Iterative phasing for fluctuation X-ray scattering

Jeffrey J. Donatelli[a,b], Peter H. Zwart[c], and James A. Sethian[a,b,1]

[a]Department of Mathematics, University of California, Berkeley, CA 94720; [b]Mathematics Department, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; and [c]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Contributed by James A. Sethian, July 14, 2015 (sent for review April 27, 2015; reviewed by Garib Murshudov)

Fluctuation X-ray scattering (FXS) is an extension of small- and wide-angle X-ray scattering in which the X-ray snapshots are taken below rotational diffusion times. This technique, performed using a free electron laser or ultrabright synchrotron source, provides significantly more experimental information compared with traditional solution scattering methods. We develop a multitiered iterative phasing algorithm to determine the underlying structure of the scattering object from FXS data.

fluctuation scattering | iterative phasing | polar Fourier transform

X-ray solution scattering of macromolecular complexes is a versatile technique, providing low-resolution structural information that, when supplemented with high-resolution crystallographic data, can result in fundamental insights into the physiological behavior and function of macromolecular machines (1). Although solution scattering has been applied successfully to many problems in the biological sciences (2, 3), the technique suffers from a significant shortcoming. Due to the 3D averaging that occurs during the X-ray snapshots, the effective information content of the data is typically only around 9–15 independent parameters (4). This lack of sufficient information in the data ultimately results in nonunique or poorly determined structural hypotheses.

To overcome these issues, it was proposed to perform the solution scattering process at timescales below rotational diffusion times (5, 6). By avoiding physical rotational averaging, speckle patterns emerge from which angular correlation functions can be computed. These so-called fluctuation X-ray scattering experiments can be performed on modern synchrotrons (7) and free electron laser sources (8, 9). The angular correlation information provided from these experiments is directly related to the imaged structure (10) and contains significantly more information than standard small- and wide-angle X-ray scattering (SAXS/WAXS) data. Although the relation from real space structure to fluctuation scattering data is straightforward, the inverse problem of determining a molecular model from fluctuation scattering data is nontrivial (8).

For 2D systems, such as macromolecules randomly oriented around a single axis (11), an analytical route for image reconstruction is available (12). However, this approach requires triple correlation measurements across the full resolution range of interest, which may not be fully accessible due to experimental limitations.

Earlier work attempting to determine 3D shape from fluctuation scattering data was based on solving two phase problems consecutively. Although this route has shown success in cases where the scattering species has helical (13) or icosahedral symmetry (14), it fails to produce a structure for the general case. Another approach was developed in which the inverse problem was solved using a reverse Monte Carlo method (8), similar to what is used to determine macromolecular envelopes from SAXS data (15–17). Although this approach allows a real space object to be reconstructed without the need for symmetry constraints, the procedure is computationally expensive and, in general, convergence can be problematic (18). Furthermore, the existing reverse Monte Carlo method assumes a binary model, where voxels either have a scattering length or do not. The breakdown of this uniform density assumption ultimately limits the resolution of features that can be reconstructed through such an approach.

Here we develop a fundamentally different approach to structure solution from fluctuation scattering, based on a multitiered iterative phasing (MTIP) algorithm. In particular, we derive a series of projection operators, which are used to iteratively modify a model to satisfy specified real space constraints and match the derived model fluctuation scattering data to external observations. This approach does not require a uniform density assumption and solves the two inherent phase problems concurrently, which allows one to obtain a reconstruction with a reduced amount of fluctuation scattering data, or even SAXS/WAXS data alone, by compensating with the extra information provided through real space constraints.

The presented method allows one to study the structure of objects that cannot be easily crystallized, while achieving higher resolution than what is possible through standard SAXS/WAXS experiments. As a result, this approach can be used to gain deeper insight into biological structure and to expand the range of characterization methods for nanostructures (19). Furthermore, the framework developed here allows one to extend many of the density modification techniques developed for crystallography to solution scattering.

Background

Here we use $\mathbf{r}$ and $\mathbf{q}$ to denote real and Fourier space variables, respectively, and designate $r = |\mathbf{r}|$ and $q = |\mathbf{q}|$. The 3D scattered intensity I of an object is given by the squared magnitude of the Fourier transform of its electron density ρ:

$$I(\mathbf{q}) = \left| \int_{\mathbb{R}^3} \rho(\mathbf{r}) e^{-2\pi i \mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \right|^2 .$$

When the object is rotated by $R \in SO(3)$, the resulting intensities become $I_R(\mathbf{q}) = I(R\mathbf{q})$. The diffraction pattern from the object with orientation R measured on a 2D detector with a beam of wavelength λ can be expressed in polar coordinates as $\mathcal{I}_R(q, \phi_I) = I_R(q, \theta(q), \phi_I)$, where $\theta(q) = \pi/2$ for a flat Ewald sphere, $\theta(q) = \arccos(q\lambda/2)$ for a curved Ewald sphere, and ϕ_I is the angle about the incident beam axis.

Significance

Fluctuation X-ray scattering is an emerging imaging technique that seeks to overcome the low data-to-parameter ratio encountered in traditional small- and wide-angle X-ray scattering methods. By acquiring a large number of ultrashort X-ray exposures on an ensemble of molecules, this technique produces a dataset that contains structural information far beyond what is obtainable from traditional solution scattering methods without requiring crystallization. However, reconstructing the underlying molecular shape from this data is challenging, as the information in each image is averaged over several molecular orientations. In this article, we introduce a flexible iterative method that can rapidly determine molecular structure from fluctuation scattering data. This allows one to visualize structural details that may be inaccessible through traditional methods.

Author contributions: J.J.D., P.H.Z., and J.A.S. designed research, performed research, analyzed data, and wrote the paper.

Reviewers included: G.M., Medical Research Council Laboratory of Molecular Biology Cambridge.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: sethian@math.berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1513738112/-/DCSupplemental.

In a fluctuation scattering experiment, one measures a series of N_{dp} diffraction patterns $\mathcal{I}_{R_n}$ from an ensemble of identical and randomly oriented particles. From these images, one computes the average angular cross-correlation function

$$C(q,q',\phi)=\frac{1}{2\pi N_{dp}}\sum_{n=1}^{N_{dp}}\int_0^{2\pi}\mathcal{I}_{R_n}(q,\phi_I+\phi)\mathcal{I}_{R_n}(q',\phi_I)d\phi_I. \quad [1]$$

If there are multiple particles per shot, then, for a sufficiently dilute system, this average converges to that of the single particle per shot case up to a zero frequency component (20).

The fluctuation scattering reconstruction problem is to determine the electron density ρ given the angular correlation data defined by Eq. **1**, along with any a priori known constraints, e.g., compact support, symmetry, or upper and lower bounds. In particular, the angular correlation data yield information about the harmonic coefficients of the intensities, which can be exploited to determine information about ρ. In the following subsections, we describe how to extract this intensity harmonic coefficient information from the angular correlation data for both the 2D and 3D versions of the fluctuation scattering reconstruction problem. See *SI Text* and Figs. S1–S6 for an introduction to circular and spherical harmonics and an overview of fluctuation scattering data.

Two-Dimensional Case. The 2D version of fluctuation scattering is obtained by assuming a flat Ewald sphere and that the particle orientations are randomly sampled from a uniform distribution of rotations about a single axis parallel to the incident beam. In this case, one can retrieve intensity information only along a flat 2D slice of the intensities, which, by the Fourier projection slice theorem, is equivalent to the squared Fourier magnitudes of the X-ray projection of ρ to a plane perpendicular to the rotation axis.* For the moment, assume that there is only one particle per shot.

A function $\rho\in L^2(\mathbb{R}^2)$ can be formally expressed in terms of its circular harmonic expansion as

$$\rho(r,\phi)=\sum_{m=-\infty}^{\infty}\rho_m(r)e^{im\phi}, \quad [2]$$

where the circular harmonic coefficients $\rho_m(r)$ are given by

$$\rho_m(r)=\frac{1}{2\pi}\int_0^{2\pi}\rho(r,\phi)e^{-im\phi}d\phi. \quad [3]$$

The circular harmonic coefficients of ρ and its Fourier transform $\hat{\rho}$ are related through the Hankel transform via

$$\hat{\rho}_m(q)=2\pi(-i)^m\int_0^{\infty}\rho_m(r)J_m(2\pi qr)r\,dr, \quad [4]$$

$$\rho_m(r)=2\pi\, i^m\int_0^{\infty}\hat{\rho}_m(q)J_m(2\pi qr)q\,dq, \quad [5]$$

where J_m is a Bessel function of order m.

We denote the circular harmonic coefficients of the angular cross-correlations in Eq. **1** by $B_m(q,q')$, which can be expressed in terms of the intensity circular harmonic coefficients I_m as

$$B_m(q,q')=I_m(q)I_m^*(q'). \quad [6]$$

We denote the circular harmonic coefficients of the angular autocorrelations, i.e., $C(q,q)$, by $B_m(q)=B_m(q,q)$. In this case, one has

$$B_m(q)=|I_m(q)|^2. \quad [7]$$

If there are K particles per shot, then, after rescaling, the B_0 data will be off by a factor of K; i.e., $B_0(q,q')=K^2I_0(q)I_0^*(q')$ and $B_m(q,q')=KI_m(q)I_m^*(q')$ for $m\neq 0$.

The cross-correlation circular harmonic coefficients form a rank one Gram matrix, which determines I_m up to a single phase factor for each m. In particular, for a discretization $\{q_j\}$ of q and a fixed m, if we form the matrix with entries $B_m(q_j,q_k)$, then its top eigenvector v_m, with corresponding eigenvalue λ_m, is related to the intensity harmonic coefficients via

$$I_m(q_j)=v_m(q_j)\sqrt{\lambda_m}e^{i\phi_m}, \quad [8]$$

where ϕ_m is a single unknown real-valued phase factor.

Three-Dimensional Case. The 3D version of fluctuation scattering allows for a curved Ewald sphere and assumes that the particle orientations are randomly sampled from a uniform distribution over $SO(3)$. For the moment, assume that there is only one particle per shot.

A function $\rho\in L^2(\mathbb{R}^3)$ can be formally expressed in terms of its spherical harmonic expansion as

$$\rho(r,\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\rho_{lm}(r)Y_l^m(\theta,\phi), \quad [9]$$

where the Y_l^m are the spherical harmonics and the spherical harmonic coefficients ρ_{lm} are given by

$$\rho_{lm}(r)=\int_0^{2\pi}\int_0^{\pi}\rho(r,\theta,\phi)Y_l^{m^*}(\theta,\phi)\sin\theta\,d\theta d\phi. \quad [10]$$

The spherical harmonic coefficients of ρ and its Fourier transform $\hat{\rho}$ are related through the spherical Hankel transform via

$$\hat{\rho}_{lm}(q)=4\pi(-i)^l\int_0^{\infty}\rho_{lm}(r)j_l(2\pi qr)r^2dr, \quad [11]$$

$$\rho_{lm}(r)=4\pi i^l\int_0^{\infty}\hat{\rho}_{lm}(q)j_l(2\pi qr)q^2dq, \quad [12]$$

where j_l is the spherical Bessel function of order l.

In the limit of an infinite number of images, one has

$$C(q,q',\phi)=\sum_{l=0}^{\infty}F_l(q,q',\phi)B_l(q,q'), \quad [13]$$

where

$$F_l(q,q',\phi)=\frac{1}{4\pi}P_l(\cos\theta(q)\cos\theta(q')+\sin\theta(q)\sin\theta(q')\cos\phi),$$

where P_l is the Legendre polynomial of order l, and

$$B_l(q,q')=\sum_{m=-l}^{l}I_{lm}(q)I_{lm}^*(q'), \quad [14]$$

where the I_{lm} are the spherical harmonic coefficients of the intensities (5). The corresponding autocorrelation data, $B_l(q)=B_l(q,q)$, yields

$$B_l(q)=\sum_{m=-l}^{l}|I_{lm}(q)|^2. \quad [15]$$

*To simplify the presentation, we abuse notation and still call the 2D projected quantity ρ and denote the slice $I(q,\pi/2,\phi)$ simply by $I(q,\phi)$.

For a fixed q and q', $B_l(q,q')$ can be retrieved by inverting the associated linear system in Eq. **13**. If there are K particles per shot, then, after rescaling, the B_0 data will be off by a factor of K; i.e., $B_0(q,q') = K^2 I_{00}(q) I_{00}^*(q')$ and $B_l(q,q') = K \sum_{m=-l}^{l} I_{lm}(q) I_{lm}^*(q')$ for $l \neq 0$.

Given a discretization $\{q_j\}$ of size N and a fixed l, the matrix B_l with entries $B_l(q_j,q_k)$ is a rank $\min(2l+1,N)$ Gram matrix with the eigendecomposition

$$B_l = V_l \Lambda_l V_l^*, \qquad [16]$$

where V_l is N by $\min(2l+1,N)$ and Λ_l has nonnegative entries. This decomposition is related to the spherical harmonic coefficients of the intensities, expressed as an N by $2l+1$ matrix I_l for each l, via multiplication by an unknown $2l+1$ dimensional unitary matrix U_l. In particular, if $2l+1 \leq N$, then

$$I_l = V_l \sqrt{\Lambda_l} U_l, \qquad [17]$$

and if $2l+1 > N$, then one obtains a similar result by zero padding; i.e., $I_l = \begin{bmatrix} V_l \sqrt{\Lambda_l} & 0 \end{bmatrix} U_l$.

Polar Fourier Transform

To obtain a reconstruction algorithm, we have to enforce several constraints in real space while matching the circular or spherical harmonic information to the measured quantities that are defined in Fourier space. Therefore, it is desirable to compute the Fourier transform and its inverse on a polar grid, where circular and spherical harmonic transforms can be computed most efficiently. Here we describe approximations to the 2D and 3D Fourier transforms and their inverses on such a polar grid. This is accomplished by computing discrete circular/spherical harmonic transforms, approximating the Hankel/spherical Hankel transforms, and then computing the inverse discrete circular/spherical harmonic transforms. We assume, after rescaling, that the support of ρ is contained within $r < 1$ and we resolve $\hat{\rho}$ up to $q < Q$.

Two-Dimensional Polar Fourier Transform. Here we use a polar grid with nodes equispaced in radii and equispaced in angle for each radius. Specifically, for $0 \leq n < N$, the real space radial node components are given by $r_n = n/N$, the Fourier space radial node components are given by $q_n = Qn/N$, and the number M_n of angular nodes for each radius can be varied.

We use a fast Fourier transform (FFT) for each radius to compute the circular harmonic coefficients and invert them with an inverse fast Fourier transform (IFFT).[†]

Note that for a smooth ρ, ρ_m has a smooth even/odd extension to $[-1,1]$ for m even/odd. Therefore, by expanding ρ_m into a cosine/sine series, we obtain spectrally accurate quadrature rules for the transforms in Eqs. **4** and **5** via[‡]

$$\hat{\rho}_m(q_n) \approx (-i)^m \sum_{n'=0}^{N-1} \rho_m(r_{n'}) w_{m,n,n'},$$

$$\rho_m(r_n) \approx i^m Q^2 \sum_{n'=0}^{N-1} \hat{\rho}_m(q_{n'}) w_{m,n,n'},$$

where for m even

$$w_{m,n,n'} = \frac{4\pi}{N} \sum_{k=0}^{N-1} c_k c_{n'} \cos\left(\pi k \frac{n'}{N}\right) \int_0^1 \cos(\pi k x) J_m\left(2\pi Q \frac{n}{N} x\right) x \, dx,$$

and for m odd

$$w_{m,n,n'} = \frac{4\pi}{N} \sum_{k=0}^{N-1} c_k c_{n'} \sin\left(\pi k \frac{n'}{N}\right) \int_0^1 \sin(\pi k x) J_m\left(2\pi Q \frac{n}{N} x\right) x \, dx,$$

where c_a is $1/2$ for $a = 0$ and is 1 otherwise.

Three-Dimensional Polar Fourier Transform. Here we use a polar grid with nodes whose radial components are equispaced, as described in the 2D case. For the nth radial component, the nodes lie on L_n inclination angles, located at the arccosines of the Gauss–Legendre quadrature nodes, and for each inclination angle we have $2L_n - 1$ equispaced azimuthal angles.

We compute the spherical harmonic transform and its inverse for each radius via the Gauss–Legendre approach, e.g., ref. 21.

Note that for a smooth ρ, ρ_{lm} has a smooth even/odd extension to $[-1,1]$ for l even/odd. Therefore, by expanding ρ_{lm} into a cosine/sine series, we obtain spectrally accurate quadrature rules for the transforms in Eqs. **11** and **12** via

$$\hat{\rho}_{lm}(q_n) \approx (-i)^l \sum_{n'=0}^{N-1} \rho_{lm}(r_{n'}) w_{l,n,n'},$$

$$\rho_{lm}(r_n) \approx i^l Q^3 \sum_{n'=0}^{N-1} \hat{\rho}_{lm}(q_{n'}) w_{l,n,n'},$$

where for l even

$$w_{l,n,n'} = \frac{8\pi}{N} \sum_{k=0}^{N-1} c_k c_{n'} \cos\left(\pi k \frac{n'}{N}\right) \int_0^1 \cos(\pi k x) j_l\left(2\pi Q \frac{n}{N} x\right) x^2 \, dx,$$

and for l odd

$$w_{l,n,n'} = \frac{8\pi}{N} \sum_{k=0}^{N-1} c_k c_{n'} \sin\left(\pi k \frac{n'}{N}\right) \int_0^1 \sin(\pi k x) j_l\left(2\pi Q \frac{n}{N} x\right) x^2 \, dx.$$

Multitiered Iterative Phasing

Iterative phasing techniques were originally introduced to solve the phase retrieval problem, which is to reconstruct an object from its intensities while enforcing any applicable constraints. These techniques are based on iteratively applying a combination of projection operators to enforce constraints in both real and Fourier space until convergence is achieved. If a sufficient number of constraints are enforced, then, in general, ab initio reconstruction is possible via iterative phasing. For example, this can be accomplished when one samples the intensity data at twice the Nyquist rate and enforces the support constraint implied by this oversampling (22).

Here we extend these iterative phasing techniques to a multitiered algorithm that can reconstruct an object from its fluctuation scattering data, which requires one to determine both intensity and phase information. In particular, we seek a function ρ whose intensity harmonic coefficients match the constraints imposed by the correlation data in Eqs. **6** and **7** for the 2D case and in Eqs. **14** and **15** for the 3D case, while satisfying any applicable real space constraints, such as being supported within a given region, nonnegativity, upper bounds, and/or symmetry. However, in practice, noise and/or an insufficient number of images may limit the amount of information obtainable from the correlation data; i.e., one may be able to accurately determine B_m and B_l only for $|m| < m_{\max}$ and $l < l_{\max}$, or may be able to retrieve only autocorrelation data instead of the full cross-correlation data.

Now we define several projection operators used to enforce the previously mentioned constraints.[§]

[†]Any circular/spherical harmonic coefficients that cannot be resolved on the grid are set to zero.

[‡]The integrals in the definition of the quadrature weights are precomputed to high precision via adaptive quadrature.

[§]If there are K particles per shot, then we assume here that the B_0 data has been rescaled by $1/K$, relative to the remaining data. Alternatively, if K is unknown, then this scaling factor could be allowed to float.

Real Space Projectors. Given a support set S, the support projection operator P_S projects a function to have support contained in S:

$$(P_S\rho)(\mathbf{r}) = \begin{cases} \rho(\mathbf{r}), & \text{if } \mathbf{r} \in S, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, one can extend this definition to enforce nonnegativity, $P_{S+}\rho(\mathbf{r}) = \max(P_S\rho(\mathbf{r}), 0)$; an upper bound τ, $P_{S\tau}\rho(\mathbf{r}) = \min(P_S\rho(\mathbf{r}), \tau)$; or both, $P_{S+\tau}\rho(\mathbf{r}) = \min(P_{S+}\rho(\mathbf{r}), \tau)$.

To enforce symmetry given by a point group G, we define the associated orthogonal projection operator P_G, which projects a function to the closest function in the L^2 norm that is invariant under G. In 2D, this projection can be represented in terms of circular harmonic coefficients by

$$(P_G\rho)_m(r) = \begin{cases} \rho_m(r) + \frac{a}{2}(\rho_{-m}(r) - \rho_m(r)), & \text{if } m \equiv 0 (\text{mod } k), \\ 0, & \text{otherwise,} \end{cases}$$

where G has a k-fold rotational element, and a is 1 if G has a reflection and is 0 otherwise. In 3D, this projection can be represented in terms of spherical harmonic coefficients by

$$(P_G\rho)_{lm}(r) = \frac{1}{|G|} \sum_{\mathcal{O} \in G} \sum_{m'=-l}^{l} (\det(\mathcal{O}))^l D_{lmm'}(\det(\mathcal{O})\mathcal{O}) \rho_{lm'}(r),$$

where the $D_{lmm'}$ are the Wigner D-matrix elements.

Two-Dimensional Correlation Projectors. Given the circular harmonic coefficient magnitude data $B_m(q)$ in Eq. **7**, we define the autocorrelation projection operator P_A, which projects a function I to the closest function in the discrete L^2 norm[¶] whose circular harmonic magnitudes match B_m. This projection can be expressed in terms of circular harmonic coefficients by

$$(P_A I)_m(q) = \frac{I_m(q)}{|I_m(q)|} \sqrt{B_m(q)}.$$

If $I_m(q) = 0$, then the quotient is replaced by 1. Additionally, we allow the coefficients to float for $|m| > m_{\max}$; i.e., $(P_A I)_m = I_m$.

If cross-correlation data are available, then recall that $v_m\sqrt{\lambda_m}$ is equal to the intensity harmonic coefficients up to a single phase factor for each m. We define the cross-correlation projection operator P_C, which projects a function I to the closest function in the discrete L^2 norm whose circular harmonics coefficients satisfy Eq. **8**. This projection can be expressed in terms of circular harmonic coefficients by

$$(P_C I)_m(q) = v_m(q)\sqrt{\lambda_m}\, \frac{\sum_j I_m(q_j) v_m^*(q_j) q_j}{\left|\sum_j I_m(q_j) v_m^*(q_j) q_j\right|}.$$

If the denominator is 0, then the quotient is replaced by 1. Again, we set $(P_C I)_m = I_m$ for $|m| > m_{\max}$.

Three-Dimensional Correlation Projectors. Given the spherical harmonic coefficient data B_l, we define the spherical harmonic autocorrelation projection operator P_A, which projects a function I to the closest function in the discrete L^2 norm[#] whose spherical harmonic coefficients satisfy Eq. **15**. This projection can be expressed in terms of spherical harmonic coefficients by

$$(P_A I)_{lm}(q) = \frac{I_{lm}(q)}{\sqrt{\sum_{m=-l}^{l} |I_{lm}(q)|^2}} \sqrt{B_l(q)}.$$

If the denominator is 0, then the quotient in the above definition is replaced by $1/\sqrt{2l+1}$. Additionally, we allow the coefficients to float for $l > l_{\max}$; i.e., $(P_A I)_{lm} = I_{lm}$.

Recall that angular cross-correlation data determine the spherical harmonic coefficients up to multiplication by a unitary matrix via Eq. **17**. Therefore, the l-th order spherical harmonic coefficients of the function closest to I that match the cross-correlation data are given in matrix form by $V_l\sqrt{\Lambda_l}U_l$, where U_l solves the unitary Procrustes problem,

$$\min_{U_l \in \mathrm{U}(2l+1)} \left\| D\left(I_l - V_l\sqrt{\Lambda_l}U_l\right)\right\|_F,$$

where $U(2l+1)$ is the set of unitary $2l+1$ dimensional matrices; $\|\cdot\|_F$ is the Frobenius norm; $D = \mathrm{diag}(q_0, \ldots, q_{N-1})$ weights the norm to match the discrete L^2 norm; and we use zero padding if $2l+1 > N$. In particular, given the singular value decomposition $\sqrt{\Lambda_l}V_l^* D^2 I_l = \mathcal{U}_l \Sigma_l \mathcal{V}_l^*$, the minimizer is given by $\mathcal{U}_l\mathcal{V}_l^*$. This result allows us to define the spherical harmonic cross-correlation projection operator P_C, which projects a function I to the closest function in the discrete L^2 norm whose spherical harmonic coefficients satisfy Eq. **14**. This projection can be expressed in terms of spherical harmonic coefficients by

$$(P_C I)_l = V_l\sqrt{\Lambda_l}\mathcal{U}_l\mathcal{V}_l^*,$$

where $(P_C I)_l$ is the N by $2l+1$ matrix of spherical harmonic coefficients of $P_C I$. Again, we set $(P_C I)_l = I_l$ for $l > l_{\max}$.

Fluctuation Scattering Operator. We now define the fluctuation scattering operator F, which given a function ρ, projects its intensities to match the angular correlation data, projects the resulting quantities to the nearest nonnegative function, and then projects ρ to the closest function whose Fourier magnitudes match the square root of the projected intensities. This operator can be expressed in the Fourier domain as

$$\widehat{(F\rho)}(\mathbf{q}) = \frac{\hat{\rho}(\mathbf{q})}{|\hat{\rho}(\mathbf{q})|} \sqrt{\max\left(\left(P_{\#}|\hat{\rho}|^2\right)(\mathbf{q}), 0\right)},$$

where, depending on what type of data is available, $P_{\#}$ is either P_A or P_C. If $\hat{\rho}(\mathbf{q}) = 0$, then the quotient in the above definition is replaced by 1.

Iterative Phasing Algorithms. Now we describe how to use these projection operators in an iterative procedure to reconstruct a function from its fluctuation scattering data while enforcing prescribed real space constraints. We start with some initial guess $\rho^{(0)}$ and iteratively apply a series of projection operators to produce the iterate $\rho^{(n)}$ after n steps.

One of the first developed iterative phasing techniques is the error-reducing (ER) method, which alternates between projecting constraints in Fourier space and in real space (23). Here we generalize ER for fluctuation scattering data, yielding[‖]

[¶]The discrete L^2 norm for the 2D polar grid is given by weighting the ℓ^2 norm by q_n/M_n.

[#]The discrete L^2 norm for the 3D polar grid is given by weighting the ℓ^2 norm by $q_n^2/(2L_n - 1)$ times the Gauss–Legendre quadrature weights associated with each inclination angle.

[‖]Due to finite precision, $\rho^{(n)}$ may become complex and should be replaced by its real part.

$$\rho^{(n+1)}(\mathbf{r}) = \left(P_{S*}F\rho^{(n)}\right)(\mathbf{r}),$$

where P_{S*} is one of P_S, P_{S+}, $P_{S+\tau}$, or $P_{S\tau}$.

One of the main disadvantages of ER is that it often gets stuck in local minima. A popular method to circumvent this issue is the hybrid input–output (HIO) method, which uses negative feedback to inhibit stagnation (24). We formulate HIO for fluctuation scattering as

$$\rho^{(n+1)}(\mathbf{r}) = \begin{cases} \left(F\rho^{(n)}\right)(\mathbf{r}), & \text{if } \left(P^c_{S*}F\rho^{(n)}\right)(\mathbf{r}) = 0, \\ \rho^{(n)}(\mathbf{r}) - \beta\left(P^c_{S*}F\rho^{(n)}\right)(\mathbf{r}), & \text{otherwise,} \end{cases}$$

where $(P^c_{S*}\rho)(\mathbf{r}) = \rho(\mathbf{r}) - (P_{S*}\rho)(\mathbf{r})$ and $\beta \in (0,1]$.

The iterative schemes listed above can be modified to enforce point group symmetry by replacing F with P_GF.

To obtain a good estimate of the true support region, we use the shrinkwrap method (25), which periodically convolves the iterate with a Gaussian of width σ and then applies a threshold ε to estimate the support; i.e., the updated support for the d-dimensional case is taken to be

$$S = \left\{ \mathbf{r} : \frac{1}{(2\pi\sigma^2)^{d/2}} \int_{\mathbb{R}^d} \rho^{(n)}(\mathbf{s}) e^{-|\mathbf{r}-\mathbf{s}|^2/2\sigma^2} d\mathbf{s} \geq \epsilon \right\}.$$

In particular, we apply several iterations of HIO, followed by a few iterations of ER, perform shrinkwrap to update the support, and repeat this process until convergence is reached. See Fig. S7 for a summary of the described iterative procedure.

We note that the solution can be reconstructed only up to translation, rotation, reflection through a line, and inversion through a point. Furthermore, a general structure might not be uniquely determined by its fluctuation scattering data (26). However, whether this nonuniqueness corresponds to small perturbations around a single solution, several clusters of solutions, or a continuum of solutions and how it depends on $m_{\max}$, $l_{\max}$, and the imposed constraints are not fully understood and likely depend on the complexity of the structure.

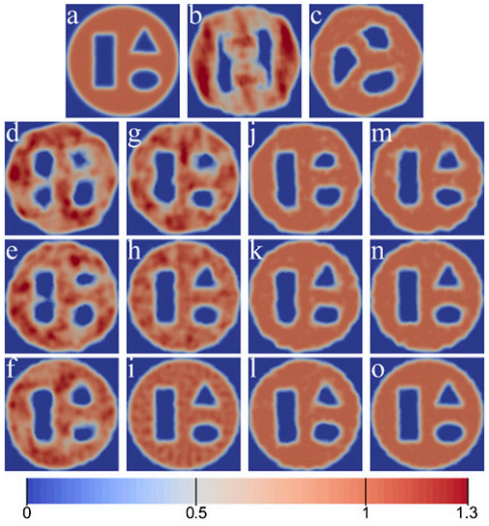

Fig. 1. Two-dimensional shape reconstructions. (*A*) Original image. (*B* and *C*) Reconstructions from SAXS data without and with upper-bound constraint. (*D–F*) Reconstructions from autocorrelation data with $m_{\max} = 4,10,20$. (*G–I*) Reconstructions from cross-correlation data with $m_{\max} = 4,10,20$. (*J–L*) Reconstructions from autocorrelation data with upper-bound constraint with $m_{\max} = 4,10,20$. (*M–O*) Reconstructions from cross-correlation data with upper-bound constraint with $m_{\max} = 4,10,20$.

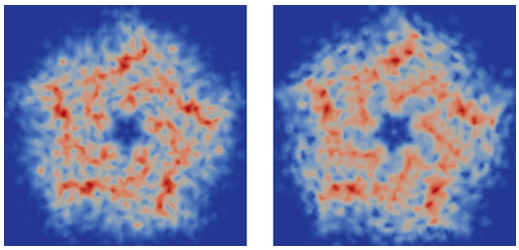

Fig. 2. Reconstruction of the X-ray projection of the pLGIC. (*Left*) Original. (*Right*) Reconstruction from cross-correlation data with $m_{\max} = 20$.

Results

Here we test the presented iterative phasing schemes by reconstructing several structures from their theoretical B_m/B_l data. In general, we enforce support and nonnegativity constraints; set $\beta = 0.5$, $\sigma = 1/N$, ε to be 2–10% of the maximum density value, $Q = N/2$; sample the fluctuation scattering data at least at twice the Nyquist rate; and initialize the procedure with the support S to be the ball of radius $1/2$ and the density $\rho^{(0)}$ to be 1 within the support and 0 outside of it. More details can be found in *SI Text*. Each reconstruction took less than 5 min on a single core of a 2.4-GHz Ivy Bridge processor.

We present reconstructions from angular correlation data, using the 2D version of the iterative phasing approach in Figs. 1 and 2. In Fig. 1, the target shape was given a uniform density of 1 within a circle; had a rectangle, a triangle, and an ellipse cut out; and was then smeared onto the grid via convolution with a Gaussian of width 0.015. We show reconstructions of the target shape from SAXS data, autocorrelation data, and cross-correlation data for $m_{\max} = 4,10,20$, both with and without upper-bound constraints, using $\tau = 1$, and set $N = 100$. In Fig. 2, we reconstruct the 2D X-ray projection of a pentameric ligand-gated ion channel (pLGIC), from Protein Data Bank entry 4NPP (27), through its symmetry axis. Here we use cross-correlation data up to a 1.4-Å resolution with $m_{\max} = 20$, enforce a fivefold symmetry axis, and set $N = 60$.

Results for the reconstruction of 3D objects from angular correlation data are shown in Figs. 3 and 4. In Fig. 3, the target shape was given a uniform density of 1 within a sphere; had a box, a cube, and an ellipsoid cut out; and was then smeared onto the grid via convolution with a Gaussian of width 0.015. We show reconstructions of the target shape from SAXS data, autocorrelation data, and cross-correlation data for $l_{\max} = 4,10,20$, with an upper-bound constraint, using $\tau = 1$, and set $N = 42$. In Fig. 4, we reconstruct the 3D electron density of pLGIC from cross-correlation data

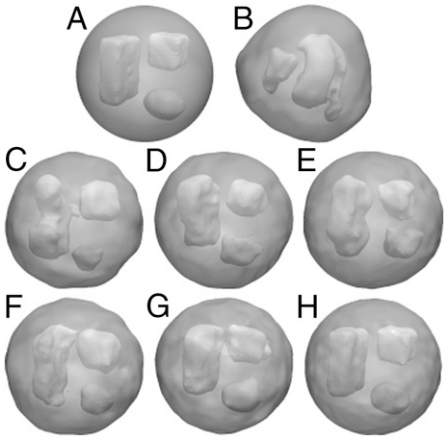

Fig. 3. Three-dimensional shape reconstructions, displayed as transparent density isosurfaces. (*A*) Original shape. (*B*) Reconstruction from SAXS data. (*C–E*) Reconstructions from autocorrelation data with $l_{\max} = 4,10,20$. (*F–H*) Reconstructions from cross-correlation data with $l_{\max} = 4,10,20$.

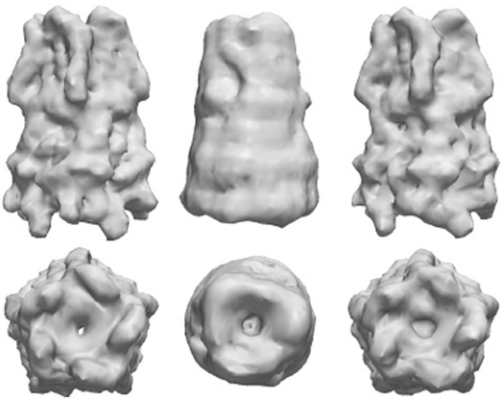

Fig. 4. Reconstructions of the pLGIC, displayed as density isosurfaces for a side and top view. (*Left*) Original. (*Center*) Average reconstruction from cross-correlation data with $I_{max} = 20$ and without a symmetry constraint. (*Right*) Reconstruction from cross-correlation data with $I_{max} = 20$ and a symmetry constraint.

up to a 4.8-Å resolution with $I_{max} = 20$, both with and without enforcing a fivefold symmetry axis, and set $N = 42$.

As shown in Figs. 1 and 3, the quality of the iterative phasing solution depends highly on the amount of experimental information and real space constraints used during the reconstruction process. In particular, the previously mentioned nonuniqueness manifests as small perturbations from the original structure, which decrease as more information is used. Nevertheless, even with a small amount of experimental information, low-resolution structure determination is still possible. However, the utilization of extra real space constraints, such as the upper density bound constraint used here, can compensate for a lack of reciprocal space information, such as the omission of cross-correlation data. In each example, the inclusion of higher-order terms allows for the recovery of finer details in the reconstructed models.

The reconstructions of pLGIC have a high degree of similarity to the target model, especially when symmetry is enforced. Without the symmetry constraint, the resulting reduction in information content leads to a lower-resolution 3D reconstruction, with spurious false details occurring below this resolution range. These false details were removed for this case by averaging over 24 aligned independent reconstructions with randomly perturbed starting densities. In contrast, utilization of the symmetry constraint yields a high-quality reconstruction without averaging. In particular, we achieved a 6-Å reconstruction with the symmetry constraint and a 12-Å reconstruction without it; see Table S1 for the reconstruction statistics.

Conclusions

We have demonstrated that it is possible to determine general structure from fluctuation X-ray scattering (FXS) data with an iterative phasing algorithm. The MTIP procedure outlined here is very flexible, as it allows for one to enforce a combination of autocorrelation, cross-correlation, or SAXS/WAXS data of various orders, depending on what information is experimentally available in each resolution shell. Furthermore, unlike other 3D solution methods for FXS, this approach does not require symmetry, is not limited to a binary model, and has a large region of convergence.

The quality of the structural models that can be obtained from this method depends on the experimental availability and accuracy of the FXS data and the amount of known prior information. At the very least, a low- to medium-resolution reconstruction is possible, but with sufficient information, high resolution can be achieved. In particular, the algorithm's performance could be improved by using additional prior real space information, such as density histograms or local smoothness constraints, or by making use of any experimentally available higher-order intensity correlations.

ACKNOWLEDGMENTS. This research was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, US Department of Energy, under Contract DE-AC02-05CH11231 and by the Division of Mathematical Sciences of the National Science Foundation and used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. J.A.S. was also supported by an Einstein Visiting Fellowship of the Einstein Foundation, Berlin. Further support originates from the National Institute of General Medical Sciences of the National Institutes of Health under Award R01GM109019. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

1. Dyer KN, et al. (2014) High-throughput SAXS for the characterization of biomolecules in solution: A practical approach. *Methods Mol Biol* 1091:245–258.
2. Graewert MA, Svergun DI (2013) Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr Opin Struct Biol* 23(5):748–754.
3. Morgan HP, et al. (2011) Structural basis for engagement by complement factor H of C3b on a self surface. *Nat Struct Mol Biol* 18(4):463–470.
4. Svergun DI, Feigin LA, Taylor GW (1987) *Structure Analysis by Small-Angle X-Ray and Neutron Scattering* (Plenum, New York).
5. Kam Z (1977) Determination of macromolecular structure in solution by spatial correlation of scattering fluctuations. *Macromolecules* 10(5):927–934.
6. Kam Z, Koch MH, Bordas J (1981) Fluctuation x-ray scattering from biological particles in frozen solution by using synchrotron radiation. *Proc Natl Acad Sci USA* 78(6): 3559–3562.
7. Pedrini B, et al. (2013) Two-dimensional structure from random multiparticle X-ray scattering images using cross-correlations. *Nat Commun* 4:1647.
8. Liu H, Poon BK, Saldin DK, Spence JC, Zwart PH (2013) Three-dimensional single-particle imaging using angular correlations from X-ray laser data. *Acta Crystallogr A* 69(Pt 4):365–373.
9. Starodub D, et al. (2012) Single-particle structure determination by correlations of snapshot X-ray diffraction patterns. *Nat Commun* 3:1276.
10. Liu H, Poon BK, Janssen AJEM, Zwart PH (2012) Computation of fluctuation scattering profiles via three-dimensional Zernike polynomials. *Acta Crystallogr A* 68(Pt 5): 561–567.
11. Saldin DK, et al. (2010) Structure of a single particle from scattering by many particles randomly oriented about an axis: Toward structure solution without crystallization? *New J Phys* 12(3):035014.
12. Kurta RP, et al. (2013) Solution of the phase problem for coherent scattering from a disordered system of identical particles. *New J Phys* 15:013059.
13. Poon HC, Schwander P, Uddin M, Saldin DK (2013) Fiber diffraction without fibers. *Phys Rev Lett* 110(26):265505.
14. Saldin DK, Poon HC, Schwander P, Uddin M, Schmidt M (2011) Reconstructing an icosahedral virus from single-particle diffraction experiments. *Opt Express* 19(18): 17318–17335.
15. Liu H, Hexemer A, Zwart P (2012) The Small Angle Scattering ToolBox (SASTBX): An open-source software for biomolecular small-angle scattering. *J Appl Cryst* 45: 587–593.
16. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879–2886.
17. Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946–2953.
18. Malmerberg E, Kerfeld CA, Zwart PH (2015) Operational properties of fluctuation X-ray scattering data. *IUCrJ* 2(Pt 3):309–316.
19. Myhra S, Riviere JC (2012) *Characterization of Nanostructures* (CRC, Boca Raton, FL).
20. Kirian RA (2012) Structure determination through correlated fluctuations in x-ray scattering. *J Phys B* 45(22):223001.
21. Schaeffer N (2013) Efficient spherical harmonic transforms aimed at pseudospectral numerical simulations. *Geochem Geophys Geosyst* 14(3):751–758.
22. Hayes MH (1982) The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform. *IEEE Trans Acoust Speech Signal Process* 30(2): 140–154.
23. Gerchberg RW, Saxton WO (1972) A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik* 35:237–246.
24. Fienup JR (1978) Reconstruction of an object from the modulus of its Fourier transform. *Opt Lett* 3(1):27–29.
25. Marchesini S, et al. (2003) X-ray image reconstruction from a diffraction pattern alone. *Phys Rev B* 68(4):140101.
26. Elser V (2011) Strategies for processing diffraction data from randomly oriented particles. *Ultramicroscopy* 111(7):788–792.
27. Sauguet L, et al. (2014) Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation. *Proc Natl Acad Sci USA* 111(3):966–971.