Implicit vs. Explicit Solvent Models for Calculating X-ray Solution Scattering Curves#

Prabhakar Ganesan,[†,‡] Jong Goo Kim,[†,‡] Jae Hyuk Lee,[†] Jeongho Kim,[§] and Hyotcherl Ihee[†,‡,*]

[†]*Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea. *E-mail: hyotcherl.ihee@kaist.ac.kr*
[‡]*Center for Nanomaterials and Chemical Reactions, Institute for Basic Science (IBS), Daejeon 305-701, Republic of Korea*
[§]*Department of Chemistry, Inha University, Incheon 402-751, Republic of Korea*
Received November 8, 2014, Accepted December 10, 2014, Published online February 20, 2015

Theoretical calculation of X-ray solution scattering curves of proteins in the solution phase is strongly influenced by solvent contributions in the form of solvent-excluded volume and hydration layer that are generally represented either implicitly or explicitly. To investigate the effect of the implicit and explicit solvent models on the calculated scattering curves, we developed a new program, **X**-ray **So**lution **S**cattering (XSoS) based on implicit (XSoS-implicit) and explicit (XSoS-explicit) solvent models. Both XSoS-implicit and XSoS-explicit can calculate X-ray solution scattering curves with high accuracy. Overall, the implicit solvent model has practical advantages over the explicit solvent model for the analysis of experimental X-ray solution scattering data.

Keywords: Explicit solvent, Implicit solvent, X-ray solution scattering, Solvent models

Introduction

X-ray solution scattering is a powerful tool for characterizing structure and dynamics of proteins.[1–7] Unlike other biophysical methods, this technique is directly sensitive to molecular structure in solution and therefore has been used to characterize a variety of structural changes in proteins such as domain orientations,[8–15] flap movements,[16,17] and folding/unfolding.[18–21] The X-ray solution scattering signal can be classified into two regions depending on the scattering angle: small-angle X-ray scattering (SAXS) and wide-angle X-ray scattering (WAXS). The two types of signals provide different but useful structural information on biomolecular complexes.[22] In general, SAXS helps to determine the size and shape of the macromolecules,[23,24] whereas WAXS gives relatively higher-resolution information on the secondary and tertiary structure of proteins.[25–27]

In principle, a solution scattering curve of a protein represents the average scattering response of an ensemble of proteins that are free to move in solution.[22,27,28] The scattering from samples of protein solution can be decomposed into (1) scattering from the solute, (2) scattering from solvent-excluded volume that is artificially introduced when subtracting scattering of pure solvent from the scattering of protein solution, and (3) scattering from a dense layer of water molecules (called hydration layer) surrounding the solute.[29–31] Therefore, calculation of X-ray solution scattering curves is significantly affected by the latter two terms, solvent-excluded volume and hydration layer.[32] In general, the solvent contributions to the solution scattering curve can be modeled using either a continuum of electron density (implicit solvent model) or explicit atomic coordinates of solvent molecules (explicit solvent model).[33–36] Between the two models, the implicit solvent model is simpler and has been used more commonly, even for the calculation of time-resolved WAXS (TR-WAXS) data.[37–39] However, it has been suggested that more sophisticated treatment of the solvent contribution by the explicit solvent model may be necessary for calculating high-resolution WAXS patterns.[40] However, the X-ray solution scattering curves calculated with the explicit solvent models are known to have inherent limitations related to the solvent atomic coordinate sets used.

In this article, to study these aspects in a systematic way and compare the advantages and disadvantages of the implicit and explicit models, we introduce our own program developed for calculating solution scattering curves using either implicit or explicit solvent model. We test the performance of the program for several proteins whose experimental scattering curves are available. In addition, using the explicit solvent model, we investigate the limitations related to the solvent atomic coordinate sets for calculating scattering curves and determine the minimum number of snapshots required to obtain a unique scattering curve that converges into the experimental X-ray solution scattering curve for a given protein structure.

Methods

All the source codes used in this work were written in C++. The source codes were compiled and executed in the Linux

This paper is dedicated to Professor Kwan Kim on the occasion of his honorable retirement.

platform on a 2.2 GHz Intel Pentium 4 processor. We named our program as XSoS and, specifically, the programs employing implicit and explicit solvent models are termed as XSoS-implicit and XSoS-explicit, respectively.

XSoS-Implicit Formalism. The theory for calculating X-ray solution scattering curves with the implicit solvent model is well established and various executable programs such as CRYSOL and FoXS have been developed.[34,41] The key differences between the XSoS-implicit and other programs (CRYSOL and FoXS) lie in the methods of orientation averaging and treating the hydration layer. For orientation averaging, XSoS-implicit uses spherical quadrature based on spiral code,[42] whereas CRYSOL and FoXS use the spherical harmonics approach and the Debye formula approach, respectively. The computational assumptions used in CRYSOL appear inadequate for calculation of wide-angle scattering, whereas the Debye approach is computationally expensive and the time–cost increases quadratically with the number of atoms in the protein. For the treatment of the hydration layer, XSoS-implicit uses the cube method, whereas CRYSOL and FoXS use the envelope function and the solvent-accessible surface, respectively. Modeling using angular envelope functions has limitations in the description of complicated shapes like those proteins having internal cavities and solvent-accessibility method requires additional procedures to calculate solvent accessibility.

Here, we review the key equations with minor modifications relevant to the XSoS-implicit program. The scattering intensity $I_p(q)$ of a protein molecule in the solution is given by the following equation,

$$I_P(q) = \left\langle |P(\mathbf{q}) - \rho_b S(\mathbf{q}) + \delta\rho\varphi(\mathbf{q})|^2 \right\rangle \quad (1)$$

Here $P(\mathbf{q})$ is the scattering amplitude of the protein molecule in vacuum, $S(\mathbf{q})$ is the scattering amplitude of the solvent-excluded volume, ρ_b is the electron density of bulk water, $\varphi(\mathbf{q})$ is the scattering amplitude of the hydration layer, $\delta\rho$ denotes the electron density contrast of the hydration layer, and $\langle\rangle$ represents the orientation average. The observed scattering amplitude of the protein molecule is related to the Fourier transform of the atomic coordinates (electron density) weighted by the X-ray atomic form factor. The scattering amplitude $P(\mathbf{q})$ of the protein in vacuum is given by the equation,

$$P(\mathbf{q}) = \sum_l e^{-i\mathbf{q}\cdot\mathbf{r}_l} f_l(q) \quad (2)$$

Here $\mathbf{q}$ is the scattering vector (or momentum transfer), of which the magnitude is given by $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength, $\mathbf{r}_l$ is the coordinate of the l-th atom and $f_l(q)$ denotes the X-ray atomic form factor of the l-th atom. The components of the vector $\mathbf{q}$ are q, θ, and φ, where θ and φ are the angles described by equally spaced points along the spiral of the unit sphere in the spherical quadrature.[35,43] The θ and φ are given by the equation,

$$\theta_j = \arccos\frac{2j-1-J}{J} \quad (3)$$

$$\varphi_j = \sqrt{\pi J}\arcsin\frac{2j-1-J}{J} \quad (4)$$

Here $j = 1, 2, \ldots, J$, where J is the total number of points distributed along the spiral of the unit sphere. The first step in the calculation of the scattering amplitude involves evaluation of the exponential dot product between the $\mathbf{q}$ vector and the atomic coordinates. Then, the exponential of the calculated dot product is multiplied by the corresponding atomic form factor and integrated to obtain the scattering amplitude. The X-ray atomic form factors are evaluated using the coefficients for analytical approximation taken from the International Tables of Crystallography, 2004. The solvent-excluded volume is estimated by placing dummy solvent atoms at all atomic positions within a macromolecule. The scattering amplitude $S(\mathbf{q})$ of the solvent-excluded volume is evaluated using the following equation,

$$S(\mathbf{q}) = \sum_l e^{-i\mathbf{q}\cdot\mathbf{r}_l} f_s(q) \quad (5)$$

The form factor of the atoms in the excluded volume is represented by Gaussian spheres, whose radii have been tabulated previously.[34] The form factor of the displaced solvent volume $f_s(q)$ is given by the equation,

$$f_S(q) = G(q)V_j \exp\left(\frac{q^2 V_j^{2/3}}{4\pi}\right) \quad (6)$$

where V_j denotes the volume of displaced solvent. The expansion factor $G(q)$ is given by,

$$G(q) = \left(\frac{V_0}{V_m}\right)\exp\left(\frac{q^2\left(V_0^{2/3} - V_m^{2/3}\right)}{4\pi}\right) \quad (7)$$

Here, V_0 represents the average atomic volume in the macromolecule and V_m represents the adjustable parameter that allows one to vary the average volume of displaced solvent per atomic group. The scattering amplitude of the excluded volume is calculated in the same manner as that of the scattering amplitude of the protein molecule except that dummy atomic form factor is used instead of the atomic form factor of solute species. In XSoS-implicit, a continuum of hydration layer in the vicinity of protein surface is evaluated using the cube method.[33,44,45] The cube method is regarded superior in calculating the scattering curves especially at higher angles.[34] Generation of the cubes for evaluating the hydration layer is facilitated by placing the protein molecule of interest inside a large parallelepiped encompassing the entire protein molecule. The parallelepiped is subsequently divided into equal-sized cubes of 2 Å edge length. The cubes that do not overlap with the protein structure and those within a distance threshold of

3 Å from the surface atoms of the protein are included for the evaluation of the hydration layer. The scattering amplitude of the hydration layer is calculated using the following equation,

$$\varphi(\mathbf{q}) = \delta\rho \sum_j \varphi(\mathbf{q}, a, b_j) \exp(i\mathbf{q}\cdot\mathbf{r}_j) \quad (8)$$

where,

$$\varphi(\mathbf{q}, a, b_j) = 8 \sin(q_x a)\sin(q_y a)\sin(q_z b_j) / q_x q_y q_z \quad (9)$$

Here $\varphi(\mathbf{q})$ is the scattering amplitude of the hydration layer, $\delta\rho$ denotes the electron density contrast of the hydration layer, $\varphi(\mathbf{q},a,b_j)$ is the scattering amplitude of j-th parallelepiped, $\mathbf{r}_j$ is the center of the j-th parallelepiped, b_j is the edge length of j-th parallelepiped, a is half of the edge length of a cube, and q_x, q_y and q_z are the components of $\mathbf{q}$ vector in three dimensions.

XSoS-Explicit Formalism. XSoS-explicit is our in-house program based on the Park formalism.[35] Here two systems are considered, system A containing the N protein molecules along with a water shell and system B containing the N water droplets corresponding to the solvent-excluded volume. The atomic coordinates of the two systems, namely protein with water shell and the solvent-excluded volume, are generated from short MD simulations snapshots. The thickness of the water shell around the protein which is an input parameter is chosen large enough that all water molecules outside and near the shell boundaries are bulk like. The scattering amplitude of the protein with the water shell is calculated using Eq. (2). Here $\mathbf{r}_l$ is the coordinate of the l-th atom contained in the protein or the water shell of thickness 7 Å. Similarly, for the scattering amplitude of the atoms contained in the water droplets, we use Eq. (2). The orientation averaging is performed using spherical quadrature based on spiral code by considering equally spaced points along the spiral of the unit sphere by utilizing Eqs. (3) and (4). The intensity is given by the equation

$$\frac{\Delta I(q)}{N} = \frac{1}{4\pi}\int d\Omega_{\mathbf{q}} D_{11}(\mathbf{q}) \quad (10)$$

where $d\Omega$ is the solid angle averaged over the vector $\mathbf{q}$ and $D_{11}(\mathbf{q})$ given by the equation,

$$D_{11}(\mathbf{q}) = |a(\mathbf{q}) - b(\mathbf{q})|^2 + \frac{1}{M}\sum_{m=1}^{M} |P^{(m)}(\mathbf{q}) - a(\mathbf{q})|^2 - \frac{M'+1}{M'(M-1)}\sum_{m=1}^{M'} |S^{(m)}(\mathbf{q}) - b(\mathbf{q})|^2 \quad (11)$$

where,

$$a(\mathbf{q}) = \frac{1}{M}\sum_{m=1}^{M} P^{(m)}(\mathbf{q}), b(\mathbf{q}) = \frac{1}{M'}\sum_{m=1}^{M} S^{(m)}(\mathbf{q}) \quad (12)$$

where M and M' are the total number of snapshots collected from MD simulation of systems A and B, respectively. In this method, the thickness of the water shell is the only adjustable input parameter and is determined based on the convergence of the scattering curves at larger q values by varying the thickness. Note that the XSoS-explicit has no parameter available for fitting experimental X-ray solution scattering data. The main difference between the XSoS-implicit and XSoS-explicit arises from the solvent atomic coordinates of the latter and the latter use the extra equations for calculating the scattering curves, namely Eqs. (10), (11), and (12).

Experimental X-ray Solution Scattering Curves. In this study, experimental X-ray solution scattering data were adopted from previous publications or retrieved from the BIOISIS, which is an open-access database of SAXS data for biological macromolecules (http://www.bioisis.net/).[4] Specifically, the experimental scattering curves for the following proteins were used: (1) lysozyme,[34] (2) malate synthase,[46] (3) superoxide dismutase from *Alvinella Pompejana*,[47] (4) xylanase,[48] (5) complement fragment C3b complex with extracellular fibrinogen binding protein (Efb) from *Staphylococcus aureus*,[49] (6) immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3,[50] (7) glucose isomerase,[51] (8) glycosyl hydrolase + C-terminus,[52] (9) ketoreductase–enoylreductase didomain,[53] and (10) dimeric PYR1–abscisic acid complex.[54] The atomic coordinates used for fitting the experimental scattering data by the corresponding theoretical scattering curve were obtained from BIOISIS. The scattering data of ubiquitin (for both protein and buffer) obtained from the AXES (**A**nalysis of **X**-ray Scattering in **E**xplicit **S**olvent)[36] web server page was also used as the experimental data.

Calculation of Scattering Curves. The capability of the XSoS-implicit program for accurately calculating the solution scattering curve was assessed by comparing with the experimental data described in the previous section. In the programs using the implicit solvent model, the theoretical scattering curve is initially calculated for a certain configuration of solution sample and is fine-tuned by adjusting the solvent-excluded volume and the electron density contrast of hydration layer[34] until the theoretical scattering curve matches the experimental curve. To account for the contribution of the solvent-excluded volume, we adjusted the average atomic radii (that are associated with V_0) while we varied $\delta\rho$ to consider the electron density contrast of hydration layer. We calculated the solution scattering curves using XSoS-implicit and compared with the scattering curves obtained with CRYSOL and FoXS.

X-ray solution scattering curves with the explicit solvent model were calculated using the XSoS-explicit program. The solvent-excluded volume and the hydration layer for various protein structures were generated by MD simulations using NAMD[55] with CHARMM22 force field.[56] The first simulation was performed for a protein embedded in a water box and the second for a pure water box without any protein to generate the solvent-excluded volume. The simulation protocol

included an equilibration step for 20 ps and a production run for 100 ps with a time step of 2 fs. The atomic coordinates of the protein structures were fixed during the MD simulations. The water molecules that lie within a threshold of 7 Å from the protein surface were considered as a hydration layer. A total of 200 snapshots was captured from the two MD runs (100 snapshots for each run), one corresponding to the hydrated protein and the other to the solvent-excluded volume. Then, theoretical scattering curves were calculated for lysozyme (PDB ID: 6LYZ) and myoglobin (PDB ID: 1WLA) based on the snapshots captured from the MD simulations using the XSoS-explicit program.

In addition, using XSoS-explicit, we examined how the solvent degrees of freedom play a role in obtaining a converged scattering curve using the explicit solvent model. Briefly, to obtain the coordinates of solvent molecules in the hydration layer, we carried out MD simulations for 2G0R structure of MbCO in water over a time period of 2 ns, and captured 2000 snapshots. In order to obtain the solvent-excluded volume, we performed an MD simulation for pure water over a time period of 2 ns and captured 2000 snapshots as well. Furthermore, we used the 2000 snapshots of pure water to examine the effect of varying solvent degrees of freedom and solvent atomic coordinate sets on the scattering curves using AXES and WAXSiS (WAXS in Solvent).[57]

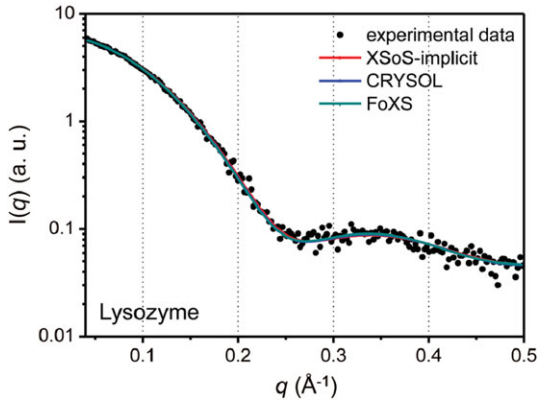

Figure 1. Accuracy of XSoS-implicit. X-ray solution scattering curve of lysozyme calculated by XSoS-implicit (red line), CRYSOL (blue line), and FoXS (green line) well match the experimental data (black dots).

Results and Discussion

Performance of XSoS for Calculation of Static Scattering Curves. The accuracy and the calculation speed of our XSoS-implicit program were inspected in comparison with CRYSOL. The static scattering curves of lysozyme calculated by XSoS-implicit, CRYSOL, and FoXS are compared in Figure 1. Also, the scattering curves for nine other proteins are compared in Figure S1 in the Supporting Information (SI). The chi-squared values that represent the agreement between experimental and theoretical scattering curves indicate that XSoS-implicit is as accurate as, or even better than, the other methods (Table 1). Also, when comparing the calculation times of XSoS-implicit, CRYSOL, and FoXS, XSoS-implicit is slower than FoXS but faster than CRYSOL for small proteins such as lysozyme and xylanase. For larger proteins, however, XSoS-implicit is slower than FoXS and CRYSOL, especially due to computations related to the cube method in the hydration layer.

For the XSoS-explicit program, we verified its accuracy and speed by comparing the calculated scattering curves of myoglobin and lysozyme with the corresponding scattering curves that were calculated by EXCESS program provided by its authors.[35] For both myoglobin and lysozyme, the scattering curves calculated by the two programs show good agreement with each other as shown in Figure 2 and Figure S2, respectively. Also, we found that the C++ source code for the XSoS-explicit program is faster by a factor of 15 than the EXCESS program written in MATLAB. For the calculation of the scattering curves of myoglobin and lysozyme, EXCESS requires the execution time of 30 h for each[35] whereas the XSoS-explicit program needs only 2 h. Thus, both

Table 1. Comparison of χ^2 values between theoretical and experimental scattering curves when using CRYSOL, XSoS-implicit, and FoXS.

Protein (ID)	q (Å^{-1})	CRYSOL	XSoS-implicit	FoXS
Lysozyme (6LYZ[a])	0.04–0.50	0.45	0.38	0.44
Malate synthase (2JQX[a])	0.03–0.50	0.88	0.82	0.93
Superoxide dismutase (APSODP[b])	0.01–0.30	3.85	4.21	4.49
Xylanase (1XYNTP[b])	0.02–0.30	0.99	0.98	1.04
C3b-fibrinogen binding protein (C3bEfP[b])	0.01–0.30	2.44	2.17	2.19
Domain of tyrosine phosphatase (LAR12P[b])	0.02–0.30	1.44	2.02	1.44
Glucose isomerase (GISRUP[b])	0.01–0.50	7.92	5.37	5.36
Glycosyl hydrolase (AT5GHP[b])	0.02–0.35	1.89	2.16	1.92
Ketoreductase-enoylreductase (ZGDWKP[b])	0.01–0.14	1.99	1.87	2.58
Absicisic acid receptor (1PYR1P[b])	0.02–0.32	2.38	2.35	1.96

[a] PDB ID.
[b] BIOISIS ID.

XSoS-implicit and XSoS-explicit exhibit high accuracy and decent calculation speed.

Need for Fitting. In general, the X-ray crystallographic protein structures used to extract the structural information from the solution scattering data are rigid due to the constraints forced by crystal contacts.[58] In the absence of such packing defects, the protein structures have significant structural fluctuations in their natural solution environment.[59] Therefore, fitting of theoretical scattering curve into experimental scattering curve is generally required to account for the uncertainty in the theoretical scattering curve calculations. However, for the implicit solvent model, the fitting is limited because the solvent-excluded volume of the individual atoms of a protein is described by uniform electron density. In the explicit solvent model, the discrepancy between the real atomic arrangement and the artificial arrangement generated by MD simulation necessitates fitting. In most cases, the methods based on the implicit solvent model use the fitting parameters of (1) solvent-excluded volume and (2) electron density contrast of hydration layer.[34,60,61] In contrast, the methods based on the explicit solvent model use the fitting parameters of (1) weighting factor for the hydration layer[62] that is empirically calibrated using the experimental scattering curves, (2) sample/buffer rescaling factor that accounts for the uncertainty in the measurement of transmitted and incident intensities as well as the uncertainty in the concentration of the solute volume, and (3) dark current of the detector[36] that accounts for the variability of the dark current of the detector as well as the effect of X-ray fluorescence. In order to demonstrate the need for the fitting of the theoretical scattering curves calculated by the programs to match the experimental scattering curves, we compared the experimental and theoretical scattering curves of ubiquitin in Figure 3. Here we note that fitting is not allowed inherently in the XSoS-explicit program, and therefore we instead used AXES web server, which allows fitting against experimental scattering curves, for calculating scattering curves based on the explicit solvent model.[36] The scattering curves calculated by XSoS-implicit and AXES web server were compared by fitting the common experimental scattering profile of ubiquitin.

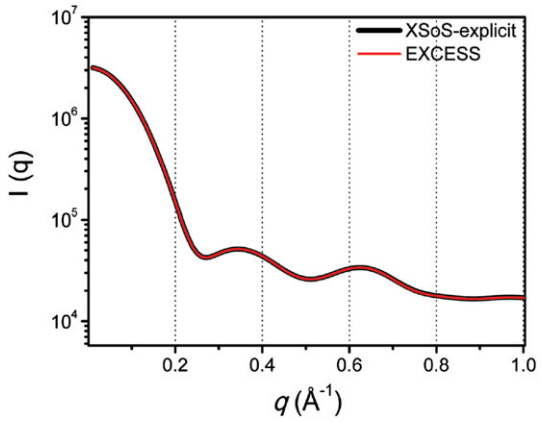

Figure 2. Accuracy of XSoS-explicit. X-ray solution scattering curve of myoglobin calculated with XSoS-implicit (black line) and EXCESS (red line) agree well with each other.

Fitting by the XSoS-implicit program was performed by adjusting the solvent-excluded volume and the electron density contrast of the hydration layer. When calculating the scattering curves with AXES web server, we utilized a total of 100 water boxes to account for the solvent-excluded volume and the hydration layer. For the fitting of the ubiquitin's experimental scattering data with AXES web server, we varied two parameters, the sample-to-buffer rescaling factor (α) and the dark current of the detector (c). As can be seen in Figure 3(a), without fitting, the calculated theoretical scattering curves show a certain level of discrepancy from the experimental curve in the low-angle region (<0.25 $Å^{-1}$) irrespective of the solvent model used. Also, the scattering curves calculated by the implicit (XSoS-implicit) and explicit (AXES webserver) solvent models deviate from each other in the low-angle region. However, once the fitting is applied, (1) the

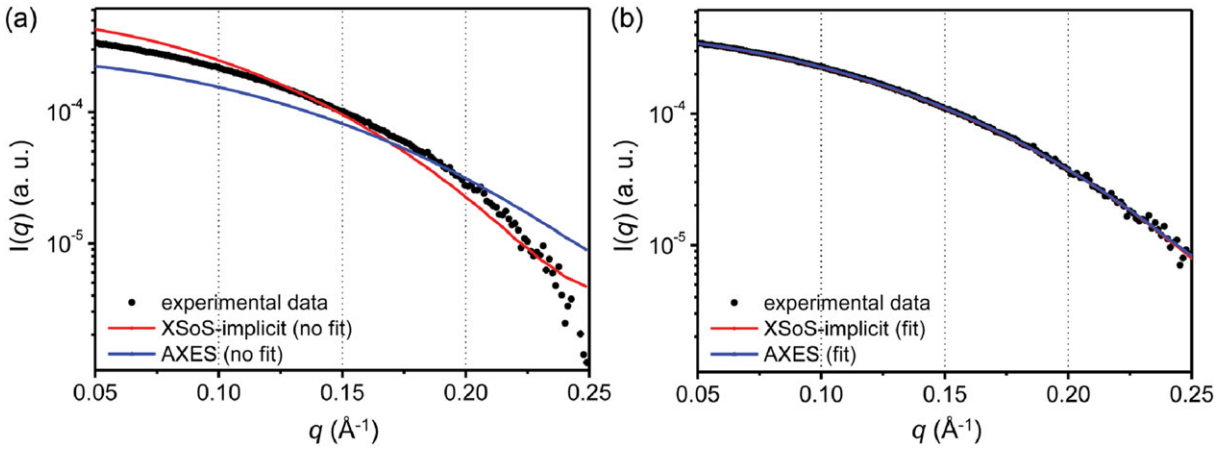

Figure 3. Fitting of experimental scattering data. (a) Raw curves from calculations using XSoS-implicit (red line) and AXES (blue line) exhibit significant discrepancy from the experimental curve (black points). (b) In contrast, the curves refined by fitting show good agreement with the experimental curve.

agreement between the theoretical and the experimental curves as well as (2) the agreement between the theoretical scattering curves calculated by XSoS-implicit and AXES improve significantly (Figure 3(b)). Quantitatively, the chi-squared values between the theoretical scattering curves from XSoS-implicit and the experimental curve are 10.30 and 3.76 before and after the fitting, respectively. The chi-squared value between the theoretical scattering curve from AXES and the experimental curve decreases from 14.97 to 3.44 after the fitting. These results indicate that fitting is critical to generate the theoretical scattering curve that matches the experimental curve.

Convergence of Scattering Curves Calculated by Explicit Solvent Models. We calculated theoretical scattering curves of a protein structure using XSoS-explicit while varying the solvent atomic coordinate sets. As can be seen in Figure 4(a), we found that the shape of the static scattering curve changes sensitively depending on the solvent atomic coordinate sets if the number of MD snapshots is within the default limit, which is 100 atomic coordinate sets each for the solvent-excluded volume and the hydration layer. According to previous studies, the effect of the solvent atomic coordinates can be subdued by averaging the ensembles obtained from the MD simulations.[35,36] Hence, we calculated the scattering curves with an increasing number of MD snapshots to reduce undesirable effects of the solvent atomic coordinate sets on the scattering curves (Figure 4(b)). In particular, we performed the MD simulation up to 2 ns while capturing a snapshot every 1 ps, giving a maximum total of 2000 snapshots. Our results indicate that the calculated scattering curve converges into a curve of unique shape when the number of solvent degrees of freedom is increased to above 500 snapshots (Figure 4(c)). The same result was further verified by calculating scattering curves using two other explicit solvent-model programs, AXES (Figure S3(a)) and WAXSiS (Figure S3(b)).

Advantages of the Implicit Solvent Model. Representative implicit solvent-model program, CRYSOL, is the most widely used program for the calculation of scattering curves and has established a database of WAXS scattering patterns of proteins, highlighting the reliability of the implicit solvent model for deciphering structural information from experimental solution scattering data.[22] In particular, the implicit solvent model does not require any extra computation to obtain atomic coordinates (e.g., MD simulation) and involves only a small number of atoms in the calculation compared to explicit solvent model, leading to substantially reduced computational load. Thus, the implicit solvent model is versatile, user-friendly, and less demanding in terms of computational load.

Limitations of the Explicit Solvent Model. The explicit solvent model has been regarded as an ideal model to realize the physiological aqueous environment of proteins in solution,

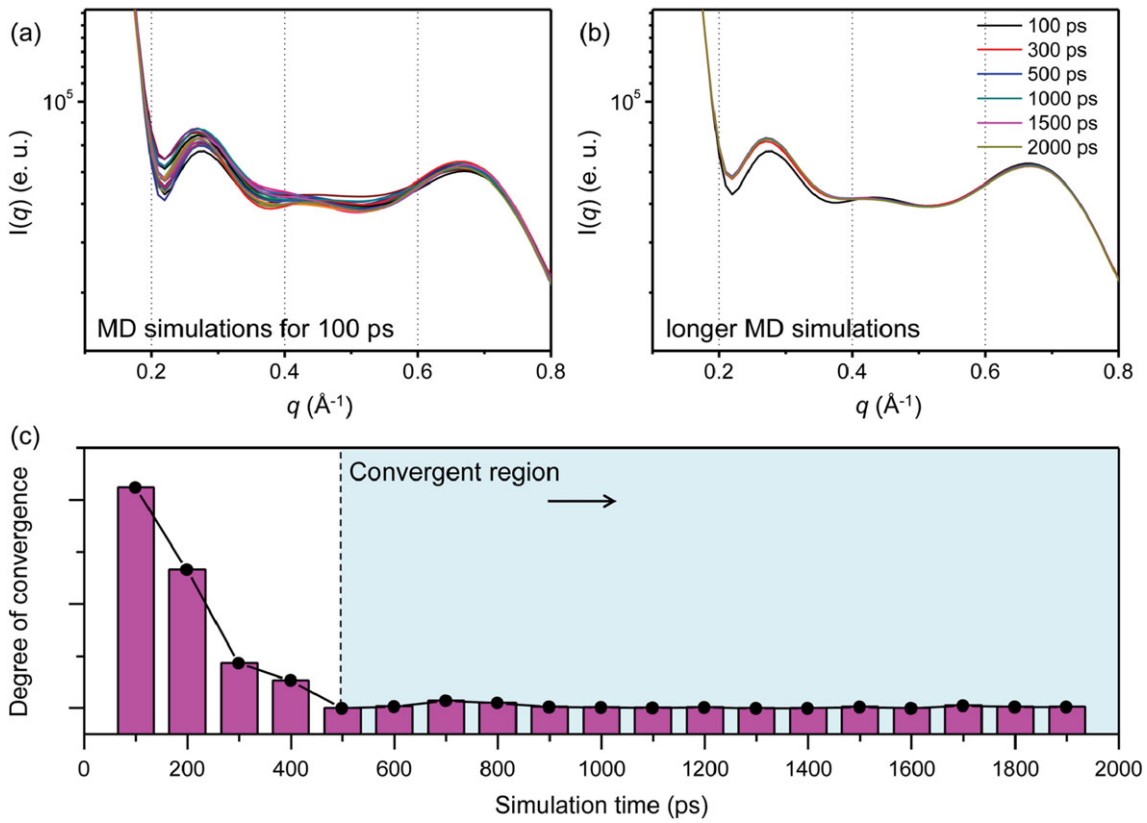

Figure 4. Convergence of explicit solvent-model scattering curves with respect to the number of MD snapshots. (a) XSoS-explicit curves calculated for 20 different atomic coordinate sets with the default number (=100) of snapshots. (b) XSoS-explicit curves calculated with an increasing number of atomic coordinate sets. (c) Degree of convergence versus the number of atomic coordinate sets.

but it has many limitations in reality. First, to obtain the atomic coordinates needed for the calculation of the scattering curves, the explicit solvent model requires snapshots from MD simulations, where the molecular vibrations are significantly under-represented.[63] Averaged atomic coordinates of water molecules sampled from an MD-simulated ensemble can introduce errors to the calculated scattering curves. Validating the results of MD simulations is yet another issue that needs to be addressed.[64] For example, the force field parameters may be stiff and not allow the exploration of full conformational landscape on available time-scales.[65] In addition, the choice of a water model in the MD simulation is critical in determining the configuration of the simulated solvent environment.[66] For example, it was suggested that the difference between the TIP3P water model and water molecules in the real solvent environment may cause variations in the shape of the scattering curve.[61,67] Also, the explicit solvent model is highly demanding in terms of computational cost because of much more atoms involved in the calculation compared to the implicit solvent model.

Conclusion

We developed our own programs to calculate the X-ray scattering curves by using both the implicit and explicit solvent models. Both XSoS-implicit and XSoS-explicit exhibit high performance in terms of calculating X-ray solution scattering curves with high accuracy. While the static scattering curve calculated by XSoS-explicit sensitively changes depending on the solvent atomic coordinate sets, we found that the calculated scattering curve converges into a curve of unique shape as long as the number of MD snapshots is over a certain limit. Overall, the implicit solvent model has practical advantages over the explicit solvent model for the analysis of experimental X-ray solution scattering data.

Acknowledgments. This work was supported by IBS-R004-G2 and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1002511). We thank Prof. Adriaan Bax and Dr. Alexander Grishaev (National Institute of Health, Bethesda) for providing the AXES programs for calculating the scattering curve. We also thank Dr. Jochen Hub (Georg-August-University Göttingen) for giving us access to the WAXSiS program.

Supporting Information. Additional supporting information is available in the online version of this article.

References

1. Y. Imamoto, H. Kamikubo, M. Harigai, N. Shimizu, M. Kataoka, *Biochemistry* **2002**, *41*, 13595.
2. B. Zagrovic, G. Jayachandran, I. S. Millett, S. Doniach, V. S. Pande, *J. Mol. Biol.* **2005**, *353*, 232.
3. C. D. Putnam, M. Hammel, G. L. Hura, J. A. Tainer, *Q. Rev. Biophys.* **2007**, *40*, 191.
4. G. L. Hura, A. L. Menon, M. Hammel, R. P. Rambo, F. L. PooleII, S. E. Tsutakawa, F. F. Jenny Jr.. , S. Classen, K. A. Frankel, R. C. Hopkins, S. Yang, J. W. Scott, B. D. Dillard, M. W. W. Adams, J. A. Tainer, *Nat. Methods* **2009**, *6*, 606.
5. D. I. Svergun, *Biol. Chem.* **2010**, *392*, 737.
6. R. P. Rambo, J. A. Tainer, *Curr. Opin. Struct. Biol.* **2010**, *20*, 128.
7. D. A. Jacques, J. Trewhella, *Protein Sci.* **2010**, *19*, 642.
8. A. Grisheav, J. Wu, J. Trewhella, A. Bax, *J. Am. Chem. Soc.* **2005**, *127*, 16621.
9. Y. Lu, E. Longman, K. G. Davis, A. Ortega, J. G. Grossmann, T. E. Michaelsen, J. G. de la Torre, S. E. Harding, *Biophys. J.* **2006**, *91*, 1688.
10. F. Förster, B. Webb, K. A. Krukenberg, H. Tsuruta, D. A. Agard, A. Sali, *J. Mol. Biol.* **2008**, *382*, 1089.
11. S. J. Caldwell, A. M. Berghuis, *Antimicrob. Agents Chemother.* **2012**, *56*, 1899.
12. C. Dumas, J. Janin, *FEBS Lett.* **1983**, *153*, 128.
13. Y. Izumi, H. Watanabe, N. Watanabe, A. Aoyama, Y. Jinbo, N. Hayashi, *Biochemistry* **2008**, *47*, 7158.
14. C. Y. Cheng, J. Yang, S. S. Taylor, D. K. Blumenthal, *J. Biol. Chem.* **2009**, *285*, 35916.
15. M. Guttman, P. Weinkam, A. Sali, K. K. Lee, *Structure* **2013**, *21*, 321.
16. L. Makowski, D. Gore, S. Mandava, D. Minh, S. Park, D. J. Rodi, R. F. Fischetti, *Biopolymers* **2011**, *95*, 531.
17. R. Dhatwalia, H. Singh, M. Oppenheimer, D. B. Karr, J. C. Nix, P. Sobrado, J. J. Tanner, *J. Biol. Chem.* **2012**, *287*, 9041.
18. M. Kataoka, Y. Goto, *Fold. Des.* **1996**, *1*, R107.
19. D. J. Segel, A. L. Fink, K. O. Hodgson, S. Doniach, *Biochemistry* **1999**, *37*, 12443.
20. S. Akiyama, S. Takahashi, T. Kimura, K. Ishimori, I. Morishima, Y. Nishikawa, T. Fujisama, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1329.
21. T. Konuma, T. Kimura, S. Matsumoto, Y. Goto, T. Fujisawa, A. R. Fersht, S. Takahashi, *J. Mol. Biol.* **2011**, *405*, 1284.
22. L. Makowski, *J. Struct. Funct. Genomics* **2010**, *11*, 9.
23. H. Fischer, M. d. O. Neto, H. B. Napolitano, I. Polikarpovb, A. F. Craievicha, *J. Appl. Cryst.* **2010**, *43*, 101.
24. B. Różycki, Y. C. Kim, G. Hummer, *Structure* **2011**, *19*, 109.
25. T. Fujisawa, A. Kostyukova, Y. Maéda, *FEBS Lett.* **2001**, *498*, 67.
26. M. Hirai, H. Iwase, T. Hayakawa, K. Miura, K. Inoue, *J. Synchrotron Rad.* **2002**, *9*, 202.
27. L. Makowski, D. J. Rodi, S. Mandava, D. Minh, B. Gore, R. F. Fischetti, *J. Mol. Biol.* **2008**, *375*, 529.
28. P. Bernadó, Y. Pérez, J. Blobel, J. Fernández-Recio, D. I. Svergun, M. Pons, *Protein Sci.* **2009**, *18*, 716.
29. D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, G. Zaccai, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267.
30. F. Merzel, J. C. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378.
31. Y. Seki, T. Tomizawa, N. N. Khechinashvili, K. Soda, *Biophys. Chem.* **2002**, *95*, 235.
32. A. S. Hyman, *Macromolecules* **1975**, *8*, 849.
33. M. Y. Pavlov, B. A. Fedorov, *Biopolymers* **1983**, *22*, 1507.
34. D. I. Svergun, C. Barberato, M. H. J. Koch, *J. Appl. Cryst.* **1995**, *28*, 768.

35. S. Park, J. P. Bardhan, B. Roux, L. Makowski, *J. Chem. Phys.* **2009**, *130*, 134114.
36. A. Grishaev, L. Guo, T. Irving, A. Bax, *J. Am. Chem. Soc.* **2010**, *132*, 15484.
37. S. Ahn, K. H. Kim, Y. Kim, J. Kim, H. Ihee, *J. Phys. Chem. B* **2009**, *113*, 13131.
38. K. H. Kim, S. Muniyappan, K. Y. Oang, J. G. Kim, D. Nozawa, T. Sato, S. Koshihara, R. Henning, I. Kosheleva, H. Ki, Y. Kim, T. W. Kim, J. Kim, S. Adachi, H. Ihee, *J. Am. Chem. Soc.* **2012**, *134*, 7001.
39. M. Cammarata, M. Levantino, M. Wulff, A. Cupane, *J. Mol. Biol.* **2010**, *400*, 951.
40. J. Bardhan, S. Park, L. Makowski, *J. Appl. Cryst.* **2009**, *42*, 932.
41. D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, *Biophys. J.* **2013**, *105*, 962.
42. A. Ponti, *J. Magn. Reson.* **1999**, *138*, 288.
43. D. Sezer, J. H. Freed, B. Roux, *J. Chem. Phys.* **2008**, *128*, 165106.
44. J. J. Müller, *J. Appl. Cryst.* **1983**, *16*, 74.
45. B. A. Fedorov, O. B. Ptitsyn, L. A. Voronin, *J. Appl. Cryst.* **1974**, *7*, 181.
46. A. Grishaev, V. Tugarinov, L. E. Kay, J. Trewhella, A. Bax, *J. Biomol. NMR* **2008**, *40*, 95.
47. D. S. Shin, M. DiDonato, D. P. Barondeau, G. L. Hura, C. Hitomi, J. A. Berglund, E. D. Getzoff, S. G. Cary, J. A. Tainer, *J. Mol. Biol.* **2009**, *385*, 1534.
48. P. A. Rambo, J. A. Tainer, *Nature* **2013**, *496*, 477.
49. H. Chen, D. Ricklin, M. Hammel, B. L. Garcia, W. J. McWhorter, G. Sfyroera, Y. Q. Wu, A. Tzekou, S. Li, B. V. Geisbrecht, V. L. Woods Jr., J. D. Lambris, *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17621.
50. B. H. Biersmith, M. Hammel, E. R. Geisbrecht, S. Bouyain, *J. Mol. Biol.* **2011**, *408*, 616.
51. D. Schneidman-Duhovny, S. J. Kim, A. Sali, *BMC Struct. Biol.* **2012**, *12*, 17.
52. M. Hammel, H. P. Fierobe, M. Czjzek, V. Kurkal, J. C. Smith, E. A. Bayer, S. Finet, V. Receveur-Bréchot, *J. Biol. Chem.* **2005**, *289*, 38562.
53. J. Zheng, D. C. Gay, B. Demeler, M. A. White, A. T. Keatinge-Clay, *Nat. Chem. Biol.* **2012**, *8*, 615.
54. N. Nishimura, K. Hitomi, A. S. Arvai, R. P. Rambo, C. Hitomi, S. R. Cutler, J. I. Schroeder, E. D. Getzoff, *Science* **2009**, *326*, 1373.
55. J. C. Philips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comp. Chem.* **2005**, *26*, 1781.
56. A. D. Mackerell Jr.. , D. Bashford, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, H. Gao, H. Guo, S. Ha, D. J. McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. W. Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B.* **1998**, *102*, 3586.
57. P. Chen, J. S. Hub, *Biophys. J.* **2014**, *107*, 435.
58. D. I. Svergun, M. Malfois, M. H. J. Koch, S. R. Wigneshweraraj, M. Buck, *J. Biol. Chem.* **2000**, *275*, 4210.
59. M. Kozak, S. Jurga, *Acta Biochim. Pol.* **2002**, *49*, 509.
60. D. Schneidman-Duhovny, M. Hammel, A. Sali, *Nucleic Acids Res.* **2010**, *38*, W540.
61. F. Poitevin, H. Orland, S. Donaich, P. Koehl, M. Delarue, *Nucleic Acids Res.* **2011**, *39*, W184.
62. S. Yang, S. Park, L. Makowski, B. Roux, *Biophys. J.* **2009**, *96*, 4449.
63. D. M. Teide, K. L. Mardis, X. Zuo, *Photosynth. Res.* **2009**, *102*, 267.
64. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesnaik, N. F. van der Vegt, H. B. Yu, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 4064.
65. K. L. Mardis, H. M. Sutton, X. Zuo, J. S. Lindsey, D. M. Tiede, *J. Phys. Chem. A* **2009**, *113*, 2516.
66. D. J. Huggins, *J. Chem. Phys.* **2012**, *136*, 064518.
67. J. J. Virtanen, L. Makowski, T. R. Sosnick, K. F. Freed, *Biophys. J.* **2011**, *101*, 2061.

