Pepsi-SAXS : an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles

Sergei Grudinin, Maria Garkavenko, Andrei Kazennov

STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles

Sergei Grudinin, Maria Garkavenko and Andrei Kazennov

Acta Cryst. (2017). D**73**, 449–464

STRUCTURAL BIOLOGY
ISSN 2059-7983

Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles

Sergei Grudinin,[a,b,c]* Maria Garkavenko[d] and Andrei Kazennov[d]

[a]Université Grenoble Alpes, LJK, F-38000 Grenoble, France, [b]CNRS, LJK, F-38000 Grenoble, France, [c]Inria, France, and [d]Moscow Institute of Physics and Technology, Dolgoprudniy, Russian Federation. *Correspondence e-mail: sergei.grudinin@inria.fr

Received 1 July 2015
Accepted 15 April 2017

Edited by S. Wakatsuki, Stanford University, USA

Keywords: *Pepsi-SAXS*; small-angle scattering; multipole expansion.

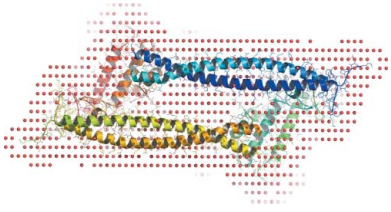
Supporting information: this article has supporting information at journals.iucr.org/d

A new method called *Pepsi-SAXS* is presented that calculates small-angle X-ray scattering profiles from atomistic models. The method is based on the multipole expansion scheme and is significantly faster compared with other tested methods. In particular, using the Nyquist–Shannon–Kotelnikov sampling theorem, the multipole expansion order is adapted to the size of the model and the resolution of the experimental data. It is argued that by using the adaptive expansion order, this method has the same quadratic dependence on the number of atoms in the model as the Debye-based approach, but with a much smaller prefactor in the computational complexity. The method has been systematically validated on a large set of over 50 models collected from the BioIsis and SASBDB databases. Using a laptop, it was demonstrated that *Pepsi-SAXS* is about seven, 29 and 36 times faster compared with *CRYSOL*, *FoXS* and the three-dimensional Zernike method in *SAStbx*, respectively, when tested on data from the BioIsis database, and is about five, 21 and 25 times faster compared with *CRYSOL*, *FoXS* and *SAStbx*, respectively, when tested on data from SASBDB. On average, *Pepsi-SAXS* demonstrates comparable accuracy in terms of χ^2 to *CRYSOL* and *FoXS* when tested on BioIsis and SASBDB profiles. Together with a small allowed variation of adjustable parameters, this demonstrates the effectiveness of the method. *Pepsi-SAXS* is available at http://team.inria.fr/nano-d/software/pepsi-saxs.

1. Introduction

Small-angle scattering is one of the fundamental techniques for the structural study of biological systems. Small-angle X-ray scattering (SAXS) is a type of small-angle scattering in which X-rays are scattered elastically from the sample and are then collected at very small angles. Compared with other structure-determination methods, SAXS experiments are very simple conceptually, and thanks to advances in instrumentation (Spilotros & Svergun, 2014), the SAXS technique, particularly solution-state SAXS, has become very popular in recent years as a complement to other methods in structural biology (Graewert & Svergun, 2013; Putnam *et al.*, 2007). SAXS also allows some of the restrictions of other experimental techniques to be overcome; for example, it is applicable to systems of all sizes, it allows the study of particles in solution, it is very fast and it destroys the sample only marginally. On the downside, SAXS can only determine the distance distribution function of the electron density at a resolution above nanometres; however, it can distinguish conformations of a protein at subnanometre resolution (Zheng & Tekpinar, 2011).

Over the years, a number of computational tools have been developed for the analysis of solution-state SAXS curves,



calculation of theoretical profiles and low-resolution reconstruction of model shapes. The most prominent of them is the *ATSAS* package developed at EMBL Hamburg (Petoukhov *et al.*, 2012). To test a structural hypothesis or to construct a model system based on a SAXS experiment, an accurate and rapid calculation of a model SAXS profile is required. The running time of a computational method depends, among other things, on the number of atoms in the model N and the number of points in the scattering curve M. Tools that directly use the Debye equation have a cost of $O(N^2)$, whereas tools that use a linear approximation to the scattering equation have a cost of $O(N)$. Generally speaking, the same type of calculation should be repeated for each point in the scattering curve, which determines the worst-case performance as $O(N^2M)$. Keeping in mind that the typical values of M and N are several thousand, this running time usually prohibits the performance of any kind of multiple model assessment. Thus, much effort has been devoted in recent years to reducing the running time of SAXS computational tools without degrading the quality of their approximations. Below, we give a brief overview of the most notable computational methods for the calculation of theoretical SAXS profiles given an atomic model as input. A deeper discussion of different computational techniques can be found elsewhere (Rambo & Tainer, 2013*b*).

The most popular method is the *CRYSOL* program developed by Svergun and coworkers (Svergun *et al.*, 1995). This method uses the theory of multipole expansions of scattering intensity initially developed by Stuhrmann (Stuhrmann, 1970*b*). The running time of the initial implementation of the method had a linear dependence on both the number of atoms in the molecule N and the number of points in the scattering curve M as $O(NM)$. A more recent version of the program, however, maps experimental scattering intensities and associated errors onto a sparser grid (Petoukhov *et al.*, 2012), thus reducing the computational cost to $O(N + M)$. The method is generally very fast, but has the major disadvantage of a simplistic representation of the hydration shell of the sample using a two-dimensional angular function (Stuhrmann, 1970*a*). The *SASSIM* method is very similar to *CRYSOL*, but the hydration shell is defined in terms of spherical harmonics and is calculated using a Lebedev grid (Merzel & Smith, 2002).

Another popular program, *FoXS*, uses a linear approximation to the Debye scattering equation, which decouples the dependency of the running time on the number of atoms N in a model and the number of points in the scattering curve M as $O(N^2 + MN)$ (Schneidman-Duhovny *et al.*, 2010, 2013). When created, the program was notably faster compared with the initial implementation of *CRYSOL* when tested on experimental curves with several thousands of points. However, the later development of *CRYSOL*, as we demonstrate below, outperforms *FoXS* for nearly all test cases.

A logical extension of the multipole expansion method is a computational scheme that uses three-dimensional Zernike polynomials for the representation of the electron density (Liu, Poon *et al.*, 2012). Here, the angular dependence of the scattering amplitudes on the scattering vector is described, similarly to that in *CRYSOL*, using spherical harmonics, but the radial dependence is expanded using a set of orthogonal functions. The computational complexity of this method is $O(N + M)$; however, the hidden time-limiting step is the computation of the three-dimensional Zernike moments. In order to calculate them, atomic models are mapped onto a three-dimensional grid, the size of which can be adjusted according to the resolution of the data.

Recently, some other linear scaling schemes have been proposed. The golden-ratio scheme described by Watson & Curtis (2013) uses Euler's formula to compute the rotationally averaged scattering intensity $I(q)$ by evaluating $I(\mathbf{q})$ in several scattering directions using the exact expression for $I(\mathbf{q})$ at a given wavevector $\mathbf{q}$. The orientations of the $\mathbf{q}$ vectors are taken from a quasi-uniform spherical grid generated by the golden ratio. The hierarchical algorithm for fast summation of the Debye equation by Gumerov *et al.* (2012) is similar to the fast multipole method (FMM) and is based on a hierarchical spatial decomposition of electron density using local harmonic expansions and translation operators for these expansions. Its computational cost is $O(N\log N)$.

Some efforts have been made to obtain a more precise description of solvation. The *AXES* method uses explicit water molecules equilibrated in a water box using molecular-dynamics (MD) simulations to accurately model the scattering amplitudes of the surface and displaced solvent (Grishaev *et al.*, 2010). Another method calculates hydration-shell intensities from MD trajectories of water molecules around a fixed protein (Park *et al.*, 2009). The *AquaSAXS* method models the non-uniform hydration shell of a protein by taking advantage of recently developed methods that compute the solvent distribution around a given solute on a three-dimensional grid, such as the Poisson–Boltzmann–Langevin formalism or the three-dimensional reference interaction-site model (Poitevin *et al.*, 2011).

Finally, to increase the speed of calculations, several coarse-grained schemes have been proposed. For example, one recent method is based on the Debye formula and a set of scattering form factors for dummy-atom representations of amino acids (Stovgaard *et al.*, 2010). The *Fast-SAXS-pro* (Yang *et al.*, 2009) algorithm uses the Debye-based approach and coarse-grained residue-level and nucleotide-level structure factors. The method explicitly takes into account the nonhomogeneous distribution within the hydration layer by assigning a different scaling factor for dummy water molecules according to their proximity to protein and DNA/RNA. Finally, the method of Zheng & Tekpinar (2011) uses a one-bead-per-residue coarse-grained protein representation coupled with the elastic network model. The hydration shell is modelled implicitly by combining each residue and its nearby implicit water molecules into a composite representation.

Here, we present *Pepsi-SAXS* (where 'Pepsi' stands for 'polynomial expansions of protein structures and interactions'), a new implementation of the multipole-based scheme proposed by Stuhrmann (Stuhrmann, 1970*b*). Overall, our method is significantly faster compared with *CRYSOL*, *FoXS* and the three-dimensional Zernike implementation in

Table 1
Crystallographic distances between heavy atoms and their attached H atoms (Allen *et al.*, 2004).

Various atomic groups typical for biological molecules are listed.

Atomic group	Distance to the H atom (Å)
C(sp^3)—H	1.099
C(sp^3)—H_2	1.092
C(sp^3)—H_3	1.059
C(sp^2)—H	1.077
C(arom)—H	1.083
O(alc)—H	0.967
O(acid)—H	1.015
N—H	1.009
N^+—H	1.033

the *SAStbx* package (Liu, Hexemer *et al.*, 2012), as we demonstrate below using a large number of test cases. Our method has the following features. Firstly, we use a very fast model for computation of the hydration shell based on a uniform grid of points. Secondly, we use the adaptive order of the multipole expansion. More precisely, according to the Nyquist–Shannon–Kotelnikov sampling theorem (Marks, 2008), we determine the required expansion order using the radius of gyration of the model's hydration shell and the value of the maximum scattering vector $\mathbf{q}_{\max}$. Thirdly, we represent the scattering intensity curve using a cubic spline interpolation, which allows us to significantly speed up the running time of our method. Finally, we introduce partial scattering intensities to rapidly fit the theoretical curve to the experimental curve using an exhaustive search in two adjustable parameters. We should also mention that we pay particular attention when deriving parameters for the form factors, especially those for charged and resonance groups.

2. Theory

Here, we follow the scattering theory initially described by H. B. Stuhrmann and later revised by D. Svergun (Stuhrmann, 1970*b*; Svergun, 1991; Svergun *et al.*, 1995). The spherically averaged scattering intensity $I(q)$ from a single molecule immersed in a solvent with bulk scattering density ρ can be written as

$$I(q) = \langle |A_a(\mathbf{q}) - \rho A_c(\mathbf{q}) + \delta\rho A_b(\mathbf{q})|^2 \rangle_\Omega, \quad (1)$$

where $A_a(\mathbf{q})$ is the scattering amplitude from the molecule in vacuum, $A_c(\mathbf{q})$ is the scattering amplitude from the excluded volume and $A_b(\mathbf{q})$ is that from the hydration shell, which is assumed to have a scattering density that differs from the bulk value by $\delta\rho$ (Svergun *et al.*, 1995). We should mention that throughout the paper we use the following definition of the scattering vector q: $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the wavelength. Owing to the spherical averaging of the intensity, it is very convenient to introduce the multipole expansion of the scattering intensities and amplitudes in the spherical coordinates system (Stuhrmann, 1970*b*). Using this expansion up to the maximum expansion order L, we can rewrite the intensity as

$$I(q) \simeq \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(q) - \rho C_{lm}(q) + \delta\rho B_{lm}(q)|^2, \quad (2)$$

where $A_{lm}(q)$, $B_{lm}(q)$ and $C_{lm}(q)$ are the expansion coefficients of the amplitudes $A_a(\mathbf{q})$, $A_b(\mathbf{q})$ and $A_c(\mathbf{q})$, respectively (Svergun *et al.*, 1995). Given the atomic coordinates of a molecule consisting of N atoms expressed in the spherical coordinate system $\mathbf{r}_i \equiv (r_i, \omega_i)$ and the corresponding form factors $f_i(q)$, we can write the vacuum scattering-amplitude expansion coefficients as

$$A_{lm}(q) = 4\pi i^l \sum_{i=1}^{N} f_i(q) j_l(qr_i) Y^*_{lm}(\omega_i), \quad (3)$$

where $j_l(qr_i)$ are the spherical Bessel functions and $Y^*_{lm}(\omega_i)$ are the complex-conjugated spherical harmonics. Similarly, given the coordinates of the hydration shell of the molecule sampled at $N_{\rm hs}$ points, its expansion coefficients can be written as

$$B_{lm}(q) = 4\pi i^l h(q) \sum_{i=1}^{N_{\rm hs}} j_l(qr_i) Y^*_{lm}(\omega_i), \quad (4)$$

where $h(q)$ is the form factor of a water molecule scaled with the ratio of the bulk water density to the density of the sampling points in the hydration shell. Finally, the excluded volume contribution can be written as

$$C_{lm}(q) = 4\pi i^l \sum_{i=1}^{N} g_i(q) j_l(qr_i) Y^*_{lm}(\omega_i), \quad (5)$$

where $g_i(q)$ are the form factors of the dummy atoms centred at the positions of molecular atoms $\mathbf{r}_i$.

2.1. Form factors and unified atomic groups

Computation of the expansion coefficients $A_{lm}(q)$, $B_{lm}(q)$ and $C_{lm}(q)$ requires knowledge of the form factors $f_i(q)$, $g_i(q)$ and $h_i(q)$. For the calculation of form factors for the individual atoms, we use the five-Gaussian approximation with coefficients taken from Waasmaier & Kirfel (1995):

$$f(q) = c + \sum_{i=1}^{5} a_i \exp(-b_i q^2). \quad (6)$$

However, structural databases such as the Protein Data Bank (PDB; Berman *et al.*, 2000) typically provide the coordinates of only non-H atoms. Therefore, it is useful to introduce unified atomic groups with the positions located at the centres of the nuclei of heavy atoms and the corresponding scattering parameters computed for the heavy atoms with covalently bonded H atoms. For example, the form factor for such a group $f_{\rm CH_n}$, with n H atoms attached to a C heavy atom, can be computed using the Debye equation as follows,

$$f_{\rm CH_n}(q)^2 = f_{\rm C}(q)^2 + n^2 f_{\rm H}(q)^2 + 2n f_{\rm H}(q) \frac{\sin(qr_{\rm H})}{qr_{\rm H}}, \quad (7)$$

where $f_{\rm C}$ and $f_{\rm H}$ are the atomic form factors for C and H atoms given by the five-Gaussian approximation (6), and $r_{\rm H}$ is the distance between C and H atoms. The distances $r_{\rm H}$ between the heavy atom and H atoms in various atomic groups typical for biological molecules are taken from Allen *et al.* (2004) and are listed in Table 1. We should note that a simpler

Table 2
Coefficients of the five-Gaussian approximation as given by (6) for different unified atomic groups.

The approximation was calculated according to (8) at 950 points for values of q in the range (0, 0.95) using the nonlinear least-squares Levenberg–Marquardt algorithm. The standard deviations of the approximations do not exceed 10^{-4} for all of the form factors. The unified atomic groups include sp^2, sp^3 and aromatic C atoms with one or several H atoms attached, carboxylate or phosphate resonance O atoms, O atoms and S atoms with an attached H atom, neutral and charged N atoms with one or several H atoms attached, and the N atoms of the guanidine group.

	C(sp^3)—H	C(sp^3)—H_2	C(sp^3)—H_3	C(sp^2)—H	C(arom)—H	O(alc)—H	O(acid)—H	O(resonance)
a_1	2.909530	3.275723	3.681341	2.909457	2.168070	0.456221	3.213280	0.688944
a_2	0.485267	0.870037	1.228691	0.484873	1.275811	3.219608	0.463019	2.929687
a_3	1.516151	1.534606	1.549320	1.515916	1.561096	0.812773	0.815724	0.416472
a_4	0.206905	0.395078	0.574033	0.207091	0.742395	2.666928	2.664450	2.606983
a_5	1.541626	1.544562	1.554377	1.541518	−6.151144	1.380927	1.384266	1.319232
b_1	13.933084	13.408502	13.026207	13.934162	12.642907	21.503498	13.383078	29.319200
b_2	23.221524	23.785175	24.131974	23.229153	18.420069	13.397134	21.362223	6.572228
b_3	41.990403	41.922444	41.869426	41.991425	41.768517	34.547137	34.531415	64.951658
b_4	4.974183	5.019072	4.984373	4.983276	1.535360	5.826620	5.823549	16.267799
b_5	0.679266	0.724439	0.765769	0.679898	−0.045937	0.412902	0.410805	0.455640
c	0.337670	0.377096	0.409294	0.338296	7.400917	0.463202	0.458919	0.537548

	N—H	N—H_2	N—H^+	N—H_2^+	N—H_3^+	(N—H) (guanine)	(N—H_2) (guanine)	S—H
a_1	1.650531	1.904157	1.426540	3.823896	1.882162	3.630164	1.792216	0.570042
a_2	0.429639	1.942536	0.426903	0.531490	1.933200	0.228310	0.724464	6.337416
a_3	2.144736	2.435585	1.878894	1.713620	2.465843	1.869734	2.347044	1.641643
a_4	1.851894	0.730512	1.608251	0.322552	0.927311	0.170550	1.903020	5.398549
a_5	1.408921	1.379728	1.200216	1.287502	1.190889	1.440894	1.313042	1.527982
b_1	10.603730	10.803702	10.652268	10.305028	10.975157	10.267139	10.830060	11.447986
b_2	6.987283	10.792421	7.017651	25.631593	10.956008	25.118086	6.846763	1.197657
b_3	29.939901	29.610479	29.878525	30.215026	29.208572	30.241288	29.579607	55.401032
b_4	10.573859	6.847755	10.619493	3.576178	6.663555	3.412776	10.800018	22.420955
b_5	0.611678	0.709687	0.631765	0.506824	0.843650	0.486644	0.720448	2.356552
c	0.510589	0.603738	0.456024	0.317728	0.597322	0.323504	0.583312	1.523944

approximation holds for practical values of the scattering vector q (Harker, 1953),

$$f_{CH_n}(q) = f_C(q) + nf_H(q)\frac{\sin(qr_H)}{qr_H}, \tag{8}$$

which can also be derived from spherical averaging of the scattering amplitudes instead of the scattering intensities.

We explicitly introduced individual form factors for the charged groups carboxylate, phosphate, guanidine and ammonium. Form factors for NH^+, NH_2^+ and NH_3^+ from guanidine and ammonium groups were approximated according to the model of the electron distribution in the ammonium ion (Banyard & March, 1961). More specifically, we modelled the central spherical charge cloud with six electrons around the N nucleus with unperturbed H electron distributions centred not on the protons but inwards along the N—H bonds at 0.76 of the N—H separation distance. We also paid particular attention to the resonance forms of the charged groups carboxylate, phosphate and guanidine. More precisely, we modelled the form factors of the resonance groups as a linear combination of the nonresonance form factors. Given the analytic form of the atomic form factors for unified atomic groups (8), we computed their five-Gaussian approximations, which were tabulated for later use. Table 2 lists the obtained coefficients.

2.2. Form factors for dummy atoms

Following Fraser *et al.* (1978), we express the form factors of the dummy atoms through the observed displaced solvent volumes V_i as

$$g_i(q) = V_i \exp(-\pi q^2 V_i^{2/3}). \tag{9}$$

Following Svergun *et al.* (1995), we introduce the effective atomic radius r_0, an adjustable parameter that scales the observed displaced solvent volumes according to

$$V_i(r_0) = \frac{4}{3}\pi r_i^3 \frac{r_0^3}{r_m^3}, \tag{10}$$

where r_i are the tabulated actual values of atomic group radii and r_m are the actual average radii of atomic groups. Changing the adjustable parameter to $\delta r \equiv r_0 - r_m$, we can expand the previous expression to the first order in δr using the Maclaurin series as

$$g_i(q, \delta r) = V_i \exp(-\pi q^2 V_i^{2/3})\left[1 + \frac{\delta r}{r_m}(3 - 2\pi q^2 V_i^{2/3})\right] + O(\delta r^2). \tag{11}$$

This equation can be further simplified to the form of expressions (12) and (13) from Svergun *et al.* (1995) as

$$g_i(q, \delta r) = V_i \exp(-\pi q^2 V_i^{2/3})\left\{1 + \frac{\delta r}{r_m}\left[3 - \left(\frac{4\pi}{3}\right)^{2/3} 2\pi q^2 r_m^2\right]\right\} + O(\delta r^2), \tag{12}$$

with the term independent of δr being the reference dummy-atoms form factor $g_i(q)$ and the term in the curly brackets being the adjustable overall expansion factor $G(q, \delta r)$. We can even simplify the second term further, dropping the dependence on q. Using the last expression, the excluded volume

amplitudes $C_{lm}(q, \delta r)$ can be adjusted through the reference values $C_{lm}(q)$ as

$$C_{lm}(q, \delta r) = C_{lm}(q)G(q, \delta r), \quad (13)$$

where the reference amplitudes $C_{lm}(q)$ are computed only once using the reference dummy-atoms form factors $g_i(q)$. To compute the excluded volumes and radii of the unified atomic groups, we used the parameters provided in Svergun *et al.* (1995).

2.3. Hydration shell

To compute the scattering contribution of the hydration shell of the molecule (2), we first constructed its grid approximation using the linked-cell approach (Artemova *et al.*, 2011). More precisely, we constructed a grid with a cell size of 3–4 Å padded by at least 12 Å in each direction, and associated each atom of the molecule with a cell in the grid. We then removed those grid cells whose centres were closer than 3 Å to any atom within the corresponding cell and its 26 direct neighbours or further than 3 Å plus the width of the shell from all of these atoms. Finally, we used the centres of the remaining cells as a grid approximation of the hydration shell. For the width of the shell, we adopted a value of 3 Å for molecules with a radius of gyration smaller than 15 Å and a value of 5 Å for molecules with a radius of gyration larger than 20 Å, and we used a linear interpolation between these values. The width value of 5 Å is somewhat larger compared with what is usually assumed to be the width of the hydration shell. However, our numerical experiments demonstrated the least overfitting of the experimental data with this value. We should mention that the actual effect of the hydration shell depends not only on its width but also on its contrast. Thus, the critical parameter that defines the potential overfitting is the product of the width of the hydration shell with its maximum contrast. In our case, this parameter equals 0.167 e Å^{-2}, which is smaller than, for example, the value of 0.180 e Å^{-2} used in *CRYSOL*. We have also experimented with a lower resolution of the grid representation by decreasing the linear density of grid points by a factor of two. This did not demonstrate any significant change in the quality of fitting of the modelled profiles to the experimental data; however, the execution time improved by about 10% on average and we thus optionally provide this possibility to the user with the `-fast` flag. Fig. 1 shows an example of our hydration-shell model.

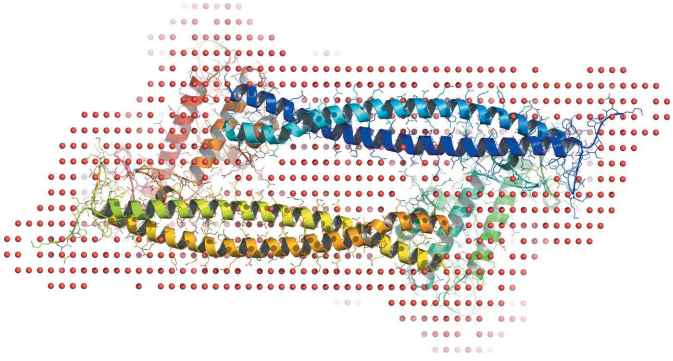


Figure 1
Grid representation of the hydration shell with a resolution of 4 Å for the SASDAW3 model from the SASBDB database. Red dots represent the positions of the sampled points in the hydration shell. The effective width of the shell in this case is 5 Å.

2.4. Adaptivity

We adapt the maximum expansion order L of the multipole expansion according to the radius of gyration of the hydration shell R_g and the maximum scattering vector of the experimental curve q_{max}. More precisely, we can estimate the value of L from the Nyquist–Shannon–Kotelnikov sampling theorem (Marks, 2008), which defines the angular resolution of encoding with complex spherical harmonics of order L to be $2\pi/L$. On the other hand, the spatial resolution of the experimental data is $R = 2\pi/q_{max}$; thus, we can relate the two resolutions using the radius of gyration R_g as

$$L = 2\pi\frac{R_g}{R} = R_g q_{max}. \quad (14)$$

This expression provides the default value of the maximum expansion order for our method. We use the same idea to approximate the radial functions in (3), (4) and (5). There, the radial basis set is given by the spherical Bessel functions of maximum order L. Therefore, we sample expansion coefficients $A_{lm}(q)$, $B_{lm}(q)$ and $C_{lm}(q)$ at $2L$ equidistant points and then use the cubic spline interpolation (Press *et al.*, 2007) to reconstruct the values of the expansion coefficients at any point q.

2.5. Fitting

If the experimental curve $I_{exp}(q)$ is provided, we adjust the two parameters δr and $\delta\rho$ such that the goodness of fit χ^2 is minimized,

$$\chi^2 = \frac{1}{N-1}\sum_j\left[\frac{I_{exp}(q_j) - cI_{theor}(q_j)}{\sigma(q_j)}\right]^2, \quad (15)$$

where N is the number of points in the experimental curve, $\sigma(q)$ are the experimental errors, $I_{theor}(q)$ is the theoretical intensity calculated according to (2) and c is the scaling factor given as in Svergun *et al.* (1995),

$$c = \left[\sum_j \frac{I_{exp}(q_j)I_{theor}(q_j)}{\sigma(q_j)^2}\right] \Big/ \left[\sum_j \frac{I_{theor}(q_j)^2}{\sigma(q_j)^2}\right]. \quad (16)$$

If the errors are not provided, we model them as $\sigma(q) = 0.01 \times I_{exp}(q)$. To speed up the calculation of the theoretical scattering-intensity curve at different values of δr and $\delta\rho$, we rewrite it as a sum of partial intensities, as shown in the Supporting Information. This allows us to reduce the computational cost of the theoretical scattering-intensity curve by a factor of $O(L^2)$. We assume the bulk scattering density ρ to be constant and equal to 334 e nm^{-3}. We then exhaustively search for the optimal values of the δr and $\delta\rho$ parameters on a grid of size 100 × 100. The values of δr are searched in the range $-0.05 \leq \delta r/r_m \leq 0.05$. This effectively means $0.95r_m \leq r_0 \leq 1.05r_m$, with a mean r_m value over our data set of 1.64 Å. The range of values of $\delta\rho$ is

Table 3
Comparison of four methods, *CRYSOL*, *FoXS*, *SAStbx* (using the three-dimensional Zernike technique and data-reduction option) and *Pepsi-SAXS*, when fitting modelled intensity profiles to experimental data collected from the BioIsis database.

For each method, we provide the value of χ and the running time measured in seconds for each of the scattering profiles. We also list the number of atoms in the models along with the average values of χ and running time.

			CRYSOL		*FoXS*		*SAStbx*		*Pepsi-SAXS*	
Structure	BioIsis ID	No. of atoms	χ	Time (s)	χ	Time (s)	χ	Time (s)	χ	Time (s)
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	6395	1.98	0.84	1.66	2.61	0.93	5.71	1.51	0.19
Abscisic acid receptor PYR1	BID_1PYR1P	2924	1.39	0.61	2.03	0.74	2.80	2.25	2.33	0.04
Rubredoxin	BID_1RBDGP	424	7.61	0.46	7.05	0.12	0.14	1.01	7.24	0.03
Superoxide reductase	BID_1SPXGP	4060	3.40	0.66	4.62	1.10	7.37	2.41	1.67	0.05
Monomeric PF1674	BID_1TSPHP	1381	6.28	0.57	7.96	0.31	21.28	1.97	6.22	0.03
Endo-1,4-β-xylanase II	BID_1XYNTP	1480	0.99	0.52	1.03	0.33	1.07	1.41	0.93	0.03
28 bp DNA	BID_28BPDD	1107	0.49	0.51	0.79	0.36	0.64	1.83	0.56	0.03
S-Adenosylmethionine riboswitch mRNA	BID_2SAMRR	2086	2.42	0.56	2.46	0.48	2.63	1.90	2.46	0.04
Superoxide dismutase	BID_APSODP	2229	3.45	0.60	3.66	0.56	6.85	1.91	3.58	0.06
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	5318	2.54	0.74	2.16	1.76	2.74	4.02	2.19	0.06
Complement C3b-Efb (from *S. aureus*)	BID_C3BEFP	12833	2.41	1.18	1.75	7.49	2.86	10.67	1.63	0.19
Complement C3b + Efb (staphylococcal)	BID_C3BSAP	12255	0.12	1.11	0.12	6.85	0.26	8.18	0.12	0.18
Glucose isomerase	BID_GIKCLP	12176	7.83	1.09	7.27	6.15	26.27	5.36	7.34	0.27
Glucose isomerase	BID_GISRUP	12176	7.99	1.09	4.69	6.15	22.42	5.70	3.38	0.28
Human regulator of chromosome condensation (RCC1)	BID_HRCC1P	3158	1.29	0.59	1.77	0.85	1.52	2.72	1.59	0.05
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1633	1.44	0.51	1.49	0.41	3.08	2.00	1.85	0.04
Lysozyme	BID_LYKCLP	1394	9.39	0.54	7.74	0.29	4.67	1.34	9.41	0.04
Hen egg-white lysozyme	BID_LYSOZP	1001	2.54	0.49	2.56	0.22	2.23	1.21	2.62	0.03
MnmE in the nucleotide-free state	BID_MNME1P	6518	0.89	0.84	0.88	2.46	0.72	4.50	0.85	0.18
E. coli MnmE–MnmG complex in the nucleotide-free state	BID_MnmEGP	16291	1.86	1.30	1.87	10.96	2.12	10.59	1.84	0.15
A. aeolicus MnmG	BID_MnmG1P	10534	1.40	0.99	1.50	5.07	2.12	8.94	1.44	0.10
A. aeolicus MnmG + tRNA	BID_MnmG2X	11184	1.70	1.03	1.71	5.66	5.71	7.32	1.75	0.19
E. coli MnmG + NbMnmG1	BID_MnmG3P	11562	1.79	1.06	2.06	5.99	1.79	7.89	2.05	0.35
E. coli MnmE–MnmG complex bound to GDP-AlF$_x$	BID_MnmGEP	23149	3.16	1.70	2.69	21.72	9.46	5.94	2.87	0.35
DNA double-strand break repair protein MRE11	BID_MRERAP	12148	0.72	1.07	1.19	6.41	3.81	7.73	1.29	0.15
Cu/Zn superoxide dismutase	BID_NMSODP	2309	1.04	0.55	1.05	0.56	0.91	1.94	0.97	0.04
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	3760	0.10	0.63	0.10	1.05	0.13	2.92	0.11	0.05
Ketoreductase-enoylreductase didomain	BID_ZGDWKP	5505	2.00	0.70	2.49	1.75	1.18	3.38	1.61	0.04
Average		6678	2.79	0.81	2.73	3.51	4.92	4.38	2.55	0.12

$0 \leq \delta\rho \leq 33.4$ e nm^{-3}. We should note that upon request from the user we allow the contrast of the hydration shell $\delta\rho$ to be slightly negative up to -15 e nm^{-3}. Indeed, as has been demonstrated by X-ray diffraction, neutron and, more recently, X-ray reflectivity studies of water–hydrophobic interfaces, there is an unambiguous and distinguishable density-depleted interfacial region near hydrophobic interfaces (Iiyama *et al.*, 1995; Mezger *et al.*, 2011; Chattopadhyay *et al.*, 2010; Uysal *et al.*, 2013). At these interfaces, the water density drops below the bulk value. There is, however, a certain controversy about the width and the density of this depletion region (Uysal *et al.*, 2013). We should admit that protein surfaces are never fully hydrophobic. Nonetheless, we allow negative $\delta\rho$ values upon request from the user. Below, we report the results for the two cases. We should also note that some experimental measurements have a systematic error in the determination of the intensity values. To account for this error, we can optionally introduce the offset constant κ and rewrite the goodness of fit as shown in the Supporting Information.

2.6. Benchmarks

We tested our methods using two benchmarks constructed from structural models with corresponding experimental SAXS profiles. We collected experimental data from two large databases dedicated to the study of biological molecules using SAXS experiments. The first database is BioIsis, which was designed by Dr Robert P. Rambo at the Lawrence Berkeley National Laboratory (Hura *et al.*, 2009). It currently contains 99 SAXS scattering profiles of biological molecules and their complexes with both known and unknown structure. The first entry in BioIsis is dated 2009. The second database is the Small Angle Scattering Biological Data Bank (SASBDB), powered by the Biological Small Angle Scattering Group at European Molecular Biology Laboratory Hamburg Outstation (Valentini *et al.*, 2015). This database contains 125 scattering profiles. The first data for SASBDB were collected in 1998. For our tests, we collected all of the experimental scattering profiles from the two databases that had corresponding atomic models. Overall, we use 28 entries from BioIsis and 23 entries from SASBDB. The models from BioIsis range from 424 to 23 149 atoms, with an average of 6676 atoms. The models from SASBDB range from 602 to 25 761 atoms, with an average of 6443 atoms.

2.7. Details of implementation

The presented method was implemented using the C++ programming language and compiled with the gcc-4.8

compiler on Linux, the clang compiler on Mac OS and the MSVC compiler on Windows systems. To speed up computations of the expansion coefficients in (3), (4) and (5), we use single instruction–multiple data (SIMD) instructions when possible. We also use multi-threaded computations for the evaluation of the expansion coefficients, as well as for the fitting procedure, if multiple CPU cores are available.

The test benchmarks were run on a MacBook Pro Mid 2015 laptop with a 2.8 GHz Intel Core i7 processor and 16 GB 1600 MHz DDR3 RAM. *Pepsi-SAXS* can optionally provide the output formatted using JSON, and change the initially guessed angular units of the experimental profile. On demand from the user, we allow negative contrast of the hydration shell using the `-neg` flag. We also provide a coarser representation of the hydration shell with the `-fast` flag, which also improves the execution time by about 10%. By default, the maximum scattering angle is set to 0.5 Å^{-1}. The user can change it using the `-ms` flag. Finally, the user can optionally require fitting of the experimental profile with constant background noise using the `-cst` flag.

3. Results

To demonstrate the speed and accuracy of the present method, we conducted seven numerical experiments using large amounts of experimental data. In the experiments, we compared the performance of *Pepsi-SAXS* with those of three widely used methods: *CRYSOL* v.2.8.2 (Svergun *et al.*, 1995; Petoukhov *et al.*, 2012), *FoXS* (Schneidman-Duhovny *et al.*, 2010) and *SAStbx* (Liu, Hexemer *et al.*, 2012). We should note that *SAStbx* provides implementations of three different methods, but we have specifically chosen the novel three-dimensional Zernike technique, with the 'data_reduct' and 'solvent_scale' options set to 'true'. We did not use more computational methods for the comparison because a recent study of the *FoXS* method (Schneidman-Duhovny *et al.*, 2013) demonstrated an advantage in speed and accuracy of *FoXS* and *CRYSOL* over other tested programs.

In the first series of tests, we aimed to compare the four methods using the data from the BioIsis database. More precisely, we measured the goodness of fit for the modelled intensities to the experimental SAXS profiles (15) and the corresponding timings. Table 3 lists the results of the tests. *Pepsi-SAXS* outperforms the other methods in running time for all of the test profiles. On average, *Pepsi-SAXS* is about seven times faster compared with *CRYSOL*, and 29 and 36 times faster compared with *FoXS* and *SAStbx*, respectively. As would be expected, for small molecules the difference in running time between *Pepsi-SAXS* and *CRYSOL* becomes

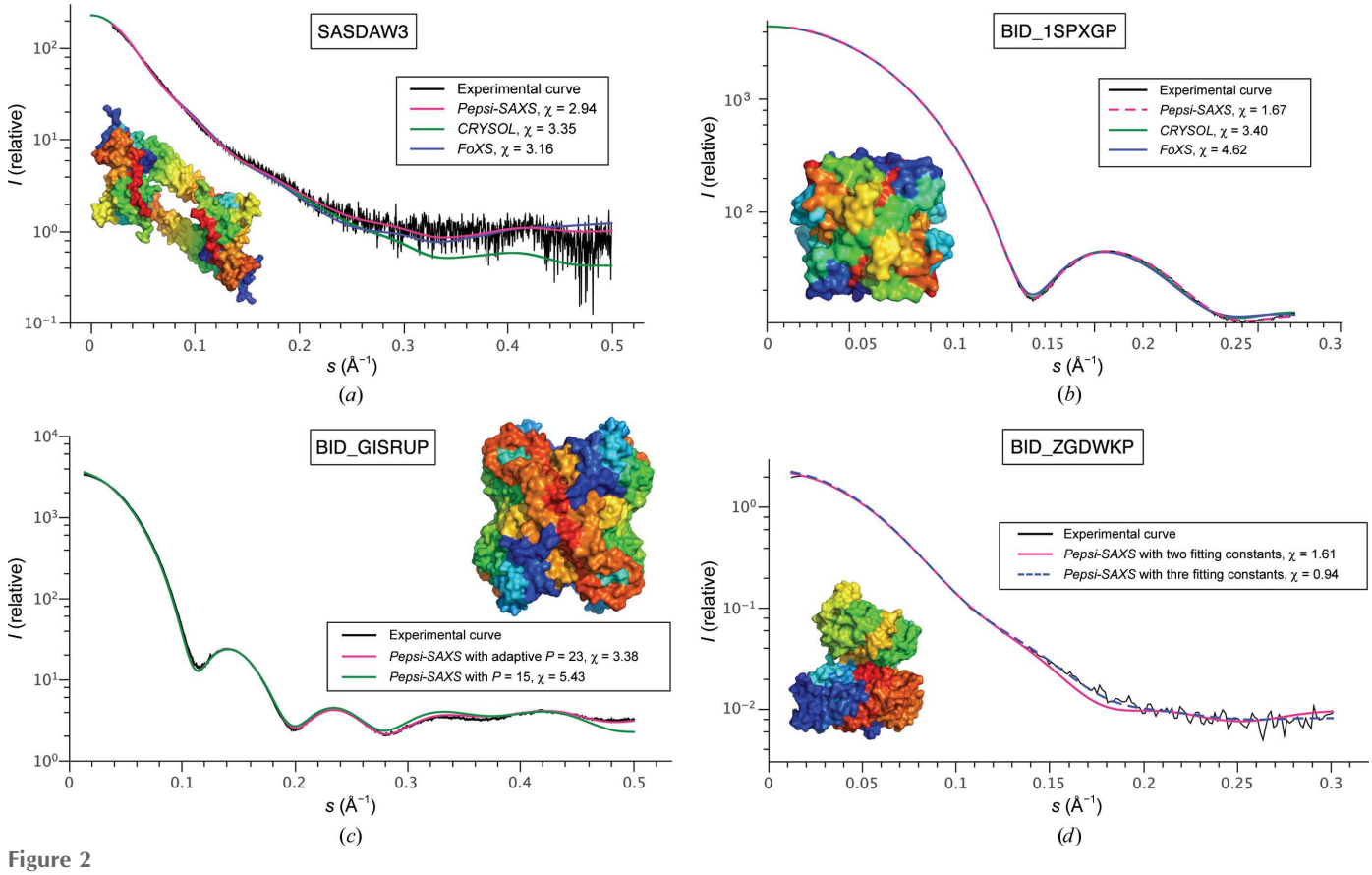


Figure 2
Comparison of modelled and experimental scattering profiles. (*a*) Comparison of *Pepsi-SAXS*, *CRYSOL* and *FoXS* on the SASDAW3 profile from the SASBDB database. (*b*) Comparison of *Pepsi-SAXS*, *CRYSOL* and *FoXS* on the BID_1SPXGP profile from the BioIsis database. (*c*) Effect of the adaptive expansion order on the model quality of *Pepsi-SAXS* applied to the BID_GISRUP profile from the BioIsis database. (*d*) Comparison of *Pepsi-SAXS* modelled scattering profiles without the subtraction (two fitting constants) and with the subtraction (three fitting constants) of the constant systematic error from the experimental data calculated for the BID_ZGDWKP model from the BioIsis database.

Table 4
Comparison of three methods, *CRYSOL*, *FoXS* and *Pepsi-SAXS*, when fitting modelled intensity profiles to experimental data with constant background noise collected from the BioIsis database.

The constant systematic error was subtracted from each of the profiles. For each method, we provide the value of χ and the running time measured in seconds for each of the scattering profiles. We also list the number of atoms in the models along with the average values of χ and running time.

			CRYSOL		*FoXS*		*Pepsi-SAXS*	
Structure	BioIsis ID	No. of atoms	χ	Time (s)	χ	Time (s)	χ	Time (s)
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	6395	1.16	0.88	1.79	2.50	1.10	0.19
Abscisic acid receptor PYR1	BID_1PYR1P	2924	1.39	0.65	2.03	0.70	2.29	0.05
Rubredoxin	BID_1RBDGP	424	6.44	0.53	7.05	0.11	7.02	0.04
Superoxide reductase	BID_1SPXGP	4060	3.24	0.70	4.63	1.06	1.66	0.06
Monomeric PF1674	BID_1TSPHP	1381	5.99	0.57	8.45	0.30	5.96	0.04
Endo-1,4-β-xylanase II	BID_1XYNTP	1480	0.95	0.59	1.03	0.32	0.93	0.04
28 bp DNA	BID_28BPDD	1107	0.44	0.48	1.25	0.34	0.51	0.04
S-Adenosylmethionine riboswitch mRNA	BID_2SAMRR	2086	2.42	0.60	2.46	0.47	2.45	0.04
Superoxide dismutase	BID_APSODP	2229	3.34	0.66	3.67	0.53	3.42	0.06
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	5318	2.36	0.76	2.17	1.68	2.11	0.06
Complement C3b-Efb (from *S. aureus*)	BID_C3BEFP	12833	2.15	1.15	2.31	7.21	1.43	0.20
Complement C3b + Efb (staphylococcal)	BID_C3BSAP	12255	0.12	1.18	0.12	6.68	0.12	0.17
Glucose isomerase	BID_GIKClP	12176	7.66	1.16	7.36	6.13	7.23	0.27
Glucose isomerase	BID_GISRUP	12176	7.64	1.16	4.73	6.16	3.37	0.28
Human regulator of chromosome condensation (RCC1)	BID_HRCC1P	3158	1.27	0.62	2.85	0.86	1.02	0.05
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1633	1.41	0.59	1.49	0.40	1.78	0.04
Lysozyme	BID_LYKClP	1394	8.02	0.62	10.92	0.29	6.92	0.05
Hen egg-white lysozyme	BID_LYSOZP	1001	2.19	0.56	2.56	0.22	2.44	0.03
MnmE in the nucleotide free state	BID_MNME1P	6518	0.79	0.89	0.89	2.51	0.78	0.20
E. coli MnmE–MnmG complex in the nucleotide-free state	BID_MnmEGP	16291	1.77	1.40	2.01	10.97	1.79	0.16
A. aeolicus MnmG	BID_MnmG1P	10534	1.39	1.04	1.60	5.07	1.43	0.10
A. aeolicus MnmG + tRNA	BID_MnmG2X	11184	1.66	1.12	1.72	5.74	1.73	0.19
E. coli MnmG + NbMnmG1	BID_MnmG3P	11562	1.75	1.12	2.10	6.05	1.97	0.35
E. coli MnmE–MnmG complex bound to GDP-AlF_x	BID_MnmGEP	23149	3.11	1.74	2.72	21.91	2.85	0.35
DNA double-strand break repair protein MRE11	BID_MRERAP	12148	0.68	1.14	1.47	6.47	1.24	0.14
Cu/Zn superoxide dismutase	BID_NMSODP	2309	0.93	0.62	1.05	0.55	0.97	0.04
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	3760	0.10	0.70	0.11	1.11	0.10	0.05
Ketoreductase-enoylreductase didomain	BID_ZGDWKP	5505	1.09	0.79	2.87	1.75	0.94	0.04
Average		6678	2.55	0.86	2.98	3.50	2.34	0.12

larger, and the difference between *Pepsi-SAXS* and *FoXS* becomes smaller. For large molecules, however, *FoXS* is significantly slower compared with *CRYSOL* and *Pepsi-SAXS*. For example, *Pepsi-SAXS* computes the scattering profile for the BID_MnmGEP model about 62 times faster compared with *FoXS*. We should mention that the reported speedup critically depends on the number of available CPU cores. Thus, we performed an additional artificial test and executed *Pepsi-SAXS* on a single CPU core. The average running time over the BioIsis data set in this case was 0.36 s, which was still several times smaller compared with *CRYSOL* and other methods.

Regarding the accuracy of the modelled profiles, on average *Pepsi-SAXS* produces scattering curves that are very similar to those computed by *CRYSOL* and *FoXS*, with approximately the same χ values, if these are computed for the same range of scattering angles. We should specifically mention that in all of the tests we have restricted the maximum scattering angles of *Pepsi-SAXS* to the default value of 0.5 Å^{-1}. This was performed for the rigorous comparison of the χ values with the results of *CRYSOL* and *FoXS*. We should also mention that in some cases of noisy experimental data *FoXS* restricts the maximum scattering angles even further, thus producing χ values that are higher on average. This, however, does not necessarily mean that the quality of the computed profiles is worse compared with the results from *Pepsi-SAXS* and *CRYSOL*. *SAStbx* generally provides a significantly worse quality of fit and is thus excluded from the detailed comparison. Figs. 2(*b*), 2(*c*) and 2(*d*) show three examples of modelled scattering profiles from the BioIsis database. More plots, together with the residuals of the scattering profiles, can be found in Supplementary Fig. S1. Generally, we can conclude that *Pepsi-SAXS* computes scattering profiles that are comparable to the other two methods. Below, we will also study the effect of the adaptive resolution in comparison with *CRYSOL* in more detail.

We should also mention that a smaller χ value achieved using a certain method for a scattering profile does not necessarily mean a better quality of the computed profile. Generally, one should be concerned about possible flexibility and conformational heterogeneity of the modelled proteins. Also, some of the models from the two benchmarks are not crystallographic structures but were produced with molecular-dynamics simulations or *MODELLER* (Webb & Sali, 2014), for example. Therefore, small values of χ for some of the models indicate potential overfitting of the experimental profiles rather than demonstrating the superiority of the fitting method. Finally, we should add that different methods use

Table 5
Comparison of four methods, *CRYSOL*, *FoXS*, *SAStbx* (using the three-dimensional Zernike technique and data-reduction option) and *Pepsi-SAXS*, when fitting modelled intensity profiles to experimental data collected from the SASBDB database.

For each method, we provide the value of χ and the running time measured in seconds for each of the scattering profiles. We also list the number of atoms in the models along with the average values of χ and running time. *SAStbx* failed for some of the profiles; the corresponding values for χ and time are marked with a dash.

			CRYSOL		*FoXS*		*SAStbx*		*Pepsi-SAXS*	
Structure	SASBDB ID	No. of atoms	χ	Time (s)	χ	Time (s)	χ	Time (s)	χ	Time (s)
Lysozyme C	SASDAC2	1001	1.21	0.53	1.22	0.23	1.01	1.33	1.24	0.07
Ubiquitin-60S ribosomal protein L40	SASDAQ2	602	3.08	0.51	3.10	0.16	3.67	1.14	3.00	0.07
Myoglobin in PBS	SASDAH2	1247	2.20	0.55	2.36	0.29	2.13	1.36	2.13	0.07
Catalase in PBS	SASDA92	16432	3.15	1.35	3.62	10.62	3.36	6.99	3.16	0.37
Methyltransferase WbdD	SASDAJ6	12662	1.22	1.11	1.04	6.94	–	–	1.35	0.12
Alcohol dehydrogenase in PBS	SASDA52	10327	3.24	1.02	4.04	4.71	–	–	2.94	0.26
Calmodulin–peptide complex	SASDAN4	1297	2.91	0.55	3.96	0.33	–	–	3.48	0.08
Psi-producing oxygenase A	SASDA45	25761	9.70	1.89	13.83	24.30	–	–	9.44	0.79
Cysteine desulfurase IscS dimer	SASDAV6	6118	1.30	0.80	1.88	2.14	–	–	1.27	0.14
Factor H CCP modules 12 to 13	SASDA25	951	1.70	0.52	1.73	0.28	1.27	1.90	1.56	0.07
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	8124	1.05	0.87	1.06	3.47	1.77	5.80	1.05	0.11
ComE–comcde complex	SASDAB7	5909	1.30	0.75	1.25	2.07	3.01	3.88	1.21	0.09
LytTR–comcde complex	SASDAC7	3605	1.18	0.63	1.17	1.15	–	–	1.26	0.06
Myomesin-1	SASDAK5	3231	1.59	0.64	1.65	1.15	2.23	3.52	1.82	0.09
IcsS, IscU and CyaY dimeric complex	SASDA27	9872	2.87	0.98	4.10	4.62	1.90	7.39	2.81	0.30
Iron–sulfur cluster assembly scaffold IscU monomer	SASDAW6	1014	1.19	0.53	1.24	0.26	–	–	1.25	0.07
Geminin–Cdt1 2:1 heterotrimer	SASDAV3	2435	1.98	0.61	2.55	0.67	–	–	2.61	0.10
Geminin–Cdt1 4:2 heterohexamer	SASDAW3	5362	3.35	0.78	3.16	1.99	3.29	5.03	2.94	0.17
Apo XMRV RT	SASDAV5	5366	1.12	0.77	1.19	1.89	0.87	5.40	1.13	0.18
XMRV RT + DNA/RNA hybrid	SASDAW5	6340	0.92	0.82	0.87	2.40	–	–	0.91	0.20
Protein CyaY monomer	SASDAX6	863	1.15	0.53	1.21	0.20	–	–	1.22	0.08
Annexin-A4	SASDAJ5	2510	5.16	0.61	5.26	0.65	3.93	2.06	5.09	0.08
Pyruvate decarboxylase	SASDAX2	17157	0.81	1.36	0.82	11.24	0.90	8.35	0.83	0.29
Average		6443	2.32	0.81	2.71	3.56	2.26	4.17	2.33	0.17

different ranges of fitting parameters, and also different models for the hydration shell, which consequently contribute differently to the potential overfitting of experimental data.

Ideally, a reference data set of native structures supplemented with experimental SAXS profiles along with non-native decoys should be established for the evaluation of SAXS algorithms. Different methods can then be tested on this data set by scoring the non-native decoys. The absence of overfitting in a SAXS method can be confirmed, for example, if the native structures have the lowest χ values among all of the scored decoys. To support our method, we should say that we use a small range of adjustable parameters compared with other methods such as *CRYSOL* and *FoXS*. Thus, we believe that *Pepsi-SAXS* does not have any significant overfitting of experimental data.

In the second series of tests, we compared the three methods, excluding *SAStbx*, on the same data from BioIsis, but this time measuring the goodness of fit for the modelled intensities to the experimental SAXS profiles and the corresponding timings for data with a constant systematic error (see equation 3 in the Supporting Information). Table 4 lists the detailed results of the tests. For all three of the methods, the running time became only marginally larger. Regarding the accuracy of the models, both *Pepsi-SAXS* and *CRYSOL* improved the value of χ by about 9% on average, while *FoXS* unexpectedly worsened the averaged value of χ. Fig. 2(*d*) shows an example of fitting for two profiles calculated by *Pepsi-SAXS* with and without the constant systematic error in the experimental curve. We can see a drastic improvement in the model when subtracting the constant noise from the experimental data.

In the third series of tests, we compared the four methods on the data from SASBDB. Here, we again first measured the goodness of fit for the modelled intensities to the experimental SAXS profiles (15) and the corresponding timings. Table 5 lists the detailed results of the tests. Similarly to the previous tests, *Pepsi-SAXS* significantly outperforms the other methods in running time. Here, on average, *Pepsi-SAXS* is about five times faster compared with *CRYSOL*, and 21 and 25 times faster compared with *FoXS* and *SAStbx*, respectively. The speedup in the running time of *Pepsi-SAXS* compared with the other methods is somewhat smaller compared with the previous tests owing to the on average higher expansion orders used here. More precisely, for SASBDB *Pepsi-SAXS* uses an average expansion order of 19, while for BioIsis it uses an order of 14. Regarding the accuracy of the modelled profiles, on average *Pepsi-SAXS*, *CRYSOL* and *FoXS* achieve the same values of χ if these are computed using the same range of scattering angles. *SAStbx* was not able to process half of the scattering profiles. It was again the slowest method among the four. Fig. 2(*a*) shows an example of modelled scattering profiles from this database. The model (SASDAW3) has a complex shape, thus we expected the quality of the *CRYSOL* modelled profile to be lower compared with profiles built with *FoXS* and *Pepsi-SAXS*. Supplementary Fig. S1 shows all of the scattering plots for SASBDB for the three methods together with the residuals of the scattering profiles.

Table 6
Comparison of three methods, *CRYSOL*, *FoXS* and *Pepsi-SAXS*, when fitting modelled intensity profiles to experimental data with constant background noise collected from the SASBDB database.

The constant systematic error was subtracted from each of the profiles. For each method, we provide the value of χ and the running time measured in seconds for each of the scattering profiles. We also list the number of atoms in the models along with the average values of χ and running time.

			CRYSOL		*FoXS*		*Pepsi-SAXS*	
Structure	SASBDB ID	No. of atoms	χ	Time (s)	χ	Time (s)	χ	Time (s)
Lysozyme C	SASDAC2	1001	1.21	0.61	1.22	0.24	1.23	0.09
Ubiquitin-60S ribosomal protein L40	SASDAQ2	602	3.07	0.58	3.10	0.16	2.99	0.09
Myoglobin in PBS	SASDAH2	1247	2.10	0.63	2.40	0.29	2.10	0.09
Catalase in PBS	SASDA92	16432	2.93	1.44	4.07	10.53	2.97	0.39
Methyltransferase WbdD	SASDAJ6	12662	1.14	1.17	1.13	6.98	1.19	0.13
Alcohol dehydrogenase in PBS	SASDA52	10327	2.54	1.11	5.41	4.72	2.46	0.28
Calmodulin–peptide complex	SASDAN4	1297	2.32	0.61	4.05	0.33	3.12	0.11
Psi-producing oxygenase A	SASDA45	25761	9.67	1.98	14.95	24.33	8.48	0.82
Cysteine desulfurase IscS dimer	SASDAV6	6118	1.28	0.85	1.93	2.13	1.27	0.15
Factor H CCP modules 12 to 13	SASDA25	951	1.42	0.59	1.74	0.27	1.25	0.09
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	8124	1.04	0.93	1.07	3.47	1.05	0.14
ComE–comcde complex	SASDAB7	5909	1.18	0.82	1.32	2.04	1.11	0.09
LytTR–comcde complex	SASDAC7	3605	1.16	0.69	1.17	1.15	1.26	0.07
Myomesin-1	SASDAK5	3231	1.57	0.72	1.66	1.16	1.64	0.10
IcsS, IscU and CyaY dimeric complex	SASDA27	9872	2.62	1.06	5.17	4.63	2.57	0.33
Iron–sulfur cluster assembly scaffold IscU monomer	SASDAW6	1014	1.19	0.61	1.26	0.26	1.25	0.09
Geminin–Cdt1 2:1 heterotrimer	SASDAV3	2435	1.94	0.62	3.13	0.65	1.80	0.12
Geminin–Cdt1 4:2 heterohexamer	SASDAW3	5362	2.94	0.88	3.19	2.04	2.90	0.19
Apo XMRV RT	SASDAV5	5366	1.11	0.85	1.21	1.94	1.10	0.21
XMRV RT + DNA/RNA hybrid	SASDAW5	6340	0.90	0.91	0.87	2.35	0.90	0.23
Protein CyaY monomer	SASDAX6	863	1.10	0.60	1.21	0.20	1.12	0.09
Annexin-A4	SASDAJ5	2510	5.09	0.67	5.28	0.63	5.06	0.11
Pyruvate decarboxylase	SASDAX2	17157	0.81	1.43	0.83	11.26	0.82	0.30

In the fourth series of tests, we again compared the three methods, excluding *SAStbx*, on the data from SASBDB and measured the goodness of fit with a constant systematic error (see equation 3 in the Supporting Information) and the corresponding timings. Table 6 lists the detailed results of the tests. As before, the running time becomes only marginally larger. Regarding the accuracy of the models, *Pepsi-SAXS* improves the value of χ by 8% on average, *CRYSOL* improves it by 6% and *FoXS* again shows no improvement in fit.

For the fifth test, we decided to compare the running time of the four methods if a user computes a scattering profile without fitting it to the experimental data. Here, we considered two scenarios: a profile with 51 points, as used to be the default option in *CRYSOL*, and a profile with 512 points, which better corresponds to modern experimental measurements. Table 7 lists the timings for all four methods run on atomic models from the BioIsis and SASBDB databases. For the 51-point profile, *Pepsi-SAXS* is on average about three times faster compared with *CRYSOL*, and 19 and 27 times faster compared with *FoXS* and *SAStbx*, respectively. With the 512-point profile, *Pepsi-SAXS*, *FoXs* and *SAStbx* increase the timings only marginally. However, the running time of *CRYSOL* depends linearly on the number of points in the scattering profile. Therefore, its timing increases about ten times.

In the sixth series of tests, we compared the effect of the adaptive choice of the multipole expansion order using data from the BioIsis and SASBDB databases. To do so, we first fixed the expansion order to the value of 15, which is used by default in *CRYSOL*, and ran *Pepsi-SAXS* in comparison with *CRYSOL*. We then chose the value of the expansion order adaptively according to (14) and ran the two programs again. Table 8 lists the details of the comparisons. As we can see from this table, using the default expansion order of 15, *Pepsi-SAXS* demonstrates a very similar quality of models compared with *CRYSOL*, with a slightly smaller value of χ. Adaptive resolution lowers the value of χ for the two methods: by about 1% for *CRYSOL* and about 2% for *Pepsi-SAXS*. We attribute the more pronounced effect of the adaptive resolution in *Pepsi-SAXS* to the different model of the hydration shell in our method. Fig. 2(*c*) shows an example of scattering profiles plotted at a different expansion order in comparison with the experimental curve for a large molecule (BID_GISRUP). We can see a pronounced difference between the curves at large values of q, which corresponds to a fine resolution in real space that is not well encoded using low-multipole expansion orders. All scattering profiles for experimental data collected from the BioIsis and SASBDB databases and computed using *Pepsi-SAXS*, *CRYSOL* and *FoXS* can be found in Supplementary Fig. S1.

Finally, in the seventh series of tests, we compared the values of two adjustable parameters, r_0 and $\delta\rho$, for the three methods, excluding *SAStbx*, on data from the BioIsis and SASBDB databases. In the case of *FoXS*, we computed the values of r_0 and $\delta\rho$ by rescaling its internal fitting parameters c_1 and c_2 as suggested by the authors (Schneidman-Duhovny *et al.*, 2010). Table 9 lists the adjustable parameters along with the mean values and the standard deviations for the three

Table 7
Comparison of four methods, *CRYSOL*, *FoXS*, *SAStbx* (using the three-dimensional Zernike technique) and *Pepsi-SAXS*, when calculating intensity profiles for models collected from the BioIsis and SASBDB databases.

No fitting to experimental data was performed. For each method, we provide two running times measured in seconds when calculating the intensity profile with 512 points and with 51 points, correspondingly. We also list the number of atoms in the models along with the average values of running times. *SAStbx* failed for some of the profiles; the corresponding values of timings are marked with dashes.

Structure	Database ID	No. of atoms	*CRYSOL*	*FoXS*	*SAXStbx*	*Pepsi-SAXS*
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	6395	3.55/0.44	2.46/2.05	5.24/5.11	0.14/0.13
Abscisic acid receptor PYR1	BID_1PYR1P	2924	1.80/0.25	0.69/0.62	2.18/2.14	0.04/0.04
Rubredoxin	BID_1RBDGP	424	0.70/0.13	0.10/0.08	0.99/0.96	0.01/0.01
Superoxide reductase	BID_1SPXGP	4060	2.25/0.31	1.04/0.97	2.33/2.26	0.05/0.05
Monomeric PF1674	BID_1TSPHP	1381	1.14/0.17	0.28/0.24	1.90/1.87	0.02/0.02
Endo-1,4-β-xylanase II	BID_1XYNTP	1480	1.16/0.18	0.31/0.26	1.39/1.38	0.02/0.02
28 bp DNA	BID_28BPDD	1107	0.97/0.16	0.35/0.19	1.71/1.69	0.02/0.03
S-Adenosylmethionine riboswitch mRNA	BID_2SAMRR	2086	1.43/0.21	0.46/0.37	1.85/1.83	0.03/0.04
Superoxide dismutase	BID_APSODP	2229	1.50/0.22	0.52/0.43	1.83/1.83	0.03/0.04
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	5318	2.86/0.38	1.71/1.50	3.90/3.97	0.09/0.09
Complement C3b-Efb (from *S. aureus*)	BID_C3BEFP	12833	6.65/0.81	7.20/6.79	10.16/10.19	0.35/0.36
Complement C3b + Efb (staphylococcal)	BID_C3BSAP	12255	6.47/0.77	6.59/6.13	8.23/8.05	0.32/0.33
Glucose isomerase	BID_GIKClP	12176	5.87/0.73	6.08/5.88	5.47/5.52	0.25/0.24
Glucose isomerase	BID_GISRUP	12176	6.09/0.73	6.12/5.90	6.01/6.10	0.25/0.24
Human regulator of chromosome condensation (RCC1)	BID_HRCC1P	3158	1.91/0.26	0.84/0.71	2.81/2.79	0.05/0.04
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1633	1.25/0.19	0.41/0.30	1.98/1.98	0.03/0.03
Lysozyme	BID_LYKClP	1394	1.13/0.17	0.28/0.24	1.38/1.29	0.02/0.02
Hen egg-white lysozyme	BID_LYSOZP	1001	0.96/0.15	0.21/0.18	1.30/1.19	0.02/0.02
MnmE in the nucleotide-free state	BID_MNME1P	6518	3.46/0.44	2.42/2.10	4.43/4.48	0.13/0.13
E. coli MnmE–MnmG complex in the nucleotide-free state	BID_MnmEGP	16291	8.42/0.98	10.76/10.11	10.33/10.87	0.55/0.55
A. aeolicus MnmG	BID_MnmG1P	10534	5.56/0.68	5.12/4.72	9.35/9.46	0.33/0.28
A. aeolicus MnmG + tRNA	BID_MnmG2X	11184	5.67/0.68	5.66/5.16	7.50/7.36	0.32/0.28
E. coli MnmG + NbMnmG1	BID_MnmG3P	11562	6.17/0.74	6.14/5.60	8.14/7.99	0.31/0.33
E. coli MnmE–MnmG complex bound to GDP-AlF$_x$	BID_MnmGEP	23149	12.29/1.39	21.79/20.63	5.80/6.03	1.84/1.29
DNA double-strand break repair protein MRE11	BID_MRERAP	12148	6.18/0.75	6.41/5.98	7.75/7.68	0.35/0.30
Cu/Zn superoxide dismutase	BID_NMSODP	2309	1.56/0.25	0.56/0.46	2.01/2.04	0.04/0.04
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	3760	2.20/0.31	1.08/0.89	3.09/2.93	0.07/0.07
Ketoreductase-enoylreductase didomain	BID_ZGDWKP	5505	2.98/0.38	1.75/1.58	3.47/3.38	0.10/0.10
Lysozyme C	SASDAC2	1001	0.97/0.16	0.21/0.17	1.28/1.27	0.02/0.02
Ubiquitin-60S ribosomal protein L40	SASDAQ2	602	0.79/0.13	0.14/0.10	1.14/1.11	0.01/0.01
Myoglobin in PBS	SASDAH2	1247	1.08/0.17	0.26/0.21	1.37/1.34	0.02/0.02
Catalase in PBS	SASDA92	16432	8.02/0.96	10.64/10.22	7.33/6.94	0.34/0.32
Methyltransferase WbdD	SASDAJ6	12662	6.70/0.80	6.94/6.49	9.23/8.85	0.39/0.38
Alcohol dehydrogenase in PBS	SASDA52	10327	5.20/0.65	4.65/4.46	5.19/5.07	0.22/0.21
Calmodulin–peptide complex	SASDAN4	1297	1.14/0.18	0.31/0.22	—/—	0.02/0.03
Psi-producing oxygenase A	SASDA45	25761	13.11/1.50	24.28/23.6	13.95/14.2	0.87/0.84
Cysteine desulfurase IscS dimer	SASDAV6	6118	3.22/0.41	2.15/1.95	3.90/3.71	0.09/0.10
Factor H CCP modules 12 to 13	SASDA25	951	0.94/0.15	0.25/0.16	1.90/1.90	0.02/0.02
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	8124	4.24/0.53	3.49/3.03	6.00/5.72	0.19/0.18
ComE–comcde complex	SASDAB7	5909	3.20/0.40	2.05/1.83	4.14/4.00	0.11/0.10
LytTR–comcde complex	SASDAC7	3605	2.16/0.28	1.13/0.86	—/—	0.07/0.07
Myomesin-1	SASDAK5	3231	2.04/0.28	1.11/0.76	3.60/3.57	0.09/0.09
IcsS, IscU and CyaY dimeric complex	SASDA27	9872	4.99/0.61	4.59/4.19	7.24/7.49	0.25/0.23
Iron–sulfur cluster assembly scaffold IscU monomer	SASDAW6	1014	0.99/0.16	0.26/0.17	—/—	0.02/0.02
Geminin–Cdt1 2:1 heterotrimer	SASDAV3	2435	1.65/0.23	0.65/0.50	2.29/2.18	0.04/0.04
Geminin–Cdt1 4:2 heterohexamer	SASDAW3	5362	3.16/0.39	1.99/1.60	4.94/4.97	0.12/0.11
Apo XMRV RT	SASDAV5	5366	3.06/0.40	1.95/1.57	5.46/5.65	0.15/0.14
XMRV RT + DNA/RNA hybrid	SASDAW5	6340	3.45/0.46	2.37/2.02	5.34/5.25	0.16/0.15
Protein CyaY monomer	SASDAX6	863	0.95/0.15	0.19/0.14	—/—	0.02/0.02
Annexin-A4	SASDAJ5	2510	1.65/0.23	0.62/0.51	2.04/1.98	0.03/0.03
Pyruvate decarboxylase	SASDAX2	17157	8.41/1.01	11.22/10.95	8.58/8.52	0.40/0.38
Average		6572.0	3.59/0.45	3.51/3.25	4.63/4.60	0.18/0.17

methods. We can see that all of the methods agree on an average value of the effective atomic radius r_0 of 1.64 Å. However, the standard deviation of this parameter in *FoXS* and *Pepsi-SAXS* is only 0.05 Å, which constitutes 3% of the average value and is several times smaller compared with the standard deviation of 0.18 Å in *CRYSOL*. We should note that if we double the width of the search window for the r_0 parameter to make it more comparable with the *CRYSOL* settings, the quality of fit to the experimental data improves only marginally (see Supplementary Table S2).

Regarding the second adjustable parameter, the contrast of the hydration shell $\delta\rho$, all of the methods provide different

Table 8
Comparison of *CRYSOL* with *Pepsi-SAXS* when using adaptive multipole expansion orders.

Experimental data were collected from the BioIsis and SASBDB databases. For each method, we provide the value of χ and the running time measured in seconds when using the default expansion order of 15 and the adaptive expansion order. We also list the number of atoms in the models and the order of the adaptive multipole expansion, along with the average values of χ and running time.

			Expansion order = 15				Adaptive expansion order				
			CRYSOL		*Pepsi-SAXS*			*CRYSOL*		*Pepsi-SAXS*	
Structure	Database ID	No. of atoms	χ	Time (s)	χ	Time (s)	Order	χ	Time (s)	χ	Time (s)
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	6395	1.98	0.82	1.55	0.13	24	1.79	1.10	1.51	0.18
Abscisic acid receptor PYR1	BID_1PYR1P	2924	1.39	0.59	2.30	0.05	10	1.39	0.51	2.33	0.04
Rubredoxin	BID_1RBDGP	424	7.61	0.45	7.42	0.03	5	7.78	0.38	7.24	0.03
Superoxide reductase	BID_1SPXGP	4060	3.40	0.63	1.67	0.06	11	3.40	0.57	1.67	0.05
Monomeric PF1674	BID_1TSPHP	1381	6.28	0.49	6.13	0.04	7	6.28	0.43	6.22	0.03
Endo-1,4-β-xylanase II	BID_1XYNTP	1480	0.99	0.50	0.93	0.04	7	0.99	0.43	0.93	0.03
28 bp DNA	BID_28BPDD	1107	0.49	0.48	0.57	0.04	11	0.49	0.46	0.56	0.03
S-Adenosylmethionine riboswitch mRNA	BID_2SAMRR	2086	2.42	0.53	2.51	0.05	9	2.40	0.46	2.46	0.03
Superoxide dismutase	BID_APSODP	2229	3.45	0.57	3.58	0.06	15	3.45	0.59	3.58	0.06
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	5318	2.54	0.71	2.23	0.08	13	2.53	0.66	2.19	0.06
Complement C3b-Efb (from *S. aureus*)	BID_C3BEFP	12833	2.41	1.15	1.57	0.16	18	2.45	1.34	1.63	0.19
Complement C3b + Efb (staphylococcal)	BID_C3BSAP	12255	0.12	1.10	0.12	0.14	18	0.12	1.28	0.12	0.17
Glucose isomerase	BID_GIKCLP	12176	7.83	1.09	6.85	0.15	23	7.86	1.47	7.34	0.26
Glucose isomerase	BID_GISRUP	12176	7.99	1.09	5.43	0.15	23	6.39	1.45	3.38	0.26
Human regulator of chromosome condensation (RCC1)	BID_HRCC1P	3158	1.29	0.59	1.63	0.06	11	1.29	0.53	1.59	0.05
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1633	1.44	0.51	1.90	0.05	10	1.43	0.48	1.85	0.03
Lysozyme	BID_LYKCLP	1394	9.39	0.53	9.42	0.05	11	9.39	0.50	9.41	0.04
Hen egg-white lysozyme	BID_LYSOZP	1001	2.54	0.49	2.60	0.04	8	2.54	0.43	2.62	0.03
MnmE in the nucleotide-free state	BID_MNME1P	6518	0.89	0.82	0.84	0.14	24	0.86	1.13	0.85	0.18
E. coli MnmE–MnmG complex in the nucleotide-free state	BID_MnmEGP	16291	1.86	1.31	1.85	0.18	16	1.85	1.38	1.84	0.15
A. aeolicus MnmG	BID_MnmG1P	10534	1.40	1.04	1.45	0.13	11	1.39	0.85	1.44	0.10
A. aeolicus MnmG + tRNA	BID_MnmG2X	11184	1.70	1.04	1.76	0.15	21	1.71	1.33	1.75	0.18
E. coli MnmG + NbMnmG1	BID_MnmG3P	11562	1.79	1.07	3.13	0.14	27	1.78	1.69	2.05	0.35
E. coli MnmE–MnmG complex bound to GDP-AlF$_x$	BID_MnmGEP	23149	3.16	1.71	2.82	0.26	19	3.20	2.19	2.87	0.35
DNA double-strand break repair protein MRE11	BID_MRERAP	12148	0.72	1.08	1.32	0.14	17	0.72	1.18	1.29	0.14
Cu/Zn superoxide dismutase	BID_NMSODP	2309	1.04	0.55	0.96	0.05	10	1.05	0.49	0.97	0.04
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	3760	0.10	0.64	0.11	0.07	13	0.10	0.59	0.11	0.05
Ketoreductase-enoylreductase didomain	BID_ZGDWKP	5505	2.00	0.72	1.62	0.07	12	2.01	0.64	1.61	0.04
Lysozyme C	SASDAC2	1001	1.21	0.53	1.24	0.08	11	1.21	0.50	1.24	0.07
Ubiquitin-60S ribosomal protein L40	SASDAQ2	602	3.08	0.51	2.99	0.08	10	3.08	0.47	3.00	0.07
Myoglobin in PBS	SASDAH2	1247	2.20	0.55	2.12	0.08	12	2.20	0.52	2.13	0.07
Catalase in PBS	SASDA92	16432	3.15	1.34	3.40	0.22	25	3.10	2.02	3.16	0.37
Methyltransferase WbdD	SASDAJ6	12662	1.22	1.11	1.23	0.17	13	1.30	1.00	1.35	0.12
Alcohol dehydrogenase in PBS	SASDA52	10327	3.24	1.03	2.95	0.17	23	3.16	1.36	2.94	0.26
Calmodulin–peptide complex	SASDAN4	1297	2.91	0.55	3.48	0.09	14	2.92	0.54	3.48	0.09
Psi-producing oxygenase A	SASDA45	25761	9.70	1.88	8.78	0.33	29	9.79	3.63	9.44	0.79
Cysteine desulfurase IscS dimer	SASDAV6	6118	1.30	0.80	1.28	0.13	20	1.30	0.91	1.27	0.14
Factor H CCP modules 12 to 13	SASDA25	951	1.70	0.51	1.56	0.08	13	1.70	0.50	1.56	0.07
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	8124	1.05	0.87	1.10	0.11	20	1.01	1.06	1.05	0.12
ComE–comcde complex	SASDAB7	5909	1.30	0.75	1.23	0.09	17	1.27	0.80	1.21	0.09
LytTR–comcde complex	SASDAC7	3605	1.18	0.64	1.29	0.07	16	1.15	0.66	1.26	0.06
Myomesin-1	SASDAK5	3231	1.59	0.63	1.76	0.08	21	1.65	0.73	1.82	0.09
IcsS, IscU and CyaY dimeric complex	SASDA27	9872	2.87	0.99	2.77	0.17	26	2.85	1.42	2.81	0.30
Iron–sulfur cluster assembly scaffold IscU monomer	SASDAW6	1014	1.19	0.56	1.25	0.09	13	1.19	0.52	1.25	0.07
Geminin–Cdt1 2:1 heterotrimer	SASDAV3	2435	1.98	0.62	2.57	0.10	17	1.99	0.64	2.61	0.10
Geminin–Cdt1 4:2 heterohexamer	SASDAW3	5362	3.35	0.77	2.85	0.13	24	3.17	1.10	2.94	0.17
Apo XMRV RT	SASDAV5	5366	1.12	0.78	1.16	0.13	23	1.10	1.00	1.13	0.18
XMRV RT + DNA/RNA hybrid	SASDAW5	6340	0.92	0.82	0.91	0.14	23	0.92	1.06	0.91	0.20
Protein CyaY monomer	SASDAX6	863	1.15	0.54	1.22	0.08	11	1.15	0.50	1.22	0.08
Annexin-A4	SASDAJ5	2510	5.16	0.60	5.08	0.10	16	5.16	0.63	5.09	0.09
Pyruvate decarboxylase	SASDAX2	17157	0.82	1.36	0.82	0.18	24	0.82	2.01	0.83	0.29
Average		6572.0	2.58	0.81	2.50	0.11	16.18	2.55	0.94	2.45	0.14

mean values. More precisely, *CRYSOL* allows variation of $\delta\rho$ between 0 and 60 e nm^{-3}, with an average of 22.4 $\pm$ 21.7 e nm^{-3}. *FoXS* allows negative values of $\delta\rho$ in the range $\sim 27 \leq \delta\rho \leq 54$ e nm^{-3}. Thus, its average $\delta\rho$ is lower compared with that computed by *CRYSOL* and equals 16.6 $\pm$ 22.2 e nm^{-3}. In our model, by default, we allow only positive values of $\delta\rho$ up to one tenth of the bulk density value of 33.4 e nm^{-3}. As a result, our mean value of the contrast of the

Table 9
Comparison of three methods, *CRYSOL*, *FoXS* and *Pepsi-SAXS*, when fitting modelled intensity profiles to experimental data for models collected from the BioIsis and SASBDB databases.

For each method, we provide the value of χ, the fitted value of the r_0 parameter and the fitted value of the contrast of the hydration shell parameter $\delta\rho$. We also list the average values of χ along with the average values and the standard deviations of the fitting parameters.

		Fitting parameters								
		CRYSOL			*FoXS*			*Pepsi-SAXS*		
Structure	Database ID	χ	r_0 (Å)	$\delta\rho$ (e nm^{-1})	χ	r_0 (Å)	$\delta\rho$ (e nm^{-1})	χ	r_0 (Å)	$\delta\rho$ (e nm^{-1})
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	1.98	1.80	30.0	1.66	1.70	14.2	1.51	1.69	10.7
Abscisic acid receptor PYR1	BID_1PYR1P	1.39	1.40	0.0	2.03	1.64	−11.0	2.33	1.58	4.3
Rubredoxin	BID_1RBDGP	7.61	1.78	10.0	7.05	1.66	−2.0	7.24	1.65	26.4
Superoxide reductase	BID_1SPXGP	3.40	1.64	10.0	4.62	1.66	9.0	1.67	1.66	13.7
Monomeric PF1674	BID_1TSPHP	6.28	1.66	0.0	7.96	1.67	−14.1	6.22	1.66	3.0
Endo-1,4-β-xylanase II	BID_1XYNTP	0.99	1.80	15.0	1.03	1.65	5.0	0.93	1.63	33.4
28 bp DNA	BID_28BPDD	0.49	1.40	75.0	0.79	1.70	54.0	0.56	1.55	33.4
S-Adenosylmethionine riboswitch mRNA	BID_2SAMRR	2.42	1.40	13.0	2.46	1.54	27.4	2.46	1.41	19.0
Superoxide dismutase	BID_APSODP	3.45	1.74	13.0	3.66	1.65	7.3	3.58	1.63	12.0
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	2.54	1.80	37.0	2.16	1.62	54.0	2.19	1.66	29.7
Complement C3b-Efb (from *S. aureus*)	BID_C3BEFP	2.41	1.40	60.0	1.75	1.63	54.0	1.63	1.65	33.4
Complement C3b + Efb (staphylococcal)	BID_C3BSAP	0.12	1.40	18.0	0.12	1.64	9.3	0.12	1.64	18.0
Glucose isomerase	BID_GIKCLP	7.83	1.40	5.0	7.27	1.66	7.6	7.34	1.64	15.7
Glucose isomerase	BID_GISRUP	7.99	1.40	7.0	4.69	1.66	3.1	3.38	1.64	19.7
Human regulator of chromosome condensation (RCC1)	BID_HRCC1P	1.29	1.54	75.0	1.77	1.67	54.0	1.59	1.69	33.4
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1.44	1.40	10.0	1.49	1.65	3.9	1.85	1.63	12.4
Lysozyme	BID_LYKCLP	9.39	1.80	0.0	7.74	1.54	−27.0	9.41	1.52	0.0
Hen egg-white lysozyme	BID_LYSOZP	2.54	1.50	10.0	2.56	1.67	5.8	2.62	1.66	33.4
MnmE in the nucleotide-free state	BID_MNME1P	0.89	1.80	25.0	0.88	1.67	11.5	0.85	1.67	17.0
E. coli MnmE–MnmG complex in the nucleotide-free state	BID_MnmEGP	1.86	1.80	3.0	1.87	1.54	−0.2	1.84	1.70	0.0
A. aeolicus MnmG	BID_MnmG1P	1.40	1.40	40.0	1.50	1.59	27.4	1.44	1.54	33.4
A. aeolicus MnmG + tRNA	BID_MnmG2X	1.70	1.80	3.0	1.71	1.67	3.1	1.75	1.66	5.0
E. coli MnmG + NbMnmG1	BID_MnmG3P	1.79	1.76	33.0	2.06	1.66	33.8	2.05	1.66	25.4
E. coli MnmE–MnmG complex bound to GDP-AlF$_x$	BID_MnmGEP	3.16	1.40	75.0	2.69	1.69	45.9	2.87	1.68	33.4
DNA double-strand break repair protein MRE11	BID_MRERAP	0.72	1.74	37.0	1.19	1.62	54.0	1.29	1.65	33.4
Cu/Zn superoxide dismutase	BID_NMSODP	1.04	1.78	35.0	1.05	1.67	43.6	0.97	1.66	32.1
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	0.10	1.80	33.0	0.10	1.66	29.7	0.11	1.67	23.0
Ketoreductase-enoylreductase didomain	BID_ZGDWKP	2.00	1.80	28.0	2.49	1.59	54.0	1.61	1.70	11.7
Lysozyme C	SASDAC2	1.21	1.66	5.0	1.22	1.65	−5.3	1.24	1.63	23.7
Ubiquitin-60S ribosomal protein L40	SASDAQ2	3.08	1.74	10.0	3.10	1.65	−0.8	3.00	1.65	28.7
Myoglobin in PBS	SASDAH2	2.20	1.78	10.0	2.36	1.67	1.1	2.13	1.66	24.0
Catalase in PBS	SASDA92	3.15	1.80	13.0	3.62	1.54	14.2	3.16	1.59	12.4
Methyltransferase WbdD	SASDAJ6	1.22	1.42	28.0	1.04	1.59	54.0	1.35	1.68	13.0
Alcohol dehydrogenase in PBS	SASDA52	3.24	1.80	0.0	4.04	1.63	−15.7	2.94	1.67	0.0
Calmodulin–peptide complex	SASDAN4	2.91	1.80	25.0	3.96	1.68	11.4	3.48	1.64	19.0
Psi-producing oxygenase A	SASDA45	9.70	1.40	52.0	13.83	1.70	22.3	9.44	1.68	33.4
Cysteine desulfurase IscS dimer	SASDAV6	1.30	1.80	65.0	1.88	1.64	54.0	1.27	1.67	33.4
Factor H CCP modules 12 to 13	SASDA25	1.70	1.80	7.0	1.73	1.67	−13.0	1.56	1.66	7.0
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	1.05	1.54	15.0	1.06	1.66	11.6	1.05	1.65	14.0
ComE–comcde complex	SASDAB7	1.30	1.80	22.0	1.25	1.64	21.3	1.21	1.62	16.0
LytTR–comcde complex	SASDAC7	1.18	1.40	33.0	1.17	1.63	30.3	1.26	1.58	19.4
Myomesin-1	SASDAK5	1.59	1.40	7.0	1.65	1.66	−0.1	1.82	1.65	9.4
IcsS, IscU and CyaY dimeric complex	SASDA27	2.87	1.80	10.0	4.10	1.54	10.3	2.81	1.62	8.0
Iron–sulfur cluster assembly scaffold IscU monomer	SASDAW6	1.19	1.80	13.0	1.24	1.54	8.0	1.25	1.57	10.7
Geminin–Cdt1 2:1 heterotrimer	SASDAV3	1.98	1.40	75.0	2.55	1.70	54.0	2.61	1.68	33.4
Geminin–Cdt1 4:2 heterohexamer	SASDAW3	3.35	1.80	0.0	3.16	1.69	−12.4	2.94	1.69	0.0
Apo XMRV RT	SASDAV5	1.12	1.46	0.0	1.19	1.54	7.7	1.13	1.66	2.3
XMRV RT + DNA/RNA hybrid	SASDAW5	0.92	1.80	3.0	0.87	1.69	6.6	0.91	1.68	8.7
Protein CyaY monomer	SASDAX6	1.15	1.78	20.0	1.21	1.65	13.5	1.22	1.63	33.4
Annexin-A4	SASDAJ5	5.16	1.80	25.0	5.26	1.68	15.8	5.09	1.67	16.7
Pyruvate decarboxylase	SASDAX2	0.81	1.40	3.0	0.82	1.62	1.6	0.83	1.61	7.3
Average		2.58	1.64 ± 0.18	22.4 ± 21.7	2.72	1.64 ± 0.05	16.6 ± 22.2	2.45	1.64 ± 0.05	18.4 ± 11.2

hydration shell $\delta\rho$ lies between those computed by *CRYSOL* and *FoXS*; however, the standard deviation is significantly lower, $\delta\rho$ = 18.4 ± 11.2 e nm^{-3}. Supplementary Table S1 demonstrates that allowing a negative contrast of the hydration shell or a larger width of the search window in the $\delta\rho$ parameter provides slightly better fits to the experimental

profiles. However, this choice of adjustable parameters might overfit the actual experimental data.

4. Discussion

In this section, we will first discuss the general considerations that affect the performance of our method, and we will then also examine the particular details of the implementation. Given the number of atoms in a model N, and the number of points in the scattering profile M, our *Pepsi-SAXS* approach achieves the asymptotically best performance, in terms of N and M, of $O(N + M)$. Methods with such a performance are generally considered as a class of linear scaling methods. However, when speaking about asymptotic complexity, we should also take into account the effect of the expansion order L, such that the complexity of our method becomes $O(L^3N + M)$. In principle, according to (14), the optimal expansion order L is proportional to the linear size of the molecular model, and thus L^3 is proportional to the number of atoms in the model N. Therefore, effectively, the asymptotic complexity of the *Pepsi-SAXS* method, as well as the three-dimensional Zernike and some other multipole expansion-based approaches, reads $O(N^2 + M)$, with the same leading term $O(N^2)$ as in the complexity of the binning techniques for the Debye equation, for example that implemented in *FoXS*. We should also note that the complexity of the recent version of *CRYSOL* for typical values of L and M is higher compared with the complexity of *Pepsi-SAXS* and the three-dimensional Zernike approach, since *CRYSOL* computes scattering amplitudes at a fixed number of scattering points. Regardless of the $O(N^2)$ asymptotic complexity of our method, it has a very low prefactor compared with Debye-based techniques, and thus it is much faster in practice. Also, *Pepsi-SAXS* and other multipole expansion-based approaches do not require the single-Gaussian approximation of the form factors, which increases the accuracy of these methods.

One of the factors that strongly affects the quality of the scattering model is the description of the hydration shell. Our method (see Fig. 1) uses a grid-based approximation, with the effective width varying from 3 to 5 Å depending on the size of the modelled molecule. The maximum allowed width of 5 Å is somewhat larger compared with the value used in *CRYSOL* and some other methods. However, we use a smaller search window for the value of the relative contrast of the hydration shell compared with *CRYSOL* and *FoXS*, for example. Also, this value agrees with more recent developments, for example, with the hydration-shell intensities computed from MD trajectories (Park *et al.*, 2009), where a width of 7 Å or larger is suggested. Finally, a recent X-ray reflectivity study suggests a hydration-gap width of up to 9 Å (Chattopadhyay *et al.*, 2010). Our representation of the hydration shell describes internal water cavities and surface pockets more accurately compared with the models built using a two-dimensional angular envelope function, for example that in *CRYSOL*. Figs. 2(*a*) and 2(*b*) demonstrate the overall superior quality of modelled scattering profiles built with *Pepsi-SAXS* for models with complex shapes. On the other hand, our hydration model is more prone to overfitting. We should mention that the shortcomings of the simplified two-dimensional angular representation of the hydration layer were discussed in the original *CRYSOL* paper (Svergun *et al.*, 1995), where it was claimed that it still correctly describes the outer hydration shell.

Another factor that influences the quality of the scattering model is the description of the excluded solvent. Our model of the excluded solvent is derived from the approximation introduced by Fraser *et al.* (1978) and closely follows the model used in *CRYSOL*. More precisely, we use the first-order Maclaurin expansion in the scaling parameter (12) of the model of Fraser and coworkers. We have also run tests using a more detailed second-order expansion approximation; however, it did not provide any notable improvement in the model quality. The zero-order q-independent approximation, surprisingly, does not change the goodness of fit either. We attribute these marginal differences between the models of the excluded solvent to a relatively high level of noise of the current profiles at large scattering angles.

The goodness of fit between the modelled intensities and the experimental profiles is influenced by the number of fitting parameters in the modelled scattering profile. The *Aqua-SAXS*, *FoXS* and *SAStbx* methods effectively use the same adjustable parameters as *Pepsi-SAXS* and *CRYSOL*. The method of Zheng & Tekpinar (2011) uses a single adjustable parameter: the relative contrast of the hydration shell. The *AXES* procedure replaces the atomic radii multiplier and the total excluded volume, which are applied to the model data, by the solvent/buffer rescaling factor and the constant offset applied to the measured data. Thanks to the partial scattering intensities (see equations 1 and 2 in the Supporting Information), the computation of the scattering profile in *Pepsi-SAXS* takes only a constant time. Thus, we are able to search for the optimal values of the adjustable parameters exhaustively on a regular grid. Some other methods, however, use minimization techniques for this purpose, such as the least-square minimization with boundaries (Petoukhov *et al.*, 2012), the Powell minimization (Grishaev *et al.*, 2010) *etc.*

We should mention that the optimization of the free parameters is very important for the quality of the resulting fit. For example, if we decrease the difference in the density of the hydration shell with respect to the bulk water, the resulting mean χ for the two benchmarks increases by more than a factor of two (see Supplementary Table S1). Alternatively, if we assume a constant difference in the density of the hydration shell, $\mathrm{d}\rho = 18$ e nm^{-1}, the obtained mean χ increases by about one third (see Supplementary Table S1). Searching for the fitting parameters less exhaustively on a grid of 10×10 gives a reasonably good fit, with a mean χ that is higher by about 4% compared with the reference value (see Supplementary Table S2). However, if we perform the search more exhaustively using a ten times denser grid of 1000×1000 (see Supplementary Table S2), the mean χ value does not change.

Regarding the adaptive choice of the multipole expansion order, Table 8 clearly demonstrates the advantage of our adaptive technique in terms of model quality. We should also

mention that, on average, the timing of *Pepsi-SAXS* with the adaptive expansion order increases only marginally with respect to the timing with the fixed order of 15 used by default in *CRYSOL*. Fig. 2(*c*) shows an example when the adaptivity significantly improves the quality of the model. It is important to mention that the the use of the cubic spline makes computations much faster, but does not affect the overall quality of fit, as can be seen from Supplementary Table S1.

We should note that the sampling theorem has been widely used in scattering studies for model-quality estimation and data reduction (Rambo & Tainer, 2013*a*; Feigin *et al.*, 1987). Thorough studies of the analysis of information content in small-angle scattering curves using the sampling theorem have been conducted by Moore (1980) and Taupin & Luzzati (1982). The sampling theorem states that the scattering-intensity profile $I(q)$ can be fully defined by its values at $d_{\max}q_{\max}/\pi$ points, where $d_{\max}$ is the maximum extension of the atomistic model. In our case, however, we use the sampling theorem to estimate the maximum expansion order of the spherical harmonics basis, $L_{\max} = d_{\max}q_{\max}/2$, and thus our estimation differs from that stated above by a factor of $\pi/2$. Strictly speaking, the estimation of the maximum expansion order has to be related to the maximum extension $d_{\max}$ rather than R_g, as is stated in (14). However, we discovered that we can relax this estimation to a typically smaller number using the radius of gyration of the solvation shell instead of the maximum extension, such that the quality of fit does not degrade, whereas the running time of *Pepsi-SAXS* becomes about three times smaller.

We have also discovered a drastic effect of form factors for the charged groups on the quality of the scattering model. For example, introducing the explicit form factors for the charged resonance guanidine groups decreased the mean value of χ by about 3%. Other form factors for the charged groups also affect the model quality favourably, although not as much as those of guanidine (see Supplementary Table S2). It is also very important to correctly account for the H atoms, which are rarely present in PDB models. Supplementary Table S2 lists the results of the modelled scattering profiles computed with the approximation that no H atoms are present in the model. As one can see, such a rude approximation leads to a resulting mean χ that is increased by about 13%.

Finally, we would like to mention that small values of χ do not necessarily indicate a good scattering model. This is because proteins in solution can be flexible or conformationally heterogeneous, their structural models can be imprecise, different fitting methods use different ranges of adjustable parameters and different models of the hydration shell *etc*. As Rambo & Tainer (2013*b*) recently suggested, a reference data set should be established for the evaluation and development of SAXS algorithms. This will allow a thorough evaluation of existing computational methods, which falls outside the scope of the current work. Nonetheless, *Pepsi-SAXS* has already been successfully applied to select putative near-native conformations of the Tb_bVSP1 complex (Jamwal *et al.*, 2015) from 10 000 models generated with the *Ensemble Optimization Method* (Tria *et al.*, 2015). The best selected conformation overlaid well with the *ab initio* model calculated using *DAMAVER* (Volkov & Svergun, 2003).

5. Conclusion

We present a new method called *Pepsi-SAXS* that calculates small-angle X-ray scattering profiles from atomistic models. Our method is based on the multipole expansion scheme and has the following distinct features. Firstly, we use a very fast model for the scattering contribution from the hydration shell based on a uniform grid of points. Secondly, we use the adaptive resolution of the multipole expansion estimated according to the Nyquist–Shannon–Kotelnikov sampling theorem. Thirdly, we represent the scattering-intensity curve using a cubic spline interpolation, which allows us to significantly speed up the running time of our method. Fourthly, we introduce partial scattering intensities to rapidly fit the modelled profiles to the experimental data using an exhaustive search in two adjustable parameters. Finally, we introduce individual form factors for charges and resonance groups, which increase the quality of the modelled scattering profiles.

To demonstrate the accuracy and efficiency of the *Pepsi-SAXS* method, we systematically validated it with seven series of tests using a large number (more than 50) of experimental profiles collected from the BioIsis and SASBDB databases. Overall, the *Pepsi-SAXS* method is significantly faster compared with the *CRYSOL*, *FoXS* and *SAStbx* (with the three-dimensional Zernike option) methods with scattering profiles that are of the same quality on average. Using a laptop, we demonstrated that *Pepsi-SAXS* is about seven, 29 and 36 times faster compared with *CRYSOL*, *FoXS* and *SAStbx*, respectively, when tested on data from the BioIsis database, and is about five, 21 and 25 times faster compared with *CRYSOL*, *FoXS* and *SAStbx*, respectively, when tested on data from SASBDB. We claim that we avoid overfitting of experimental profiles owing to a very much smaller allowed variation of the adjustable parameters.

Notably, we argue that the multipole expansion-based schemes (such as *Pepsi-SAXS*, *CRYSOL* and the three-dimensional Zernike approach) have the same quadratic dependence on the number of atoms in the model as the Debye approach when using the adaptive resolution of the multipole expansion. Thus, these schemes are only more efficient owing to a much smaller prefactor in the complexity of these methods. *Pepsi-SAXS* is available at http://team.inria.fr/nano-d/software/pepsi-saxs. We have also developed a GUI interface for *Pepsi-SAXS* that will be distributed as a plugin for the modular software platform *SAMSON* at https://www.samson-connect.net.

Acknowledgements

The authors thank Dr Adam Round from EMBL Grenoble, France for critical assessment of the method; Dr Robert P. Rambo from Diamond Light Source, Didcot, England for his help with data entries from the BioIsis database; Dr Fabrice Rastello from Inria Grenoble, France for advice on code

optimization; Andrey Krivoy from MIPT Moscow for some initial developments in the code; Dr Valentin Gordeliy from IBS Grenoble, France and FZJ Jülich, Germany and Dr Jose Teixeira from LLB CEA Saclay, France for their comments on the density of the hydration layer of a protein; and, finally, Dr David W. Ritchie from Inria Nancy, France for suggesting the *Pepsi-SAXS* acronym. This work was supported by the Agence Nationale de la Recherche (grant ANR-11-MONU-006-01). It was also partially supported by the Helmholtz–Russia Joint Research Group (Russian Foundation for Basic Research research project 13-04-91320 and Helmholtz Association of German Research Centres project ID HRJRG-401). MG and AK thank the 5TOP100 program of the Ministry for Education and Science of the Russian Federation.

References

Allen, F., Watson, D., Brammer, L., Orpen, A. & Taylor, R. (2004). *International Tables for Crystallography*, Vol. *C*, edited by E. Prince, pp. 790–811. Dordrecht: Springer. https://doi.org/10.1107/9780955360206000621.

Artemova, S., Grudinin, S. & Redon, S. (2011). *J. Comput. Chem.* **32**, 2865–2877.

Banyard, K. E. & March, N. H. (1961). *Acta Cryst.* **14**, 357–360.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Chattopadhyay, S., Uysal, A., Stripe, B., Ha, Y., Marks, T. J., Karapetrova, E. A. & Dutta, P. (2010). *Phys. Rev. Lett.* **105**, 037803.

Feigin, L., Svergun, D. I. & Taylor, G. W. (1987). *Structure Analysis by Small-angle X-ray and Neutron Scattering.* New York: Springer.

Fraser, R. D. B., MacRae, T. P. & Suzuki, E. (1978). *J. Appl. Cryst.* **11**, 693–694.

Graewert, M. A. & Svergun, D. I. (2013). *Curr. Opin. Struct. Biol.* **23**, 748–754.

Grishaev, A., Guo, L., Irving, T. & Bax, A. (2010). *J. Am. Chem. Soc.* **132**, 15484–15486.

Gumerov, N. A., Berlin, K., Fushman, D. & Duraiswami, R. (2012). *J. Comput. Chem.* **33**, 1981–1996.

Harker, D. (1953). *Acta Cryst.* **6**, 731–736.

Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L. II, Tsutakawa, S. E., Jenney, F. E. Jr, Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S., Scott, J. W., Dillard, B. D., Adams, M. W. W. & Tainer, J. A. (2009). *Nature Methods*, **6**, 606–612.

Iiyama, T., Nishikawa, K., Otowa, T. & Kaneko, K. (1995). *J. Phys. Chem.* **99**, 10075–10076.

Jamwal, A., Round, A. R., Bannwarth, L., Bryan, C. V., Belrahli, H., Yogavel, M. & Sharma, A. (2015). *J. Biol. Chem.* **290**, 30498–30513.

Liu, H., Hexemer, A. & Zwart, P. H. (2012). *J. Appl. Cryst.* **45**, 587–593.

Liu, H., Poon, B. K., Janssen, A. J. E. M. & Zwart, P. H. (2012). *Acta Cryst.* A**68**, 561–567.

Marks, R. J. II (2008). *Handbook of Fourier Analysis and its Applications.* Oxford University Press.

Merzel, F. & Smith, J. C. (2002). *Acta Cryst.* D**58**, 242–249.

Mezger, M., Reichert, H., Ocko, B. M., Daillant, J. & Dosch, H. (2011). *Phys. Rev. Lett.* **107**, 249801.

Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.

Park, S., Bardhan, J. P., Roux, B. & Makowski, L. (2009). *J. Chem. Phys.* **130**, 134114.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.

Poitevin, F., Orland, H., Doniach, S., Koehl, P. & Delarue, M. (2011). *Nucleic Acids Res.* **39**, W184–W189.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press.

Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Q. Rev. Biophys.* **40**, 191–285.

Rambo, R. P. & Tainer, J. A. (2013*a*). *Nature (London)*, **496**, 477–481.

Rambo, R. P. & Tainer, J. A. (2013*b*). *Annu. Rev. Biophys.* **42**, 415–441.

Schneidman-Duhovny, D., Hammel, M. & Sali, A. (2010). *Nucleic Acids Res.* **38**, W540–W544.

Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. (2013). *Biophys. J.* **105**, 962–974.

Spilotros, A. & Svergun, D. I. (2014). *Encyclopedia of Analalytical Chemistry*, pp. 1–34. New York: Wiley. https://doi.org/10.1002/9780470027318.a9447.

Stovgaard, K., Andreetta, C., Ferkinghoff-Borg, J. & Hamelryck, T. (2010). *BMC Bioinformatics*, **11**, 429.

Stuhrmann, H. B. (1970*a*). *Z. Phys. Chem.* **72**, 177–184.

Stuhrmann, H. B. (1970*b*). *Acta Cryst.* A**26**, 297–306.

Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 485–492.

Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.

Taupin, D. & Luzzati, V. (1982). *J. Appl. Cryst.* **15**, 289–300.

Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. (2015). *IUCrJ*, **2**, 207–217.

Uysal, A., Chu, M., Stripe, B., Timalsina, A., Chattopadhyay, S., Schlepütz, C. M., Marks, T. J. & Dutta, P. (2013). *Phys. Rev. B*, **88**, 035431.

Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. (2015). *Nucleic Acids Res.* **43**, D357–D363.

Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.

Waasmaier, D. & Kirfel, A. (1995). *Acta Cryst.* A**51**, 416–431.

Watson, M. C. & Curtis, J. E. (2013). *J. Appl. Cryst.* **46**, 1171–1177.

Webb, B. & Sali, A. (2014). *Curr. Protoc. Bioinform.* **47**, 5.6.1–5.632.

Yang, S., Park, S., Makowski, L. & Roux, B. (2009). *Biophys. J.* **96**, 4449–4463.

Zheng, W. & Tekpinar, M. (2011). *Biophys. J.* **101**, 2981–2991.