Journal of
Applied Crystallography
ISSN 0021-8898

Received 14 May 2009
Accepted 18 August 2009

SoftWAXS: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules

Jaydeep Bardhan,[a,b] Sanghyun Park[c] and Lee Makowski[a]*

[a]Biosciences Division, Argonne National Laboratory, IL, USA, [b]Department of Molecular Biophysics and Physiology, Rush University Medical Center, Chicago, IL, USA, and [c]Mathematics and Computer Science Division, Argonne National Laboratory, IL, USA. Correspondence e-mail: lmakowski@anl.gov

This paper describes a computational approach to estimating wide-angle X-ray solution scattering (WAXS) from proteins, which has been implemented in a computer program called *SoftWAXS*. The accuracy and efficiency of *SoftWAXS* are analyzed for analytically solvable model problems as well as for proteins. Key features of the approach include a numerical procedure for performing the required spherical averaging and explicit representation of the solute–solvent boundary and the surface of the hydration layer. These features allow the Fourier transform of the excluded volume and hydration layer to be computed directly and with high accuracy. This approach will allow future investigation of different treatments of the electron density in the hydration shell. Numerical results illustrate the differences between this approach to modeling the excluded volume and a widely used model that treats the excluded-volume function as a sum of Gaussians representing the individual atomic excluded volumes. Comparison of the results obtained here with those from explicit-solvent molecular dynamics clarifies shortcomings inherent to the representation of solvent as a time-averaged electron-density profile. In addition, an assessment is made of how the calculated scattering patterns depend on input parameters such as the solute-atom radii, the width of the hydration shell and the hydration-layer contrast. These results suggest that obtaining predictive calculations of high-resolution WAXS patterns may require sophisticated treatments of solvent.

1. Introduction

Decades of experimental studies of solution X-ray scattering of proteins have demonstrated that measured data contain a wealth of information about macromolecular structure and structural fluctuations (Luzzati & Tardieu, 1980). Small-angle X-ray scattering (SAXS) and wide-angle X-ray scattering (WAXS) offer information about macromolecular shape and the range of motion experienced in near *in vivo* conditions (Koch *et al.*, 2003; Vachette *et al.*, 2003). Although solution X-ray scattering does not provide enough structural information to determine atomic resolution structures, there is sufficient information that several groups have developed methods to couple solution scattering with crystallography and NMR (Tsutakawa *et al.*, 2007; Putnam *et al.*, 2007; Kojima *et al.*, 2004). SAXS offers structural information at length scales greater than about 10 Å, which is adequate to estimate macromolecular shapes (Svergun *et al.*, 1997; Chacón *et al.*, 1998) and to study interactions between proteins (Kim *et al.*, 2008; Williamson *et al.*, 2008). Large changes in molecular conformation can be resolved using SAXS (Doniach, 2001; Durchschlag *et al.*, 1991), as can time-dependent phenomena (Huxley *et al.*, 1980; Weiss *et al.*, 2005; Cammarata *et al.*, 2008; Plaxco *et al.*, 1999; Segel *et al.*, 1999; Ichiyanagi *et al.*, 2009; Ihee, 2009).

WAXS experiments provide information on structure at length scales ranging from 2 to 50 Å (Tiede *et al.*, 2002; Fischetti *et al.*, 2004; Makowski, Rodi, Mandava, Devarapalli & Fischetti, 2008), with length scales less than 10 Å commonly referred to as the wide-angle regime. Signal-to-noise ratios are much lower than those in the SAXS regime, both because the signal is one to two orders of magnitude weaker, and because the intensity of scattering from solvent increases rapidly beyond scattering angles corresponding to ∼5 Å. WAXS has only become feasible following the introduction of high-brilliance X-ray sources. The higher doses of X-rays required for WAXS necessitate careful study to ensure that protein integrity is not compromised (Fischetti *et al.*, 2003), but the resulting data have proven valuable for probing details of protein structure in solution (Zhang *et al.*, 2000; Hirai *et al.*, 2002), conformational changes due to ligand binding (Fischetti *et al.*, 2004; Rodi *et al.*, 2007) and the breadth of the populated conformational ensembles (O'Donnell *et al.*, 2007; Makowski, Rodi, Mandava, Minh *et al.*, 2008).

In response to the growing popularity of solution scattering experiments, a number of groups have introduced software to analyze experimental data or predict scattering from molecular structure (Svergun *et al.*, 1995, 1997; Tiede *et al.*, 2002; Hiragi *et al.*, 2003; Konarev *et al.*, 2003; Zhou *et al.*, 2005; Merzel *et al.*, 2007). Many of the available programs focus on using SAXS for determining low-resolution models of protein shape or domain organization (Walther *et al.*, 2000; Wriggers & Chacón, 2001; Svergun *et al.*, 2001) or for narrowing the space of possible protein folds (Zheng & Doniach, 2005; Makowski, Rodi, Mandava, Devarapalli & Fischetti, 2008). Sokolova *et al.* (2003) have introduced a database of SAXS patterns from different proteins. The development of similar tools to couple molecular modeling with WAXS is more difficult; one of the most important challenges for accurately predicting scattering is the task of estimating the scattering from water in the immediate vicinity of protein (Svergun *et al.*, 1995; Soda *et al.*, 1997; Seki *et al.*, 2002). Whereas for low-resolution experiments (SAXS) the water molecules in the hydration layer may not evidence significant structure, this is not necessarily true at higher resolutions. Krack *et al.* (2002) have compared several reasonably accurate approaches to predicting scattering from pure water, and other recent work suggests that the prediction problem is relatively tractable (Sorenson *et al.*, 2000; for a recent review, see Head-Gordon & Hura, 2002).

Software developed by Svergun and collaborators (see, for example, Semenyuk & Svergun, 1991; Svergun, 1992; Svergun *et al.*, 1995, 1997) has played important roles in growing the capabilities and popularity of SAXS for studying proteins (Putnam *et al.*, 2007). One of their most important contributions has been the development of the program *CRYSOL* (Svergun *et al.*, 1995). *CRYSOL* employs spherical harmonics to allow the rapid prediction of scattering from atomic coordinate sets (Lattman, 1989), and in addition allowed Svergun *et al.* to address the problem of hydration-layer water. Spherical harmonics can also be used to make the inverse problem tractable (Svergun *et al.*, 1997). Spherical harmonics are not necessarily an optimal approach, however, because some structures are difficult or impossible to represent. Svergun *et al.* noted that, at angles higher than $q \simeq 4\ \mathrm{nm}^{-1}$, a method based on cubes (Pavlov & Fedorov, 1983) would be expected to be more accurate than *CRYSOL*, because a cube representation of the excluded volume and the hydration layer can capture finer structural details. In this work, we explore precisely such a comparison, using a program we have developed, called *SoftWAXS*.

Our primary goal in developing *SoftWAXS* is to provide a flexible computational tool for estimating WAXS patterns such that discrepancies between predicted and experimental scattering can be assigned to one of four sources: experimental error; the effect of structural flexibility (that is, the breadth of the ensemble); inadequacy of the scattering model [representing the solvent as a continuum (Svergun *et al.*, 1995; Park *et al.*, 2009)]; and real differences between the structural model used for the calculation and the structure of the protein in solution. It is also possible that the experimental data contain systematic errors that have not yet been identified. As an example, Park *et al.* (2009) recently showed that a common approach to remove solvent scattering can introduce errors and that an alternative based on molecular dynamics (MD) can systematically improve agreement between experimental data and theoretical prediction. A second discrepancy that we expect to encounter is that the model of a rigid protein is inadequate for high-resolution scattering predictions, and that conformational flexibility must be taken into account. We further anticipate that an important class of discrepancies will be those that illustrate inadequacies of the scattering models employed in prediction; at the present time, even calculating X-ray scattering from pure water is a theoretical challenge (Krack *et al.*, 2002). Finally, we expect in many cases that WAXS may provide important evidence of differences between the structure of the protein in solution and the structure represented by the atomic coordinates used in the calculation. For such differences to be credibly interpreted, we need to demonstrate that discrepancies cannot be accounted for by other effects.

Three key findings emerge from our development of *SoftWAXS*. First, it appears that the treatment of the volume excluded to solvent by the protein is a more subtle problem than has been thought. Our calculations of scattering using a cube method, even after extensive validation, exhibit significant deviations from the popular sum-of-atomic-volumes approach employed in *CRYSOL* (Svergun *et al.*, 1995) and by other groups (*e.g.* Tiede *et al.*, 2002). Our second finding compounds the first, as we find that, although the variation of atomic radii can have a significant impact on the calculated scattering, it does not necessarily resolve differences between theory and experiment. Third and finally, by comparing *SoftWAXS* calculations with scattering patterns calculated from all-atom, explicit-solvent MD simulations (Park *et al.*, 2009), it becomes clear that continuum-density representation of solvent may, in the general case, be problematic for predicting wide-angle scattering.

The following section, §2, reviews the mathematical model for estimating WAXS scattering of proteins, and §3 presents the numerical techniques used in *SoftWAXS* to implement this model. Computational results in §4 demonstrate that the numerical algorithm has been implemented correctly and illustrate that there exist important differences between scattering patterns calculated using *SoftWAXS* and the approach employed in *CRYSOL* (Svergun *et al.*, 1995). A brief discussion in §5, with a description of important areas for future work, concludes the paper.

2. Theory

We now present a brief description of the mathematical model used to estimate scattering from the atomic coordinates of a molecular solute. Several recent reviews have presented more comprehensive discussions of the scattering model (Koch *et al.*, 2003; Vachette *et al.*, 2003).

Experimental measurements of solution X-ray scattering capture the squared magnitude of the Fourier transform of the

electron density in the sample volume, averaging over both time and the molecular orientation as solutes tumble freely in solution. We write the time of exposure as T, the scattering vector as $\mathbf{q}$ and the scalar magnitude of the vector as q, where $q = 4\pi \sin(\theta/2)/\lambda$ and θ is the scattering angle. The intensity at a scattering vector $\mathbf{q}$ at a given time t is written as

$$I(\mathbf{q}, t) = \left| \int \exp(-i\mathbf{q} \cdot \mathbf{r}) \rho(\mathbf{r}, t)\, \mathrm{d}\mathbf{r} \right|^2, \tag{1}$$

where $\rho(\mathbf{r}, t)$ is the electron density as a function of position $\mathbf{r}$ and time t. The actual measured intensity at q is then

$$I(q) = (1/T) \int_0^T \langle I(\mathbf{q}, t) \rangle_\Omega \, \mathrm{d}t, \tag{2}$$

where T is the duration of exposure and the subscripted angle bracket denotes an average over the solid angle of q (that is, over all $\mathbf{q}$ such that $||\mathbf{q}|| = q$).

In most approaches, and in *SoftWAXS*, the Fourier transform of the excess electron density $\mathcal{F}(\mathbf{q})$ is decomposed into a sum of three components as

$$\mathcal{F}(\mathbf{q}, t) = A(\mathbf{q}, t) + B(\mathbf{q}, t) + C(\mathbf{q}, t), \tag{3}$$

where A represents the solute electron density, B models the removal of solvent from the solute volume and C approximates the perturbation of the solvent electron density near the solute–solvent interface (Svergun *et al.*, 1995).

Calculations of protein solution scattering often treat the protein as rigid and the solvation shell as a featureless continuum. This allows all three terms of equation (3) to be regarded as constant with respect to time, which simplifies the calculations considerably. These assumptions are reasonable for SAXS, which provides information on length scales greater than about 10 Å. At this length scale, water is not expected to display long-range ordering. For larger q, this may not be the case. Furthermore, because the time averaging during measurement does not commute with the squaring of the Fourier-transformed electron density, one expects to have to account for, at the very least, the correlations of solvent motion around the solute.

If the solute is modeled as a collection of N spherically symmetric scatterers, corresponding to atoms or chemical groups, then the contribution from each scatterer is symmetric in $\mathbf{q}$, and by superposition

$$\mathcal{F}(\mathbf{q}) = \sum_{j=1}^{N} f_j(q) \exp(-i\mathbf{q} \cdot \mathbf{r}_j), \tag{4}$$

where $\mathbf{r}_j$ is the center for the jth scatterer and $f_j(q)$ is the corresponding scattering factor if the scatterer was located at the origin. For spherically symmetric scatterers, the orientational average can be evaluated analytically to give the Debye formula

$$I(q) = \sum_{j=1}^{N} \sum_{k=1}^{N} f_j(q) f_k(q) \frac{\sin(q r_{jk})}{q r_{jk}}, \tag{5}$$

where r_{jk} is the distance between scatterers j and k. In *SoftWAXS*, as in other work (Svergun *et al.*, 1995; Tiede *et al.*, 2002), the solute-atom electron densities are presumed to be spherically symmetric. It should be noted, however, that recent work has shown that accurate calculation of X-ray scattering from water requires a more detailed description of the electron density (Sorenson *et al.*, 2000; Krack *et al.*, 2002).

3. Numerical algorithm

3.1. Spherical averaging

The *SoftWAXS* approach to calculating $I(q)$ centers around calculating the orientational average using *numerical quadrature*, estimating the integral over the domain Ω (the solid angle) by computing a weighted sum of function values taken at certain points in the domain. The quadrature rules for calculating the orientational average take the form

$$I(q) = \frac{1}{4\pi} \int_\Omega \left| [A(\mathbf{q}) + B(\mathbf{q}) + C(\mathbf{q})] \right|^2 \mathrm{d}\Omega \tag{6}$$

$$\simeq \frac{1}{4\pi} \sum_{j=1}^{N} \alpha_j \left| A(\mathbf{q}_j) + B(\mathbf{q}_j) + C(\mathbf{q}_j) \right|^2. \tag{7}$$

An approximation with N points is called an N-point quadrature rule, and $\mathbf{q}_j$ and α_j are referred to as the jth quadrature point and weight, respectively. Results in §4.1 illustrate how large N must be for estimating WAXS patterns from proteins.

3.2. Solute scattering

The solute-atom scattering $A(\mathbf{q})$ can be calculated using equation (4) with the sum taken over all atoms in the solute:

$$A(\mathbf{q}) = \sum_{j=1} f_j(q) \exp(-i\mathbf{q} \cdot \mathbf{r}_j). \tag{8}$$

This approach implicitly assumes that the solute is rigid. However, it has been noted that it can be important to account for solute flexibility at least in an average way, in order to achieve agreement between calculated and measured scattering (Tiede *et al.*, 2002). To this end, Tiede *et al.* use displacement parameters (also known as a Debye–Waller factor or B factor) from crystal structures. Files in the Protein Data Bank format (PDB; Berman *et al.*, 2000) contain a per-atom data field for displacement parameters. Every atomic contribution is then convolved with a Gaussian of width proportional to the atom's B factor. Denoting the B factor of atom j by B_j, we have

$$A(\mathbf{q}) = \sum_{j=1}^{M} f_j(q) \exp(-B_j q^2/16\pi^2) \exp(-i\mathbf{q} \cdot \mathbf{r}_j). \tag{9}$$

Displacement parameters are not necessarily an ideal representation (they assume isotropic fluctuations and may be affected by crystal contacts), but provide a useful approach that has been shown to improve agreement with experiment (Tiede *et al.*, 2002).

SoftWAXS allows the use of either equation (8) or (9) according to the interests of the user.

3.3. Excluded-volume scattering

The excluded-volume scattering $B(\mathbf{q})$ can be calculated using either of two different models depending on the user's interests. The first approach is to add the Gaussian excluded volumes of Fraser *et al.* (1978) (see Fig. 1*a*). This is the dummy-atom approach used in *CRYSOL* (Svergun *et al.*, 1995) and discussed by Tiede *et al.* (2002). In *SoftWAXS*, hard spheres can also be used in place of Gaussians, with the sphere radii defined such that the total excluded volume remains unchanged from the Fraser *et al.* total excluded volume. Fig. 1(*a*) is an illustration of the sum-of-hard-spheres approach to computing the excluded volume for two spheres.

In the second approach, the excluded volume is defined by a surface that separates the interior of the solute from the solvent exterior. The solute–solvent interface can be defined as the van der Waals (vdW) surface, the solvent-accessible surface (SAS) or the solvent-excluded (also known as the molecular) surface (Richards, 1977; Connolly, 1983). The excluded volume is then said to be the space inside the defined surface. The volume enclosed by the vdW surface is merely the union of spheres with atomic radii (Fig. 1*b*). The SAS and molecular surface are defined by rolling a probe sphere, which approximates a water molecule, around the spheres' union.

We illustrate the sum-of-Gaussians and union-of-spheres approaches to defining the excluded volume using the protein lysozyme [PDB accession number 6lyz (Diamond, 1974)]. In Fig. 2(*a*) is an illustration of a line segment whose endpoints are atoms on opposite sides of the protein, and the two excluded-volume functions are plotted in Fig. 2(*b*). The sum-of-Gaussians approach leads to surprisingly large nonphysical variations as a function of position along the line segment. This heterogeneity should have minimal impact on calculated scattering at small q, but may result in erroneous predictions at wide angles, which capture details on the same length scale as the variations in the excluded-volume function.

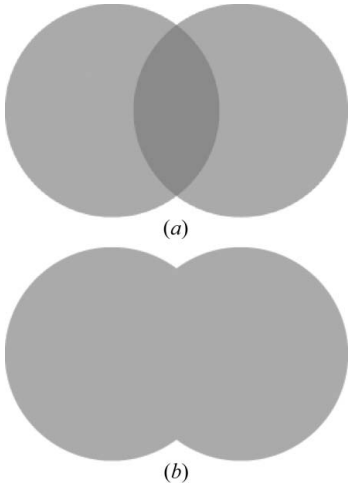

Figure 1
Two approaches to modeling the excluded volume using atomic or group contributions. (*a*) The sum-of-excluded-volume functions. (*b*) The union-of-excluded-volume functions. For clarity, in these figures the atomic contributions have been modeled as hard spheres rather than Gaussians.

Fig. 3(*a*) is a diagram illustrating an excluded-volume surface and a hydration-layer surface; the region between these surfaces is defined to be the hydration layer. For a given set of atomic coordinates, the excluded-volume and hydration-layer surfaces are defined using software developed for boundary-element simulation of molecular electrostatics (Altman *et al.*, 2009). The program *MSMS* (Sanner *et al.*, 1996; Sanner, 1996) is used to calculate a set of planar triangles that approximate the appropriate surfaces. The density of triangle vertices per unit area can be specified by the user.

The scattering from the excluded volume and hydration shell are calculated using hierarchical Fourier transforms. A bounding cube is first defined that surrounds all of the boundary elements for a given surface. This cube is then recursively subdivided into smaller cubes using what is known as an *octree decomposition*. Cubes that intersect no boundary elements are not subdivided further. Fig. 3(*b*) is a schematic illustrating the recursion process. A cube's *children* are

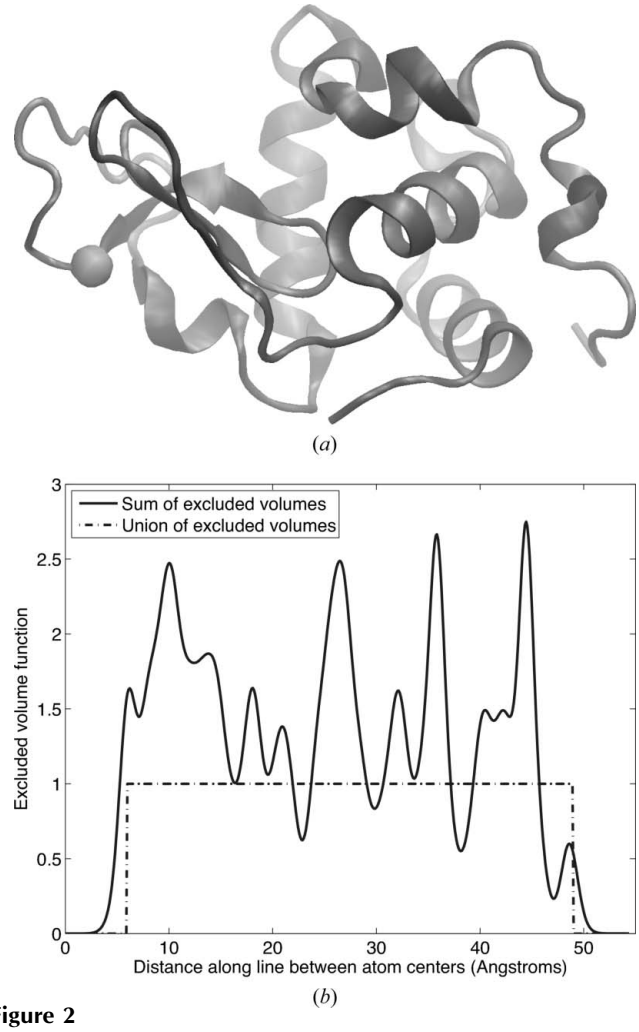

Figure 2
(*a*) Illustration of the atoms that define the line segment along which we plot the sum-of-excluded-volume function. (*b*) Sum of excluded volumes along the indicated line segment. Distances are in Å. The atomic excluded-volume functions are defined as in *CRYSOL* (Svergun *et al.*, 1995). Along the plotted axis, the ASP66:CA atom is at 10^5 Å on this line and the ALA122:N atom is at 44.0 Å.

defined to be the eight cubes inside it that make it up; such a cube is said to be the *parent* to these eight cubes. The recursion is halted at a user-specified depth. A cube that has no children (whether because it intersects no panels or because it is at the maximum recursion depth) is called a *leaf* cube. §4 contains results demonstrating WAXS patterns calculated using different octree depths.

Because all the surfaces are closed, every leaf cube is either inside or outside the appropriate surface. Whether a cube is inside or outside is determined by the cube center's relationship to the nearest triangles approximating the surface. The Fourier transform of the associated volume is then easily calculated as the sum of Fourier transforms of the leaf cubes that are inside the surface,

$$B(\mathbf{q}) = \rho_{\rm w} \sum_j \mathcal{F}(\mathbf{q}; h_j) \exp(-i\mathbf{q}\cdot\mathbf{r}_j), \tag{10}$$

where $\rho_{\rm w}$ is the electron density of water (taken in this work to be 0.334 e Å^{-3}), j denotes the current leaf cube, h_j denotes the length of a cube edge and $\mathbf{r}_j$ its center, and $\mathcal{F}(\mathbf{q}; h_j)$ is the Fourier transform of a cube of edge length h_j and centered at the origin, evaluated at $\mathbf{q}$:

$$\mathcal{F}(\mathbf{q}; h_j) = \int_{-h_j/2}^{+h_j/2}\int_{-h_j/2}^{+h_j/2}\int_{-h_j/2}^{+h_j/2} \exp(-i\mathbf{q}\cdot\mathbf{r}')\,{\rm d}x'\,{\rm d}y'\,{\rm d}z'. \tag{11}$$

3.4. Hydration-layer scattering

SoftWAXS can calculate hydration-shell scattering when the hydration layer is defined as the region between the two surfaces (Fig. 3). In this case, a calculation similar to that in equation (10) is easily performed, and the current treatment follows that of *CRYSOL* (Svergun *et al.*, 1995): the electron density in the hydration layer is taken to be a uniform value different from that of bulk water. However, the means by which the excluded-volume and hydration-layer scattering are calculated offer the possibility of using more complex models.

In the current implementation, *SoftWAXS* computes the Fourier transform of the whole region inside the outermost surface, with a weight of $\Delta\rho_{\rm w}$ in place of $\rho_{\rm w}$, where $\Delta\rho_{\rm w}$ is the change in electron density from bulk water to that assumed in the hydration shell. The weight for the excluded volume is then modified to be $\rho_{\rm w} - \Delta\rho_{\rm w}$. This approach is faster than computing the hydration-layer and excluded-volume terms separately, because the relatively thin hydration layer would have to be represented using many small cubes at the interior surface.

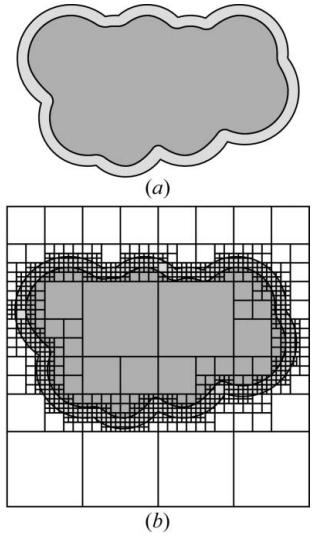

Figure 3
The *SoftWAXS* hierarchical-cube approach to estimating excluded-volume and hydration-layer scattering. (*a*) Definition of the excluded volume and hydration layer. (*b*) An illustration of the recursive octree decomposition.

4. Results

The algorithm presented in the previous section for estimating $I(q)$ given a protein structure, under the assumption of rigidity, has been validated extensively. We first present a series of analytically solvable cases to demonstrate how the calculated scattering varies with several *SoftWAXS* input parameters. In particular, we vary $n_{\rm sph}$, the number of points used to evaluate the spherical average; $\rho_{\rm vert}$, the density of vertices used to approximate the excluded-volume surface; and $d_{\rm oct}$, the depth of the octree.

4.1. Spherical averaging

We first demonstrate the correct implementation and numerical performance of the numerical quadrature approach to calculating spherical averages, using hen egg white lysozyme (PDB accession 6lyz) as a test problem. Fig. 4 contains plots of the atomic contributions to scattering when using the analytical Debye formula, *CRYSOL* (Svergun *et al.*, 1995) and

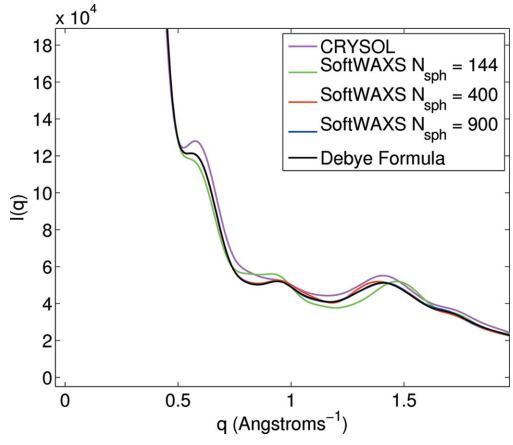

Figure 4
Calculations of the atomic form-factor contribution to scattering for lysozyme [PDB accession code 6lyz (Diamond, 1974)]. Calculations have been performed using the Debye formula, *CRYSOL* and *SoftWAXS* using 144, 400 or 900 quadrature points to evaluate the spherical averages. The curve for *SoftWAXS* using 900 quadrature points is indistinguishable on this scale from the curve obtained using the Debye formula. The scattering angle q is in Å^{-1}. Units on the ordinate axis are arbitrary intensity units.

SoftWAXS. The *SoftWAXS* atomic form-factor scattering has been calculated using equation (8) with different numbers of quadrature points to demonstrate the spherical averaging procedure's convergence with increasing numbers of points. All of the *SoftWAXS* numerical approaches provide excellent accuracy to approximately 0.5 Å^{-1}, and the 144-point quadrature rule begins to provide poor accuracy beyond this resolution. The 400-point and 900-point accuracy are maintained out to about 1.5 and 1.8 Å^{-1}, respectively. For large proteins or to examine scattering beyond 1.8 Å^{-1}, more than 900 quadrature points may be required. All *SoftWAXS* results presented in the remainder of this paper use 900-point quadrature unless explicitly noted otherwise.

Fig. 5 contains plots of the excluded-volume and hydration-shell scattering using the same numbers of quadrature points. Results are plotted on a semilog scale to improve the visibility of minor differences between calculated intensities. For comparison, *CRYSOL* results have been plotted also, though we emphasize that discrepancies between *CRYSOL* and *SoftWAXS* excluded-volume scattering are largely attributable to differences in the methods' treatment of excluded volume (see Figs. 5 and 6). It is worth noting that the *CRYSOL* hydration shell appears larger than that employed in *SoftWAXS*. We attribute this difference to the size of the probe sphere used to define the excluded-volume and hydration-shell surfaces.

In Figs. 4 and 5, discrepancies between the *CRYSOL* and *SoftWAXS* atomic contributions are likely due to the former's group treatment of H atoms; in *SoftWAXS*, all H atoms are explicitly accounted for, having been added using molecular mechanics software. We have used *VMD* (Humphrey *et al.*, 1996) and the *CHARMM* parameter set (Brooks *et al.*, 1983; MacKerell *et al.*, 1998). The addition of H atoms to PDB files does represent an extra step in the computation, particularly in comparison with *CRYSOL* which does not require them. However, it seems likely that at high angles the placement of H atoms will be a necessary step in the accurate calculation of scattering.

More quadrature points are needed at higher angles because the transformed density varies more rapidly as a function of the solid angle. For the excluded-volume scattering, a Gibbs-like ringing phenomenon is observed when the octree is only allowed to recurse to five levels, because this depth corresponds to cube lengths of the order of 0.5 Å, which are captured at the highest angles (smallest length scales).

Use of quadrature also enhances computational efficiency for most proteins. The time required to compute scattering using the Debye formula scales quadratically with the number of scatterers, as one accounts for all pairs of atoms. In contrast, as we have found that 900 quadrature points suffice to perform spherical averaging numerically, the time scales linearly with the number of scatterers, and thus the numerical averaging procedure is faster for macromolecules with more than approximately 900 atoms. Lysozyme, for example, has more than 1900 atoms when H atoms are included explicitly. On a 2.4 GHz Intel MacBook laptop, the time required to compute scattering using the *SoftWAXS* numerical averaging method is approximately 2 min, whereas the pure Debye formula requires 10 min. Using *CRYSOL* with spherical harmonics up to order 50 leads to a computation time of about 90 s; we expect that its speed is (like *SoftWAXS*) attributable to the avoidance of the pairwise computation of the Debye formula. Thus, on the basis of computation time *SoftWAXS* is not dramatically slower than existing tools for predicting scattering intensities.

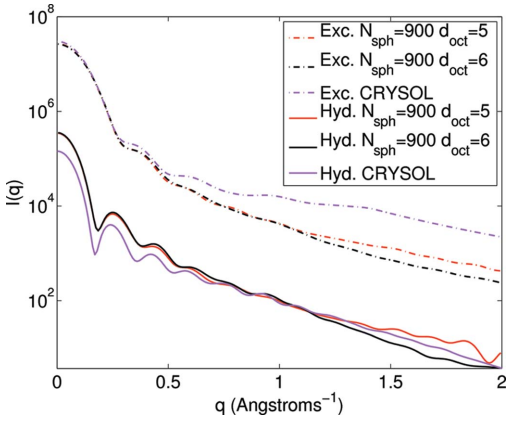

Figure 5
Excluded-volume and hydration-layer contributions to scattering for lysozyme. Calculations have been performed with *SoftWAXS* using 900 quadrature points to evaluate the spherical averages. The results using an octree depth of 5 clearly show artifacts at wide angles ($q > 1$ Å^{-1}) due to the use of coarser cubes. Units on the ordinate axis are arbitrary intensity units. Excluded-volume contributions to $I(q)$ are as follows: for 5-level octree, $I(0.005) = 2.68 \times 10^7$; for 6-level octree, $I(0.005) = 2.64 \times 10^7$; for *CRYSOL*, $I(0) = 3.01 \times 10^7$.

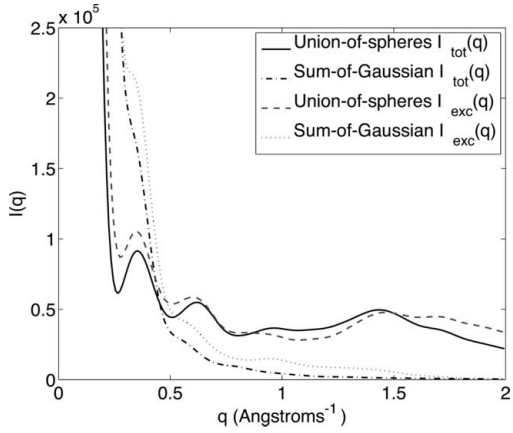

Figure 6
Calculated scattering of lysozyme, using either the union-of-hard-spheres approach (with the surface defined as the solvent-excluded surface using a probe sphere of radius 1.4 Å) or the sum-of-Gaussians method (Svergun *et al.*, 1995). The hard-sphere radii have been set to the optimal radii for fitting to lysozyme experimental data. The volumes for the atoms are: C 44.60, O 7.24, N 1.44, H 4.19 Å^3. Using this approach the excluded-volume contribution is $I(0.005) = 3.06 \times 10^7$. See §4.3.4 for more details. The sum-of-Gaussians calculation employed the same atomic excluded volumes as used in *CRYSOL*: C 16.44, O 9.13, N 2.49, H 5.15 Å^3 (Fraser *et al.*, 1978; Svergun *et al.*, 1995). The excluded-volume contribution is $I(0.005) = 3.41 \times 10^7$.

4.2. Hierarchical transform

4.2.1. Non-overlapping spheres. The problem of multiple non-overlapping hard spheres furnishes a simple, analytically solvable test problem to ensure that the orientational averaging and hierarchical transforms are evaluated correctly. Our test system contains a 3 Å-radius sphere centered at $(0, 0, 0)$, a 7 Å-radius sphere centered at $(13, 0, 0)$ and a 5 Å-radius sphere centered at $(0, 8, 0)$. Fig. 7 is a plot of the analytical scattering intensity and the intensity calculated numerically using *SoftWAXS* and the hierarchical transform. The small discrepancies at $q = 1.6$ Å are greatly exaggerated by the use of the semilog plot.

4.3. Treatment of volume excluded to solvent

As described in §2, a common approach to estimating the excluded-volume scattering of an actual macromolecule is to sum the excluded volumes associated with each atom or atomic group (Svergun *et al.*, 1995; Tiede *et al.*, 2002). In this section we compare this approach to *SoftWAXS* calculations that estimate the excluded volume as the union of excluded volumes.

We present first the relatively simple case of two spheres and analyze how the excluded-volume scattering varies as the sphere separation is varied. Figs. 1(*a*) and 1(*b*) are graphical representations of the difference between the sum-of-excluded-volume and union-of-excluded-volume approaches for the problem of two overlapping spheres. For clarity, hard spheres are represented rather than the usual Gaussians (Fraser *et al.*, 1978). In Fig. 1(*a*), the region in which the two spheres' excluded-volume functions overlap is darker than the non-overlapping regions. The overlap region is counted twice in the sum-of-volumes approach, and only once in the union-of-volumes approach. The use of hard spheres rather than Gaussians exaggerates the effect but, as seen in Fig. 6, using the sum of Gaussians results in contributions to the intensity substantially different from the union-of-spheres model.

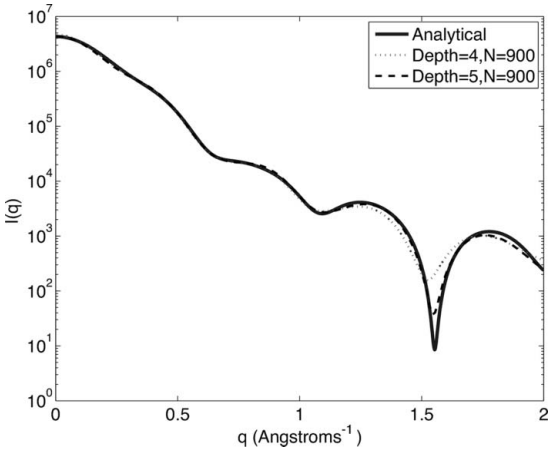

Figure 7
Calculation of the excluded-volume scattering due to three spheres using analytical methods and the numerical methods in *SoftWAXS*. The scattering intensities calculated using 400-point and 900-point numerical quadrature for spherical averaging are indistinguishable at this scale. Using an octree depth of 5, the *SoftWAXS* calculation is essentially indistinguishable from the analytical solution except at the minimum at $q = 1.5$ Å^{-1}, where the semilog scale provides adequate resolution to observe the minute discrepancy.

4.3.1. Lysozyme. The sum-of-excluded-volumes approach double counts regions of the solute interior not only for the two-sphere problem of Fig. 1, but for macromolecules as well. We illustrate this point using hen egg white lysozyme (PDB accession number 6lyz), defining a line from the α-carbon of ASP 66 to the backbone nitrogen of ALA 122 (Fig. 2*a*) and explicitly calculating the sum of excluded volumes along this line (Fig. 2*b*). To illustrate how far this function extends beyond the atoms themselves, it is plotted 10 Å beyond each of the atoms used to define the line. In this calculation, we have used the radii defined by Fraser *et al.* (1978), which are the same as employed by *CRYSOL* (Svergun *et al.*, 1995). The group volumes have not been fitted to match experimental excluded-volume measurements.

4.3.2. Surface definition. The boundary of the excluded volume can be defined in several ways. One approach uses a hard-sphere model for the atoms or atomic groups, takes the union of the spheres, and uses the resulting boundary as the solute–solvent boundary; this is known as the solvent-accessible surface (Lee & Richards, 1971). It is well known that the solvent-accessible surface can have deep, narrow valleys and sharp cusps, which, in the context of defining a solvent dielectric constant for molecular electrostatics, have a non-physical interpretation. An alternative definition, designed to avoid these non-physical surfaces, is called the solvent-excluded or molecular surface (Richards, 1977; Connolly, 1983). The solvent-excluded surface is generated by rolling a probe sphere of user-specified radius around the union of spheres of van der Waals radii, and taking the surface to be the points of closest approach of the probe sphere to the union of spheres. For defining solvent–solute boundaries in electrostatics calculations, most commonly for solving the Poisson or Poisson–Boltzmann equation (Sharp & Honig, 1990), the probe radius is usually taken to be 1.4 Å (approximately that of water). Except where explicitly mentioned otherwise, all calculations reported in the remainder of the paper used a probe sphere of radius 1.4 Å, in order to have the calculated surface match the molecular surface, avoiding some of the most pathological kinds of singularities associated with solvent-accessible surfaces; the hydration-layer surface is defined similarly, but separated by 3 Å from the excluded-volume surface, as in *CRYSOL* (Svergun *et al.*, 1995). We reiterate that this surface definition is not precisely the union-of-spheres definition; however, it should approximate the volume excluded to solvent electrons more closely than the sum-of-Gaussians definition.

Fig. 6 contains plots of the total and excluded-volume scattering from lysozyme, using both the union-of-spheres and the sum-of-Gaussians approaches. In these calculations, no hydration shell was employed (Tiede *et al.*, 2002) and the *SoftWAXS* union-of-spheres radii used were those found to optimize agreement with experimental data (see §4.3.4); the sum-of-Gaussians radii were taken from *CRYSOL*, which employed the Fraser–MacRae–Suzuki (FMS) radii (Fraser *et*

al., 1978; Svergun *et al.*, 1995). For clarity, we present the total excluded volumes for each atom rather than radii (because the union-of-spheres model consists of hard spheres and the sum-of-Gaussians does not).

The volumes for the union-of-spheres model are C 44.60, O 7.24, N 1.44, H 4.19 Å^3. The sum-of-Gaussians calculation employed the volumes as follows: C 16.44, O 9.13, N 2.49, H 5.15 Å^3 (Fraser *et al.*, 1978; Svergun *et al.*, 1995). The locations of the peaks and troughs are in close correspondence, though the ratios of the peak and trough heights are less so. However, the scattering from either model, when using the volumes from the other, are quite different (data not shown). These volumes appear to be unrealistic and we discuss this point in §5.

4.3.3. Hydration-shell parameters. In this section we explore the impact of hydration-layer scattering on WAXS by systematically varying hydration-shell thickness and contrast, not to argue for a particular set of parameters but to demonstrate the flexibility of *SoftWAXS* for exploring scattering models. Svergun and collaborators have demonstrated that the water molecules immediately surrounding macromolecules, known as hydration-shell or solvation-shell waters, contribute significantly to the overall measured intensity (Svergun *et al.*, 1995; Merzel & Smith, 2002). Svergun *et al.* reported that calculated SAXS patterns appear insensitive to the particular choices of the hydration-layer thickness d and contrast $\rho^{\rm hyd}$ so long as the product $d\rho^{\rm hyd}$ remains constant. To explore whether this holds for WAXS patterns as well, we calculated scattering using hydration layers of multiple thicknesses at different contrasts while holding constant the product. Fig. 8 contains plots of the resulting scattering intensities. In the *CRYSOL* software package, the standard parameters are $d = 3$ Å, $\rho^{\rm hyd} = 0.03$ e Å^{-3}. Varying the hydration-layer thickness and contrast simultaneously, it is clear that WAXS patterns are sensitive to individual variations in the hydration-layer parameters.

4.3.4. Parameterizing atomic radii for *SoftWAXS* scattering calculations. To determine how strongly the atomic radii impact the computed scattering profiles, an exhaustive search over C, O and N radii was conducted for lysozyme, myoglobin and cytochrome c. The best radii were determined by computing the χ^2 between predicted and measured scattering profiles for each set of radii (after allowing a linear scaling), and taking for each protein the set that generated the smallest χ^2 as the optimal fit radii. All H radii were fixed at 1 Å. Denoting the C, N and O radii by $R_{\rm C}$, $R_{\rm N}$ and $R_{\rm O}$, the search space was defined by 1.5 Å $\leq R_{\rm C} \leq$ 1.9 Å, 0.95 Å $\leq R_{\rm O} \leq$ 1.2 Å, 0.7 Å $\leq R_{\rm N} \leq$ 1 Å. In this search space, the three proteins produced the same optimum with $R_{\rm C} = 1.9$ Å and $R_{\rm O} = 1.2$ Å, with essentially no dependence on $R_{\rm N}$. Enlarging the search space by increasing the upper limit on $R_{\rm C}$ produced inconsistent results for the different proteins (data not shown). For myoglobin and lysozyme, $R_{\rm C}$ increased to the new upper bound for $R_{\rm C}$ and the optimal O radius decreased; for cytochrome c, however, the optimal radius stayed below 2.0 Å.

Fig. 9(*a*) contains plots of the experimental lysozyme scattering as well as *CRYSOL*-predicted scattering, with and without fitting of parameters to match experimental data (Svergun *et al.*, 1995). Allowing the atomic radii to be fitted clearly improves the agreement with experiment (Svergun *et al.*, 1995; Tiede *et al.*, 2002), but such a refinement procedure, because it is repeated independently for each protein, lacks a clear physical picture. Our exhaustive search in *SoftWAXS*, on the other hand, represents an attempt to find a general set of radii that might be satisfactory over different proteins and yet have a meaningful interpretation (*i.e.* the exhaustive search was an attempt at parameterization based on the three-protein sample set).

Fig. 9(*b*) contains plots of the scattering patterns calculated during the lysozyme search (allowing $R_{\rm C}$ to vary up to 1.9 Å) as well as experimentally measured data. In this plot, the experimental data have been scaled to match the set of calculated intensities. Although the search produces a breadth of different patterns, it is clear that qualitative differences between calculation and experiment remain over the entire space of possible radii. Furthermore, agreement with experiment does not substantially improve when the search space is expanded (data not shown). Fig. 9(*c*) contains plots of the experimental data and the scattering calculated using the optimal *SoftWAXS* radii, where the calculated scattering has been scaled to minimize the χ^2 compared with experiment. The estimates of experimental error are computed as described in earlier work (Rodi *et al.*, 2007), as the standard deviation of multiple (usually seven) X-ray exposures. Figs. 10(*a*) and 10(*b*) contain analogous plots of cytochrome c experimental data, and predicted scattering with and without fitting to the data. We omit the myoglobin results, which resemble those of lysozyme. The χ^2 metric appears to be less than ideal as a metric to compare intensities, because the

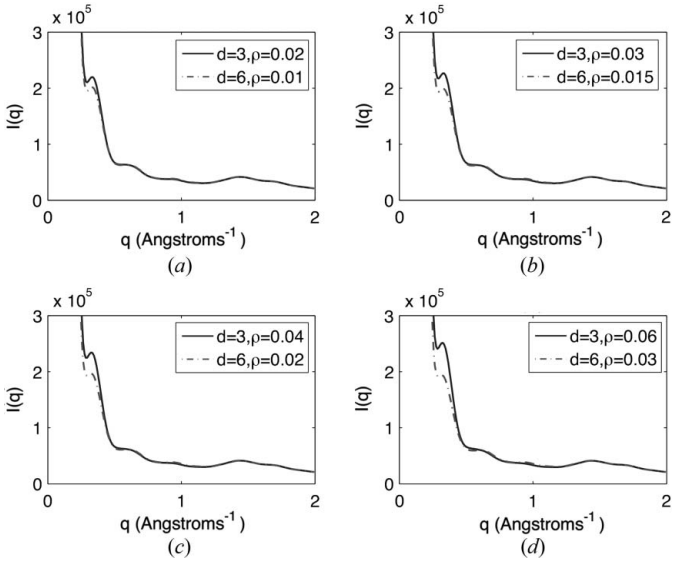

Figure 8
Calculations of WAXS patterns are sensitive to both hydration-layer thickness d and contrast $\rho^{\rm hyd}$ for $q < 0.5$ Å^{-1}. For larger q the parameters, when varied within reasonable limits, have relatively little effect on the pattern. In each plot the products of hydration-layer density (relative to bulk water) and thickness are the same for the two patterns. (*a*) $d\rho^{\rm hyd} = 0.06$ e Å^{-2}, (*b*) $d\rho^{\rm hyd} = 0.09$ e Å^{-2}, (*c*) $d\rho^{\rm hyd} = 0.12$ e Å^{-2}, (*d*) $d\rho^{\rm hyd} = 0.18$ e Å^{-2}.

scattering over the angles of interest takes such a wide range of magnitudes and the error bars (not shown) are small relative to these magnitudes up to very high angles. Finding an appropriate q-dependent scaling to aid the χ^2 calculation might significantly aid in parameterization.

One interesting result of our preliminary search has been the observation that the ratio of small-angle to wide-angle intensity is often mispredicted by a factor of two to three using the Debye formula and FMS radii. This result can be seen using both *CRYSOL* and *SoftWAXS*. Fitting the radii in *SoftWAXS* and *CRYSOL* appears to mitigate this discrepancy significantly, though not to eliminate it entirely. It is possible that protein–protein interactions are responsible for a portion of this scaling problem, though the explicit-solvent-based

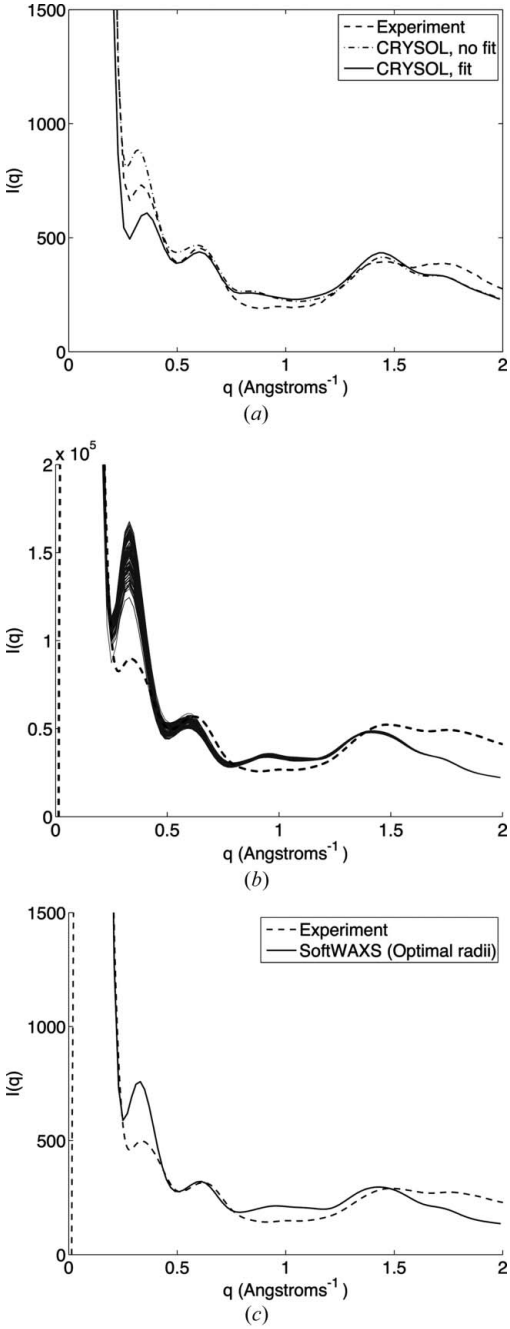

Figure 9
A comparison of lysozyme scattering from experiment, *CRYSOL* and *SoftWAXS*. (*a*) Scattering calculated using *CRYSOL* [both with and without fitting to experiment (Svergun *et al.*, 1995)]. (*b*) Scattering patterns calculated using *SoftWAXS* while varying the C, O and N radii; see text for details. For comparison, the experimental data have also been plotted (dashed line). (*c*) Scattering from experiment and the optimal radii.

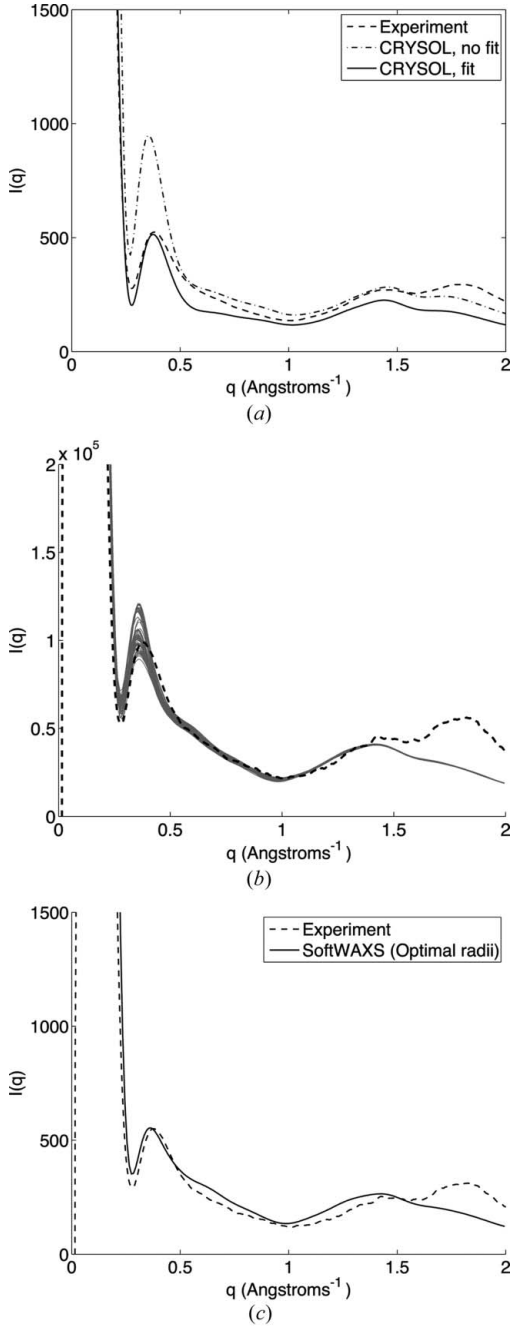

Figure 10
A comparison of cytochrome c scattering from experiment, *CRYSOL* and *SoftWAXS*. (*a*) Scattering calculated using *CRYSOL* [both with and without fitting to experiment (Svergun *et al.*, 1995)]. (*b*) Scattering patterns calculated using *SoftWAXS* while varying the C, O and N radii; see text for details. For comparison, the experimental data have also been plotted (dashed line). (*c*) Scattering from experiment and the optimal radii.

results of Park *et al.* (2009) do not suffer from the same problem; this suggests that protein–protein interactions may play only a minor role.

5. Discussion

In this paper, we have described the implementation and validation of the computer program *SoftWAXS*, the ultimate goal being the development of a computational tool for interpreting wide-angle X-ray scattering data taken from protein solutions in terms of protein structure and structural fluctuations. *SoftWAXS* employs the same mathematical model for estimating scattering as employed by other groups (Svergun *et al.*, 1995; Tiede *et al.*, 2002), and uses a set of numerical techniques that allow higher accuracy with respect to the model. In particular, in *SoftWAXS* orientational averages are computed numerically, rather than analytically. This design decision permits us finer-grained control over the definition of the excluded volume and hydration layer, at the cost of longer computational time. Another key aspect of our implementation is the use of a modified cube method, based on octrees, for representing the components of the scattering which in this model are treated as continua. Our results confirm the assertion by Svergun *et al.* that, for large q, cube methods should allow better accuracy than representations based on spherical harmonics (Svergun *et al.*, 1995).

Our efforts to calculate accurate intensity patterns center on the goal of maximizing the amount of structural information that can be obtained from solution X-ray scattering experiments. We want to ensure that our WAXS calculations are as accurate as possible so that SAXS/WAXS can be used to characterize changes in structure and structural ensembles as the protein is subjected to different conditions of temperature (Hirai *et al.*, 1999), crowding (Makowski, Rodi, Mandava, Minh *et al.*, 2008), pH or ligand binding (Rodi *et al.*, 2007). Other motivators include the possibility of classifying protein folds from solution scattering experiments (Makowski, Rodi, Mandava, Minh *et al.*, 2008) and following time-resolved changes in molecular structure (Cammarata *et al.*, 2008). Once the calculation of WAXS patterns from molecular structure reaches a sufficiently mature state, coupling of X-ray solution scattering into modeling appears to be a promising next step in its evolution towards helping us learn about biomolecular structure and function in solution (Förster *et al.*, 2008; Gorba *et al.*, 2008; Kojima *et al.*, 2004; Putnam *et al.*, 2007).

The treatment of the excluded volume as a union of spheres, rather than as a sum, is a key difference in our approach. We have demonstrated that the radii of Fraser *et al.* (1978) introduce systematic errors that may be important in interpretation of WAXS data. We were initially surprised to see that scattering patterns calculated using the union definition were so different from those calculated using the popular sum definition. The union-of-volumes excluded-volume definition necessitates the development of new parameterizations for the estimation of WAXS patterns, and in this paper we have illustrated several of the parameterization issues that appear most relevant. Using the solvent-accessible or solvent-excluded surface as the definition of the excluded-volume boundary, as we have done in this work, creates first of all the need to define the radii of the solute atoms, and second the need to define the radius of the probe sphere rolled around the solute atoms. Optimizing the choice of radii on the basis of fit to data does not appear to be a viable strategy as the optimal radii appear to depend on the protein being studied, and the free variations of these parameters may well hide interesting phenomena. Although in the present study we have performed a search using a limited number of proteins, a search employing a larger number of proteins is a subject of ongoing work.

The definition of the hydration layer is similarly problematic. One subject for future work is the formation of a continuum model for the hydration layer based on MD simulations with explicit solvent. It is possible that these simulations may suggest new strategies for developing a hydration-layer model based on physical features specific to the protein under study, such as hydration-layer thickness or contrast. The *SoftWAXS* approach is flexible enough that it is possible to model non-uniform electron densities in the hydration layer. This can be implemented by taking every leaf cube in the hydration shell, estimating the density at each of a number of control points in the leaf cube, and then computing the Fourier transform of a function that interpolates between the control-point densities. Testing and validation of non-uniform hydration-shell densities are the subject of ongoing work. Furthermore, calculations based on MD simulations suggest that there is an upper bound to the accuracy of WAXS calculations that employ continuum theory to model the excluded volume and hydration layer (Park *et al.*, 2009). This bound appears to arise owing to solvent–solvent correlations in the electron density. In effect, the continuum model of the solvent performs the time averaging before the magnitude is squared, whereas the experimental measurement performs the time averaging outside of the magnitude squaring (and subsequent orientational averaging). The fact that these operations do not commute makes it impossible, in general, for a continuum-solvent model to completely match experiment even if all other modeling were exact. Thus, the solute–solvent cross-terms may be playing a significant role in the observed discrepancies, especially at wide angles. For calculation of WAXS patterns it may be important to explicitly include surrounding water molecules and furthermore to appropriately sample the configuration space.

Park *et al.* (2009) demonstrated recently that explicit-solvent MD simulations can match experimental WAXS patterns extremely well, if *excess intensity* is used as a basis for comparison. Lysozyme and myoglobin were kept artificially rigid in the simulations reported in that work. Rigidity is not an unreasonable approximation for these small proteins, as numerous prior studies have shown that calculated scattering matches suprisingly well to experiment for these proteins under the assumption of rigidity. However, it is to be expected that flexibility must be accounted for in estimating scattering from larger monomeric proteins, multi-domain proteins,

oligomeric proteins such as hemoglobin, or proteins with a significant proportion of intrinsically disordered residues (Dunker *et al.*, 2001; Dyson & Wright, 2005).

Finally, the atomic volumes employed for the scattering calculations of Fig. 6 warrant comment because they appear unrealistic. For example, in both parameter sets the C seems to be too large and the N seems too small. The radii that Fraser *et al.* suggested be used were parameterized before the advent of high-resolution crystal structures, and it is theoretically possible that the radii reflect some ill-conditioning in the parameterization process. On the other hand, the union-of-spheres radii were calculated using an exhaustive search and comparison against the experimental WAXS pattern for lysozyme. It should be noted that the two approaches to parameterization give consistent results: in both models, the atoms' excluded volumes are ranked in the same order. Although the physical interpretation of this consistency is not clear, the unrealistic radii appear to provide more evidence that the continuum-solvent model is generally inadequate for WAXS calculations.

6. Conclusions

We have implemented an accurate numerical approach for calculating WAXS patterns of proteins. Our results suggest that, at wide angles, the excluded-volume contribution to scattering cannot be reliably estimated using a continuum representation of water density. Thus, high-quality estimates of WAXS patterns ought to incorporate explicit-solvent detail.

This work was supported by a EUREKA grant from the National Institutes of Health (grant No. R01-GM085648). JB gratefully acknowledges partial support from a Wilkinson Fellowship in Scientific Computing funded by the Mathematical, Information, and Computational Sciences Division Subprogram of the Office of Advanced Scientific Computing Research, Office of Science, US Department of Energy, under contract DE-AC02-06CH11357. SP gratefully acknowledges support from a Director's Fellowship from Argonne National Laboratory. We would also like to acknowledge the insightful suggestions made by an anonymous referee.

References

Altman, M. D., Bardhan, J. P., White, J. K. & Tidor, B. (2009). *J. Comput. Chem.* **30**, 132–153.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.
Cammarata, M., Levantino, M., Schotte, F., Anfinrud, P. A., Ewald, F., Choi, J., Cupane, A., Wulff, M. & Ihee, H. (2008). *Nat. Methods*, **5**, 881–887.
Chacón, P., Morán, F., Díaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys. J.* **74**, 2760–2775.
Connolly, M. L. (1983). *J. Appl. Cryst.* **16**, 548–558.
Diamond, R. (1974). *J. Mol. Biol.* **82**, 371–391.
Doniach, S. (2001). *Chem. Rev.* **101**, 1763–1778.
Dunker, A. K. *et al.* (2001). *J. Mol. Graph. Model.* **19**, 26–59.
Durchschlag, H., Zipper, P., Wilfing, R. & Purr, G. (1991). *J. Appl. Cryst.* **24**, 822–831.
Dyson, H. J. & Wright, P. E. (2005). *Nat. Rev. Mol. Cell Biol.* **6**, 197–208.
Fischetti, R. F., Rodi, D. J., Gore, D. B. & Makowski, L. (2004). *Chem. Biol.* **11**, 1–20.
Fischetti, R. F., Rodi, D. J., Mirza, A., Irving, T. C., Kondrashkina, E. & Makowski, L. (2003). *J. Synchrotron Rad.* **10**, 398–404.
Förster, F., Webb, B., Krukenberg, K. A., Tsuruta, H., Agard, D. A. & Sali, A. (2008). *J. Mol. Biol.* **382**, 1089–1106.
Fraser, R. D. B., MacRae, T. P. & Suzuki, E. (1978). *J. Appl. Cryst.* **11**, 693–694.
Gorba, C., Miyashita, O. & Tama, F. (2008). *Biophys. J.* **94**, 1589–1599.
Head-Gordon, T. & Hura, G. (2002). *Chem. Rev.* **102**, 2651–2670.
Hiragi, Y., Sano, Y. & Matsumoto, T. (2003). *J. Synchrotron Rad.* **10**, 193–196.
Hirai, M., Arai, S. & Iwase, H. (1999). *J. Phys. Chem. B*, **103**, 549–556.
Hirai, M., Iwase, H., Hayakawa, T., Miura, K. & Inoue, K. (2002). *J. Synchrotron Rad.* **9**, 202–205.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Huxley, H. E., Faruqi, A. R., Bordas, J., Koch, M. H. J. & Milch, J. R. (1980). *Nature* (*London*), **284**, 140–143.
Ichiyanagi, K., Sato, T., Nozawa, S., Kim, K. H., Lee, J. H., Choi, J., Tomita, A., Ichikawa, H., Adachi, S., Ihee, H. & Koshihara, S. (2009). *J. Synchrotron Rad.* **16**, 391–394.
Ihee, H. (2009). *Acc. Chem. Res.* **42**, 356–366.
Kim, S. J., Dumont, C. & Gruebele, M. (2008). *Biophys. J.* **94**, 4924–4931.
Koch, M. H. J., Vachette, P. & Svergun, D. I. (2003). *Q. Rev. Biophys.* **36**, 147–227.
Kojima, M., Timchenko, A. A., Higo, J., Ito, K., Kihara, H. & Takahashi, K. (2004). *J. Appl. Cryst.* **37**, 103–109.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Krack, M., Gambirasio, A. & Parrinello, M. (2002). *J. Chem. Phys.* **117**, 9409–9412.
Lattman, E. E. (1989). *Proteins*, **5**, 149–155.
Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Luzzati, V. & Tardieu, A. (1980). *Annu. Rev. Biophys. Bioeng.* **9**, 1–29.
MacKerell, A. D. Jr *et al.* (1998). *J. Phys. Chem. B*, **102**, 3586–3616.
Makowski, L., Rodi, D. J., Mandava, S., Devarapalli, S. & Fischetti, R. F. (2008). *J. Mol. Biol.* **383**, 731–744.
Makowski, L., Rodi, D. J., Mandava, S., Minh, D. D. L., Gore, D. B. & Fischetti, R. F. (2008). *J. Mol. Biol.* **375**, 529–546.
Merzel, F., Fontaine-Vive, F. & Johnson, M. R. (2007). *Comput. Phys. Commun.* **177**, 530–538.
Merzel, F. & Smith, J. C. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 5378–5383.
O'Donnell, J. L., Zuo, X., Goshe, A. J., Sarkisov, L., Snurr, R. Q., Hupp, J. T. & Tiede, D. M. (2007). *J. Am. Chem. Soc.* **129**, 1578–1585.
Park, S., Bardhan, J. P., Roux, B. & Makowski, L. (2009). *J. Chem. Phys.* **130**, 134114.
Pavlov, Yu. M. & Fedorov, B. A. (1983). *Biopolymers*, **22**, 1507–1522.
Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. (1999). *Nat. Struct. Biol.* **6**, 554–556.
Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Q. Rev. Biophys.* **40**, 191–285.
Richards, F. M. (1977). *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
Rodi, D. J., Mandava, S., Gore, D. B., Makowski, L. & Fischetti, R. F. (2007). *J. Biomol. Screen.* **12**, 994–998.
Sanner, M. F. (1996). *Molecular Surfaces Computation*, http://www.scripps.edu/~sanner/html/msms_home.html.

Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). *Biopolymers*, **38**, 305–320.
Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1999). *J. Mol. Biol.* **288**, 489–499.
Seki, Y., Tomizawa, T., Khechinashvili, N. N. & Soda, K. (2002). *Biophys. Chem.* **95**, 235–252.
Semenyuk, A. V. & Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 537–540.
Sharp, K. A. & Honig, B. (1990). *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301–332.
Soda, K., Miki, Y., Nishizawa, T. & Seki, Y. (1997). *Biophys. Chem.* **65**, 45–53.
Sokolova, A. V., Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 865–868.
Sorenson, J. M., Hura, G., Glaeser, R. M. & Head-Gordon, T. (2000). *J. Chem. Phys.* **113**, 9149–9161.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
Svergun, D. I., Volkov, V. V., Kozin, M. B., Stuhrmann, H. B., Barberato, C. & Koch, M. H. J. (1997). *J. Appl. Cryst.* **30**, 798–802.
Tiede, D. M., Zhang, R. & Seifert, S. (2002). *Biochemistry*, **41**, 6605–6614.
Tsutakawa, S. E., Hura, G. L., Frankel, K. A., Cooper, P. K. & Tainer, J. A. (2007). *J. Struct. Biol.* **158**, 214–223.
Vachette, P., Koch, M. H. J. & Svergun, D. I. (2003). *Methods Enzymol.* **374**, 584–615.
Walther, D., Cohen, F. E. & Doniach, S. (2000). *J. Appl. Cryst.* **33**, 350–363.
Weiss, T. M., Narayanan, T., Gradzielski, M. C. W., Panine, P., Finet, S. & Helby, W. I. (2005). *Phys. Rev. Lett.* **94**, 038303.
Williamson, T. E., Craig, B. A., Kondrashkina, E., Bailey-Kellogg, C. & Friedman, A. M. (2008). *Biophys. J.* **94**, 4906–4923.
Wriggers, W. & Chacón, P. (2001). *J. Appl. Cryst.* **34**, 773–776.
Zhang, R., Thiyagarajan, P. & Tiede, D. M. (2000). *J. Appl. Cryst.* **33**, 565–568.
Zheng, W. & Doniach, S. (2005). *Protein Eng. Des. Sel.* **18**, 209–219.
Zhou, J., Deyhim, A., Krueger, S. & Gregurick, S. K. (2005). *Comput. Phys. Commun.* **170**, 186–204.