS curves of a similar RNA sequence, also in RbCl and $SrCl_2$, were computed by MD simulation and reported earlier, however in the relative scale.[38] In contrast to comparisons on the full SAXS profiles of B-DNA, where the PSE-n closures are not as good as the KH closure in terms of matching with the experiment near the $q = 0$ region, the PSE-n closures give better results when compared with ASAXS data. Earlier work has shown that PSE-n gives better results than KH for ion distributions around nucleic acids.[26] Since the ASAXS profile is a complicated sum of ion-solute, ion-water, and ion-ion terms, it is not surprising that none of the calculated ASAXS curves from RISM fit the experimental data. Due to weaker site-site interactions, KH closure places ions farther from the solute, leading to a more rapid decay of the ASAXS curve than expected from the PSE2 and PSE3 closures.

This method adds to the growing body of work validating sophisticated models of ion interactions with nucleic acids, by introducing absolute calibration. The absence of a scale factor provides a very stringent test of computational assumptions. Further analysis, including comparisons to MD simulations, will be presented elsewhere.

V. CONCLUSIONS

The availability of atomic coordinates of biological macromolecules is necessary but not sufficient for computing SAXS profiles for comparison with experiment. In addition to the underlying solute, SAXS provides important information into how the solute modifies the bulk solvent. Here, we describe a method for computing the solvent contribution that

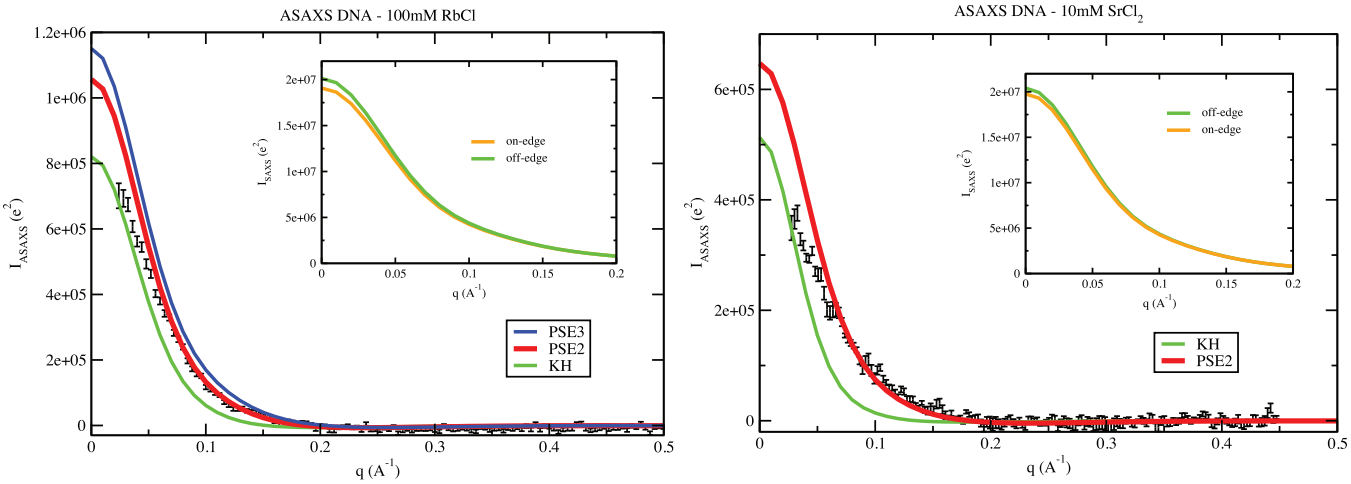

FIG. 10. ASAXS signals of Rb^+ (left) and Sr^{2+} (right) around the 25-bp duplex DNA computed from RISM–SAXS coupled with different closures, with the experimental profiles shown as error bars. The insets show the calculated on-edge and off-edge SAXS profiles in which only the atomic scattering factors of the cations are varied. The ASAXS signal is obtained by subtracting the on-edge from the off-edge.

uses the 3D-RISM model to describe the solvent environment. These integral equation models are far from perfect, but provide estimates of useful accuracy that agree better with experiment for a number of test cases than do the predictions of simple competing models, and rival the results of much more expensive molecular dynamics simulations. The 3D-RISM is particularly attractive for cases where there are both ions and water in the environment, since there are few existing implicit models that describe both, and equilibration of ion densities in MD simulations can be difficult to achieve. As an example, we report here one of the first simulations of anomalous scattering for mobile counterions near DNA. Results are in good agreement with the observed profiles, and support the experimental estimates of excess ion parameters reported earlier.

The basic analysis described here uses a single structure to describe the solute biomolecule. Even relatively rigid biomolecules may have solute conformational fluctuations that can affect scattering profiles in the wide-angle region beyond $q \approx 0.3\text{Å}^{-1}$. A very simple approach models these fluctuations in the same way as do atomic displacement parameters (or B-factors) in crystallography. This model provides some insight, but ignores differences between the crystal and solution environment, and fails to include the effects of correlated fluctuations that affect solution scattering but not the intensities of Bragg peaks in crystallography. Averaging over snapshots from MD simulations offers one way to investigate such effects, but more work in this area is needed. We consider here only the "forward" challenge of estimating SAXS profiles based on an input structure; the "inverse" problem of constructing a structure or ensemble consistent with a given profile is more challenging, and is generally problem-specific. Our computation is fast enough (requiring a few minutes for the examples considered here) to allow one to average over many solute configurations, or to use SAXS results (perhaps in combination with other restraints) to construct ensembles of configurations consistent with the data.

Analysis of the $q = 0$ limits, and comparison to experimentally calibrated profiles, allows one to count the numbers of excess ions and waters in the vicinity of biomolecules. These in turn can be used to test the accuracy of computations (including finding the limitations of the RISM model used here), and to complement other ion counting experiments. These counts are related to partial molar volumes and contributions to osmotic pressure, and offer insights into molecular interactions and function. A full analysis of the expected uncertainties in these estimates is beyond the scope of this paper, but would include an analysis of experimental calibration errors and methods to distinguish between ion and water contributions. A preliminary example, of duplex DNA in NaCl/water, suggests that the excess number of waters surrounding the duplex can be estimated with a precision of 1–2 water molecules per base pair, and that the force fields and RISM models used here tend to overestimate the number of excess waters. (This tendency only affects the scattering curves for $q < 0.05$ Å^{-1}, and generally good results are obtained for higher scattering angles.)

The characterization of the solvent perturbation used here relies on a thermally–averaged density profile, and appears to be only appropriate for $q < 1.5$ Å^{-1}. At wider angles, fluctuations in the solvent densities (and not just the average density) become important, and a different type of theory is needed. (At high angles, errors in the 3D-RISM description of pure water may also be a factor limiting the application of this model.) Nevertheless, this range of scattering angles covers a large fraction of reported scattering profiles, and our model should be of considerable use. The programs used here will be incorporated into the *AmberTools* suite of programs, available at `http://ambermd.org`.

VI. EXPERIMENTAL DETAILS

A. DNA sample preparation

To prepare double-stranded DNA molecules for SAXS measurements, single-stranded molecules (one strand has a sequence GCATCTGGGCTATAAAGGGCGTCG and the other strand is its complement) were purchased from Integrated DNA Technologies (Coralville, IA). The single strands were annealed using standard procedures used in previous studies[43] and were dialyzed extensively using spin dialysis 10 kDa cutoff columns (Amicon Ultra-0.5 mL, Millipore, Billerica, MA) in buffered salt solutions of either 100 mM NaCl, 100 mM RbCl, or 10 mM $SrCl_2$ with 1 mM Na-MOPS pH 7.0. All reagents were purchased from Sigma-Aldrich (St. Louis, MO). The concentration of the DNA samples were determined from UV absorbance measurements at 260 nm using the hypochromicity-corrected extinction coefficient of the base-paired duplex, $\epsilon_{ds} = 397\,558\,\text{L mol}^{-1}\,\text{cm}^{-1}$.[44] To ensure an accurate measure of concentration, we either performed inline UV absorption measurements using a fiber-coupled spectrometer (Avaspec, Avantes, Broomfield, CO) during SAXS experiments or measured the concentration of the sample immediately after SAXS data collection using a standard benchtop spectrometer (Cary 50, Agilent, Santa Clara, CA).

B. SAXS data collection and processing

DNA SAXS data were taken at the Cornell High Energy Synchrotron Source (CHESS). For DNA in 100 mM NaCl solution, the data were taken at the G1 beamline with x-ray energy of 10.5 keV. To get a large range in scattering angles ($0.008 < q < 0.95$ Å^{-1}), the experimental curve was constructed by combining two distinct data sets measured at different times, one at low angle ($q < 0.26$ Å^{-1}) and another at wider angle. In both cases, the liquid samples were held in an in-vacuum quartz capillary with 2 mm diameter and 10 μm thickness (Hampton Research, Aliso Viejo, CA) and oscillated through the beam using a computer-controlled syringe pump to prevent radiation damage.[45] Scattering images were recorded from both the sample and the buffer background with a photon-counting area detector (Pilatus 100 K, Dectris, Baden, Switzerland). For the low-angle region, the sample to detector distance was 1.7 m and scattering profiles were taken in four exposures of 60 s each from samples with DNA concentration of 0.05 mM. For the wider-angle region, the sample to detector distance was 0.45 m, the scattering profiles were taken in 40 exposures of 1s each and the DNA concentration was 0.44 mM. To account for the variability in the x-ray

intensity between exposures, the data were normalized using either the photocurrent from a PIN diode integrated into the beamstop, for the low angle measurements, or a by measuring the intensity counts going through a semi-transparent beamstop made of Molybdenum foils (Goodfellow, Coraopolis, PA), for the higher angles.

The ASAXS measurements were done at the CHESS C-line station as described in our previous work[46,47] with the following experimental upgrades: a Pilatus 100 K photon-counting area-detector was used for collecting scattered images, the liquid samples was placed in a 3 mm diameter quartz capillary (Hampton Research, Aliso Viejo, CA) and oscillated during x-ray exposures to prevent radiation damage, a semi-transparent beamstop consisting of a stack of Molybdenum foils (Goodfellow, Coraopolis, PA) was mounted inside the flight-tube and a quartz flow cell with a 0.1-mm path length (Starna Cells, Atascadero, CA) was placed in-line with the x-ray capillary to facilitate accurate DNA concentration measurements using a fiber-coupled UV spectrometer (Avantes, Broomfield, CO). The sample to detector distance was 0.96 m ($0.024 < q < 0.5$ Å^{-1}). The ASAXS data shown here are from 0.15 mM DNA samples in either 100 mM RbCl or in 10 mM $SrCl_2$. To generate ASAXS curves, SAXS profiles were taken at 2 different energies below the elemental K absorption edge of Rb and Sr, respectively. For Rb^+, the energies used were 15.093 keV (far from edge, off-edge) and 15.191 keV (near edge, on-edge) and for Sr^{2+}, we used 15.997 keV (off-edge) and 16.097 keV (on-edge). Data were taken for DNA samples and background buffer by cycling between energies every 4 min (4 exposures of 60 s each), with a total exposure time of 32 min. A Rontec 1001 X-flash detector was used to record the x-rays elastically scattered by the beamstop for normalization of the SAXS profiles.

A silver behenate powder diffraction standard (d-spacing = 58.376 Å) was used to convert the detector coordinates to momentum transfer ($q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle, and λ is the x-ray wavelength). The silver behenate scattering pattern and the location of the beamstop-attenuated direct beam (if applicable) determined the location of $q = 0$ within the detector image. We used azimuthal integration to turn the image to a one-dimensional scattering profile using data analysis code written in MATLAB (MathWorks, Natick, MA). The reported scattered intensity was determined by subtracting the averaged buffer background scattering from the averaged scattering of the DNA samples. The scattering intensity was calibrated to absolute intensity per molecule (in units of electron^2) as described below.

C. Absolute calibration using water as reference

To facilitate accurate comparison between calculations and SAXS data, the scattered intensity was placed on an absolute scale using liquid water as calibrant, as in Ref. 48. The SAXS pattern of pure double-distilled deionized water (Barnestead Nanopure, Thermoscientific, Waltham, MA) was measured by taking SAXS profiles of pure water and subtraction of a similarly normalized background curve (an empty clean sample holder). This method accounts for the experimental parameters that affects scattering intensity like size of the x-ray beam, sample holder thickness and detector performance. The scattered intensity of water is nearly constant in the experimental q-range and can be extrapolated to the $q \to 0$ value for the absolute calibration.

The absolute scattered intensity of the sample (in terms of the macroscopic scattering cross-section, $d\Sigma/d\Omega$) is related to the water value

$$\left.\frac{d\Sigma(q)}{d\Omega}\right|_{\text{sample}} = \frac{I(q)_{\text{sample\_norm}}}{I(q\to 0)_{\text{water\_norm}}} \left.\frac{d\Sigma}{d\Omega}\right|_{\text{water}}. \quad (22)$$

Here, $I(q)_{\text{sample\_norm}}$ is the measured normalized scattering intensity of the DNA sample and $I(q \to 0)_{\text{water\_norm}}$ is the measured normalized intensity of water extrapolated to $q = 0$. We prefer to convert the sample scattering intensity $I(q)_{\text{sample}}$ to the absolute intensity per molecule (in units of electron^2), by dividing with the macromolecular concentration (number density n_{sample}) and the square of the classical electron radius $r_0 = 0.28179 \times 10^{-12}$ cm per electron

$$I(q)_{\text{sample}}[electron^2] = \frac{1}{n_{\text{sample}} r_0^2} \left.\frac{d\Sigma(q)}{d\Omega}\right|_{\text{sample}} = \frac{1}{n_{\text{sample}} r_0^2} \cdot \frac{I(q)_{\text{sample\_norm}}}{I(q\to 0)_{\text{water\_norm}}} \left.\frac{d\Sigma}{d\Omega}\right|_{\text{water}}. \quad (23)$$

The forward x-ray scattering cross-section of a liquid is proportional to the isothermal compressibility

$$\frac{d\Sigma}{d\Omega} = \rho^2 k_B T \chi_T, \quad (24)$$

where ρ is the scattering length density, and χ_T is the (temperature-dependent) isothermal compressibility, which has been determined accurately for water using speed of sound measurements.[49] For liquid water at 23C, $\chi_T = 4.55 \times 10^{-10}$ Pa^{-1} and $d\Sigma/d\Omega = 0.0164$ cm^{-1}. Therefore in practical terms, we can calculate the calibrated intensity using

$$I(q)_{\text{sample}}[electron^2] = \frac{1}{c_{\text{sample}}} \cdot \frac{1}{N_A r_0^2} \cdot \frac{I(q)_{\text{sample\_norm}}}{I(q\to 0)_{\text{water\_norm}}} \left.\frac{d\Sigma}{d\Omega}\right|_{\text{water}}, \quad (25)$$

where N_A is Avogadro's number and c_{sample} is the measured concentration of the DNA sample.

VII. COMPUTATIONAL DETAILS

We take two proteins – lysozyme and myoglobin – and a 25-bp duplex DNA as test cases for validating the RISM–SAXS method. (Additional tests are reported in Table I.) The coordinates for the proteins are taken from Protein Data Bank with PDB ID 1WLA and 6LYZ for Mb and Lys, respectively. The server w3DNA[40] is used to build the duplex DNA to B and B' forms. The DNA sequence is GCATCTGGGC-TATAAAAGGGCGTCG.

A. RISM calculations

All calculations are performed using the `rism1d` and `rism3d.snglpnt` codes from *AmberTools*.[18] We use Amber ff12SB force field for describing the DNA and

proteins. (Since there is no histidine-bound heme group parameter for Amber force field, we use the cysteine-bound heme parameter for Cytochrome P450 taken from Shahrokh *et al.*[50]) Monovalent ions (alkali, halide) parameters are taken from Joung–Cheatham ion model.[51] Sr^{2+} ion is taken from Li *et al.* (we here report the IOD set results as we find that there are no difference between SAXS calculation using these three sets).[52] The water model used in this study is cSPC/E;[18] we also did some calculations on cTIP3P water, but found that the SAXS profiles are not sensitive such a change.[18] First, the 1D-RISM is carried out with only the solvent (water + ion if any) to obtain the solvent susceptibility $\chi_{\alpha\beta}^{VV}$ which contains all the information about the bulk solvent. This will be subsequently used for 3D-RISM to compute the solvent structure around a solute of choice. Thus, one needs to perform only one 1D-RISM step, and use the resulting χ^{VV} for all subsequent 3D-RISM calculations which are at the same condition (salt concentration, temperature, pressure ...). The output from RISM program is $g(\mathbf{r})$ for each atomic sites in solvents (for instance, Hw and Ow in water). These distribution functions reflect the excess or deficit of each solvent site relative to bulk concentration around the solute in real space, and can be directly used to compute SAXS profiles.

The modified direct inversion of the iterative subspace solver (MDIIS)[53] was used to iteratively solve the RISM equations to a residual tolerance of 10^{-12} and 10^{-5} for 1D and 3D-RISM, respectively, at 298.15 K. For 1D-RISM, the 0.025 Å grid spacing is used with 16 384 and 32 768 grid points for pure water and 100 mM NaCl solution, respectively. With more diluted solutions (10 mM $SrCl_2$, for example), the grid points are doubled until we get the results converged. For 3D-RISM, a 3D grid with 0.5 Å grid spacing is used with the buffer region of 20 Å for proteins and 40 Å for DNA in 100 mM NaCl and up to 80 Å for 10 mM $SrCl_2$.

For simple and neutral protein in water (such as lysozyme, myoglobin), the RISM calculation takes 13 s (using 1.0 Å grid spacing, 20 Å buffer), SAXS calculations take ~5 min ($q = 1$ Å^{-1}) on a conventional desktop. For complex system (DNA in NaCl/water), a much bigger and finer box (0.5 Å grid spacing, buffer 40 Å) is required to obtain good ion distribution, RISM takes 20 min and SAXS calculations take an additional ~1 h using 16 CPU cores.

B. Calculation of x-ray scattering

We summarize here the derivation from Park *et al.*[16] to compute X-ray scattering curve from MD simulations. The electron density of the system $\tilde{A}(\mathbf{r})$ is separated into contribution from the solute plus its hydration shells $\tilde{A}_1(\mathbf{r})$, and the bulk solvent $\tilde{A}_0(\mathbf{r})$ that is not in the hydration shells

$$\tilde{A}(\mathbf{r}) = \tilde{A}_1(\mathbf{r}) + \tilde{A}_0(\mathbf{r}). \tag{26}$$

The intensity is the Fourier transform of correlations in this electron density (where $\langle\rangle$ denotes an ensemble average),[1]

$$\langle|A(\mathbf{q})|^2\rangle = \int [\langle\tilde{A}_0(\mathbf{r})\tilde{A}_0(\mathbf{r}')\rangle + \langle\tilde{A}_1(\mathbf{r})\tilde{A}_1(\mathbf{r}')\rangle + \langle\tilde{A}_1(\mathbf{r})\tilde{A}_0(\mathbf{r}')\rangle + \langle\tilde{A}_0(\mathbf{r})\tilde{A}_1(\mathbf{r}')\rangle]e^{-i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}d\mathbf{r}d\mathbf{r}'. \tag{27}$$

In the "blank," we separate $\tilde{B}(\mathbf{r})$ into contribution of the water droplet $\tilde{B}_1(\mathbf{r})$ and the rest $\tilde{B}_0(\mathbf{r})$ (where the water droplet is all the water within the grid where $\tilde{A}_1(\mathbf{r})$ is non-zero), and thus similarly to Eq. (26) we have

$$\tilde{B}(\mathbf{r}) = \tilde{B}_1(\mathbf{r}) + \tilde{B}_0(\mathbf{r}) \tag{28}$$

and an equation for $\langle|B(\mathbf{q})|^2\rangle$ analogous to Eq. (27).

For the bulk solvent regions, we can write

$$\langle\tilde{A}_0(\mathbf{r})\rangle = \langle\tilde{B}_0(\mathbf{r})\rangle, \tag{29}$$

$$\langle\tilde{A}_0(\mathbf{r})\tilde{A}_0(\mathbf{r}')\rangle = \langle\tilde{B}_0(\mathbf{r})\tilde{B}_0(\mathbf{r}')\rangle. \tag{30}$$

Write the cross term as

$$\langle\tilde{A}_1(\mathbf{r})\tilde{A}_0(\mathbf{r}')\rangle = \langle\tilde{A}_1(\mathbf{r})\rangle\langle\tilde{A}_0(\mathbf{r}')\rangle + \alpha(\mathbf{r},\mathbf{r}'), \tag{31}$$

where $\alpha(\mathbf{r},\mathbf{r}')$ is the correlation between these two points $\mathbf{r}$ and $\mathbf{r}'$. Similarly for pure solvent system

$$\langle\tilde{B}_1(\mathbf{r})\tilde{B}_0(\mathbf{r}')\rangle = \langle\tilde{B}_1(\mathbf{r})\rangle\langle\tilde{B}_0(\mathbf{r}')\rangle + \beta(\mathbf{r},\mathbf{r}'). \tag{32}$$

With a big enough hydration shell, we can set $\alpha(\mathbf{r},\mathbf{r}') = \beta(\mathbf{r},\mathbf{r}')$, since the solvent in the A_0 or B_0 region will be far from the solute and is little perturbed by it.

The intensity is now computed as the difference between a sample containing the solvent and the corresponding region in the pure solvent

$$I(q) = \langle|A(\mathbf{q})|^2\rangle - \langle|B(\mathbf{q})|^2\rangle. \tag{33}$$

Substituting Eqs. (27), (31), and (32) into Eq. (33), and using the fact that $\alpha(\mathbf{r},\mathbf{r}') = \beta(\mathbf{r},\mathbf{r}')$ yields

$$I(\mathbf{q}) = \int [\langle\tilde{A}_1(\mathbf{r})\tilde{A}_1(\mathbf{r}')\rangle - \langle\tilde{B}_1(\mathbf{r})\tilde{B}_1(\mathbf{r}')\rangle + \langle\tilde{A}_1(\mathbf{r})\rangle\langle\tilde{A}_0(\mathbf{r}')\rangle + \langle\tilde{A}_0(\mathbf{r})\rangle\langle\tilde{A}_1(\mathbf{r}')\rangle - \langle\tilde{B}_1(\mathbf{r})\rangle\langle\tilde{B}_0(\mathbf{r}')\rangle - \langle\tilde{B}_0(\mathbf{r})\rangle\langle B_1(\mathbf{r}')\rangle]e^{-i\mathbf{q}(\mathbf{r}-\mathbf{r}')}d\mathbf{r}d\mathbf{r}' \tag{34}$$

or

$$I(\mathbf{q}) = [\langle A_1(\mathbf{q})A_1^*(\mathbf{q})\rangle - \langle B_1(\mathbf{q})B_1^*(\mathbf{q})\rangle] + [\langle A_1(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle]\langle B_0^*(\mathbf{q})\rangle + \langle B_0(\mathbf{q})\rangle[\langle A_1^*(\mathbf{q})\rangle - \langle B_1^*(\mathbf{q})\rangle], \tag{35}$$

which is Eq. (18) in Park *et al.*[16] From Eq. (28) we have

$$\langle B_0(\mathbf{q})\rangle = \langle B(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle, \tag{36}$$

where $\langle B(\mathbf{q})\rangle = \int \langle\tilde{B}(\mathbf{r})\rangle e^{-i\mathbf{q}\cdot\mathbf{r}}d\mathbf{r}$ is the Fourier transform of the shape of the entire scattering volume. In MD simulation and RISM calculation, this volume reaches infinity and thus $\langle B(\mathbf{q})\rangle = 0$ everywhere except at $q = 0$, where its value is the number of electrons in that volume. The $q = 0$ point is regarded as a singularity. To make our scattering curve continuous at $q = 0$, we assume $\langle B(0)\rangle = 0$, too. Hence Eq. (36) can be rewritten

$$\langle B_0(\mathbf{q})\rangle = -\langle B_1(\mathbf{q})\rangle, \tag{37}$$

which is essentially the Babinet's principle. Substitute that into Eq. (35) we have

$$I(\mathbf{q}) = [\langle A_1(\mathbf{q})A_1^*(\mathbf{q})\rangle - \langle B_1(\mathbf{q})B_1^*(\mathbf{q})\rangle] - [\langle A_1(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle]\langle B_1^*(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle[\langle A_1^*(\mathbf{q})\rangle - \langle B_1^*(\mathbf{q})\rangle]. \quad (38)$$

From this we obtain a working formula for the total intensity

$$I(\mathbf{q}) = [\langle A_1(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle]^2 + [\langle |A_1(\mathbf{q})|^2\rangle - |\langle A_1(\mathbf{q})\rangle|^2] - [\langle |B_1(\mathbf{q})|^2\rangle - |\langle B_1(\mathbf{q})\rangle|^2]. \quad (39)$$

In RISM, only the ensemble-averaged distribution of water around the solute is obtained and there is no information about the time-dependent fluctuations of $A_1(\mathbf{q})$ and $B_1(\mathbf{q})$, so that the second and third terms are not accounted for by the RISM theory. For RISM–SAXS, we use the following formula:

$$I(\mathbf{q}) = [\langle A_1(\mathbf{q})\rangle - \langle B_1(\mathbf{q})\rangle]^2. \quad (40)$$

1. Effects of density fluctuations

To see the effect of the density fluctuations on the SAXS curve, Fig. 11 illustrates the effects of ignoring the second and the third terms, using a MD simulation of Lys. It is obvious that the density fluctuations do not affect the low angle region (near $q = 0$), only moderate and high angle regions. At high angle ($q > 1.5$ Å^{-1}), the fluctuation is in the same order with the first term, even making the intensity negative (near $q = 2$ Å^{-1}).

2. Effect of grid fineness

We check the convergence for SAXS computation as a function of the grid spacing in 3D-RISM (Fig. 12). The scattering profiles at 0.5 and 1 Å grid spacing show negligible differences, while at 2 Å discrepancies start showing up. Note that this is not to say that the scattering technique is able to probe the difference down to 2 Å but only show that at least a 1 Å grid spacing is needed for distribution function computed in 3D-RISM to converge.

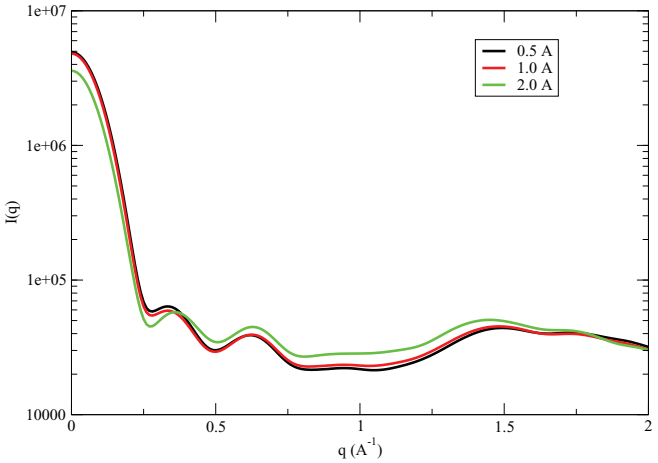

FIG. 12. Effect of grid spacing of RISM calculation onto the scattering profile of Lys.

3. Effects of thermal disorder

Even in the simplest case where the solute adopts a known, single and relatively rigid conformation, the scattering profiles are still affected by small thermal fluctuations of the solute. Modeling these variations as a Debye–Waller factor, as in Eq. (12), sometimes improves comparisons between predicted and experimental scattering profiles.[13,54] As pointed out by Moore,[55] the B-factor is only a rough guide to thermal disorder in solution, at least because B-factors are usually obtained from crystallography and are not necessarily comparable to solution scattering. In addition, correlated thermal motions (not modeled by Debye–Waller factors) contribute to scattering in solution. Fortunately, these effects are often minor, although more study is warranted. Figure 13 shows the effect of incorporating these effects into the calculation of Lys profiles via the B-factor as described in Eq. (12). Only

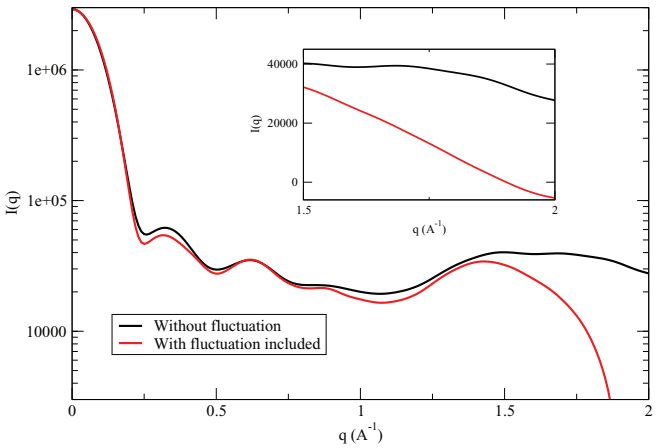

FIG. 11. SAXS of lysozyme computed by MD simulation with (red) and without (black) the fluctuation. The inset shows two curves in linear scale at high angle.

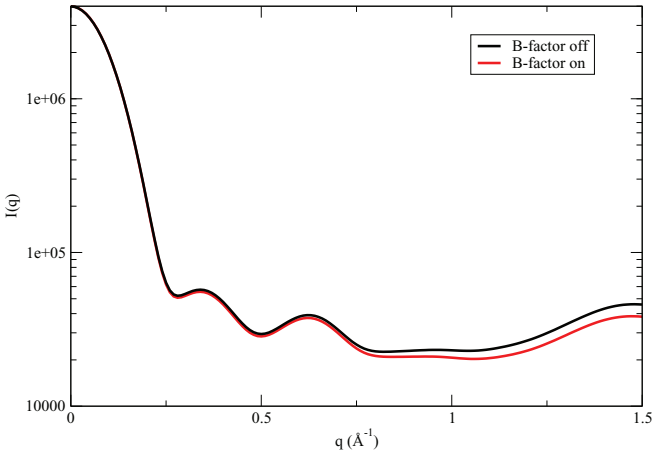

FIG. 13. Effect of thermal fluctuation onto scattering of Lys. Black and red lines are scattering curves of Lys computed by RISM–SAXS with and without B-factor included, respectively.

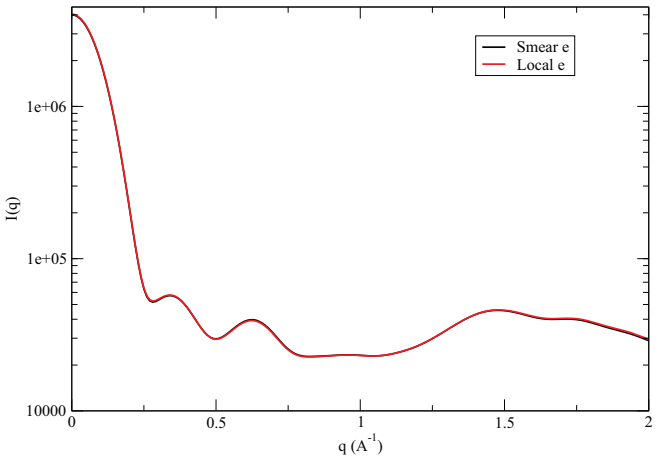

FIG. 14. Computed scattering profiles for lysozyme comparing the model of Eq. (15) (red) with a model that distributes the electron density of each water molecule onto 27 adjacent cubes (black).

the moderate and high angle, but not the low angle region as expected, are impacted by thermal fluctuation. The lowest scattering angle at which the thermal fluctuation effect is significant can be computed from Eq. (12), and is inversely related to the average B-factor.[55]

4. Computation of electron densities in the CUBE method

The RISM model provides the density of each type of solvent component (hydrogen and oxygen of waters, plus ions) on a three-dimensional grid surrounding the solute. This needs to be converted to an electron density representation in order to compute SAXS profiles. In the simplest model (described above), atomic densities at each grid point are simply multiplied by the number of electrons in each component. In fact, the distribution of electrons around the nucleus spreads out beyond a single cube. We performed calculations where we redistribute the electrons of water over 26 neighboring cells based on the spatial distribution of an electron around a single water molecule, and find the SAXS profiles are not changed up to $q = 2.0$ Å^{-1}, see Fig. 14. This is expected since at even high angle region, the resolution in real space is still not fine enough to look at electron in atom (for example, $q = 2.0$ Å^{-1} corresponds to $r \approx 3$ Å).

ACKNOWLEDGMENTS

We thank Peter Moore for many helpful discussions. This work was supported by NIH Grant Nos. GM103297 (D.A.C.) and GM085062 (L.P.) and is also based upon research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation and the National Institutes of Health/National Institute of General Medical Sciences under NSF Award No. DMR-1332208, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by Award No. GM-103485 from the National Institute of General Medical Sciences, National Institutes of Health. We thank members of the Pollack group and the CHESS staff, particularly Arthur Woll and Ken Finkelstein, for experimental assistance.

[1]M. H. J. Koch, P. Vachette, and D. I. Svergun, Quart. Rev. Biophys. **36**, 147–227 (2003).
[2]C. E. Blanchet and D. I. Svergun, Annu. Rev. Phys. Chem. **64**, 37–54 (2013).
[3]C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, Quart. Rev. Biophys. **40**, 191–285 (2007).
[4]D. I. Svergun and M. H. J. Koch, Rep. Prog. Phys. **66**, 1735 (2003).
[5]L. Pollack, Annu. Rev. Biophys. **40**, 225–242 (2011).
[6]R. Fraser, T. MacRae, and E. Suzuki, J. Appl. Cryst. **11**, 693–694 (1978).
[7]M. Y. Pavlov and B. A. Fedorov, Biopolymers **22**, 1507–1522 (1983).
[8]S. R. Hubbard, K. O. Hodgson, and S. Doniach, J. Biol. Chem. **263**, 4151–4158 (1988).
[9]D. Svergun, C. Barberato, and M. H. J. Koch, J. Appl. Cryst. **28**, 768–773 (1995).
[10]D. Schneidman-Duhovny, M. Hammel, and A. Sali, Nucl. Acids Res. **38**, W540–W544 (2010).
[11]A. Grishaev, L. Guo, T. Irving, and A. Bax, J. Am. Chem. Soc. **132**, 15484–15486 (2010).
[12]F. Poitevin, H. Orland, S. Doniach, P. Koehl, and M. Delarue, Nucl. Acids Res. **39**, W184–W189 (2011).
[13]J. Bardhan, S. Park, and L. Makowski, J. Appl. Cryst. **42**, 932–943 (2009).
[14]J. J. Virtanen, L. Makowski, T. R. Sosnick, and K. F. Freed, Biophys. J. **101**, 2061–2069 (2011).
[15]F. Merzel and J. C. Smith, Acta Crystl. D **58**, 242–249 (2002).
[16]S. Park, J. P. Bardhan, B. Roux, and L. Makowski, J. Chem. Phys. **130**, 134114 (2009).
[17]A. Kovalenko and F. Hirata, J. Chem. Phys. **110**, 10095–10112 (1999).
[18]T. Luchko, S. Gusarov, D. R. Roe, C. Simmerling, D. A. Case, J. Tuszynski, and A. Kovalenko, J. Chem. Theory Comput. **6**, 607–624 (2010).
[19]F. Hirata, "Theory of molecular liquids," in *Molecular Theory of Solvation* (Kluwer Academic Publishers, 2004), Chap. 1.
[20]D. Chandler and H. C. Andersen, J. Chem. Phys. **57**, 1930–1937 (1972).
[21]B. M. Pettitt and P. J. Rossky, J. Chem. Phys. **77**, 1451–1457 (1982).
[22]T. Luchko, I. S. Joung, and D. A. Case, "Integral equation theory of biomolecules and electrolytes," in *Innovations in Biomolecular Modeling and Simulations* (Royal Society of Chemistry, London, 2012), Chap. 4.
[23]I. S. Joung, T. Luchko, and D. A. Case, J. Chem. Phys. **138**, 044103 (2013).
[24]J. Perkyns and B. M. Pettitt, J. Chem. Phys. **97**, 7656–7666 (1992).
[25]S. M. Kast and T. Kloss, J. Chem. Phys. **129**, 236101 (2008).
[26]G. Giambasu, T. Luchko, D. Herschlag, D. York, and D. Case, Biophys. J. **106**, 883–894 (2014).
[27]K. Berlin, N. A. Gumerov, D. Fushman, and R. Duraiswami, J. Appl. Cryst. **47**, 755–761 (2014).
[28]V. Lebedev and D. Laikov, Dokl. Math. **59**, 477–481 (1999).
[29]O. Treutler and R. Ahlrichs, J. Chem. Phys. **102**, 346–354 (1995).
[30]Y. Bai, M. Greenfeld, K. J. Travers, V. B. Chu, J. Lipfert, S. Doniach, and D. Herschlag, J. Am. Chem. Soc. **129**, 14981–14988 (2007).
[31]R. Das, T. T. Mills, L. W. Kwok, G. S. Maskel, I. S. Millett, S. Doniach, K. D. Finkelstein, D. Herschlag, and L. Pollack, Phys. Rev. Lett. **90**, 188103 (2003).
[32]G. S. Manning, J. Chem. Phys. **51**, 924–933 (1969).
[33]G. S. Manning, Quart. Rev. Biophys. **11**, 179–246 (1978).
[34]V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, Biophys. J. **93**, 3202–3209 (2007).
[35]S. Kirmizialtin, A. R. J. Silalahi, R. Elber, and M. O. Fenley, Biophys. J. **102**, 829–838 (2012).
[36]I. Besseova, P. Banas, P. Kuhrova, P. Kosinova, M. Otyepka, and J. Sponer, J. Phys. Chem. B **116**, 9899–9916 (2012).
[37]M. Feig and B. M. Pettitt, Biophys. J. **77**, 1769–1781 (1999).
[38]S. Kirmizialtin, S. A. Pabit, S. P. Meisburger, L. Pollack, and R. Elber, Biophys. J. **102**, 819–828 (2012).
[39]J. J. Howard, G. C. Lynch, and B. M. Pettitt, J. Phys. Chem. B **115**, 547–556 (2011).
[40]G. Zheng, X.-J. Lu, and W. K. Olson, Nucl. Acids Res. **37**, W240–W246 (2009).
[41]J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774–777 (1951).
[42]D. I. Svergun, J. Appl. Cryst. **25**, 495–503 (1992).

[43]S. A. Pabit, X. Qiu, J. S. Lamb, L. Li, S. P. Meisburger, and L. Pollack, Nucl. Acids Res. **37**, 3887–3896 (2009).
[44]A. V. Tataurov, Y. You, and R. Owczarzy, Biophys. Chem. **133**, 66–70 (2008).
[45]S. S. Nielsen, M. Møller, and R. E. Gillilan, J. Appl. Cryst. **45**, 213–223 (2012).
[46]S. A. Pabit, K. D. Finkelstein, and L. Pollack, "Using anomalous small angle x-ray scattering to probe the ion atmosphere around nucleic acids," in *Methods in Enzymology*, edited by D. Herschlag (Academic Press, 2009), Chap. 19, Vol. 469.
[47]S. A. Pabit, S. P. Meisburger, L. Li, J. M. Blose, C. D. Jones, and L. Pollack, J. Am. Chem. Soc. **132**, 16334–16336 (2010).
[48]D. Orthaber, A. Bergmann, and O. Glatter, J. Appl. Cryst. **33**, 218–225 (2000).
[49]G. S. Kell, J. Chem. Eng. Data **15**, 119–122 (1970).
[50]K. Shahrokh, A. Orendt, G. S. Yost, and Thomas E. Cheatham, J. Comput. Chem. **33**, 119–133 (2012).
[51]I. S. Joung and T. E. Cheatham, J. Phys. Chem. B **112**, 9020–9041 (2008).
[52]P. Li, B. P. Roberts, D. K. Chakravorty, and K. M. Merz, J. Chem. Theory Comput. **9**, 2733–2748 (2013).
[53]A. Kovalenko, S. Ten-no, and F. Hirata, J. Comput. Chem. **20**, 928–936 (1999).
[54]D. M. Tiede, R. Zhang, and S. Seifert, Biochemistry **41**, 6605–6614 (2002).
[55]P. Moore, Biophys. J. **106**, 1489–1496 (2014).