Beyond the Crystallization Paradigm: Structure Determination from Diffraction Patterns from Ensembles of Randomly Oriented Particles

H. C. Poon[1] and D. K. Saldin[1]

[1] *Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI 53211*

We amplify on the priciples of the method we have recently proposed for recovering an oversampled diffraction pattern of a single particle from measured diffraction patterns from multiple particles in orientations related by rotation about an axis parallel to the incident radiation. We propose an alternative method of phasing a reference resolution ring by means of a non-negativity constraint on the diffraction intensities, point out the need for caution about enantiomeric ambiguities in the reconstruction of a diffraction pattern from its angular correlations, and show that converged correlations may be deduced by appropriate averaging of even very noisy data.

PACS numbers: 61.05.cp,87.64.Bx,82.53.Ps

I. INTRODUCTION

In tribute to John Spence, the present paper is an attempt to clarify some of the principles of a method we have recently developed in collaboration with him and others[1,2], on the extraction of structural information about a particle from diffraction patterns of randomly oriented copies.

As in our previous paper[2], we concentrate on a special case, where the particles are related by a random angle of rotation about an axis parallel to an incident x-ray beam. A possible application might be to the determination of the projected structure of membrane proteins *situ*, where at least for the so-called ion channel proteins, it may be argued that the different embedded molecules are likely to be related by just such rotations since their internal pores (ion channels) need to remain perpendicular to the membranes for their biological function.

For about a century, the foremost method for determining the atomic-scale structure of molecules has been x-ray crystallography. In this technique, x-rays are directed into a sample containing a large number of identical molecules periodically arranged in identical orientations in a crystal. The main problem with this method is that not all molecules of interest form crystals. This is particularly true of a certain class of biomolecules of particular importance to life, namely the so-called membrane proteins, which tend to be embedded in cell membranes and, for example, control the traffic of metal ions between the inside and outside of a living cell.

A particularly noteworthy example is the so-called K-channel protein, which controls the transmission of electrical signals between neighboring neurons in the process of neurotransmission. A crucial part of this process is the passage of K ions through cell membranes. This ionic transport happens through the central pore of the K-channel protein, seen in a projection perpendicular to it as a central hole (Fig. 1). This particular membrane protein was crystallized[3], and the elucidation of its crystallographic structure was recognised by the award of a Nobel Prize in 2003 to Roderick MacKinnon. The method we describe promises the possibility of structure determination of other membrane proteins which may not so easily be crystallized.

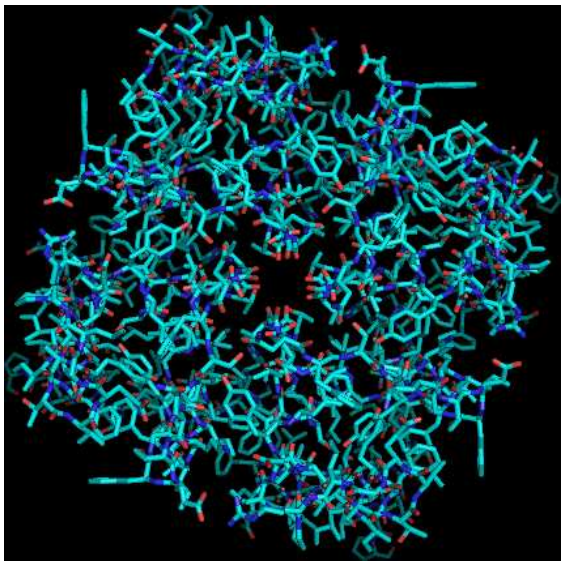


FIG. 1: Stick model of the structure of the K-channel membrane protein, viewed in a direction parallel to the central pore that forms the actual ion channel through a membrane.

The method maintains the advantages of scattering from multiple copies of the molecule to reduce the flux on each molecule (thereby reducing radiation damage) in contrast to proposals for so-called "diffract-and-destroy" experiments on single molecules in a molecular beam and radiation from an x-ray free electron laser (XFEL)[4]. In contrast to crystallography, the scattered radiation is not concentrated in Bragg spots, but forms a completely diffuse angular distribution. This actually helps the process of structure determination as the powerful methods of diffraction microscopy[5] may be brought to bear on the scattering distribution, enabling a reconstruction of a real-space image by an iterative algorithm that alternately satisfies constraints in real and reciprocal space. Of course, for a correlation function to be a useful quantity for the reconstruction of information about a single

particle, pairs of photons contributing to the correlation function need to have scattered off the same particle. This limits the scattered intensities to which such methods are applicable to those which have at least two photons scatterd per particle. However, this is not a serious limitation for protein structure determination by the proposed "diffract-and-destroy" experiment with XFEL radiation, as the best estimates (e.g.[6]) for the number of detected photons per mesured diffraction pattern from a typical protein are about 100.

The basic principles of reconstructing the diffraction pattern of a single particle from those of many identical ones related by rotation about an axis parallel to the incident beam are illustrated in Figs. 2 and 3. First consider, for the sake of illustration, a resolution ring containing just two prominent peaks, a_1 and a_2, of intenisty value 1, with zero intensities elsewhere, as indicated in the outer rings in Fig. 2. Now consider another identical intensity distribution rotated relative to the original by an angle $\Delta\phi$, whose peaks are denoted by $a_{1'}$ and $a_{2'}$ on the inner rings of Fig. 2. The angular autocorrelation of the intensity of the original ring is obtained by treating each intensity distribution as a vector with the intensities of each pixel as its components and taking the scalar products of the two vectors as a function of $\Delta\phi$. Let us denote this function by $C_2(\Delta\phi)$. The values of C_2 are progressively generated as $\Delta\phi$ is varied, as may be seen in the right-hand columns of each panel. For each of the illustrated values of $\Delta\phi$ the values of C_2 are progressively 0, 1, 2, 1, 0.

Now consider the same resolution ring formed by simultaneous illumination of the same particle and a copy rotated relative to it by an arbitrary angle, ψ. The particle in the original orientation will give the same two (blue) peaks, denoted by a_1 and a_2 in Fig. 3. Its rotated copy will give rise to two identical peaks (red) rotated relative to those from the first particle, as denoted by b_1 and b_2. The inner rings contain copies of this intensity distribution rotated relative to this one by (a varying) angle $\Delta\phi$, with copies of the above peaks denoted by $a_{1'}$, $a_{2'}$, $b_{1'}$, and b'_2. respectively. Forming the autocorrelation of this more complicated intensity distribution yields that shown in the right-hand columns of the panels in Fig. 3. The value of C_2 corresponding to the value $\Delta\phi$ illustrated in panel 3(a) is zero, as in panel 2(a) as there is no overlap between any of the non-zero pixels in the two copies of the resolution ring intensity. The situation is different when non-zero pixels overlap. It will be seen that for the value of $\Delta\phi$ illustrated by panel 3(b), there are two overlaps of non-zero pixels, unlike the corresponding single-particle case (2(b)), where there there is only one. From the other panels in Fig. 3, it will be seen that the values of C_2 in this case are progressively 0, 2, 4, 2, 0. Comparing with the C_2 distribution of Fig. 2 shows it to be identical to the former except for a multiplying factor of 2 (because of the double overlaps).

It is not hard to see that N particles rotated relative to the first by different random angles will also give rise to

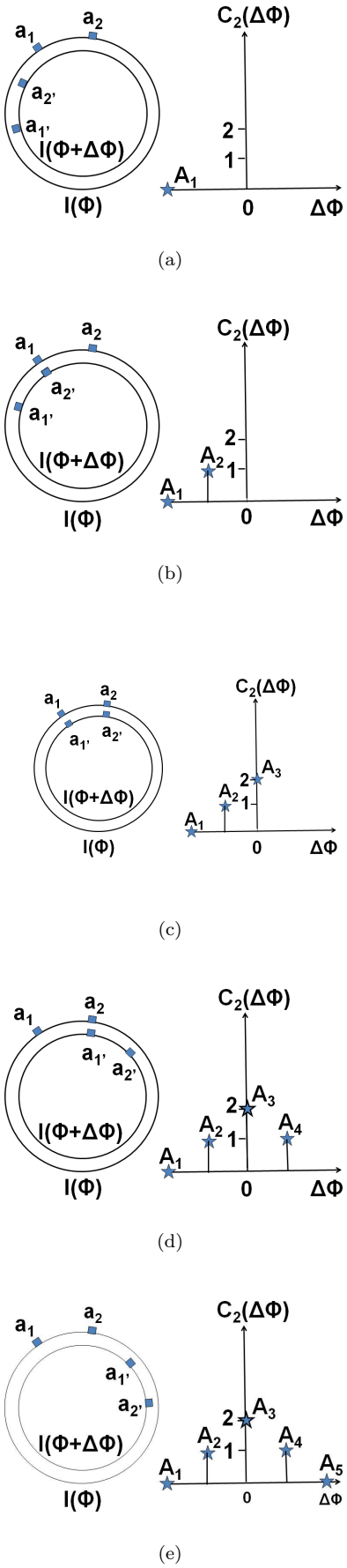
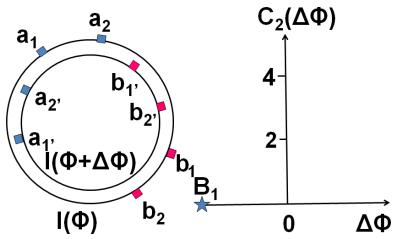
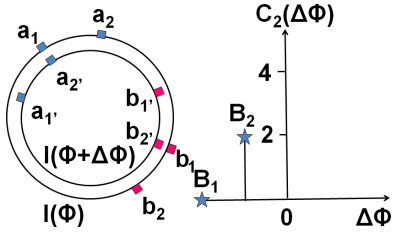


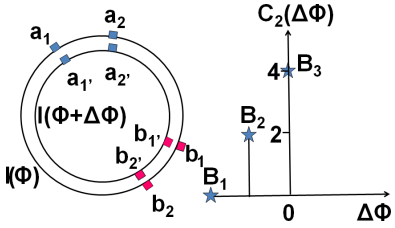
FIG. 2: Construction of the angular autocorrelation $C(\Delta\phi)$ of a single two-peak resolution ring of the diffraction pattern of a single particle, by multiplying by copies with varying relative angle $\Delta\phi$.



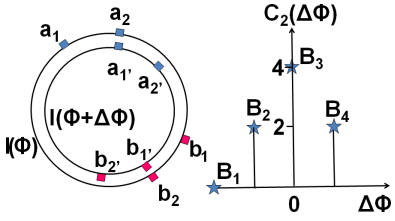
(a)



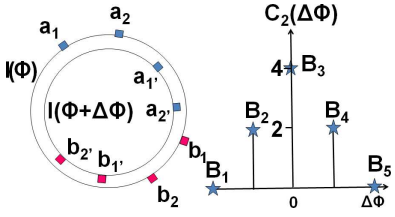
(b)



(c)



(d)



(e)

FIG. 3: Same as Fig. 2 except that the resolution ring has contributions (blue and red) from two particles of random relative orientation. The autocorrelation $C_2(\Delta\phi)$ is the same as in Fig. 2 except scaled by a factor of 2.

the identical autocorrelation C_2, only scaled this time by N. This is the key result we will exploit. *The autocorrelation of any number of particles rotated by random angles relative to each other is the same (apart from an unimportant scaling factor proportional to the number of particles) as that of a diffraction pattern of a single particle.* Thus if it were possible to recover the original angular distribution from the autocorrelation, it must be possible to recover a single-particle diffraction pattern from such a multi-particle one (which in the limit of a large number of particles will resemble a set of rings with no apparent angular distribution). [Strictly speaking, it may be pointed out with reference to Fig. 2 that additional peaks of C_2 occur when, for example the inner intensity peaks $a_{1'}$ and $a_{2'}$ overlap with the outer peaks b_1 and b_2, or when the inner peaks $b_{1'}$ and $b_{2'}$ overlap with the outer peaks a_1 and a_2. However, these overlaps make contributions of order 1 to C_2 at angles $\Delta\phi$ determined by the random interparticle orientations, ψ, unlike the main contributions illustrated in Fig. 2, which coincide for contributions from all particles at $\Delta\phi$=0, and consequently have intensity N. In the limit of a large number of particles N of random orientations, the latter are the only significant contributors to the correlation function C_2.]

If information were available of just the autocorrelation functions of the resolution rings, the intensity distribution of each of the rings would be recovered with random relative angles (this is inevitable since the recovery of the original inensity from the autocorrelations is somewhat akin to a square root operation, which in general has multiple solutions. In this case the multiple solutions are different angular orientations of the intensity distribution). The recovery of a full 2D diffraction pattern (up to an overall rotation) is only possible if one also exploits the angular cross-correlations of the intensities across different resolution rings, as we shall see in the nest section.

II. THEORY

The position of a molecule in an x-ray beam contributes only to unmeasured phases of its scattering amplitude. The measured *intensities* of a diffraction pattern are not sensitive to the position of a molecule a whole. However, a change in the orientation of a molecule is immediately manifested in an observable change in the appearance of a diffraction pattern. For simplicity we restrict ourselves in this paper to molecules differing in orientations about an axis paralell to the direction of an incoming x-ray beam. Even this restricted geometrical variation may apply to molecular ensembles of interest, for example ensembles of K-channel proteins within a flat cell membrane.

We consider here a case where the molecular orientations assumed random only about a single axis, which is assumed parallel to the incident x-ray beam[1,2]. A possi-

ble application is e.g. a set of a membrane bound protein molecules randomly oriented about an axis normal to a cell membrane in a plane perpendicular to the incident beam[2]. If the x-rays are incident on a ensemble of such molecules, randomly oriented about an axis paralell to the incident beam, provided enough molecules are illuminated, the resulting composite diffraction pattern has the appearance of a SAXS pattern, i.e. one with obvious variation with the magnitude q of the scattering vector, but no apparent angular variation for a given q[1,2]. However, this first impression is misleading. If one forms an angular pair correlation function, $C_2(q,q';\Delta\phi)$, from the measured data, where

$$C_2(q,q';\Delta\phi) = \left\langle \frac{1}{N_\phi}\sum_j I'(q,\phi_j)I'(q',\phi_j+\Delta\phi)\right\rangle_{DP}, \quad (1)$$

N_ϕ is the number of azimuthal angles ϕ_j at which the intensities are measured, the angular brackets denote an average over diffraction patterns (DP),

$$I'(q,\phi_j) = I(q,\phi_j) - I_{SAXS}(q), \quad (2)$$

$$I'(q',\phi_j+\Delta\phi) = I(q',\phi_j+\Delta\phi) - I_{SAXS}(q'), \quad (3)$$

and $I_{SAXS}(q)$ is defined as the angular average of the intensities over a particular resolution shell q. This subtraction of the mean value $I_{SAXS}(q)$ exposes the strong angular dependence of the remainder, I'. Consequently, the correlation function, C_2 acquires a strong dependence on $\Delta\phi$. As a function of the angle ϕ, the diffraction pattern intensities on resolution ring q may be represented by the circular harmonic expansion

$$I'(q,\phi) = \sum_{m\neq 0} I_m(q)\exp(im\phi), \quad (4)$$

the omission of the term $m=0$ in the sum being due to the subtraction of the angular average (or SAXS) term. The corresponmding intensity on resolution ring q' at angle $\phi+\Delta\phi$ is likewise given by

$$I'(q',\phi+\delta\phi) = \sum_{m\neq 0} I_m(q')\exp(im(\phi+\Delta\phi)) \quad (5)$$

Substituting (4) and (5) into

$$C_2(q,q';\Delta\phi) = \int I'(q,\phi)I'(q',\phi+\Delta\phi)d\phi \quad (6)$$

its possible to show that[2]

$$C_2(q,q';\Delta\phi) = N_p\sum_{m\neq 0} I_m^*(q)I_m(q')\exp(im\Delta\phi). \quad (7)$$

Since $C_2(q,q';\Delta\phi)$ is real, it is also equal to the complex conjugate of the RHS, and one may also write

$$C_2(q,q';\Delta\phi) = N_p\sum_{m\neq 0} I_m(q)I_m^*(q')\exp(-im\Delta\phi). \quad (8)$$

The angular Fourier transform

$$B_M(q,q) = \frac{1}{N_\phi}\sum_{\Delta\phi} C_2(q,q,\Delta\phi)e^{\pm iM\Delta\phi} \quad (9)$$

of the autocorrelation functions yield the square moduli $|I_M(q)|^2$ (apart from a scaling factor of N_p) regardless of the sign of the exponent in (9). This is not true of the Fourier transforms of the cross-correlations $C_2(q,q';\phi)$ for $q\neq q'$. As we will see, the values of these Fourier transforms do depend on the sign of the exponent in (9), a fact that has consequences for calculating the correct registries of the intensities of the different resolution rings q, as we point out in subsection IID.

The result (9) is just an expression of the well-known result that the Fourier transform of an autocorrelation of a function is the square modulus of the Fourier transform of that function. The fact that there are contributions from N_p particles of random orientations just gives a scaling factor of N_p since the autocorrelations, averaged over many diffraction patterns, tend towards an orientationally independent quantity otherwise characteristic of a single particle. Thus, just taking the square roots of the quantities $B_M(q,q)$ enables the determination of the magnitudes of the circular-harmonic expansion coefficients (up to an unimportant scaling factor) via

$$|I_M(q)| \propto \sqrt{B_M(q,q)} \quad (10)$$

The reconstruction of the intensity distribution of a single particle can then be performed via the equation

$$I(q,\phi) = \sum_M I_M(q)\exp(iM\phi) \quad (11)$$

where the $I_M(q)$ are a set of complex coefficients. The moduli $|I_M(q)|$ may be determined by the angular Fourier transforms (9), but a full determination of these coefficients requires a determination of their phases. Once a single particle diffraction pattern is reconstructed, the electron density of the identical objects giving rise to the composite diffraction pattern may be found if the second phase problem, namely that of finding the phases associated with the amplitudes of the single-particle diffraction pattern may be found.

Thus, the reconstruction of a single-particle diffraction pattern from its angular correlations bears a remarkable similarity to the conventional phase problem of crystallography. In the latter case, the intensities $I(\mathbf{q}) = |A(\mathbf{q})|^2$ are known, but the aim is to find the complex amplitudes $A(\mathbf{q})$, whereas in the present phase problem $|I_M(q)|^2$ may be found from the measured angular autocorrelations, but the aim is to find the complex coefficients $I_M(q)$. Since powerful phasing algorithms have been developed recently for finding $A(\mathbf{q})$ by iteratively satisfying constraints in real and reciprocal space, it has been speculated[7] whether an adaptation of such an algorithm may also solve the phase problem of the complex expansion coefficients $I_M(q)$. Our recent paper[2]

showed that this was indeed possible, by alternately satisfying constraints in "correlation space" and reciprocal space. In the following three subsections we describe alternative approaches to solving this phase problem, all of which have shown some success.

Previous proposals for finding these phases of the $I_M(q)$ coefficients have involved the evaluation and optimization of so-called triple angular correlations[1,8]. However, it has been found in practice[2] that the construction of converged triple correlation functions required the averaging of data from many more measured diffraction patterns than required to construct converged angular pair correlation functions. In addition, such a method generally requires a relatively slow global optimization technique, such as simulated annealing or a genetic algorithm. Thus, a method of phasing which dispenses with the need for triple correlations would be preferable from the point of view of an experimentalist, as it would require less beam time at a central x-ray source, and hence involve a less expensive experiment. A method which avoids the need for a slow optimization method would be advantageous from the point of view the speed of data analysis.

A. Iterative Phasing Algorithm for the Phases of the Circular Harmonic Expansion Coefficients

We first examine the method for determining the phases of the $I_M(q)$ coefficients which dispenses with the need to evaluate triple correlations as proposed earlier[1,9]. At any particular iteration, n, say, the connection to the reciprocal space intensities is given by the equation

$$I^{(n)}(q',\phi) = \sum_M |I_M^{(n)}(q')| \exp i\psi_M^{(n)}(q') \exp(iM\phi), \quad (12)$$

where $\psi_M^{(n)}(q')$ is the phase of the complex number $I_M^{(n)}(q)$. The algorithm proposed[2] was to begin with $|I_M^{(0)}(q')|$ defined as $|I_M^{(exp)}(q')|$, from the square roots of the angular Fourier transform of the angular autocorrelation function of the diffracted intensities of resolution ring q', at the start (the 0-th iteration). Random values are assigned to the phases $\psi_M^{(0)}(q')$ at this iteration. At each iteration, the intensities $I^{(n)}(q',\phi)$ were evaluated with the current phases. These are, of course, a set of intensities in reciprocal space. At this point a "flipping" algorithm is applied to these intensities, according to the prescription of Oszláyi and Süto[10]. That is, the signs of the intensities below a relatively small fraction (say 4%) of the maximum value of $I(q',\phi)$ were flipped (i.e., if positive the signs were changed to negative and *vice versa*). These constitute what Fienup called *object domain operations*. From the resulting set of intensities, $I^{'(n)}(q',\phi)$, say, new circular harmomic expansion coefficients, $I_M^{'(n)}(q')$ were calculated via

$$I_M^{'(n)}(q') = \frac{1}{2\pi}\int I^{'(n)}(q',\phi)\exp(-iM\phi)d\phi. \quad (13)$$

Now constraints in "correlation space" were applied to give improved estimates of the phases by constraining the magnitudes of circular harmonic expansion coefficients to remain at $|I_M^{(exp)}(q')|$ and by defining

$$\psi^{(n+1)} = \arg(I_M^{'(n)}(q')) \quad (14)$$

and the procedure repeated to convergence. This gave an estimate of the complex circular harmonic expansion coefficients $I_M(q')$ for this particular resolution ring. It should be noted that, in principle, this phasing algorithm needs be performed only on a single resolution ring since the coefficients $I_M(q)$ for all other resolution rings q may be found from the equation

$$B_M(q,q') = \frac{1}{N_\phi}\sum_{\Delta\phi} C_2(q,q')\exp(-iM\Delta\phi) = N_p I_M(q) I_M^*(q'). \quad (15)$$

Since the LHS is a quantity that may be found from experiment, and the magnitudes $|I_M(q)|$ of all the circular-harmonic expansion coefficients may be determined, the phases of these coefficients corresponding to the same values of M but different values of q are not independent. In fact, there are only $M_{max}-1$ independent phases that need to be determined (where M_{max} is the maximum value of the azimuthal quantum number used in the calculations). Thus, in principle, this entire phasing problem for the complete 2D diffraction pattern from a single particle may be solved by solving that for a single resolution ring.

B. Use of a Positivity Constraint on the Intensities

The above "flipping" algotithm requires a relatively large dynamic range of the intensities of resolution ring q' (obviously in a case where there is less than about a 20:1 ratio between the highest and lowest values, none of the intensities will be be flipped, and the whole basis of the algorithm will be inoperative). For such cases, we experimented with applying the only obvious constraint on the diffracted intensities, namely that they be non-negative. Remarkably, we found that such a simple constraint sufficed for this particular phasing algorithm. To do this, we defined a cost function consisting of the absolute value of the sum of the negative parts of the intensities $I(q',\phi)$ on resolution ring q' for the current values of the phases $\psi_m(q')$, by evaluating these intensities using Eq. (12), and minimized this cost function with respect to these phases. For the reconstructions reported in section III we used a simulated annealing algorithm[11] to perform

the required global optimization as a function of these parameters. As before, the phases $\psi_(q)$ of the intensity expansion coefficients $I_m(q)$ of the other resolution rings were estimated from the calculated cross-correlations via Eq.(15).

C. Exploiting Hilbert Transform Relations

If the intensities $I(q,\phi)$ of the resolution rings are at least two-fold degenerate (as would be the case due to Friedel's Law for a flat Ewald sphere, or more generally for particles with at least two-fold rotational symmetry about the incident-beam axis) the initial estimates of the phases obtained by either of the above methods may be refined by exploiting Hilbert transform relations between real and imaginary parts of a "causal function" (e.g.[12]).

The real and imaginary parts $x_r(t)$ and $x_i(t)$, respectively, of the Fourier transforms of a causal function of the form $X(\omega)$ where $X(\omega)$ is zero for $\omega < 0$ are known to be related by Hilbert transforms[12]. For our application. we idenitfy $X(\omega)$ with the function $J(q,\phi)$ for each resolution ring q (defined below). Of course, $I(q,\phi)$ is usually a continuous function of ϕ for $\phi = -\pi$ to $+\pi$ for each value of q. However the two-fold repeating nature of this function means that, in general, all its distinct values cover only half the range, say from 0 to $+\pi$. If we define a function

$$J(q,\phi) = \begin{cases} I(q,\phi), & \phi \geq 0 \\ 0, & \phi < 0 \end{cases} \tag{16}$$

its angular Fourier transform is

$$\begin{aligned} J_M(q) &= \int_{-\pi}^{\pi} J(q,\phi) \exp(iM\phi) d\phi \\ &= \begin{cases} I_M/2, & M \text{ even} \\ 0, & M \text{ odd} \end{cases} \end{aligned} \tag{17}$$

i.e. it will be equal to half the Fourier transform of the function of interest, $I(q,\phi)$, for even values of M, and be equal to zero for odd values of M. Since, by construction, $J(q,\phi)$ is a causal function with respect to ϕ, not only are the real and imaginary parts of $J_M(q)$ related by the Hilbert transforms, so are the real and imaginary parts of $I_M(q)$, at least for even M (corresponding relations for odd M are irrelevant, as Friedel's Law implies that $I_M(q) = 0$ for odd M).

At least for a diffraction pattern with C2 symmetry, as here, this relationship may be exploited to refine the phases obtained with the flipping algorithm. (Friedel's Law guarantees C2 symmetry for a flat Ewald sphere. In the present case the symmetry of the molecular projection guarantees this even for a curved Ewald sphere.) The (real) intensities $I(q,\phi)$ of each resolution ring q play the role of the function $X(\omega)$, while the real and imaginary parts of its complex Fourier coefficients $I_M(q)$ may be identified with $x_r(t)$ and $x_i(t)$, respectively. Although $I(q,\phi)$ is not strictly a causal function, if the diffraction pattern has C2 symmetry, it is a two-fold redundant function, which implies $I(q,\phi+\pi) = I(q,\phi)$. Thus the non-zero Fourier coefficients $I_M(q)$ may be found by taking $I(q,\phi) = 0$ for e.g. negative ϕ (assuming ϕ is defined from $-\pi$ to π). This means that the $I_M(q)$'s may be related to Fourier transforms of a causal function, and thus their real and imaginary parts are related by Hilbert transforms. In the phasing of the coefficients $I_M(q)$ we are faced with a situation where both the amplitudes and phases (and therefore the real and imaginary parts) of the $I_M(q)$'s are known for $M = M'$, while only the amplitudes are known for $M \neq M'$. By constraining the known portions of the real and imaginary parts of $I_M(q)$ an iterative algorithm that repeatedly relates the real and imaginary parts of $I_M(q)$ allows the recovery of the real and imaginary parts of all the $I_M(q)$ coefficients on convergence. We used this method to refine the initial estimates of the phases of the $I_M(q)$ coefficients found by the flipping algorithm above.

It should be emphasized that the Hilbert transform method above can only be used if the diffraction data on a resolution ring q have (at least) a two-fold redundancy. This condition always follows from Friedel's Law if the data may be regarded as lying on a flat region of the Ewald sphere, an assumption that becomes less valid for higher-resolution data. Such a redundancy also follows if each scattering object has (at least) two-fold rotation symmetry about the incident-beam direction.

For a general object and higher-resolution data, the assumptions that justify the use of a Hilbert transform method may not be valid, and this method cannot be used, so one is reliant on the accuracy of one the previous two methods.

D. A Word of Caution

It is well known that measurements of diffraction intensities cannot distinguish between different molecular enantiomorphs. Right-handed and left-handed versions of a chiral molecule give rise to identical diffraction patterns. The process of recovering an angular intensity distribution on a resolution ring in a diffraction pattern from its angular correlation suffers from a similar ambiguity.

Given that the only firm constraint from the experimentally measured autocorrelations for ring q', say, is on $I_M^*(q')I_M(q')$, a solution that coverges on $I_M^*(q')$ is just as likely as one which converges on the correct $I_M(q')$. The two solutions are not identical, but related by a simple transformation. The quantity

$$\begin{aligned} I_M(q') &= \sum_M I_M^*(q') \exp(iM\phi) \\ &= \sum_M I_M(q') \exp(-iM\phi) \end{aligned} \tag{18}$$

which is related by inversion symmtery with respect to the angular coordinate compared with one calculated

from the expression

$$I_M(q') = \sum_M I_M(q') \exp(iM\phi). \tag{19}$$

Unfortunately, when the angular Fourier transforms $B_M(q, q')$ of the cross-correlations are used to determine the expansion coefficients $I_M(q)$ of other resolution shells q, the two solutions do not give rise to 2D diffraction patterns which are so simply related.

To see this, note that value of the angular Fourier transform

$$B_M(q, q') = \frac{1}{N_\phi} \sum_{\Delta\phi} C_2(q, q'; \Delta\phi) e^{\pm iM\Delta\phi} \tag{20}$$

for $q \neq q'$ actually does depend on the sign of the exponent in (20), yielding $N_p I^*_M(q) I_M(q')$ if the negative sign is used and $N_p I_M(q) I^*_M(q')$ from the positive sign. Thus if the solution $I_M(q')$ is recovered by the phasing algorithm for the resolution ring q', it is necessary to calculate $B_M(q, q')$ with the negative sign in (20) in order to recover the correct coefficients $I_M(q)$ for the other resolution rings q. On the other hand, if the phasing algorithm for ring q' yields the solution $I^*_M(q')$ (equally likely) it is necessary to calculate $B_M(q, q')$ using the positive sign in the exponent in (20).

Of course, as with all such phasing algorithms, it is not possible to know in advance when it will yield the correct $I_M(q')$ coefficients or their complex conjugates. Our arbitray choice of sign of the exponent to calculate the Fourier transforms (9) breaks the symmetry, and makes only one of these solutions a valid one. In practice, we suggest trying to reconstruct the 2D single-particle diffraction pattern from whatever solution is found for $I_M(q')$ from the phasing algorithm for resolution ring q'. The calculated values of $B_M(q, q')$ for arbitrary choice of sign in (20) may then used to find the values of the expansion coefficients $I_M(q)$ for the other resolution rings q. If the resulting 2D diffraction pattern does not "look right", we suggest that $I_M(q')$ be replaced by $I^*_M(q')$ and the procedure repeated. One of the two reconstructed 2D diffraction pattern will be the correct one apart possibly for an overall angular inversion.

III. ILLUSTRATIVE EXAMPLE

The K-channel protein forms a channel for the microtransport of K ions through a cell membrane e.g. in the process of neurotransmission[3]. To a good approximation, the ion channel has to remain perpendicular to the membrane for it to perform its function. However, the different K-channel molecules in a given membrane may have random angles of orientation about the membrane normal. Our previous paper[2] showed how the angular correlations of a diffraction pattern of an ensemble of such molecules can allow the calculation of the magnitudes $|I_m(q)|$ of the circular harmonic expansion coefficients of a single-particle diffraction pattern, and how an iterative phasing algorithm, supplemented by the Hilbert transform method above may be used to find their phases. This allowed a reconstruction of an oversampled diffraction pattern from a single molecule, from which the projected electron density of the molecule was calculated by a conventional phasing algorithm.

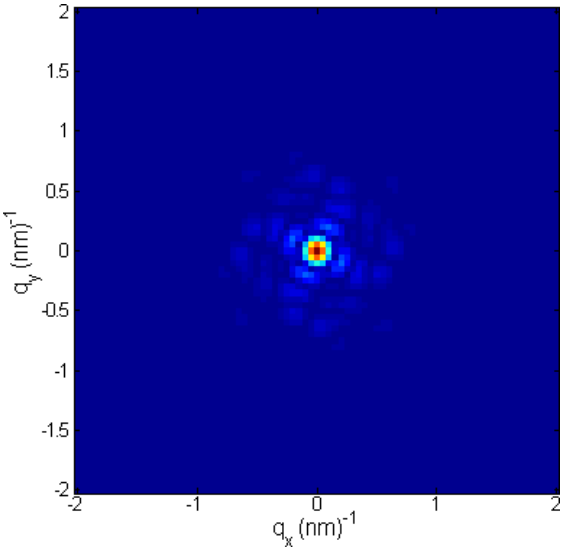


FIG. 4: Amplitudes of a simulated diffraction pattern from x-rays incident down the central pore of a single K-channel protein molecule up to about 3 Å resolution.

In our present paper, we show that similar results may be obtained by the enforcement of a positivity constraint on the diffraction intensities with a simulated annealing algorithm.

Fig. 4 illustrates the expected diffraction pattern of such a molecule, in an orientation with incident x-ray beam parallel to its central pore, as simulated with the structure data in entry 3e8f of the Protein Data Bank (PDB), using the usual structure factor fomula

$$F(q_x, q_y) = \sum_j f_j(q_x, q_y) e^{i(q_x x_j + q_y y_j)} \tag{21}$$

where $f_j(q_x, q_y)$ is the form factor of an atom j whose coordinates projected onto a plane perpendicular to the incident beam are (x_j, y_j), and (q_x, q_y) are the corresponding 2D reciprocal-space coordinates (with maximum vales of $|q_x|$ and and $|q_y|$ of 2 Å^{-1}, or about 3 Å resolution). The projected electron density (Fig. 5) to the resolution of the diffraction pattern was then computed from the inverse Fourier transform of $|F(q_x, q_y; \omega_0)|$, with phases computed by an iterative phasing algorithm[10,13].

If this is compared with, the projection of a stick model of this molecule, Fig. 1 (from the PDB atomic coordinates), it will be seen that, at this resolution, not only is the correct shape of the molecule revealed, but even the projection of the central pore and the X-shaped region somewhat denuded of atoms.

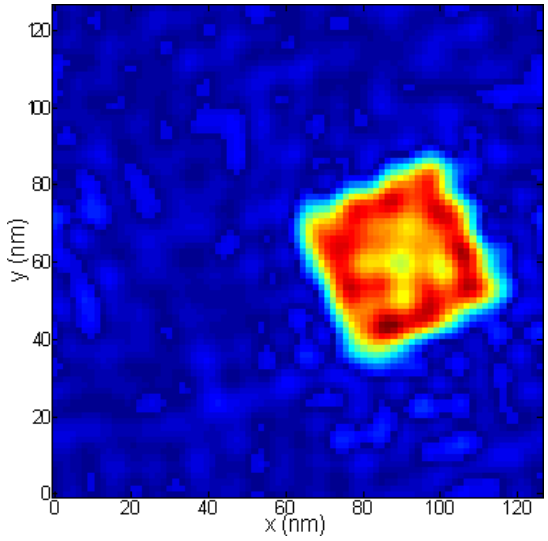


FIG. 5: Electron density of the K-channel protein projected in the direction parallel to its central pore as calculated from a Fourier transform of the scattering amplitudes of Fig. 1 with phases from an iterative phasing algorithm.

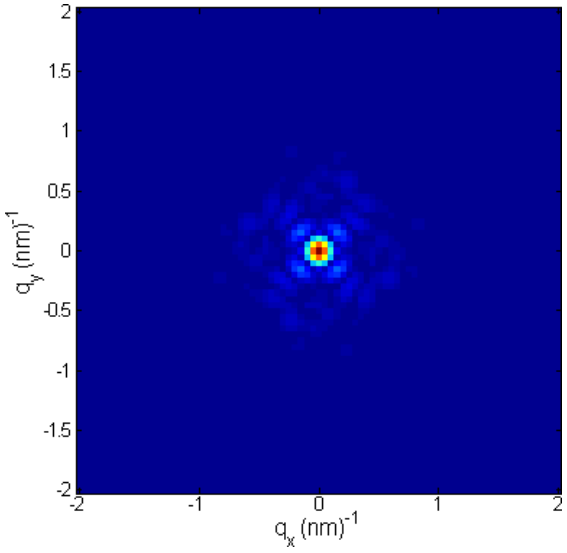


FIG. 7: Amplitudes a diffraction pattern reconstructed from the circular harmonic expansion coefficients $I_m(q)$ extracted from the average of angular correlations of 1000 diffraction pattern of the form of Fig. 6.

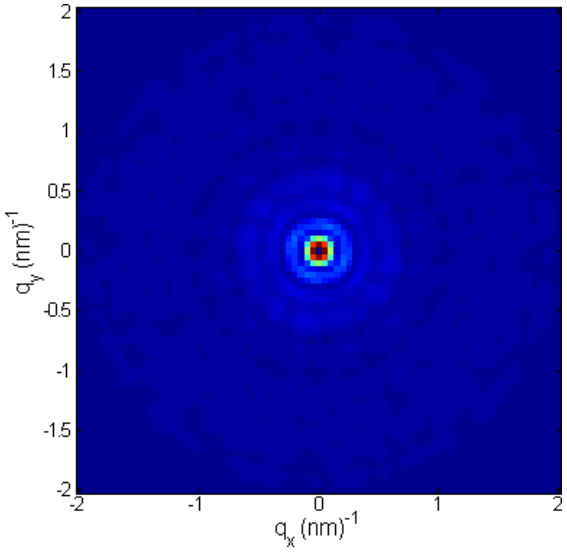


FIG. 6: Amplitudes a typical diffraction pattern from 10 K-channel protein molecules in random orientations about the central pore.

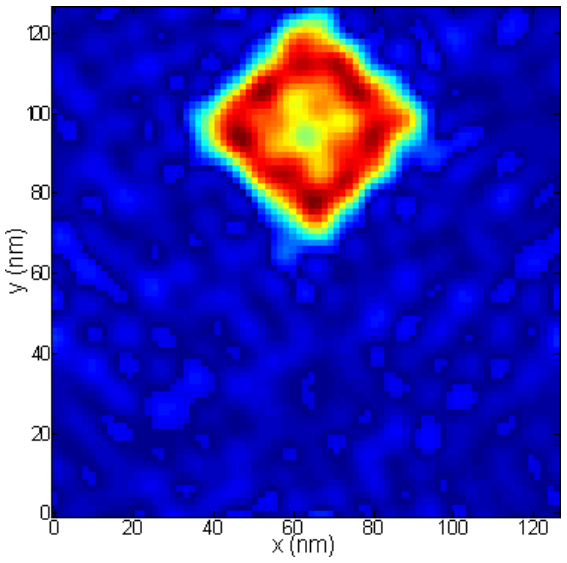


FIG. 8: Same electron density as in Fig. 5, except that projected electron density is reconstructed from the diffraction pattern of Fig. 7.

A typical diffraction pattern is measured from 10 particles in random orientations is shown in Fig. 6. In order to extract useful structural information from such multiparticle diffraction patterns, we computed the average of angular correlations of 1000 such diffraction patterns. We then reconstructed a single-particle diffraction pattern from these averaged angular correlations according to the theory of the last section. The result is shown in Fig. 7. Comparing this with the single-particle diffraction pattern (Fig. 4) shows them to be remarkably similar.

The projected electron density reconstructed from these reconstructed intensities by an iterative phasing algorithm is shown in Fig. 8. This shows that the quality of the reconstruction of Fig. 7 is high enough for a second phasing algorithm to recover the main features of the projected electron density of the molecule, including the square-shaped projection, the central pore, and the X-shaped feature of lower electron density.

IV. ROBUSTNESS OF PAIR CORRELATIONS TO RANDOM NOISE

When scattering off an ensemble of a relatively small number of molecules, the scattered intensities on diffraction patterns are likely to be very low, and thereby to be strongly affected by random noise. It is a legitimate question whether single-particle diffraction patterns may be reconstructed from angular correlations of such noisy data, averaged over a large number of measured diffraction patterns.

Finding the average of pair correlations involves taking products of intensities at corresponding pairs of pixels on different diffraction patterns and averaging them. The theory above shows that, for noise-free data, such averaged correlations tend towards those of a single particle. May the same conclusion may be made of noisy data from a real experiment?

The key is the fact the mean of the product of two independent random variables is the product of the means. To be more specific, assume that the probability of a measured photon count n at any particular pixel is given by a probability function of the form $P(n, \lambda)$, where λ is the expected photon count, i.e. where

$$\lambda = \sum_{n=0}^{\infty} nP(n, \lambda). \tag{22}$$

Note that, although the commonly-assumed (e.g.[6,14]) Poisson distribution

$$P(n, \lambda) = \lambda^n e^{-\lambda}/n! \tag{23}$$

fits this requirement, the argument does not depend on the precise form of $P(n, \lambda)$.

The (mean-subtracted) angular correlations have contributions only from similarly oriented particles. (Scattering from differently oriented particles contributes only to the mean (SAXS) value, which is subtracted off before forming the correlations.) If λ_1 and λ_2 are the expected photon counts at two pixels from a given molecular orientation, the mean value of the products of the actual photon counts due to all particles of similar orientation (to within experimental resolution) is

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} nmP(n, \lambda_1)P(m, \lambda_2)$$
$$= \{\sum_{n=0}^{\infty} nP(n, \lambda_1)\}\{\sum_{m=0}^{\infty} mP(m, \lambda_2)\} = \lambda_1\lambda_2, \tag{24}$$

which is the product of the expected values of the contributions to the pair correlations from particles of that orientation.

It should be noted that these results are valid even if there is some correlation between the means λ_1 and λ_2. Indeed, this is the very correlation being sought. The present analysis suggests that such correlations may be deduced even from very noisy experimental data such as those expected from scattering of radiation from an XFEL by a typical protein molecule. Of course, the fact that the signal being sought is the correlated scattering from a single particle requires that the number of detected scattered photons on a single-particle diffraction pattern to be at least two. However, as pointed out in the Introduction, the number of detected photons in XFEL scattering by a typical protein is expected to be significantly greater than this limit.

V. DISCUSSION

The present paper expands on some of the principles underlying the method recently proposed[1,2] for the determination of the structure of a molecule from diffraction patterns of ensembles of randomly oriented ones, with a possible application to difficult-to-crystallize membrane proteins.

An established method of extracting structural information from randomly oriented molecules is that of small angle x-ray scattering (SAXS)[15–17]. Such experiments are usually performed with radiation of pulse length longer than the rotational diffusion time, τ, of the molecules. Under such circumstances, the signal from each molecule will be its rotational average. Variations in a SAXS signal are greatest in the region of very small magnitude scattering wave vector q, where the signal is sensitive largely to the overall shape of the molecule. Although sophisticated theoretical methods have been developed[15,16] for extracting even anisotropic details of the molecular shape, such methods are inevitably limited by the sparsity of information extractable from experiment since scatterd intensities are a function of the single scalar variable of magnitude, q, of the scattering vector[18]. The information is contained in approximately $N = \Delta q/\delta q$ measurable features (where Δq is the range of q and δq the expected width of a feature in the $I(q)$ curve). δq is expected to have a magnitude given by the Shannon sampling criterion $\delta q = \pi/L$ where L is a typical linear dimension of the molecule. Under typical condition of a SAXS experiment, N is typically of the order 10. In principle, much more information is available in the small fluctuations of intensity in diffraction rings of constant q, which are becoming much more accessible to experiment with the recent advent of much brighter x-ray sources.

This proposed method overcomes this limitation by extracting also hidden information contained in the angular correlations of a SAXS pattern, previously regarded as containing only radial variations of intensity. The method straddles a middle line between conventional crystallography, for which it is necessary to scatter off a large number of perfectly aligned molecules in a crystal, and proposals for single-molecule structure determination using extremely brilliant radiation from an x-ray free electron laser (XFEL) in a so-called "diffract-and

destroy" experiment. By simultaneously scattering off several copies of the molecule, the number of scattered photons per detector pixel is increased substantially compared to single-molecule diffraction experiments. Yet it is possible reconstruct diffraction intensities from a single particle, spread out over the whole diffraction pattern, and not confined to just Bragg spots, as in conventional crystallography, thus allowing the powerful phasing methods of diffraction microscopy to directly reconstruct the real-space structure of a single particle. By analyzing not only the radial variation of the diffraction intensities as in the technique of small-angle x-ray scattering (SAXS), but also the hidden angular variations revealed by angular correlation fuctions[19], information about not only the shape of a molecule, but even of its internal structure, may readily be found. Yet, at the same time, the relaxation of the condition of molecular alignment permits application to molecules, such as many membrane proteins, which are resistant to crystallization.

We describe in this paper several algorithms that have been successfully used for extracting the phases of the circular harmonic expansion coefficients, $I_m(q)$, characterizing a single particle diffraction pattern, without the need to evaluate the so-called angular triple correlations[1,9], converged values of which require the measurement of a much larger number of diffraction patterns, and hence a more expensive experiment.

The results are also relevant also to the kind of "diffract and destroy" approach to biomolecular structure determination[4,6,20] with an x-ray free electron laser (XFEL), where measurements of diffraction from single molecules have been proposed. In this case, an average of the angular correlation functions over many such very weak diffraction patterns allows a building up of a more statistically reliable signal from which it has been proposed that the diffraction volume of an individual molecule may be built up[21]. Since the number of values of the angular correlation functions do not grow as the data of more diffraction patterns are allowed to contribute the the average, this would also be expected to provide an efficient method of data reduction of the millions of diffraction very weak diffraction patterns that are typically measured in such experiments. In addition, the SAXS background, which arises from uncorrelated scattering by *different* particles, and which needs to be subtracted out in the present treatment, is entirely absent, as is scattering from membrane or solution atoms.

We have also suggested that the method may be applicable also to x-ray fluorescence patterns from a specific atomic species buried deeply within large molecules[22] (e.g. the Fe atom in haemoglobin) in solution. Such patterns would be most sensitive to intramolecular scattering, less to that from its surroundings. This may allow the reconstruction of the 3D diffraction volume of an individual molecule from that of many randomly oriented ones, and hence of the molecular structure, without too much interference from solvent (or membrane) scattering.

VI. CONCLUSIONS

Recent advances in technology, such as fast column read-out area detectors, brighter sources, shorter pulses, and zone-plate focusing, have dramatically improved the possibility of making the kind of sensitive measurements needed to successfully record the minute angular intensity fluctuations. Wochner et al.[19] have suggested the use of such measurements to reveal hidden symmetries of short-range order in amorphous materials. For scattering from an ensemble of randomly oriented particles, it has been suggested[1,2] that the angular correlations of intensity may even form the basis of the structure determination of identical randomly oriented particles, thus potentially allowing a relaxation of a requirement for crystallization, while still allowing the sharing of radiation dose amongst different particles, thus greatly reducing the dose per particle. This reduction could have a significant impact on structure determination of radiation sensitive particles, such as biomolecules. The present paper clarifies some aspects of the theory of fluctuation scattering, particularly its use for the structure determination of randomly oriented particles.

Acknowledgments

We thank Marius Schmidt for help with generating an atomic model of the K-channel membrane protein, and John Spence and Henry Chapman for helpful discussions. We acknowledge support from DOE Grants DE-SC0002141 and DE-FG02-06-ER46277 and the Research Growth Initiative of the University of Wisconsin-Milwaukee.

[1] D. K. Saldin, V. L. Shneerson, M. Howells, S. Marchesini, H. N. Chapman, M. Bogan, D. Shapiro, R. A. Kirian, U. Weierstall, K. E. Schmidt, and J. C. H. Spence, New J. Phys. **12**, 035014 (2010).
[2] D. K. Saldin, H. C. Poon, V. L. Shneerson, M. Howells, H. N. Chapman, R. A. Kirian, K. E. Schmidt, and J. C. H. Spence, Phys. Rev. B **81**, 174105 (2010).
[3] D. A. Doyle, J. M. Cabral, R. A. Pfeutzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon, Science **280**, 69 (1998).
[4] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, Nature **406**, 752 (2000).
[5] S. Marchesini, Rev. Sci. Instr., **78**, 049901 (2007).
[6] R. Fung, V. L. Shneerson, D. K. Saldin, and A. Ourmazd, Nature Physics **5**, 64 (2009).
[7] J. C. H. Spence, private communication (2009).
[8] Z. Kam, Macromolecules **10**, 927 (1977).
[9] Z. Kam, J. Thoer. Biol. **82**, 15 (1980).

[10] G. Oszlànyi and A. Süto, Acta Cryst. A **60**, 134 (2004).
[11] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[12] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing* (Prentice-Hall, Upper Saddle River, 1999).
[13] G. Oszlànyi and A. Süto, Acta Cryst. A **61**, 147 (2005).
[14] N.-T. D. Loh and V. Elser, Phys. Rev. E **80**, 026705 (2009).
[15] H. B. Stuhrmann, Acta Cryst. A **26**, 297 (1970).
[16] D. I. Svergun and H. B. Stuhrmann, Acta Cryst. A **47**, 736 (1991).
[17] V. L. Shneerson and D. K. Saldin, Acta Cryst. A **65**, 128 (2009),
[18] D. I. Svergun and M. H. Koch, Curr. Opin. Struct. Biol., **12**, 654, (2002).
[19] P. Wochner, C. Gutt, T. Autenreith, T. Demmer, V. Bugaev A. D. Ortiz, A. Duri, F. Zontone, G. Greubel, and H. Dosch, PNAS **106**, 11511 (2009).
[20] V. L. Shneerson, A. Ourmazd, and D. K. Saldin, Acta Cryst. A **64**, 303 (2008).
[21] D. K. Saldin, V. L. Shneerson, R. Fung, and A. Ourmazd, J. Phys.: Condens. Matter **21**, 134014 (2009).
[22] G. Faigel, G. Bortel, C. S. Fadley, A. S. Simionovici, and M. Tegze, X-Ray Spectrometry **36**, 3 (2007).