Hierarchical $O(N)$ Computation of Small-Angle Scattering Profiles and their Associated Derivatives

Konstantin Berlin[*], Nail A. Gumerov[†], David Fushman[‡], and Ramani Duraiswami[§]

University of Maryland, College Park

May 07, 2012

Abstract

Fast algorithms for Debye summation, which arises in computations performed in crystallography, small/wide-angle X-ray scattering (SAXS/WAXS) and small-angle neutron scattering (SANS), were recently presented in Gumerov et al. (J. Comput. Chem., 2012, **33**, 1981).

[*]Equal first author, Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, and Institute for Advanced Computer Studies; all at the University of Maryland, College Park, MD; email:kberlin@umd.edu web:https://sites.google.com/site/kberlin/

[†]Equal first author, Institute for Advanced Computer Studies, University of Maryland, College Park, MD; and at Fantalgo, LLC, Elkridge, MD; email: gumerov@umiacs.umd.edu; web:http://www.umiacs.umd.edu/users/gumerov

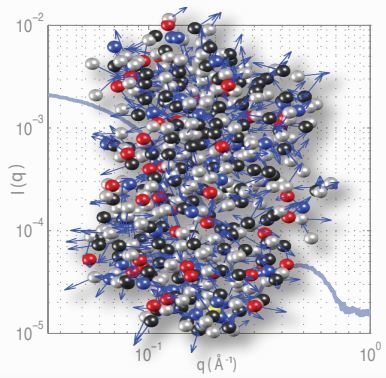
[‡]Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, and Institute for Advanced Computer Studies, all at University of Maryland, College Park, MD; email:fushman@umd.edu; web:http://www.chem.umd.edu/research/facultyprofiles/davidfushman

[§]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD; and at Fantalgo, LLC; email: ramani@umiacs.umd.edu; web:http://www.umiacs.umd.edu/users/ramani

The use of these algorithms can speed up computation of scattering profiles in macromolecular structure refinement protocols. However, these protocols often employ an iterative gradient-based optimization procedure, which then requires derivatives of the profile with respect to atomic coordinates. An extension to one of the algorithms is presented which allows accurate, $O(N)$ cost computation of the derivatives along with the scattering profile. Computational results show orders of magnitude improvement in computational efficiency, while maintaining prescribed accuracy. This opens the possibility to efficiently integrate small-angle scattering data into structure determination and refinement of macromolecular systems.

Keywords: Small Angle Scattering (SAS), Wide Angle Scattering (WAS), Jacobian, Gradient

■



SAXS based structure refinement protocols iteratively adjust the posited positions of atoms in a structure via a gradient based constrained optimization to ensure that experimental scattering profile data is matched. This requires computation of the profile and its gradient with respect to atomic coordinates at every step. The fast algorithms proposed in this paper reduce these to a cost that is linear in the number of atoms for any specified accuracy.

1 Introduction

Accurate characterization of biomolecular structures in solution is required for understanding their biological function and for therapeutic applications. Small-angle scattering (SAS) of X-rays and

neutrons can indirectly measure the distribution of interatomic distances of a molecule in solution, providing a set of structural restraints [1]. As a result, solution SAS studies have become increasingly used in structural biology, with a broad range of applications including structure refinement of biological macromolecules and their complexes [2, 3, 4, 5], analysis of conformational ensembles and flexibility in solution [6, 7, 8, 9], and high-throughput structural studies [10, 11].

In order to use SAS experimental data as an atomic-level structural restraint, the SAS profile needs to be predicted *ab initio* from the assumed molecular atomic structure, which requires computing all-pairs interactions of the atoms in the molecule (also referred to as an N-body problem). In addition, this computation must be performed numerous times in an iterative structure refinement algorithm [2, 12] or for a high-throughput structural analysis.

Several approximation methods have been proposed to speed up this computation [13, 14, 12, 15]. However, depending on the size of the molecule, the approximations can introduce significant errors [16]. Recently we developed and demonstrated a hierarchical harmonic expansion method, based on the fast multipole method (FMM), which has superior asymptotical performance than previously proposed approximation methods, while maintaining any prescribed accuracy [16]. This large speedup in computation has the potential to significantly improve integration of SAS into structure refinement protocols, like Xplor-NIH [17] and HADDOCK [18], where the scattering profile needs to be computed thousands of times during iterative structure refinement.

A potential problem with integrating SAS data into a structure refinement protocol is the need to accurately and quickly compute the gradient ("force") of the SAS profile, in addition to the actual SAS profile [19]. In order to save on computation time it was suggested in [20] to approximate the low-frequency part of the SAS profile using Taylor expansion. The drawback of this approach is that significant structural information contained in the higher-frequency region of the scattering profile is ignored. Furthermore, the low-frequency profile region becomes smaller as the molecule size increases, since this approximation depends on the magnitude of the product of the wavenumber and molecule size. An alternative approach was suggested in [3], where

the gradient was approximated by taking the derivative of the scattering amplitude explicitly, and then averaged over a large number of orientations. However, since no error guarantees are given, this approximation might result in an incorrect gradient direction, significantly slowing down any derivative-based structure optimization method. In addition to the computational complexity of the derivative computation, macromolecules can be combined with a solvation layer, to improve the accuracy of the SAS profile prediction [12]. These additional water molecules, especially the large number of atoms that result from discretizing a continuous water model, can significantly increase the computation time needed for estimation of the SAS profile and its derivatives [21, 15, 22].

Here we propose an extension to the method proposed in [16] to achieve an accurate and fast method for the simultaneous computation of the full scattering profile and associated gradients. The method completely avoids computationally intractable all-to-all atom summations. We show that it is possible to quickly and accurately compute the gradient of the SAS profile using the algorithm developed here so that the computation only slightly adds to the overall computation time compared to just a straight computation of the profile in [16]. Furthermore, leveraging the computational advantages of the hierarchical approach, we demonstrate that the profile and derivatives for a macromolecule solvated in a very dense water grid can be computed many orders of magnitude faster than by current methods.

2 Method

The SAS profile value of an N atom system such as a macromolecule and associated solvent molecules can be computed as

$$I(\mathbf{r}_i, \ldots, \mathbf{r}_N; q) = \left\langle \left| A\left(\mathbf{q}\right) \right|^2 \right\rangle_\Omega , \tag{1}$$

for $q = q_1 \dots q_Q$, where

$$A(\mathbf{q}) = \sum_{j=1}^{N} f_j(q) \exp(i\mathbf{q} \cdot \mathbf{r}_j) \tag{2}$$

is the scattering amplitude from the particle *in vacuo*, $q = |\mathbf{q}|$ is the scattering wavenumber ($q = (4\pi/\lambda)\sin\vartheta$, 2ϑ is the scattering angle, λ the wavelength), $\mathbf{r}_j = [x_j, y_j, z_j]$ is the position of the jth atom, $f_j(q)$ is the form factor of the jth atom, and $\langle\cdot\rangle_\Omega$ indicates averaging of $\mathbf{q}$ over a sphere, to reflect averaging over all possible molecular orientations in solution or powder.

Equation (1) can be analytically averaged, such that

$$\left\langle |A(\mathbf{q})|^2 \right\rangle_\Omega = \sum_{i=1}^{N} f_i(q) \sum_{j=1}^{N} f_j(q) s(qr_{ij}) = \sum_{i=1}^{N} f_i(q) \Psi_i, \tag{3}$$

where s is the sinc function,

$$s(qr_{ij}) = \frac{\sin(q\|\mathbf{r}_i - \mathbf{r}_j\|_2)}{q\|\mathbf{r}_i - \mathbf{r}_j\|_2}, \tag{4}$$

Ψ_i is defied as in Eq. 3, and $\|\cdot\|_2$ is the Euclidean norm.

In order to refine macromolecule structure, uniquely defined by its atomic positions $[\mathbf{r}_i, \dots, \mathbf{r}_N]$, against experimentally collected scattering data, the discrepancy between the experimental scattering, $I^{exp}(q)$ and the predicted scattering, $I(\mathbf{r}_i, \dots, \mathbf{r}_N; q)$, for $q = q_1 \dots q_Q$, is minimized. To resolve the structural degeneracy of SAS, additional energy terms coming from NMR, X-ray crystallography, other structural methods, and force fields restraints may also be used. This refinement is typically formulated as an energy minimization problem,

$$\min_{\mathbf{r}_i, \dots, \mathbf{r}_N} \chi^2(\mathbf{r}_i, \dots, \mathbf{r}_N) + \Lambda(\mathbf{r}_i, \dots, \mathbf{r}_N), \tag{5}$$

where

$$\chi^2(\mathbf{r}_i, \dots, \mathbf{r}_N) = \sum_{k=1}^{Q} \|I(\mathbf{r}_i, \dots, \mathbf{r}_N; q_k) - I^{exp}(q_k)\|_2^2 = \|\mathbf{d}\|_2^2 = \sum_{k=1}^{Q} d_k^2, \tag{6}$$

is the discrepancy of between the predicted and the experimental scattering profile, and $\Lambda(\mathbf{r}_i, \dots, \mathbf{r}_N)$

is the weight of all the other structural restraints, as well as potential regularization terms. In the sequel we abbreviate $I(\mathbf{r}_i, \ldots, \mathbf{r}_N; q_k)$ as I_k.

The cost of computing the profile I in Eq. ((1)) requires the sums in Eq. ((3)), which has $O(N^2)$ computational cost, if done directly. A widely used computer program, CRYSOL [13], uses truncated harmonic expansions to speed up the computations. In [16] it was shown that without an associated error bound, the harmonic expansion methods are likely to give incorrect results, especially at large q. Two fast algorithms, based on hierarchical decomposition of the spatial domain, local spherical expansion and translation, were presented in the paper. One of these algorithms, is used in this paper to develop an extended algorithm, which can for a slight increase in cost, compute the derivatives of the profile with respect to the Cartesian atomic coordinates.

2.1 Computation of Derivatives

We focus on ϵ-accurate computation of the gradient of $I(q)$, sometimes referred to as the "force" of $I(q)$, which is required for a local convex optimization of Eq. (5) using a Newton-type minimization, such as done in Xplore-NIH (as well as other programs). Computation of the Hessian is typically avoided, either by an iterative approximations of the Hessian using a quasi-Newton method, or in the case of least-squares, an approximation based on the Jacobian [23].

Assuming the derivatives of Λ are already computed, based on the linearity of differentiation, in order to compute the gradient of Eq. (5) we only need to compute the Jacobian of $I(q)$,

$$\mathbf{J} = \begin{pmatrix} \frac{\partial I_1}{\partial x_1} & \frac{\partial I_1}{\partial y_1} & \frac{\partial I_1}{\partial z_1} & \cdots & \frac{\partial I_1}{\partial x_N} & \frac{\partial I_1}{\partial y_N} & \frac{\partial I_1}{\partial z_N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial I_Q}{\partial x_1} & \frac{\partial I_Q}{\partial y_1} & \frac{\partial I_Q}{\partial z_1} & \cdots & \frac{\partial I_Q}{\partial x_N} & \frac{\partial I_Q}{\partial y_N} & \frac{\partial I_Q}{\partial z_N} \end{pmatrix}, \tag{7}$$

where $\mathbf{J}$ is a $Q \times 3N$ matrix. From the Jacobian, the gradient can be computed as

$$\nabla \chi^2 = 2\mathbf{J}^T\mathbf{d}, \tag{8}$$

and the Hessian can be approximated as

$$H \approx 2\mathbf{J}\mathbf{J}^T + 2\lambda\mathbf{I}, \tag{9}$$

where $\mathbf{I}$ is the identity matrix, for some $\lambda \in [0, 1]$, in the case of a least-squares method. In case of an iterative quasi-Newton minimization of equation Eq. (5), only equation Eq. (8) is needed, and the Hessian is iteratively updated.

Without loss of generality for $\frac{\partial I_k}{\partial y_i}$ and $\frac{\partial I_k}{\partial z_i}$,

$$\frac{\partial I_k}{\partial x_i} = 2f_i(q_k)\Psi_i' = 2f_i(q_k)\sum_{j=1}^{N} f_j(q_k)s'(q_k r_{ij}), \tag{10}$$

where $s'(q_k r_{nj})$ and Ψ_i' are the partial derivatives with respect to x_i. The derivative $s'(q_k r_{nj})$ can be analytically computed as

$$s'(q_k r_{ij}) = (x_i - x_j)\left[\frac{\cos(q_k r_{ij})}{r_{ij}^2} - \frac{\sin(q_k r_{ij})}{q_k r_{ij}^3}\right], \tag{11}$$

for $r_{ij} \neq 0$. Therefore, direct computation of $\mathbf{J}$ (and $\nabla\chi^2$) has complexity of $O(N^2Q)$.

The N^2 computational complexity can be avoided if we first expand Ψ_i using an expansion in terms of spherical basis functions $R_n^m(\mathbf{r})$ as

$$\Psi_i = \sum_{m=-n}^{m=n} R_n^m(\mathbf{r}_i)B_n^m(\mathbf{r}_j) + \epsilon_p, \tag{12}$$

and

$$B_n^m(\mathbf{r}_j) = 4\pi\sum_{j=1}^{N} f_j(q)R_n^{-m}(\mathbf{r}_j) \tag{13}$$

for all i, where R_n^m is a regular spherical basis function

$$R_n^m(\mathbf{r}) = j_n(qr)Y_n^m(\mathbf{s}) = j_n(qr)Y_n^m(\theta,\varphi), \quad \mathbf{r} = r\mathbf{s}. \tag{14}$$

Functions $R_n^m(\mathbf{r})$ here are given in spherical coordinates (r,θ,φ), $\mathbf{r} = r(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$, where $j_n(qr)$ are spherical Bessel functions, while $Y_n^m(\theta,\varphi)$ are the normalized spherical harmonics of degree n and order m,

$$Y_n^m(\theta,\varphi) = (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)e^{im\varphi}, \tag{15}$$
$$n = 0,1,2,...; \qquad m = -n,...,n.$$

where $P_n^{|m|}(\mu)$ are the associated Legendre functions consistent with that in [24], or Rodrigues' formulae

$$P_n^m(\mu) = (-1)^m\left(1-\mu^2\right)^{m/2}\frac{d^m}{d\mu^m}P_n(\mu), \quad n \geqslant 0, \quad m \geqslant 0, \tag{16}$$
$$P_n(\mu) = \frac{1}{2^n n!}\frac{d^n}{d\mu^n}\left(\mu^2-1\right)^n, \quad n \geqslant 0,$$

where $P_n(\mu)$ are the Legendre polynomials.

Here p and ϵ_p are the cutoff degree and associated error respectively. This expansion can be quickly computed for Ψ_i, for all $i = 1,\ldots,N$, using the hierarchical algorithm in $O(p^3\log(p) + N\log N)$ time [16]. However, unlike the basic case of straight computation of $I(q)$, a downward pass is now required because individual $R_n^m(\mathbf{r}_i)$ values also need to be computed for the derivatives.

Once Eq. (12) is computed, taking the derivative of Eq. (12) yields

$$\Psi_i' = \sum_{n=0}^{p-1}\sum_{m=-n}^{m=n} R'^m_n(\mathbf{r}_i)B_n^m(\mathbf{r}_j) + \hat{\epsilon}_p, \tag{17}$$

where the derivative of the spherical function, R'^m_n, has been derived in [26], and is given in the Supplementary Information section. This derivative can be written as a weighted linear combination of two to four adjacent spherical basis functions (depending on the particular spatial derivative needed), where the weights are independent of $\mathbf{r}_i$, and accordingly can be expressed as the product of a sparse matrix with the expansion coefficients. Therefore,

$$\Psi'_i = \sum_{n=0}^{p-1} \sum_{m=-n}^{m=n} R^m_n(\mathbf{r}_i)\alpha^m_{(j)n} + \hat{\epsilon}_p, \tag{18}$$

where $\alpha^m_{(j)n}$ can be quickly computed by a sparse $O(p^2)$ matrix multiplication, and is the same for all Ψ'_i. As noted in the original reference, the error bound for derivative computations require slightly higher truncation numbers, which are one or two higher than those for the expansion of Ψ.

Thus the overall complexity of computing $\mathbf{J}$, given that the harmonic expansion has already been computed for all Ψ_i for all q, is $O(p^2NQ)$. The proper value of p for a specific value of q is proportional to qD, where D is the diameter of the molecule (largest distance between any two atoms) [16]. Therefore, the overall complexity of the gradient computation is $O(Q^3D^2N)$, for an arbitrary error bound.

3 Results

The computational complexity of the hierarchical algorithm is dependent on the atomic density of the molecule. This complexity also extends to the computation of the Jacobian, since since it is directly dependent on speed of the spherical expansion of $I(q)$. To specifically quantify the advantages of our approach for biological macromolecules, we compare the computational speed of our algorithm to direct all-to-all computation of eqs. (3,10) ("Direct"), vs. spherical harmonic expansion method like CRYSOL [13], the properly error truncated version, "Middleman" [16], and the hierarchical method ("Hierarchical," which combines the above expressions for the derivative with

the algorithm in [16]) for various sized random and actual macromolecules, solvated at different water layer densities.

3.1 Derivatives

We ran the benchmarks on a Dual Quad-Core Intel Xeon X5560 CPU @ 2.80GHz 64bit Linux machine with 24GB ECC DDR3 SDRAM. All programs were compiled using the Intel 11.1 compiler using the "-parallel -O3 -ipo" flags. We timed the computational time for generating a SAS profile made out of 50 uniformly spaced evaluations of the profile on the range $0.01 \leq q \leq 0.5$ Å^{-1}, for a randomly generated molecule with atomic density of $d = 0.02$ Å^{-3}. The timing and accuracy results for the randomly generated molecules are presented in Fig. 1.

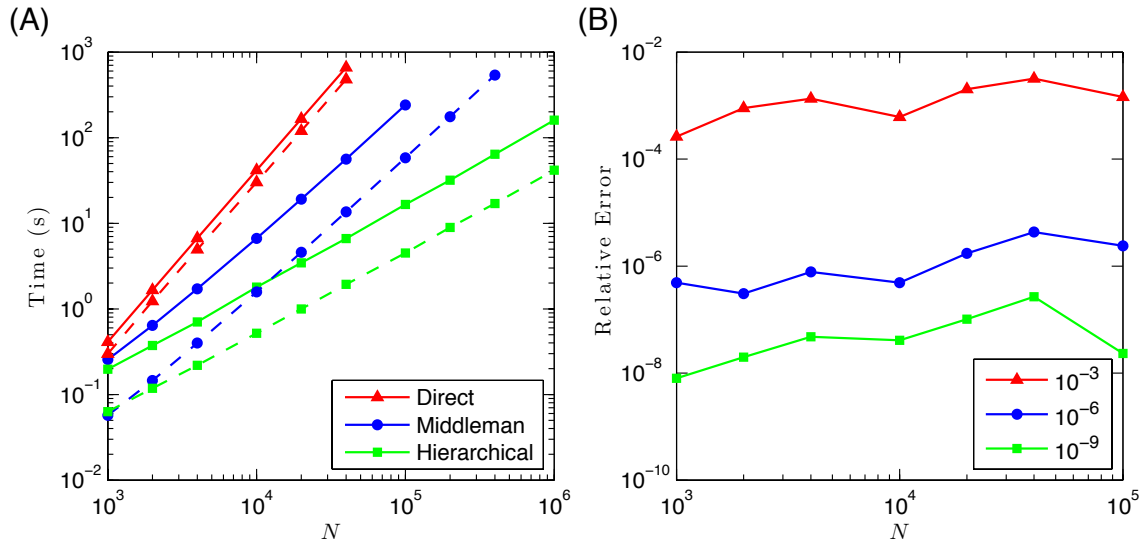


Figure 1: Results for randomly generated proteins, timing for uniformly distributed 50 point scattering profile, $0 < q \leq 0.5$ Å^{-1}. (A) Computation time for just the SAS profile (dashed), and the SAS profile with Jacobian (solid). (B) Maximum relative error in the Jacobian, as measured by $\max_i \|\mathbf{J}_{i,*} - \hat{\mathbf{J}}_{i,*}\|_2 / \|\hat{\mathbf{J}}_{i,*}\|_2$, where $\hat{\mathbf{J}}$ is the true Jacobian, and $\mathbf{J}_{i,*}$ is the ith row of $\mathbf{J}$.

Fig. 1A shows that computation of the derivatives does not change the asymptotic properties of any of the algorithms. For the hierarchical method, the computation of the Jacobian increases the computation time by about 3.5. This increase is partially because of the addition of a downward

pass in our hierarchical method. The hierarchical method, with or without Jacobian computation, is faster than the CRYSOL-like "Middleman" algorithm, as well as direct computation. Fig. 1B shows that the accuracy of the Jacobian computation is within an order of magnitude of the prescribed accuracy for $I(q)$.

3.2 Water Layer

We have demonstrated the computational improvement of our algorithm, compared to the direct and middleman algorithm, at approximate atomic density of proteins. However, one of the principal advantages of our hierarchical method is that we decouple the number of atoms from molecule's diameter in the computation. This gives us a large computational advantage over previous approaches when computing SAS profile of dense objects. We now show that for macromolecules solvated in a dense water grid layer, such that the overall atomic density increases without significantly increasing the molecule's diameter, the speedup of our hierarchical algorithm becomes even larger.

We choose two proteins, archaeal peroxiredoxin (PDB 2E2G) and diubiquitin (PDB 1AAR), that have very different atomic density (distribution of inter-atomic distances) to illustrate the speedup the algorithm provides. Archaeal peroxiredoxin has a large empty cavity in the center (see Fig. 2). The number of atoms for PDB 1AAR increases from 2575 atoms, with no water layer, to several orders of magnitude larger 548462 atoms for 0.5 Å water layer grid. For PDB 2E2G the number of atoms increases from 39609, to 543832 atoms for a slightly coarser minimum 1.0 Å water layer grid. Good speedup of our algorithm compared to previously published algorithms can be seen in Fig. 3.

Fig. 3 demonstrates that not only is our hierarchical approach significantly faster, the speedup gets progressively larger at higher water density.

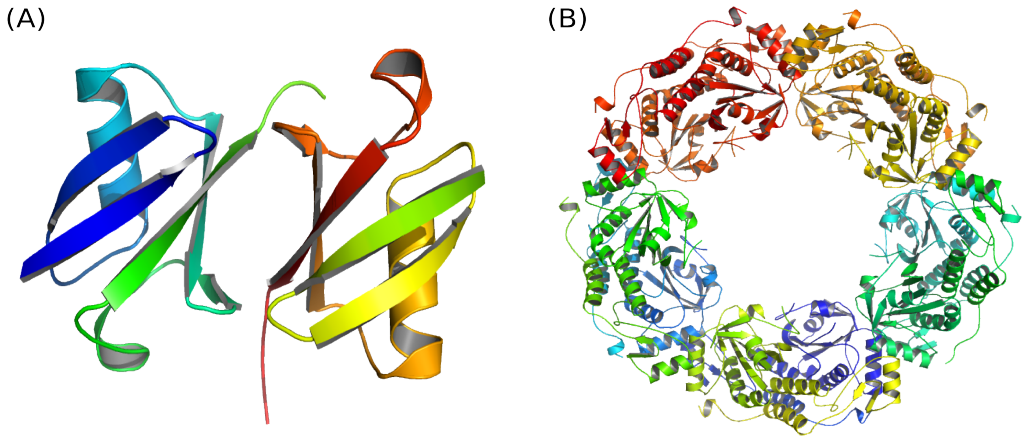


Figure 2: Cartoon representation of (A) PDB 1AAR, (B) PDB 2E2G.

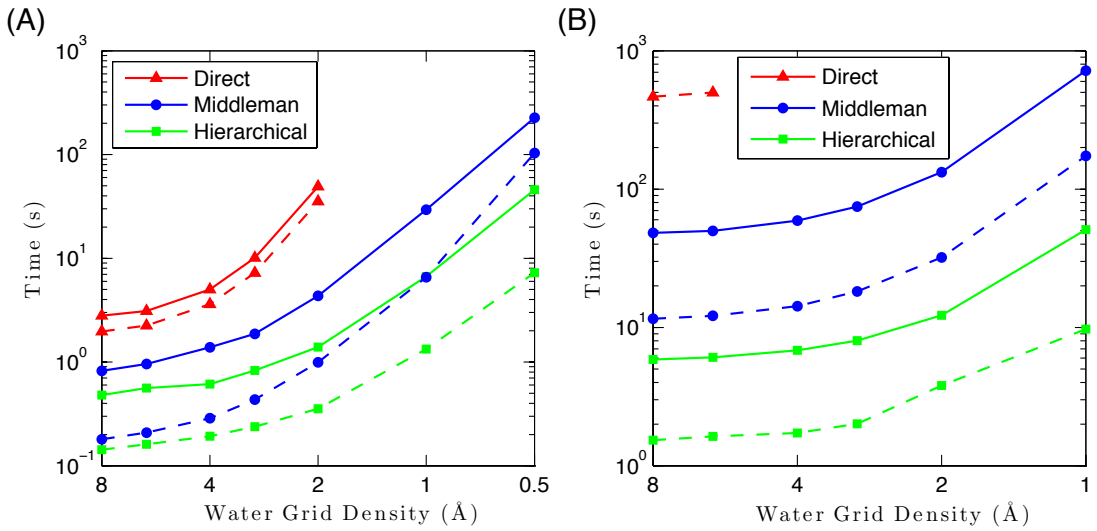


Figure 3: Timing results for computation of the profile (just SAS profile is dashed line, and SAS profile with the Jacobian is solid line), with an 8 Å water layer, at different water grid densities (inter-atomic distances). (A) PDB 1AAR, (B) PDB 2E2G.

4 Discussion

The computational advantage of the algorithm presented for large and dense molecules can be understood in terms of how it relates to the three main alternatives for computing $I(q)$: (i) direct all-to-all summation of Eq. (3); (ii) numerical integration or averaging of Eq. (1), as suggested in

[2, 14, 15]; and (iii) analytical expansion of Eq. (2) [13, 27]. As discussed in [16], approach (i) has $O(N^2)$ complexity, making it intractable for larger macromolecules, while (ii,iii) approaches are bounded by the bandwidth of Eq. (2), requiring either $O(q^2D^2)$ function evaluations or basis functions in order to accurately approximate the scattering profile, resulting in $O(q^2D^2N)$ computational complexity. Additionally, a spherical harmonic series expansion of $A(\mathbf{q})$ is not sparse, so it is not possible to accurately approximate Eq. (3) using non-uniform sampling of the sphere.

When considering large and/or dense molecules, where D or N is large, all the three approaches become computationally intractable. However, a large speedup is possible if we compute an approximate series expansion of Eq. (2) faster than in $O(q^2D^2N)$ time. That is the main idea behind our hierarchical method, where only a small expansion is performed locally, where D is small, and then translated hierarchically onto progressively larger domains. This grouping of adjacent atoms together is somewhat similar to the intuitive idea of grouping of atoms by their secondary structure as suggested in [27]. However, because grouping is done at multiple hierarchical levels and combined together using proven theoretical bounds, this paradigm allows much faster computation, can be made arbitrarily accurate, and can better accommodate various macromolecular shapes (including cavities).

As the result of the hierarchical grouping of atoms, our algorithm has $O((qD)^3 \log(qD) + N \log N)$ complexity for each $I(q)$ evaluation, which is a significant theoretical improvement for large molecules and dense water models, because it separates the size of the domain, D, from the number of atoms in the macromolecule, N. In practice, this allows us to accurately compute $I(q)$ for extremely large systems made of millions of atoms or even when the water density is high enough to well approximate a continuous water model. As a byproduct of this computation, we also compute the $O(q^2D^2)$-sized series expansion of Eq. (3). Once this expansion is computed, the derivatives of the minimization function can be quickly computed by taking the derivative of each basis function.

Even though the hierarchical method is faster than direct spherical harmonic expansions of

$I(q)$, it is not required for computing the Jacobian, since only the final series expansion is needed to compute the derivative. This implies that programs like CRYSOL, which are based on direct series expansion of Eq. (2), can be easily adapted to compute derivatives using Eq. (18), as long as the series expansion is properly truncated according to a proper error bound.

The hierarchical approach is not limited to strict spatial subdivision of macromolecules, and can be adapted to use various other groupings, including secondary structures, or other rigid components, further speeding up the computation. For certain applications the derivatives for the water molecules might be required. Our hierarchical approach allows us to take advantage of this case by computing a downward pass only on the required atomic domain, if necessary. We note that this algorithm is a proof of principle, and thus the performance can be significantly improved, either by improving the translation operators or by removing complex number multiplications in several instances.

5 Conclusion

We developed and demonstrated a fast method for computation of SAS profile, and its associated derivatives with respect to atomic coordinates of the molecule. The method is based on our previously developed hierarchical expansion method, and is able to compute the SAS profile and associated derivatives to within arbitrary ϵ accuracy, while being an order of magnitude faster than current method. The algorithm's computational advantage is even larger for high density atom distribution, which can occur when discretizing continuous water models. This opens up the possibility to efficiently integrate small-angle scattering data into the existing protocols for structure determination and refinement of macromolecular systems.

6 Acknowledgments

This study has been partially supported by the New Research Frontiers Award of the Institute of the Advanced Computer Studies of the University of Maryland, by Fantalgo, LLC., and by NIH grant GM065334 to D.F.

A Appendix: Derivatives of R_n^m

The derivation of partial derivatives for $R_n^m(\mathbf{r})$, where $\mathbf{r} = (x, y, z)$ is given in [28, 26]. For completeness, The main result of the derivation is given below.

$$\frac{\partial R_n^m(\mathbf{r})}{\partial x} = \frac{1}{2}\left[b_{n+1}^{-m-1} R_{n+1}^{m+1}(\mathbf{r}) - b_n^m R_{n-1}^{m+1}(\mathbf{r}) + b_{n+1}^{m-1} R_{n+1}^{m-1}(\mathbf{r}) - b_n^{-m} R_{n-1}^{m-1}(\mathbf{r})\right], \tag{A19}$$

$$\frac{\partial R_n^m(\mathbf{r})}{\partial y} = \frac{1}{2}\mathbf{i}\left[b_{n+1}^{-m-1} R_{n+1}^{m+1}(\mathbf{r}) - b_n^m R_{n-1}^{m+1}(\mathbf{r}) - b_{n+1}^{m-1} R_{n+1}^{m-1}(\mathbf{r}) + b_n^{-m} R_{n-1}^{m-1}(\mathbf{r})\right], \tag{A20}$$

$$\frac{\partial R_n^m(\mathbf{r})}{\partial z} = a_{n-1}^m R_{n-1}^m(\mathbf{r}) - a_n^m R_{n+1}^m(\mathbf{r}), \tag{A21}$$

where

$$b_n^m = \begin{cases} \sqrt{\frac{(n-m-1)(n-m)}{(2n-1)(2n+1)}} & \text{for } 0 \le m \le n, \\ -\sqrt{\frac{(n-m-1)(n-m)}{(2n-1)(2n+1)}} & \text{for } -n \le m < 0, \\ 0 & \text{for } n < |m|, \end{cases} \tag{A22}$$

and

$$a_n^m = \begin{cases} \sqrt{\frac{(n+1+m)(n+1-m)}{(2n+1)(2n+3)}} & \text{for } n \ge |m|, \\ 0 & \text{for } n < |m|, \end{cases} \tag{A23}$$

References

1. Koch, M. H. J.; Vachette, P. and Svergun, D. I., *Quarterly Reviews of Biophysics*, 2003, **36**(02), 147–227.

2. Grishaev, A.; Tugarinov, V.; Kay, L. E.; Trewhella, J. and Bax, A., *Journal of Biomolecular NMR*, 2008, **40**(2), 95–106.

3. Grishaev, A.; Ying, J.; Canny, M.; Pardi, A. and Bax, A., *Journal of Biomolecular NMR*, 2008, **42**(2), 99–109.

4. Grishaev, A.; Wu, J.; Trewhella, J. and Bax, A., *Journal of the American Chemical Society*, 2005, **127**(47), 16621–16628.

5. Pons, C.; D'Abramo, M.; Svergun, D. I.; Orozco, M.; Bernadó, P. and Fernández-Recio, J., *Journal of Molecular Biology*, 2010, **403**(2), 217–230.

6. Bernadó, P.; Modig, K.; Grela, P.; Svergun, D. I.; Tchorzewski, M.; Pons, M. and Akke, M., *Biophysical Journal*, 2010, **98**(10), 2374–2382.

7. Datta, A.; Hura, G. and Wolberger, C., *Journal of Molecular Biology*, 2009, **392**(5), 1117–1124.

8. Jehle, S.; Vollmar, B. S.; Bardiaux, B.; Dove, K. K.; Rajagopal, P.; Gonen, T.; Oschkinat, H. and Klevit, R. E., *Proceedings of the National Academy of Sciences*, 2011.

9. Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M. and Svergun, D. I., *Journal of the American Chemical Society*, 2007, **129**(17), 5656–5664.

10. Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole Ii, F. L.; Tsutakawa, S. E.; Jenney Jr, F. E.; Classen, S.; Frankel, K. A.; Hopkins, R. C. and others, , *Nature Methods*, 2009, **6**(8), 606–612.

11. Grant, T. D.; Luft, J. R.; Wolfley, J. R.; Tsuruta, H.; Martel, A.; Montelione, G. T. and Snell, E. H., *Biopolymers*, 2011, **95**(8).

12. Grishaev, A.; Guo, L.; Irving, T. and Bax, A., *Journal of the American Chemical Society*, 2010, **132**(44), 15484–15486.

13. Svergun, D. I.; Barberato, C. and Koch, M. H. J., *Journal of Applied Crystallography*, 1995, **28**(6), 768–773.

14. Bardhan, J.; Park, S. and Makowski, L., *Journal of Applied Crystallography*, 2009, **42**(5), 932–943.

15. Poitevin, F.; Orland, H.; Doniach, S.; Koehl, P. and Delarue, M., *Nucleic Acids Research*, 2011, **39**(suppl 2), W184–W189.

16. Gumerov, N. A.; Berlin, K.; Fushman, D. and Duraiswami, R., *Journal of Computational Chemistry*, 2012, **33**(25), 1981–1996.

17. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N. and Marius Clore, G., *Journal of Magnetic Resonance*, 2003, **160**(1), 65–73.

18. Dominguez, C.; Boelens, R. and Bonvin, A., *NMR-based docking of protein-protein complexes*, 2003, **125**, 51.

19. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S. and others, , *Acta Crystallographica Section D: Biological Crystallography*, 1998, **54**(5), 905–921.

20. Gabel, F.; Simon, B. and Sattler, M., *European Biophysics Journal*, 2006, **35**(4), 313–327.

21. Stumpe, M. C.; Blinov, N.; Wishart, D.; Kovalenko, A. and Pande, V. S., *The Journal of Physical Chemistry B*, 2010, **115**(2), 319–328.

22. Virtanen, J. J.; Makowski, L.; Sosnick, T. R. and Freed, K. F., *Biophysical journal*, 2011, **101**(8), 2061–2069.

23. Boyd, S. and Vandenberghe, L., *Convex Optimization*, Cambridge university press, 2004.

24. Abramowitz, M. and Stegun, I. A., *National Bureau of Standsards, Applied Mathematics Series*, 1970, **55**.

25. Gumerov, N. A.; Berlin, K.; Fushman, D. and Duraiswami, R. A hierarchical algorithm for fast Debye summation with applications to small angle scattering Technical report, University of Maryland, 2011.

26. Gumerov, N. A. and Duraiswami, R., *Fast Multipole Methods For the Helmholtz Equation in Three Dimensions*, Elsevier Science, 2004.

27. Liu, H.; Morris, R. J.; Hexemer, A.; Grandison, S. and Zwart, P. H., *Acta Crystallographica Section A: Foundations of Crystallography*, 2012, **68**(2), 0–0.

28. Gumerov, N. A. and Duraiswami, R., *SIAM Journal on Scientific Computing*, 2004, **25**(4), 1344–1381.