Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments

Dina Schneidman-Duhovny,[†*] Michal Hammel,[‡] John A. Tainer,[§¶] and Andrej Sali[†*]

[†]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California at San Francisco; [‡]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [§]Department of Molecular Biology, Skaggs Institute of Chemical Biology, The Scripps Research Institute, La Jolla, California; and [¶]Life Sciences Division, Department of Molecular Biology, Lawrence Berkeley National Laboratory, Berkeley, California

ABSTRACT A major challenge in structural biology is to characterize structures of proteins and their assemblies in solution. At low resolution, such a characterization may be achieved by small angle x-ray scattering (SAXS). Because SAXS analyses often require comparing profiles calculated from many atomic models against those determined by experiment, rapid and accurate profile computation from molecular structures is needed. We developed fast open-source x-ray scattering (FoXS) for profile computation. To match the experimental profile within the experimental noise, FoXS explicitly computes all interatomic distances and implicitly models the first hydration layer of the molecule. For assessing the accuracy of the modeled hydration layer, we performed contrast variation experiments for glucose isomerase and lysozyme, and found that FoXS can accurately represent density changes of this layer. The hydration layer model was also compared with a SAXS profile calculated for the explicit water molecules in the high-resolution structures of glucose isomerase and lysozyme. We tested FoXS on eleven protein, one DNA, and two RNA structures, revealing superior accuracy and speed versus CRYSOL, AquaSAXS, the Zernike polynomials-based method, and Fast-SAXS-pro. In addition, we demonstrated a significant correlation of the SAXS score with the accuracy of a structural model. Moreover, FoXS utility for analyzing heterogeneous samples was demonstrated for intrinsically flexible XLF-XRCC4 filaments and Ligase III-DNA complex. FoXS is extensively used as a standalone web server as a component of integrative structure determination by programs IMP, Chimera, and BILBOMD, as well as in other applications that require rapidly and accurately calculated SAXS profiles.

INTRODUCTION

Atomic resolution modeling and validation of macromolecular assemblies in solution by small-angle x-ray scattering (SAXS) has become a key tool in structural biology (1–4). Combining data from solution scattering with atomic resolution structures holds great promise for addressing how specific complexes and conformational changes drive biological processes. High-accuracy and precision SAXS experiments with advanced instrumentation (5) lead to more confident assignment of the conformational state(s) of a given sample. Notwithstanding instrumentation developments, the accurate and rapid calculation of a model SAXS profile is in turn essential for accurate and rapid development of a structural hypothesis. Applications of SAXS profile calculation from atomic models include modeling of perturbed conformations relative to available crystal structures (6,7), determination of biologically relevant states from the crystal (8), binding of small molecules (9), structural characterization of protein flexibility (10,11), assembly of complexes from subunit structures (12–14), fold recognition (15), and comparative modeling (16,17).

Several methods are available to calculate SAXS profiles from atomic models. They differ in the calculation of the interatomic distances, in the treatment of the hydration layer, and in background adjustment (1,18,19). Interatomic distances can be computed explicitly (20) by applying the Debye formula (21), using spherical harmonics for rapid calculation (18,22,23) or based on a hierarchical spatial decomposition (24). Another approach to speed up the calculation is Monte Carlo sampling of the distances in the model (25). Coarse graining that combines several atoms in a single scattering center is also used to speed up profile calculation (26–28). However, with experimental data of higher accuracy, precision, and range, there is a greater need for high-accuracy rapid computations of SAXS profiles to interpret experimental data for informative results. By using explicit atom distances (20) and water models to account for the effect of solvent (18), superior fits between experimental high-resolution structures and SAXS data are obtained (3). The explicit representation of the molecule is particularly useful for multidomain flexible assemblies, which frequently adopt highly anisometric shapes (18).

A SAXS experiment involves determining both the scattering profile of a solute in a buffer and a buffer on its own; a scattering profile is an intensity of a sample as a function of spatial frequency. The experimental scattering profile of a solute with its ordered hydration layer is that obtained as the difference between these two profiles. The theoretical SAXS profile is also computed as the difference between the scattering of a single solute molecule with its ordered hydration layer and the scattering of the excluded volume (i.e., the scattering of the bulk solvent in the excluded volume of the solute with its ordered hydration layer)

Submitted September 10, 2012, and accepted for publication July 11, 2013.

*Correspondence: sali@salilab.org or dina@salilab.org

Editor: Lois Pollack.

0006-3495/13/08/0962/13 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.07.020

(22,29,30). However, accurate approximation of the excluded volume is a challenging task, because the volume varies significantly depending on the atomic radii. Additionally, the protein crystal packing and cryo-cooling used in x-ray crystallography may reduce the protein volume in the crystal compared to that in solution (31). Thus, adjustment of the excluded volume is required for optimal matching of the experimental SAXS profile.

Although solute atoms dominate the scattering signal at high angles, the scattering from the ordered solvent atoms in the hydration layer must also be considered (32). Strongly bound water molecules are known to fill surface grooves and channels, stabilizing their structures and smoothing the excluded volume (33). The first hydration layer with an average density ~10% larger than that of the bulk solvent was observed using x-ray and neutron scattering (34). Computationally, the hydration shell has been modeled by explicitly placing water molecules on the surface (18,26,28,35–37). Alternatively, implicit models surround the solute particle by a continuous envelope representing the hydration shell of 3 Å with a density that can differ from the free solute (22) or use voxelized representation of the protein and the hydration layer (23,38–40). In implicit models, the density of the hydration layer can be adjusted for optimal fitting to the experimental SAXS profile. Modeling the hydration layer improves the accuracy of the calculated scattering profiles, although the contribution from the ordered hydration layer is several orders of magnitude lower than the scattering from the solute or the excluded volume (40).

Here, we developed fast open-source x-ray scattering (FoXS) to take advantage of high-quality data and enable us to better distinguish between proposed atomic models. FoXS provides an efficient method with explicit computation of interatomic distances, accurate estimation of the excluded volume, and an implicit model of the first hydration layer. FoXS was tested on 14 cases with available high-quality experimental SAXS profiles and crystallographic atomic structures. We also compared FoXS to four state-of-the-art programs, including CRYSOL (22), AquaSAXS (39), the Zernike polynomials-based method (23), and Fast-SAXS-pro (28) for accuracy and running times. In addition, we performed an in-depth comparison of the excluded volume and hydration layer modeling to the most commonly used program CRYSOL. To assess the relevance of adjusting the hydration layer density parameter, we performed contrast variation experiments and found that FoXS can accurately represent the changes in the density of the hydration layer.

METHODS

Profile computation

The Debye formula (21) is used for accurate computation of SAXS profiles,

$$I(q) = \sum_{i=1}^{N}\sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}}, \quad (1)$$

where the intensity, $I(q)$, is a function of the momentum transfer $q = (4\pi \sin\theta)/\lambda$, in which 2θ is the scattering angle and λ is the wavelength of the incident x-ray beam. The value $f_i(q)$ is the atomic form factor, d_{ij} is the distance between atoms i and j, and N is the number of atoms in the molecule.

Form factors

In FoXS, the form factor $f_i(q)$ takes into account the displaced solvent as well as the hydration layer,

$$f_i(q) = f_i^v(q) - C_1(q) f_i^s(q) + c_2 s_i f^w(q), \quad (2)$$

where $f^v(q)$ is the atomic form factor in vacuo, $f^s(q)$ is the form factor of the dummy atom that represents the displaced solvent represented by the Gaussian function (29), s_i is the fraction of solvent-accessible surface of the atom i (41), and $f^w(q)$ is the water form factor. The surface is calculated using a probe with radius of 1.8 Å. The third term in Eq. 2 implicitly accounts for the first hydration layer by placing a water molecule with the same coordinates as the solvent-exposed solute atom. The function $C_1(q)$ is used to adjust the total excluded volume of the atoms and is equivalent to the $G(s)$ function in CRYSOL (22) (Eq. 13),

$$C_1(q) = c_1^3 \exp\left[\frac{-\left(\frac{4\pi}{3}\right)^{3/2} q^2 r_m^2 \left(c_1^2 - 1\right)}{4\pi}\right], \quad (3)$$

where c_1 is the scaling of the atomic radius (default value = 1.0) and r_m is the average atomic radius of the molecule. During fitting, we allow 5% variance of the radius $0.95 \le c_1 \le 1.05$. For comparison, CRYSOL allows ~11% variance in the atomic radius ($1.4 \le r_m \le 1.8$, r_m value is ~1.62 for proteins). The parameter c_2 is used to adjust the difference between the densities of the hydration layer and the bulk water (default value = 0.0). During fitting, the value of c_2 can vary from 0 to 4.0, reflecting an estimate of up to four water molecule neighbors for an exposed solute atom. This threshold is comparable to the average number of water molecules within 3 Å of each other in the TIP3P water box. The hydration shell density for $c_2 = 4.0$ is 0.388e/Å3. The density of the hydration layer around the protein can, in principle, be lower than that of bulk water, depending on the amount of surface charge. Therefore, we also allow c_2 to be negative ($-2.0 \le c_2 \le 4.0$). The hydration shell density for $c_2 = -2.0$ is 0.307e/Å3.

Profile fitting

The computed profile is fitted to a given experimental SAXS profile by minimizing the χ-function with respect to c, c_1, and c_2,

$$\chi = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{I_{\exp}(q_i) - cI(q_i)}{\sigma(q_i)}\right)^2}, \quad (4)$$

where $I_{\exp}(q)$ is the experimental profile, $\sigma(q)$ is the experimental error of the measured profile, M is the number of points in the profile, and c is the scale factor. The minimal value of χ is found by enumerating c_1 and c_2 ($0.95 \le c_1 \le 1.05$ and $-2.0 \le c_2 \le 4.0$), in steps of 0.005 and 0.05, respectively, and performing the linear least-squares minimization to find the value of c that minimizes χ for each c_1, c_2 combination. To avoid profile computation for each combination of c_1 and c_2, the multiplication of two

form factors from Eq. 1, $f_i(q)f_j(q)$, is divided into six parts independent of the c_1 and c_2 values:

$$\begin{aligned} f_i(q)f_j(q) \;=\; & \underbrace{f_i^v(q)f_j^v(q)}_{1} + C_1^2(q)\underbrace{f_i^s(q)f_j^s(q)}_{2} \\ & - C_1(q)\underbrace{\left(f_i^v(q)f_j^s(q) + f_i^s(q)f_j^v(q)\right)}_{3} \\ & + c_2\underbrace{f^w(q)\left(s_j f_i^v(q) + s_i f_j^v(q)\right)}_{4} \\ & - C_1(q)c_2\underbrace{f^w(q)\left(s_j f_i^s(q) + s_i f_j^s(q)\right)}_{5} \\ & + c_2^2\underbrace{\left(f^w(q)\right)^2 s_i s_j}_{6}. \end{aligned} \tag{5}$$

Profiles are computed separately using Debye formula for each of the six parts and summed up during parameter enumeration.

Form factors and distances approximation

The computational time for the Debye formula is proportional to $N^2|q|$, where N is the number of atoms and $|q|$ is the number of points in the computed profile (typically $|q| > 100$). To allow for rapid calculation of many profiles, form factors and distances are approximated as follows.

A form factor $f_i(q)$ is approximated by $f_i(0)E(q)$, where $f_i(0)$ is the form factor at $q = 0$ and $E(q)$ is an approximation function (13). The atomic form factor in vacuo is simply the number of electrons in the atom. The form factor of the dummy atom is ρV_i, where ρ is the electron density of the solvent and V_i is the solvent volume displaced by the atom. The form factors of hydrogen atoms are added to the covalently bound heavy atom. $E(q)$ is approximated by a Gaussian function e^{-bq2}, where b is obtained from the full form factor calculation (13).

The distances between atoms are binned into a distance histogram, where each bin specifies the number of atom pairs with distance $r \pm \Delta$. However, instead of counting atom pairs at distance $r \pm \Delta$, their zero form factors are summed up into six form factor histograms, $F(r)$, independent of c_1 and c_2 values (Eq. 5). The bin size of 0.5 Å is much smaller than the highest resolution of the data (at the maximal q of 1 Å^{-1}, the resolution is 2πÅ).

These approximations allow for a rapid evaluation of the Debye formula, without sacrificing accuracy. Once the distance histogram $F(r)$ is computed, it is converted into reciprocal space:

$$I(q) \;=\; E^2(q)\sum_r F(r)\frac{\sin(qr)}{qr}. \tag{6}$$

For additional speed-up, a squared-distance histogram is used in practice; squared interatomic distance is computed first and the square root is computed for r during conversion to reciprocal space in Eq. 7. These approximations reduce the computational complexity from $N^2|q|$ to N^2 only.

Additional program options

The program has additional options that may be useful in different applications:

χ^2_{free} for assessing model data agreement

Analogous to R_{free} in x-ray crystallography, χ^2_{free} uses a cross-validation scheme, where only a fraction of randomly selected data points are used for calculating the χ-score. The value χ^2_{free} is taken as a median over multiple rounds of selections of data points (42).

Profile offset

A constant can be added to the experimental data to minimize χ:

$$\chi \;=\; \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{I_{\exp}(q_i) - cI(q_i) + a}{\sigma(q_i)}\right)^2}. \tag{7}$$

This offset accounts for possible systematic errors in the experimental data due to mismatched buffers. The optimal value of the constant is found using the linear least-squares minimization.

Background correction for high-resolution datasets

FoXS provides optional background correction for high-resolution SAXS data ($q_{\max} > 0.3$ Å^{-1}) with small noise (43). Briefly, the profile is transformed into

$$I_{\text{corrected}}(q) \;=\; \frac{I(q)}{[1 + G_1q^2 + (G_1/4 + G_2/12)q^4]}, \tag{8}$$

where $G_1 = B/C$ and $G_2 = 3(4AC\text{-}B^2)/C^2$. To determine the values of A, B, and C, the SAXS profile is transformed to

$$\sum_{q>0.2} q^2 I(q)$$

as a function q^2; A, B, and C are the coefficients of the parabola that fits this function.

Residue-level coarse-graining

The profile is computed using one point per residue, resulting in significantly faster but less accurate calculation. The option can be used for initial filtering of unlikely models, followed by more accurate atomic resolution profile calculation.

Log-scale fitting

This option allows the user to fit the profiles using logarithms of intensity values and their standard deviations.

Single form factor profile computation

This option allows estimating the profile of shapes represented by a bead model.

Benchmark

FoXS was tested on 14 experimental datasets and corresponding high-resolution structures (Table 1). The cases vary in size (from 424 to 12,833 atoms), in shape (globular versus elongated), and in molecule type (11 proteins and three nucleic acids). Atomic structures fit the profiles well in all cases (Fig. 1), including three cases with high-resolution profiles ($q_{\max} = 0.5$ Å^{-1}) of glycosyl hydrolase (44), glucose isomerase, and superoxide dismutase (45). The application of FoXS in combination with minimal ensemble search (MES) (11) was tested on the XLF-XRCC4 complex (46) and flexible human ligase III-DNA assembly (47).

Data collection

All SAXS data was collected at the Advanced Light Source SIBYLS beamline (BL12.3.1) (5) at Lawrence Berkeley National Laboratory (Berkeley, CA), except for glycosyl hydrolase data that was collected at the European Synchrotron Radiation Facility (Grenoble, France) beamline ID02 (48). The datasets are available from BIOISIS.net (Table 1). The SAXS data for contrast variation experiments have been collected using two standard proteins, lysozyme (14 kDa) and glucose isomerase (174 kDa). Proteins at the concentration 5 mg/mL were first dialyzed in buffer containing

TABLE 1 Benchmark results (χ-values and running times in seconds) for FoXS, CRYSOL, AquaSAXS, the Zernike polynomials-based method, and Fast-SAXS-pro

			FoXS		CRYSOL		AQUASAXS		Zernike method		Fast-SAXS-pro	
Molecule	PDB	BIOISIS-ID	χ	Time	χ	Time	χ	Time	χ	Time	χ	Time
Rubredoxin (6)	1BQ9	1RBDGP	7.05	0.14	7.60	2.37	10.50	80.0	15.24	6.73	39.23	5.07
Lysozyme (this work)	2VB1	LYSOZP	1.89	0.22	1.77	2.79	1.77	107.0	4.53	30.67	4.09	20.52
28 bp DNA	—	28BPDD	1.70	0.45	1.57	2.76	—	—	4.52	13.42	1.94	51.83
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3 (67)	3PXJ	LAR12P	1.16	0.39	1.24	3.02	1.37	118.0	5.11	31.05	3.44	58.55
S-adenosylmethionine riboswitch mRNA (56)	2GIS	2SAMRR	2.28	0.44	2.25	3.16	2.11	135.0	2.47	20.36	3.69	82.29
Superoxide dismutase (45)	1HL5	APSOD	1.69	0.51	1.74	3.51	2.54	187.0	11.50	31.22	17.76	105.38
Abscisic acid receptor PYR1 (6)	3K3K	1PYR1P	2.03	0.69	1.39	3.65	2.93	117.0	5.06	31.32	4.06	147.77
Glycosyl hydrolase + C-terminus (44)	1EDG	AT5GHP	1.30	0.78	1.04	4.07	1.10	190.0	2.23	42.43	2.46	204.95
P4-P6 RNA Ribozyme domain (56)	1GID	1P4P6R	4.39	1.00	2.81	3.79	3.51	170.0	5.35	32.88	10.28	212.11
Superoxide reductase (5)	1DQK	1APXGP	4.69	1.12	3.55	4.16	7.96	141.0	9.95	38.78	17.66	181.71
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain (57)	3T7E	ATG7CP	2.16	1.70	2.54	4.69	2.53	172.0	3.87	50.86	5.05	339.31
DNA double-strand break repair protein MRE11 (58)	3AV0	MRERAP	1.19	7.28	0.72	7.89	1.14	351.0	1.80	112.46	6.76	1746
Glucose isomerase (this work)	2G4J	GISRUP	4.69	6.69	7.99	8.10	10.31	514.0	24.74	184.00	25.33	2088
Complement C3b + Efb-C (59)	—	C3BEFP	1.62	8.18	2.51	8.56	1.77	494.0	2.48	230.29	9.24	1860
Average			2.70	2.11	2.77	4.47	3.81	213.54	7.06	61.18	10.79	507
Median			1.96	0.74	2.01	3.72	2.53	170.00	4.80	32.10	5.91	165

References in the first column indicate the sources of experimental SAXS profiles. Benchmark cases are ordered according to the number of atoms in the structure.

50 mM Tris pH 7.6. Dialyzed proteins were aliquoted into the 50-μL fractions that were diluted with 50 μL of different NaCl or KCl salt solutions giving final salt concentration varying between 25 and 1000 mM. The corresponding buffer blanks for SAXS experiments were prepared identically by diluting 50-μL aliquots of the dialyzing buffer with salt solutions. SAXS data was collected and analyzed as described in Hura et al. (5). These datasets are also available from BIOISIS.net with IDs LNaClP, LYKClP, GNaClP, and GIKClP for lysozyme and glucose isomerase in NaCl and KCl, respectively.

RESULTS

FoXS is fast and accurate

FoXS has produced accurate fits for the 14 benchmark cases (Fig. 1, Table 1). The average χ-value was 2.70 and the average running time was 2.11 s. Next, we have tested the accuracy of the form factors and distance histogram approximations by performing profile calculation using the Debye equation directly. There was no significant difference in the accuracy of the fits as measured by χ-values (see Fig. S1 in the Supporting Material). However, the average running time was significantly longer (13 min). We have also tested whether increasing the range of values for c_1 and c_2 parameters can result in improved fits. Setting the ranges to $0.85 \leq c_1 \leq 1.15$ and $-4.0 \leq c_2 \leq 8.0$ did improve the fits to the average χ-value of 2.64 (compared to 2.70). However, because the physical model does not justify these ranges, we did not modify the default values. We have also tested the effect of the probe radius for calculation of atomic solvent accessibility. The average χ-value was 2.72, 2.71, 2.7, and 2.7 for the radii of 1.4, 1.6, 1.8, and 2.0 Å, respectively.

Comparison with other programs

FoXS results were compared to the four state-of-the-art programs broadly used for profile computation, including CRYSOL (ATSAS package 2.4.1-6, EMBL Hamburg, Hamburg Germany; http://www.embl-hamburg.de/biosaxs/crysol.html), AquaSAXS (Delarue Group, Institut Pasteur, Paris, France; http://lorentz.dynstr.pasteur.fr/aquasaxs.php), the Zernike polynomials-based method, and Fast-SAXS-pro (http://yanglab.case.edu/software.html). All programs were assessed for accuracy as reflected by the χ-value and for efficiency as recorded by running times. The running times were measured on a standard PC computer (Intel Xeon 2.27 GHz) for FoXS, CRYSOL, Zernike polynomials-based method, and Fast-SAXS-pro. Because the download version for AquaSAXS is not available, the running times were taken from the server log. All the runs were performed with default parameters for all programs. By default, FoXS produces a profile sampled with $|q| =$ 500 data points, which is comparable to the number of points in the experimental profiles. The average number of points in our benchmark profiles is 600. In CRYSOL the default is 51 data points. For comparison, we increased the number of data points in CRYSOL to the maximum possible value of 256. AquaSAXS also produces 50–100 profile points by default; however, increasing profile resolution resulted in significantly longer running times, therefore the default profile sampling was used. Fast-SAXS-pro requires $\log^{10}$(intensity) for the input experimental profile and also outputs the $\log^{10}$(intensity) in the output profile. We have therefore converted the experimental profile to $\log^{10}$(intensity) for input to Fast-SAXS-pro. Because Fast-SAXS-pro

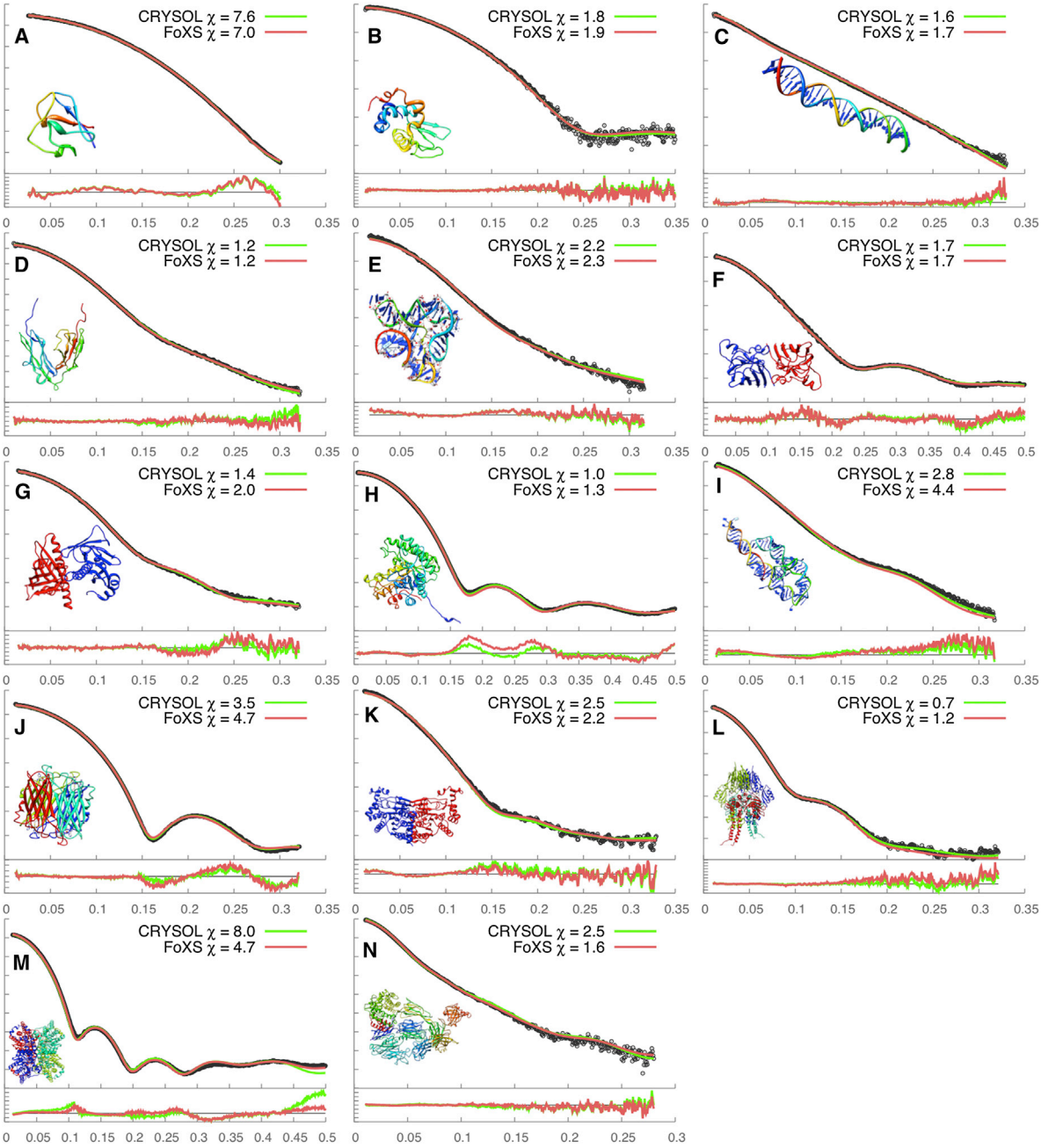

FIGURE 1 Fit and residual plots for FoXS and CRYSOL: experimental data (*dark gray*), FoXS (*red*), and CRYSOL (*green*). The fit plots are for $q(\text{Å}^{-1})$ (x axis) versus log-intensity (y axis). The residual plots are for $q(\text{Å}^{-1})$ (x axis) versus experimental intensity divided by computed intensity (y axis). The cases are ordered by the number of atoms as in Tables 1 and 2.

does not report the χ-value, we have converted the $\log^{10}$(intensity) theoretical profile back to normal intensity and used FoXS χ-minimization protocol to compute the χ-value. A scattering profile calculated by FoXS, CRYSOL, AquaSAXS, and the Zernike polynomials-based method was fit to the experimental curve by adjusting the excluded volume and hydration layer density parameters as well as the scaling factor separately for each case in the benchmark.

In contrast, a scattering profile from Fast-SAXS-pro was calculated by adjusting only the scaling factor. Overall, FoXS had the lowest average χ-value (2.70) and the shortest average running time (2.11 s) for the 14 benchmark cases (Table 1 and Fig. 1, and see Fig. S2 and Fig. S4). CRYSOL produces a similar range of χ-values (average $\chi = 2.77$) with slightly longer running times (4.47 s), especially for smaller proteins. AquaSAXS, and the Zernike polynomials-based method, have higher χ-values and significantly longer running times (Table 1). Fast-SAXS-pro has the highest χ-values (average $\chi = 2.77$) and the longest running times (507 s). Due to the coarse-graining of the molecules,

the profiles computed by Fast-SAXS-pro often do not fit the data at high q values ($q > 0.25$) (see Fig. S4, *A*, *D*, *F–I*, and *M*). Moreover, there is often a significant deviation from the experimental profile at low q values (see Fig. S4, *F*, and *K–N*). These results indicate that obtaining a good fit to the data requires a full atom model as well as fitting excluded volume and hydration layer density parameters, in addition to the scaling factor. Because χ-values may vary due to profile experimental noise and are not directly comparable between different experimental datasets, we have also compared the median χ-value for the 14 benchmark cases among the five programs. FoXS had the lowest median χ-value (1.96), followed by CRYSOL (2.01) and AquaSAXS (2.53).

FoXS accurately accounts for the excluded volume contribution

Whereas the accuracy and efficiency of the programs can be measured by χ-values and running times, respectively, most of the programs include a few free adjustable parameters, such as c_1 and c_2 in FoXS. Next, we compared CRYSOL and FoXS in terms of optimal adjustable parameters.

FoXS c_1 parameter is equivalent to r_0/r_m in CRYSOL, where r_m is the average atomic radius of the molecule (typically 1.62 Å) and r_0 is enumerated for best fit of the data ($1.4 < r_0 < 1.8$). The FoXS c_2 parameter is comparable to $\delta\rho$ in CRYSOL and can be converted to $\delta\rho$ by taking the c_2 value of 4.0 to correspond to the density of 0.388e Å^{-3}. Both programs produce similarly good fits to the data. In eight of the fourteen cases, the χ-values are within 0.5 units of each other; in three cases, FoXS produces a better fit and in the remaining three cases CRYSOL produces a better fit (Table 2, Fig. 1).

The closer look at the fitting parameter that allows scaling of the atomic radius shows large variation in CRYSOL. However, the values of c_1 parameter in FoXS are much closer to the ideal value of 1.0. In FoXS, the values vary between 0.95 and 1.05 with a standard deviation of 0.025, whereas in CRYSOL c_1 significantly varies between 0.865 and 1.265 with a standard deviation of 0.12 (Table 2, Fig. 2 *a*). This variation indicates that CRYSOL significantly adjusts the excluded volume variable to achieve a better fit of the computed profile to the experimental data. On the other hand, the c_2 parameter correlates well with $\delta\rho$ in CRYSOL (Table 2, Fig. 2 *b*), indicating that the adjustment of the hydration layer density is similar for the two programs.

In the case of glucose isomerase data with $q_{max} = 0.5$ Å^{-1}, FoXS produces a significantly better fit, especially in the high q region. The discrepancy for the high q range is due to the large size of glucose isomerase, which makes it difficult for the default number of multipoles to accurately represent the density of the molecule. Similar discrepancies are also observed for other tested programs using multipole expansion (see Fig. S1, Fig. S2, and Fig. S3). Increasing the order of harmonics from a default value of 15–50 improves the fit from $\chi = 8.0$ to $\chi = 6.4$, although the running time increases from 8 to 25 s. For comparison, the default FoXS run takes 7 s and the χ-value is 4.7.

FoXS is significantly faster than CRYSOL for average size structures below 10,000 atoms (Fig. 2 *c*). The programs have similar performance for structures with ~15,000 atoms. However, for this size of macromolecule, CRYSOL requires higher-order harmonics to obtain more accurate spherical averaging. In summary, CRYSOL is slower than FoXS across the whole range of macromolecules that are typically studied by SAXS.

Importance of excluded volume and hydration layer fitting

To demonstrate that FoXS fitting of excluded volume and hydration layer does not result in over-fitting of the experimental profile, we test the ability of the χ-score to

TABLE 2 Benchmark results for FoXS and CRYSOL

Molecule	PDB	No. of atoms	q_{max}	FoXS χ	CRYSOL χ	FoXS c_1	CRYSOL c_1	FoXS c_2	CRYSOL $\delta\rho$
Rubredoxin	1BQ9	424	0.30	7.05	7.60	1.02	1.10	−0.15	0.01
Lysozyme	2VB1	1001	0.35	1.89	1.77	1.00	0.87	1.91	0.02
28 bp DNA	—	1209	0.33	1.70	1.57	1.05	1.11	3.28	0.05
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	3PXJ	1624	0.32	1.16	1.24	1.04	1.10	−0.33	0.01
S-adenosylmethionine riboswitch mRNA	2GIS	2058	0.32	2.28	2.25	0.95	1.12	0.88	0.01
Superoxide dismutase	1HL5	2220	0.50	1.69	1.74	1.02	0.88	−0.18	0.00
Abscisic acid receptor PYR1	3K3K	2924	0.32	2.03	1.39	1.02	0.87	−0.82	0.00
Glycosyl hydrolase + C terminus	1EDG	3035	0.50	1.30	1.04	1.02	1.04	0.05	0.01
P4-P6 RNA ribozyme domain	1GID	3409	0.32	4.39	2.81	1.05	1.21	4.00	0.07
Superoxide reductase	1DQK	4060	0.32	4.69	3.55	1.03	1.00	0.68	0.01
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	3T7E	5318	0.33	2.16	2.54	1.00	1.11	4.00	0.04
DNA double-strand break repair protein MRE11	3AV0	12,148	0.32	1.19	0.72	1.00	1.07	4.00	0.04
Glucose isomerase	2G4J	12,176	0.50	4.69	7.99	1.02	0.87	0.23	0.01
Complement C3b + Efb-C	—	12,833	0.28	1.62	2.51	1.01	0.87	4.00	0.05

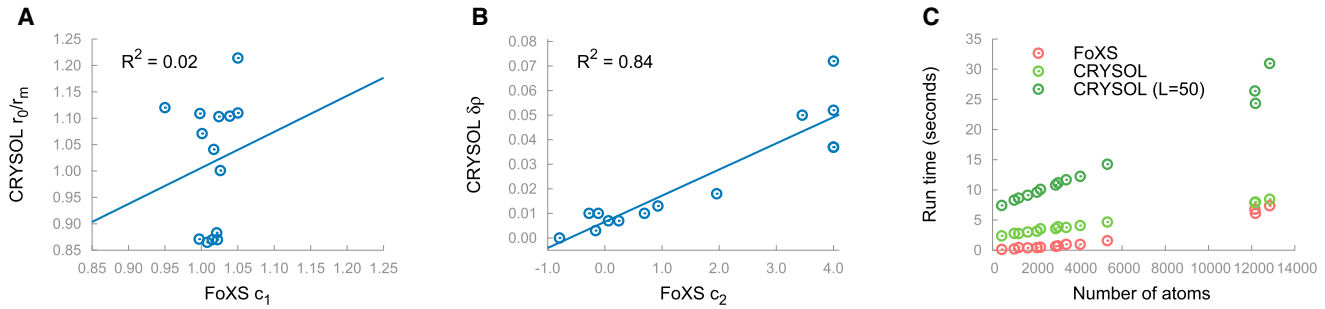

FIGURE 2 Comparison of FoXS and CRYSOL adjustable parameters and running times. (*A*) CRYSOL r_0 versus c_1. (*B*) CRYSOL $\delta\rho$ versus FOXS c_2. (*C*) Running times of FoXS (*red*) versus CRYSOL (*green*) and CRYSOL L50 (*dark green*, order of harmonics = 50).

discriminate structural models based on their accuracy. Ideally, the fitting scores would correlate strongly with the accuracy of the model over a broad range of accuracy (i.e., C_α RMSD from 0 to 10 Å). We compare this correlation for the default FoXS protocol, which fits the excluded volume and hydration layer as well as the scaling factor, to the correlation for FoXS′ that fits only the scaling factor, but fixes the excluded volume and hydration layer ($c_1 = 1.0$, $c_2 = 0.0$). Each dimeric structure in our benchmark (superoxide dismutase, abscisic acid receptor PYR1, ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain, and DNA double-strand break repair protein MRE11) was separated into monomers. Dimeric models were generated using docking by PatchDock or SymmDock (49).

We have examined the correlation of χ-scores with RMSD for models with RMSD $<$ 10 Å (Fig. 3, *a–d*). For FoXS, in all cases except for the abscisic acid receptor PYR1, the Pearson cross-correlation coefficient of χ-values with RMSD is in the range from 0.42 to 0.48. In contrast, for FoXS′, no significant correlation is observed. For abscisic acid receptor PYR1, the correlation coefficient for FoXS is lower (0.24) and similar to that for FoXS′; it is possible that there are additional conformations in solution, because upon binding of abscisic acid, PYR1 adopts a more compact closed conformation (6). In addition, we have extracted all glucose isomerase structures from the PDB and computed the correlation between the resolution of the structure and the values of χ-scores (Fig. 3 *e*). Although there is a significant correlation for the default FoXS, there is no correlation for FoXS′. For comparison, the CRYSOL program, which also adjusts the excluded volume and hydration layer, has a cross-correlation comparable to FoXS for superoxide dismutase, abscisic acid receptor PYR1, and DNA double-strand break repair protein MRE11 cases, and no significant correlation for ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain and glucose isomerase cases.

Validation of hydration layer model by contrast variation experiments

SAXS profiles for glucose isomerase and lysozyme at the concentration of ~2.5 mg/mL were collected under different ionic strengths using NaCl or KCl concentrations varying between 0 and 1000 mM. Each profile was fitted to high-resolution structures (PDB:2GLK and PDB:2VB1) with FoXS and CRYSOL. We examined changes in the density of water layer as represented by fitting parameters c_2 for FoXS and $\delta\rho$ for CRYSOL (Fig. 4). To simplify comparison between FoXS and CRYSOL c_2 values were converted to $\delta\rho$ by fixing c_2 to 4.0, corresponding to the density of 0.388e Å^{-3}.

For glucose isomerase, the values of $\delta\rho$ do not change across different salt concentrations, indicating that the density of hydration layer is not varying. For lysozyme, the $\delta\rho$-value increases from -0.007 to 0.03e Å^{-3} for the higher salt concentrations and indicates an increase in the hydration layer density at higher salt concentrations. We observed similar changes in hydration layer parameters for both FoXS and CRYSOL.

The difference between glucose isomerase and lysozyme is presumably associated with different dynamics of the hydrogen-bonding network of water molecules near the protein surface. Glucose isomerase is a highly charged protein with a highly dynamic hydration shell (50) and ions accumulating away from the hydration layer. In contrast, lysozyme has only a few exposed charged residues, whereas the surface residues are more hydrophobic with polar backbone. Addition of ions to this kind of protein/water binary system can cause reorganization of the hydrogen bonds, resulting in an increased hydrogen bond lifetime (51). For this type of protein surface, increasing ionic strength is associated with incorporation of ions into the first hydration layer (51), causing higher electron density of the first hydration layer. This incorporation of salt ions and the resulting higher density of the first hydration layer explain the increase in c_2 at high salt concentrations for lysozyme.

Furthermore, we show that FoXS parameter c_2 can successfully mimic the hydration layer as determined by water molecules in high-resolution crystal structures. We used sub-Ångstrom resolution x-ray structures for both lysozyme and glucose isomerase that contain positions of explicit water molecules up to the 6 Å distance from the protein surface (lysozyme PDB:2VB1 with 170 waters; glucose isomerase PDB:2GLK with 1705 waters). The water molecules

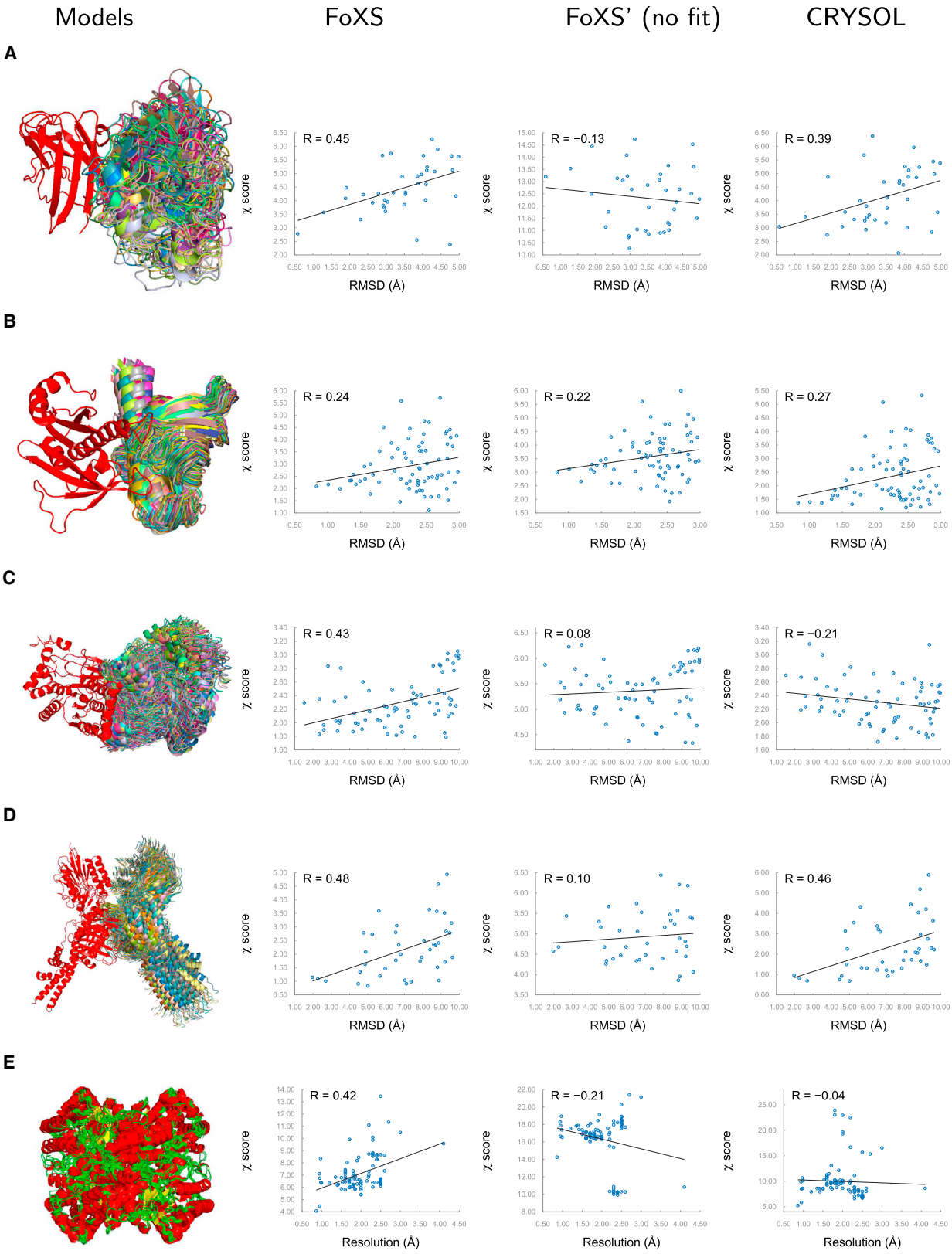

FIGURE 3 Cross correlation of χ-scores with model accuracy. (*A*) Superoxide dismutase, (*B*) abscisic acid receptor PYR1, (*C*) ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain, (*D*) DNA double-strand break repair protein MRE11, and (*E*) glucose isomerase. For each of the five cases, a set of models is shown, followed by the χ-scores versus model accuracy for FoXS, FoXS′, and CRYSOL.

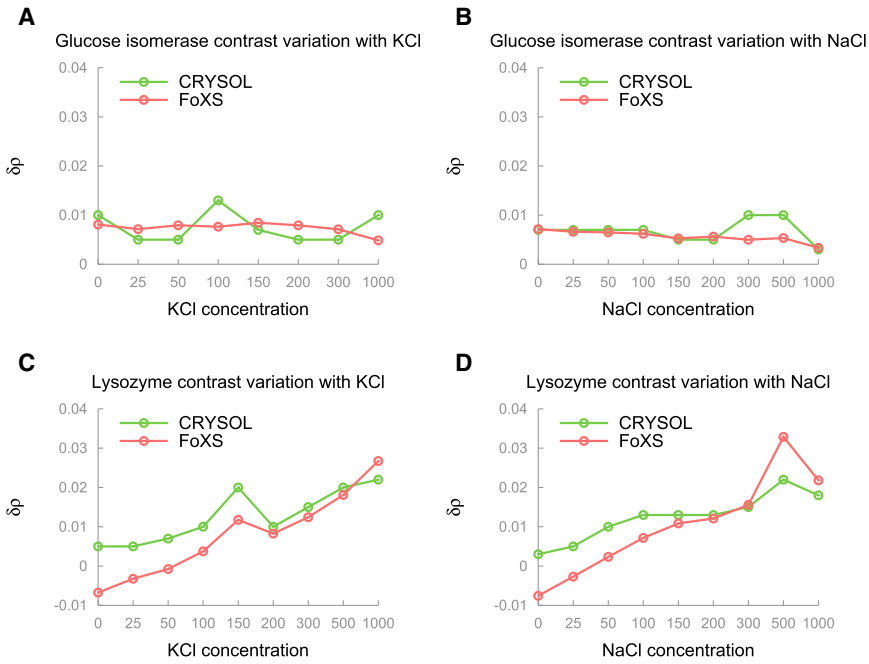

FIGURE 4 Salt concentration versus hydration layer density ($\delta\rho$) for FoXS (*red*) and CRYSOL (*green*). (*A* and *B*) Glucose isomerase with KCl and NaCl, respectively. (*C* and *D*) Lysozyme with KCl and NaCl, respectively.

were used explicitly by FoXS as additional atoms, and hydration layer modeling was turned off ($c_2 = 0$). We performed FoXS calculations with varied thickness of the hydration shell using different cutoffs for the protein surface-crystal water distances. The distance to a closest protein atom defines the distance of a crystal water molecule to the protein surface.

FoXS calculated the fit to experimental data including all salt concentrations using crystal waters, starting from distance 0 (no hydration layer) to 5.5 Å (see Table S1 and Table S2 in the Supporting Material). For both tested proteins, the fits using crystal waters are similar in accuracy to the fits using the FoXS hydration layer model. In addition to the similar accuracy of the fits for implicit and explicit hydration layers, the χ-values obtained for both models are similar. The minimal χ-value obtained for crystal water models over the range of distances (columns 4–8, *bold-green*) is comparable with the χ-value for the FoXS-modeled hydration layer (column 2, *bold*). The dependence of the c_2 parameter on the salt concentration is similar to that of the crystal water layer thickness. The optimal thickness of the crystal water layer for lysozyme increases with higher salt concentrations (Fig. 5 *a*). For glucose isomerase, the optimal crystal water distance of 2.8 Å gives the best fit for all salt concentrations (Fig. 5 *c*), further confirming that the hydration layer does not change with an increase in salt concentration. In addition, we show that the optimal thickness of crystal water layer correlates well with the c_2 values of the implicit hydration layer for both lysozyme (Fig. 5 *b*) and glucose isomerase (Fig. 5 *d*). These results confirm that the FoXS c_2 parameter efficiently mimics the changes in the hydration layer and that the hydration layer needs to be included for accurate fitting to the experimental SAXS profiles.

Application of FoXS to modeling macromolecular flexibility

SAXS profiles are informative about macromolecular flexibility. Previously, we developed a SAXS-based modeling technique describing dynamic systems in terms of ensembles of structures (11). Here, we take advantage of FoXS as a fast and accurate SAXS calculator, and combine it with the minimal ensemble search (MES). We apply the approach to describe the transient states of human XLF-XRCC4 filaments and the intrinsic flexibility of human ligase III-DNA complex.

The XLF-XRCC4 complex forms filaments in the crystal lattice (8). SAXS was used to determine the transient character and filament formation in solution. Theoretical profiles were computed for various filament lengths using FoXS, followed by an ensemble analysis using MES (11). The selected combination of various filament lengths achieved a significant improvement in the fit to the experimental data. The best fit obtained with the single filament length had the χ-value of 5.1, whereas the minimal ensemble of three different filament lengths reduced the χ-value to 1.7 (Fig. 6, *a* and *b*) (46).

In our second example, we present FoXS-MES application for the identification of the conformational flexibility of the human ligase III-DNA assembly. The crystal structure of the ligase III-DNA complex assumed a ring-like structure with the ligase III wrapped around a DNA molecule. This closed conformation fit experimental SAXS data very

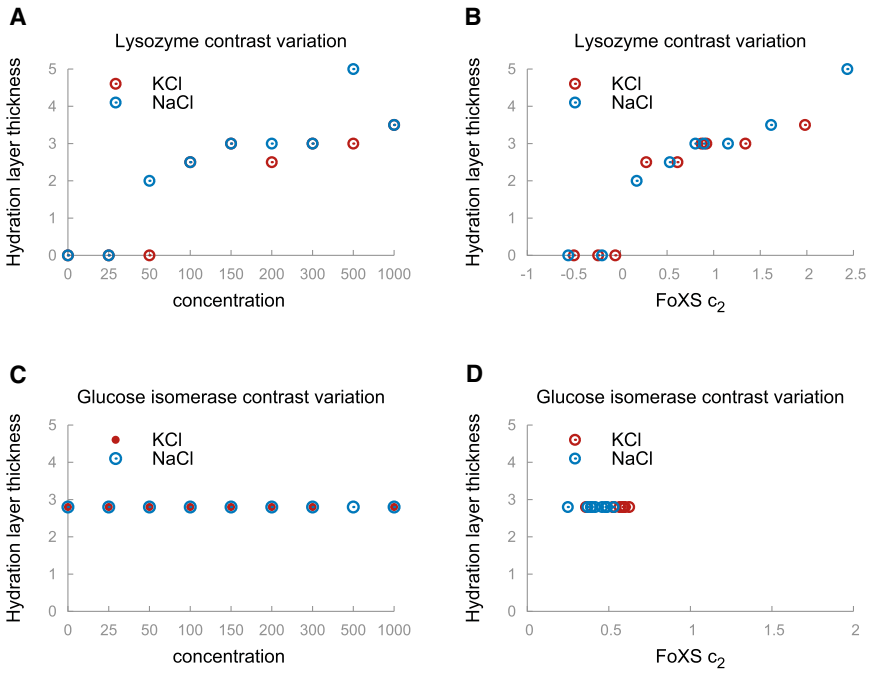

FIGURE 5 The thickness of hydration layer based on crystal waters versus concentration for (*A*) lysozyme and (*C*) glucose isomerase. The thickness of hydration layer based on crystal waters versus hydration layer density (FoXS c_2) for (*B*) lysozyme and (*D*) glucose isomerase. (*Blue*) NaCl; (*red*) KCl.

poorly ($\chi = 12.9$). To match the data, the mixture of open and closed states was selected by MES ($\chi = 5.1$, Fig. 6, *c* and *d*) (47). The significant improvement of FoXS-MES fit in comparison to single FoXS fit in the entire q range indicates the flexibility of the molecule.

FoXS allows simultaneous analysis of multiple models against experimental data. Because MES (11) is part of the FoXS suite presented here, the FoXS-MES combination becomes a powerful tool to identify the flexibility (52) as well as structural and compositional heterogeneity (8) of the studied system.

Software availability

FoXS is a part of the open source IMP package (http://salilab.org/imp). FoXS is also available for download and as a web-server from http://salilab.org/foxs, and as a part of UCSF Chimera visualization software (53). MES application for ensemble analysis is a part of the FoXS web-server and also available for download from SIBYLS http://bl1231.als.lbl.gov/. The model decoy dataset used is available from http://modbase.compbio.ucsf.edu/foxs/decoys.zip.

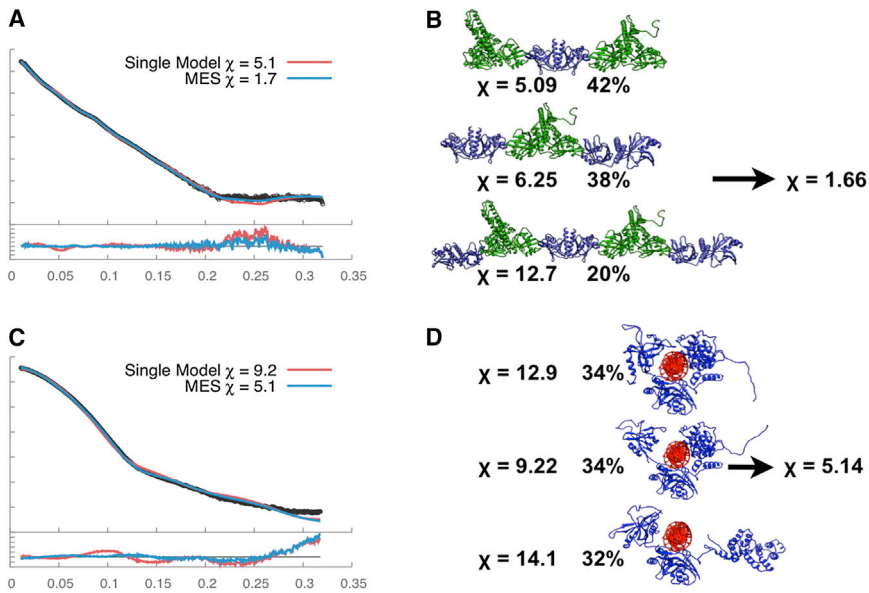

FIGURE 6 Finding the minimal ensemble consistent with a SAXS profile. Ensemble fit (*blue*) of the SAXS data (*black*) versus single conformation fit (*red*) for (*A*) XLF-XRCC4 filaments and (*C*) ligase III-DNA complex. (*B* and *D*) Conformations and their weights in the minimal ensemble.

DISCUSSION

Major challenges in structural biology include efficiently and accurately determining the structures of proteins and their assemblies in solution (2,5,54). Here, we present the development of software for rapid, accurate, and user-friendly calculation of SAXS profiles from atomic resolution models, enabling comparison of sequence-based structures to experimental SAXS profiles. Calculating SAXS profiles from atomic models is nontrivial. High-resolution SAXS datasets produced by current beamlines require higher accuracy in theoretical SAXS profile calculators for correct interpretation of experimental data. There are several useful methods available, differing in the spherical averaging algorithms, the approximation of the excluded volume, and the treatment of the hydration layer. Multiple approximations, such as significant adjustment of excluded volume and hydration layer density, may diminish the ability of the method to identify a near-native model in a set of multiple candidates based on a SAXS profile. The FoXS algorithm explicitly computes all interatomic distances for accurate spherical averaging, estimates excluded volume scattering without significant variance in the protein volume, and models the first hydration layer based on the atomic solvent-accessible areas. The results on a benchmark of 14 cases show that this approach is more accurate and faster than the best-in-class and most widely used method, CRYSOL (Figs. 1 and 2 and Table 2), as well as three additional often-used programs (AquaSAXS, the Zernike polynomials-based method, and Fast-SAXS-pro (Table 1, see Fig. S1, Fig. S2, and Fig. S3).

Adjustment in the hydration layer density is required for accurate SAXS fitting

Contrast variation experiments confirm the importance of adjusting the hydration layer in the SAXS calculation (Fig. 4). Although explicit water layer experiments could provide a more-accurate representation at higher computational cost, the theoretical hydration layer calculated in FoXS successfully mimics the data derived from contrast variation experiments, as well as an explicit water layer defined by waters from high-resolution structures (55). The modeling of hydration layer is important for anisotropic structures that are the main candidates for solution structure determination. The results from the benchmark presented here show that larger adjustment of the hydration layer density correlates with the anisometry of the system, which may contribute to the intrinsic flexibility (Table 2, Fig. 2 *A*). Both FoXS and CRYSOL require high density hydration layers for accurate fitting of the P4P6 RNA Ribozyme domain (56), 28-bp DNA, the ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain (57), DNA double-strand break repair protein MRE11 (58), and complement C3b-Efb-C (59). The datasets in our experimental benchmark were considered to reflect systems adopting single conformations under physiological salt concentrations. However, the anisometric or elongated structures possess intrinsic elasticity. This conformational fluctuation around well-defined conformations alters experimental SAXS profiles and needs to be considered during the SAXS fitting (1,18). To simplify testing and comparison of different programs, we fit our experimental benchmark with a single conformer. However, high c_2 values observed for P4P6, 28bp DNA, ATG7, MRE11+ATP, and C3b-Efb-C reflect flexibility of the elongated macromolecules rather then high density hydration values.

FoXS is an open-source method with multiple user-friendly interfaces

SAXS profile calculation from the structural model is a necessary component of the SAXS toolbox. As an open-source method, it can be simply integrated with other modeling methods using C++ or PYTHON interfaces. Indeed, the FoXS method is a module of the integrative modeling platform (IMP) software suite (60,61), allowing SAXS profiles to be used simultaneously with other experimental datasets supported by IMP, such as electron microscopy and various proteomics data. FoXS can also be accessed through UCSF Chimera visualization software (53), where SAXS calculations can be performed in combination with other tasks, such as comparative modeling with MODELER (62). We have also developed a freely available web service for profile calculation, sparing the user of any software installation (63).

FoXS-MES as a tool for ensemble analysis

The FoXS web server allows profile fitting and simultaneous analysis of a collection of high-resolution models against an experimental SAXS profile with MES (11). This combination of methods is important for identification of partially populated solution states that are crucial to biological function. SAXS measurements provide a collective snapshot of the thermodynamic ensemble. It is likely that accurate modeling of solution structures requires an ensemble of different conformational states. The two test cases presented here, the transient XLF-XRCC4 complex and flexible ligase III, clearly demonstrate that the interpretation of the scattering profile with the population-weighted thermodynamic ensemble is crucial for accurate description of solution states.

CONCLUSIONS

Because the quality and throughput of SAXS experimental data has recently improved dramatically, there is a corresponding sharply increased demand for more accurate and faster computation of SAXS profiles (64). Here, superior

fits between experimental high-resolution structures and SAXS data are obtained by using explicit all-atom distances and a hydration layer model to account for the effect of solvent. The explicit representation of the hydrated macromolecule is particularly useful for multidomain-flexible assemblies, which frequently adopt highly anisometric shapes. The FoXS algorithm explicitly computes all interatomic distances that include the first solvation layer based on the atomic solvent-accessible areas. In addition to accuracy, rapid calculation speed is critical for SAXS-based structural modeling tools, which often perform multiple SAXS calculations for large pools of conformers or assemblies (5). FoXS capabilities have proven valuable for defining the archaeal motor mechanism (65), for validating model-data resolution limits (42), and for comprehensive SAXS structural comparison maps revealing structural similarities (66), suggesting FoXS will prove important for emerging biology plus advanced SAXS metrics and analysis techniques. In fact, we show here that FoXS is faster than other best-in-class programs. Because FoXS is available through a web server, it enables uploading and simultaneous analysis of a collection of atomic coordinate input files against experimental data. In combination with the MES (11), the user is provided with tools to identify the flexibility as well as the conformational and compositional heterogeneity of the studied system.

SUPPORTING MATERIAL

Two tables and four figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)00805-9.

D.S.-D. has been funded by the Weizmann Institute's Advancing Women in Science Postdoctoral Fellowship. We also acknowledge support from National Institutes of Health grant Nos. R01 GM083960, U54 RR022220, and PN2 EY016525, and Rinat (Pfizer) Inc. (to A.S.) as well as the Lawrence Berkeley National Lab IDAT program and National Institutes of Health grant No. MINOS R01GM105404 (to J.A.T.).

REFERENCES

1. Putnam, C. D., M. Hammel, …, J. A. Tainer. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40:191–285.
2. Rambo, R. P., and J. A. Tainer. 2010. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle x-ray scattering. *Curr. Opin. Struct. Biol.* 20:128–137.
3. Hammel, M. 2012. Validation of macromolecular flexibility in solution by small-angle x-ray scattering (SAXS). *Eur. Biophys. J.* 41:789–799.
4. Jacques, D. A., and J. Trewhella. 2010. Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci.* 19:642–657.
5. Hura, G. L., A. L. Menon, …, J. A. Tainer. 2009. Robust, high-throughput solution structural analyses by small angle x-ray scattering (SAXS). *Nat. Methods.* 6:606–612.
6. Nishimura, N., K. Hitomi, …, E. D. Getzoff. 2009. Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science.* 326:1373–1379.
7. Lamb, J. S., B. D. Zoltowski, …, L. Pollack. 2009. Illuminating solution responses of a LOV domain protein with photocoupled small-angle x-ray scattering. *J. Mol. Biol.* 393:909–919.
8. Hammel, M., M. Rey, …, J. A. Tainer. 2011. XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *J. Biol. Chem.* 286:32638–32650.
9. Fenton, A. W., R. Williams, and J. Trewhella. 2010. Changes in small-angle x-ray scattering parameters observed upon binding of ligand to rabbit muscle pyruvate kinase are not correlated with allosteric transitions. *Biochemistry.* 49:7202–7209.
10. Bernadó, P., E. Mylonas, …, D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle x-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.
11. Pelikan, M., G. L. Hura, and M. Hammel. 2009. Structure and flexibility within proteins as identified through small angle x-ray scattering. *Gen. Physiol. Biophys.* 28:174–189.
12. Petoukhov, M. V., and D. I. Svergun. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 89:1237–1250.
13. Förster, F., B. Webb, …, A. Sali. 2008. Integration of small-angle x-ray scattering data into structural modeling of proteins and their assemblies. *J. Mol. Biol.* 382:1089–1106.
14. Perkins, S. J., and A. Bonner. 2008. Structure determinations of human and chimeric antibodies by solution scattering and constrained molecular modeling. *Biochem. Soc. Trans.* 36:37–42.
15. Zheng, W., and S. Doniach. 2002. Protein structure prediction constrained by solution x-ray scattering data and structural homology identification. *J. Mol. Biol.* 316:173–187.
16. Zheng, W., and S. Doniach. 2005. Fold recognition aided by constraints from small angle x-ray scattering data. *Protein Eng. Des. Sel.* 18:209–219.
17. dos Reis, M. A., R. Aparicio, and Y. Zhang. 2011. Improving protein template recognition by using small-angle x-ray scattering profiles. *Biophys. J.* 101:2770–2781.
18. Grishaev, A., L. Guo, …, A. Bax. 2010. Improved fitting of solution x-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J. Am. Chem. Soc.* 132:15484–15486.
19. Schneidman-Duhovny, D., S. J. Kim, and A. Sali. 2012. Integrative structural modeling with small angle x-ray scattering profiles. *BMC Struct. Biol.* 12:17.
20. Zuo, X., G. Cui, …, D. M. Tiede. 2006. X-ray diffraction "fingerprinting" of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* 103:3534–3539.
21. Debye, P. 1915. Scattering of x-rays [Zerstreuung von Röntgenstrahlen]. *Annalen der Physik.* 351:809–823.
22. Svergun, D., C. Barberato, and M. H. J. Koch. 1995. CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *j. Appl. Cryst.* 28:768–773.
23. Liu, H., R. J. Morris, …, P. H. Zwart. 2012. Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. *Acta Crystallogr. A.* 68:278–285.
24. Gumerov, N. A., K. Berlin, …, R. Duraiswami. 2012. A hierarchical algorithm for fast Debye summation with applications to small angle scattering. *J. Comput. Chem.* 33:1981–1996.
25. Tjioe, E., and W. T. Heller. 2007. ORNL_SAS: software for calculation of small-angle scattering intensities of proteins and protein complexes. *J. Appl. Cryst.* 40:782–785.
26. Yang, S., S. Park, …, B. Roux. 2009. A rapid coarse residue-based computational method for x-ray solution scattering characterization

of protein folds and multiple conformational states of large protein complexes. *Biophys. J.* 96:4449–4463.

27. Stovgaard, K., C. Andreetta, …, T. Hamelryck. 2010. Calculation of accurate small angle x-ray scattering curves from coarse-grained protein models. *BMC Bioinformatics.* 11:429.
28. Ravikumar, K. M., W. Huang, and S. Yang. 2013. Fast-SAXS-pro: a unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes. *J. Chem. Phys.* 138:024112.
29. Fraser, R. D. B., T. P. MacRae, and E. Suzuki. 1978. An improved method for calculating the contribution of solvent to the x-ray diffraction pattern of biological molecules. *J. Appl. Cryst.* 11:693–694.
30. Lattman, E. E. 1989. Rapid calculation of the solution scattering profile from a macromolecule of known structure. *Proteins.* 5:149–155.
31. Fraser, J. S., H. van den Bedem, …, T. Alber. 2011. Accessing protein conformational ensembles using room-temperature x-ray crystallography. *Proc. Natl. Acad. Sci. USA.* 108:16247–16252.
32. Perkins, S. J. 2001. X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modeling fits. *Biophys. Chem.* 93:129–139.
33. Kuhn, L. A., M. A. Siani, …, J. A. Tainer. 1992. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J. Mol. Biol.* 228:13–22.
34. Svergun, D. I., S. Richard, …, G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95:2267–2272.
35. Hubbard, S. R., K. O. Hodgson, and S. Doniach. 1988. Small-angle x-ray scattering investigation of the solution structure of troponin C. *J. Biol. Chem.* 263:4151–4158.
36. Grossmann, J. G., Z. H. Abraham, …, S. S. Hasnain. 1993. X-ray scattering using synchrotron radiation shows nitrite reductase from *Achromobacter xylosoxidans* to be a trimer in solution. *Biochemistry.* 32:7360–7366.
37. Fujisawa, T., T. Uruga, …, T. Ueki. 1994. The hydration of Ras p21 in solution during GTP hydrolysis based on solution x-ray scattering profile. *J. Biochem.* 115:875–880.
38. Bardhan, J., S. Park, and L. Makowski. 2009. SoftWAXS: a computational tool for modeling wide-angle x-ray solution scattering from biomolecules. *J. Appl. Cryst.* 42:932–943.
39. Poitevin, F., H. Orland, …, M. Delarue. 2011. AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models. *Nucleic Acids Res.* 39(Web Server issue):W184–W189.
40. Virtanen, J. J., L. Makowski, …, K. F. Freed. 2011. Modeling the hydration layer around proteins: applications to small- and wide-angle x-ray scattering. *Biophys. J.* 101:2061–2069.
41. Connolly, M. L. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science.* 221:709–713.
42. Rambo, R. P., and J. A. Tainer. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 496:477–481.
43. Ciccariello, S., J. Goodisman, and H. Brumberger. 1988. On the Porod law. *J. Appl. Cryst.* 21:117–128.
44. Hammel, M., H. P. Fierobe, …, V. Receveur-Bréchot. 2005. Structural basis of cellulosome efficiency explored by small angle x-ray scattering. *J. Biol. Chem.* 280:38562–38568.
45. Shin, D. S., M. Didonato, …, J. A. Tainer. 2009. Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. *J. Mol. Biol.* 385:1534–1555.
46. Hammel, M., Y. Yu, …, J. A. Tainer. 2010. XLF regulates filament architecture of the XRCC4·ligase IV complex. *Structure.* 18:1431–1442.
47. Cotner-Gohara, E., I. K. Kim, …, T. Ellenberger. 2010. Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry.* 49:6165–6176.
48. Hammel, M., H. P. Fierobe, …, V. Receveur-Bréchot. 2004. Structural insights into the mechanism of formation of cellulosomes probed by small angle x-ray scattering. *J. Biol. Chem.* 279:55985–55994.
49. Schneidman-Duhovny, D., Y. Inbar, …, H. J. Wolfson. 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33(Web Server issue):W363–W367.
50. Russo, D., G. Hura, and T. Head-Gordon. 2004. Hydration dynamics near a model protein surface. *Biophys. J.* 86:1852–1862.
51. Russo, D. 2008. The impact of kosmotropes and chaotropes on bulk and hydration shell water dynamics in a model peptide solution. *Chem. Phys.* 345:200–211.
52. Burke, J. R., G. L. Hura, and S. M. Rubin. 2012. Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control. *Genes Dev.* 26:1156–1166.
53. Yang, Z., K. Lasker, …, T. E. Ferrin. 2012. UCSF Chimera, MODELER, and IMP: an integrated modeling system. *J. Struct. Biol.* 179:269–278.
54. Perry, J. J., E. Cotner-Gohara, …, J. A. Tainer. 2010. Structural dynamics in DNA damage signaling and repair. *Curr. Opin. Struct. Biol.* 20:283–294.
55. Kuhn, L. A., C. A. Swanson, …, E. D. Getzoff. 1995. Atomic and residue hydrophilicity in the context of folded protein structures. *Proteins.* 23:536–547.
56. Rambo, R. P., and J. A. Tainer. 2010. Improving small-angle x-ray scattering data for structural analyses of the RNA world. *RNA.* 16:638–646.
57. Taherbhoy, A. M., S. W. Tait, …, B. A. Schulman. 2011. Atg8 transfer from Atg7 to Atg3: a distinctive E1-E2 architecture and mechanism in the autophagy pathway. *Mol. Cell.* 44:451–461.
58. Williams, G. J., R. S. Williams, …, J. A. Tainer. 2011. ABC ATPase signature helices in Rad50 link nucleotide state to Mre11 interface for DNA repair. *Nat. Struct. Mol. Biol.* 18:423–431.
59. Chen, H., D. Ricklin, …, J. D. Lambris. 2010. Allosteric inhibition of complement function by a staphylococcal immune evasion protein. *Proc. Natl. Acad. Sci. USA.* 107:17621–17626.
60. Webb, B., K. Lasker, …, A. Sali. 2011. Modeling of proteins and their assemblies with the integrative modeling platform. *Methods Mol. Biol.* 781:377–397.
61. Russel, D., K. Lasker, …, A. Sali. 2012. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* 10:e1001244.
62. Sali, A., and T. L. Blundell. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.
63. Schneidman-Duhovny, D., M. Hammel, and A. Sali. 2010. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* 38(Web Server issue, Suppl):W540–W544.
64. Rambo, R. P., and J. A. Tainer. 2013. Super-resolution in solution x-ray scattering and its applications to structural systems biology. *Annu. Rev. Biophys.* 42:415–441.
65. Reindl, S., A. Ghosh, …, J. A. Tainer. 2013. Insights into FlaI functions in archaeal motor assembly and motility from structures, conformations, and genetics. *Mol. Cell.* 49:1069–1082.
66. Hura, G. L., H. Budworth, …, J. A. Tainer. 2013. Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods.* 10:453–454.
67. Biersmith, B. H., M. Hammel, …, S. Bouyain. 2011. The immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR adopt an unusual horseshoe-like conformation. *J. Mol. Biol.* 408:616–627.