IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Structure of isolated biomolecules obtained from ultrashort x-ray pulses: exploiting the symmetry of random orientations

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2009 J. Phys.: Condens. Matter 21 134014

(http://iopscience.iop.org/0953-8984/21/13/134014)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 129.252.86.83
The article was downloaded on 29/05/2010 at 18:49

Please note that terms and conditions apply.

J. Phys.: Condens. Matter **21** (2009) 134014 (11pp) doi:10.1088/0953-8984/21/13/134014

Structure of isolated biomolecules obtained from ultrashort x-ray pulses: exploiting the symmetry of random orientations

D K Saldin, V L Shneerson, R Fung and A Ourmazd

Department of Physics, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI 53211, USA

Received 14 October 2008
Published 12 March 2009
Online at stacks.iop.org/JPhysCM/21/134014

Abstract

Amongst the promised capabilities of fourth-generation x-ray sources currently under construction is the ability to record diffraction patterns from individual biological molecules. One version of such an experiment would involve directing a stream of molecules into the x-ray beam and sequentially recording the scattering from each molecule of a short, but intense, pulse of radiation. The pulses are sufficiently short that the diffraction pattern is that due to scattering from identical molecules 'frozen' in random orientations. Each diffraction pattern may be thought of as a section through the 3D reciprocal space of the molecule, of unknown, random, orientation. At least two algorithms have been proposed for finding the relative orientations from just the measured diffraction data. The 'common-line' method, also employed in 3D electron microscopy, appears not best suited to the very low mean photon count per diffraction pattern pixel expected in such experiments. A manifold embedding technique has been used to reconstruct the 3D diffraction volume and hence the electron density of a small protein at the signal level expected of the scattering of an x-ray free electron laser pulse from a 500 kD biomolecule. In this paper, we propose an alternative algorithm which raises the possibility of reconstructing the 3D diffraction volume of a molecule *without determining the relative orientations of the individual diffraction patterns*. We discuss why such an algorithm may provide a practical and computationally convenient method of extracting information from very weak diffraction patterns. We suggest also how such a method may be adapted to the problem of finding the variations of a structure with time in a time-resolved pump–probe experiment.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The expected advent, in the next few years, of fourth-generation x-ray sources of ultrashort pulses of hard x-rays through the x-ray free electron lasers (XFELs) currently under construction in the US, Japan, and Europe (Normille 2006) gives rise to the very real possibility of a recording of a diffraction pattern of a single molecule from a single ultrashort radiation pulse. The photoionization caused by the intense pulse is expected to cause a Coulomb explosion of the molecule in about 50 fs (Neutze *et al* 2000). However, since the pulse duration may be significantly shorter, the possibility exists of recording a meaningful diffraction pattern of the molecule while it is still in something like its original state.

The need to determine the structure of an individual biomolecule stems from the main limitation of biomolecular x-ray crystallography as it is currently practised, namely that not all biomolecules can be crystallized. Indeed, some 40% of biomolecules do not crystallize, and many cannot easily be purified. Although more than $\frac{1}{2}$ million proteins have been sequenced, the structures of less than 10% have been determined (Protein Data Bank, http://www.pdb.org). Thus, the ability to determine the structure of individual biological molecules—without the need for crystallization—would constitute a significant breakthrough.

 © 2009 IOP Publishing Ltd Printed in the UK

Since a single molecule does not have translational periodicity like a crystal, its diffraction pattern is expected to be continuous, or diffuse, allowing the possibility of *oversampling* the pattern in comparison to the usual Bragg sampling interval in reciprocal space, $\Delta q = 2\pi/L$, where L is a linear dimension of the molecule. As pointed out by Miao *et al* (1999) this may allow the determination of the phases associated with the measured intensities, and hence of the molecular electron density, by an iterative phasing algorithm (e.g. Fienup 1978, 1982; Oszlányi and Süto 2004, 2005).

Of course, even in traditional x-ray crystallography, in general a 3D structure cannot be determined from a single diffraction pattern alone. Many diffraction patterns from many tilts of the crystal relative to the x-ray beam are needed. This is necessary since a single diffraction pattern provides information about only a limited slice through the 3D reciprocal space of the molecule, knowledge of all of which is needed for a successful 3D structure determination[1].

In the proposed single-molecule crystallography, obviously a single XFEL pulse can likewise produce information only about one slice through the reciprocal space of the molecule. It is proposed that this limitation be overcome by directing a beam of identical hydrated protein molecules into the x-ray beam by electrospraying or via Rayleigh-droplet formation (Fenn 2002, Spence *et al* 2005). The molecular density is controlled to be small enough that there is unlikely to be more than one molecule in the beam at a given time. Each diffracted x-ray pulse would generate a diffraction pattern of a molecule in a particular (random) orientation (see figure 1).

The collection of diffraction patterns from a complete ensemble of randomly oriented molecules would then, in principle, contain enough information for performing the 3D structure determination. A problem is that, unlike for the case of a crystal, where the orientations are controlled by a goniometer, and thus known, the individual molecules have random and initially *unknown* orientations. It has been proposed (e.g. Huldt *et al* 2003) that the relative orientations of the diffraction patterns in 3D reciprocal space may be determined through their 'common lines', in analogy with work on projected electron microscope images in cryo-electron microscopy (Frank 2006). The first demonstration of such an algorithm for simulated noise-free x-ray diffraction patterns from a set of 480 random molecular orientations was by Shneerson *et al* (2008). This paper also demonstrated that the resulting *oversampled* (Miao *et al* 1999) 3D intensity distribution may be inverted to recover the 3D molecular electron density by means of an iterative phasing algorithm.

An investigation of the effectiveness of the algorithm for reduced mean photon counts (MPC) per pixel and resulting Poisson noise was also described in that paper. It was found that the method ceased to be effective for an MPC per pixel of less than about 10. This is a far cry from the MPC per pixel of about 4×10^{-2} in the high-resolution part of the diffraction pattern expected from a 500 kD molecule exposed to a single XFEL pulse (Shneerson *et al* 2008). At such levels of detected signal, it is difficult to imagine how it would be possible to even identify common lines on each of the diffraction patterns.

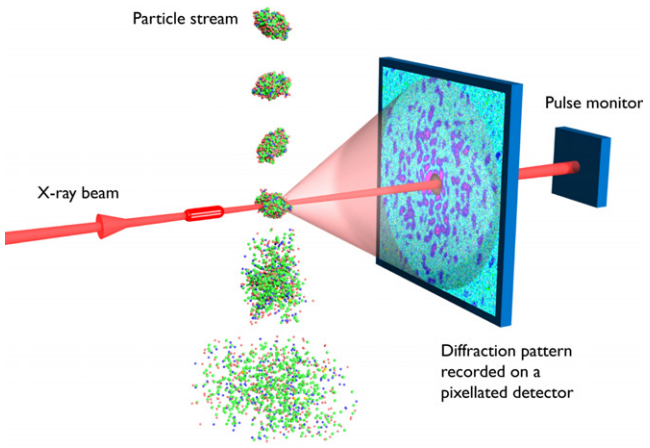

Figure 1. Schematic diagram of the proposed single-molecule diffraction experiment with a x-ray free electron laser (XFEL). (Graphic reproduced from Gaffney and Chapman (2007), with kind permission. Copyright 2007 AAAS.) Each molecule disintegrates about 50 fs after illumination with the XFEL pulse (Neutze *et al* 2000). However, if the pulse is significantly shorter than this, one may expect a single-pulse diffraction pattern to be characteristic of the original atomic configuration of the molecule. A new molecule is then illuminated by another pulse to give rise to another diffraction pattern, albeit from a different (random) molecular orientation, and then another and so on. It is assumed that the resulting set of a large number of very weak diffraction patterns contains enough information for the reconstruction of the 3D molecular structure.

The diffraction pattern pixels marking the positions of common lines consist of a small fraction of all the data of a diffraction pattern. Thus, it may be argued that use of a common-line method for finding the relative orientations of diffraction patterns is rather inefficient. A much more effective method of identifying the relative orientations in 3D reciprocal space of very weak diffraction patterns has been proposed by Fung *et al* (2008). In this technique, each diffraction pattern is represented by a p-dimensional vector (where p is the number of pixels in each diffraction pattern). The magnitudes of the components of this vector are just the numbers of photons in each of the pixels. If the ends of such vectors are plotted in the p-dimensional space, they would be expected to trace out a three-dimensional manifold in the p-dimensional space (where $p \gg 3$) subject to noise. The reason is that, for a given molecule, there are just three *latent variables* associated with each of the diffraction patterns, the three Euler angles specifying their spatial orientations in 3D. It was shown by Fung *et al* that it is possible to find their relative orientations from the ensemble of diffracted photons by the manifold embedding technique known as generative topographic mapping (Bishop 1998). By combining diffraction patterns with MPCs of about 4×10^{-2}, related by rotation about two mutually perpendicular axes (each perpendicular also to the incident beam direction), it was shown how the

[1] However, it should be noted that a single diffraction pattern from a bundle of fibers from beam incidence normal to the fiber axis sometimes suffices for structure determination, since, in this case, the 3D diffraction volume may be generated by sweeping this pattern about a line through its center parallel to the fiber axis. The most famous example is the so-called 'Maltese cross' diffraction pattern (Franklin and Gosling 1953, Watson and Crick 1954) from which the essential elements of the DNA structure were deduced.

3D reciprocal space of the sample molecule may be filled out sufficiently densely to allow the reconstruction of the electron density of the small protein chignolin (PDB entry 1UAO) using an iterative phasing algorithm. In principle, the same method may be used for diffraction patterns of completely random orientations of the $SO(3)$ group.

In the present paper, we propose an alternative approach to structure determination from the same set of simulated diffraction patterns. We examine the possibility that it may not be necessary to determine the relative orientations of the individual diffraction patterns in order to reconstruct the 3D reciprocal-space distribution. Instead, we focus on the possibility of extracting, from the *ensemble* of measured diffraction patterns, the one feature they all have in common, namely that each represents a 2D section (not necessarily planar for a curved Ewald sphere) through a single 3D reciprocal-space distribution of scattered intensity from a single molecule or molecular ensemble. In particular, we will show that the coefficients of a shell-by-shell spherical harmonic expansion of the 3D intensity distribution may be determined from cross-correlations between the intensities at different pixels in the ensemble of measured diffraction patterns.

Hence, the method proposed bears some similarity to that of fluctuation x-ray scattering, as proposed by Kam (1978), where diffraction patterns from a short x-ray pulse are measured for a set of randomly oriented proteins in solution. In a usual x-ray experiment on molecules in solution, the random orientations of the molecules result in a scattering signal which depends only on the magnitude q of the scattering vector, $\mathbf{q} = \mathbf{k} - \mathbf{k}_s$, where $\mathbf{k}$ is the wavevector of the incident x-rays, and $\mathbf{k}_s$ is that of the scattered x-rays. (We define $|\mathbf{k}| = 2\pi/\lambda$, where λ is the x-ray wavelength.) The resulting variation $I(q)$ of the scattered intensity is the usual small angle x-ray scattering (SAXS) signal (e.g. Svergun and Stuhrmann 1991). Kam suggested that, if the diffraction patterns arise from scattering of short-pulse radiation, deviations from the SAXS signal may be observable. He suggested that measurements of the cross-correlations amongst these fluctuations may enable the reconstruction of the structure of the dissolved protein molecules. An experimental difficulty is that the fluctuations sought are a small fraction of the total measured signal.

We point out in this paper that measurements of diffraction patterns from individual molecules, as proposed for an XFEL experiment, allow much more direct access to the intensity correlations, without the background isotropic (SAXS) signal, which may itself nevertheless be reconstructed if required as a part of a much larger database of useful signals. Despite the fact that the intensity correlations are difficult to measure accurately on an individual diffraction pattern (due to the low MPC per pixel), they may be measured to arbitrary accuracy by summing the signal over the large number of diffraction patterns expected to be measured in an XFEL experiment. What is more, the size of the array of these cross-correlations, which forms the input to the reconstruction algorithm, does not increase as more diffraction patterns from random molecular orientations are added. This keeps the algorithm computationally tractable however large the measured data set.

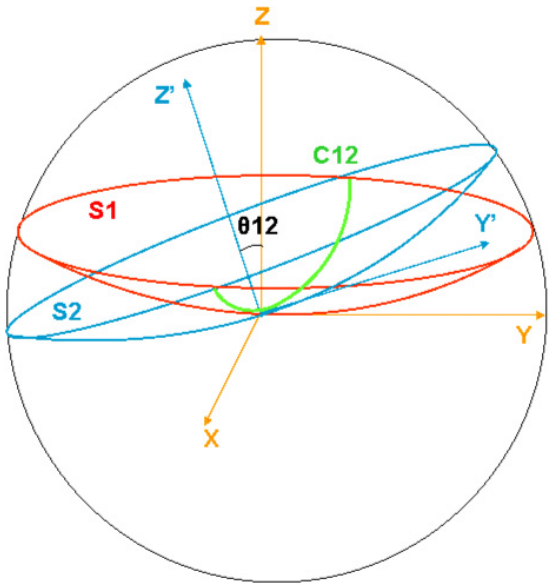

Figure 2. Construction of a 3D diffraction volume from Ewald spheres of random orientation. Two Ewald spheres are shown, one (S1) due to beam incidence antiparallel to the Z axis, and one (S2) due to beam incidence antiparallel to the Z' axis. The orientation of each Ewald sphere is specified by the set of three Euler angles (ϕ, θ, ψ). The difference between the θ Euler angles of the two spheres is θ_{12}. The line of intersection of the two Ewald spheres (the *common line*) is specified by C12.

2. Theory

The starting point of the theory is the expression of the distribution of the scattered intensity I over the 3D molecular reciprocal space (represented by the polar coordinates, q, θ, and ϕ) as the spherical harmonic expansion

$$I(q, \theta, \phi) = \sum_{lm} I_{lm}(q) Y_{lm}(\theta, \phi) \tag{1}$$

where $Y_{lm}(\theta, \phi)$ is a spherical harmonic.

A diffraction pattern represents a section through the reciprocal space of a molecule in a given orientation. For example, if one imagines this reciprocal space labeled by a set of three Cartesian axes, X, Y, and Z (see figure 2), and we take the x-ray incidence direction to be opposite to that of the Z axis, then the diffraction intensities are those that would lie on a portion of the Ewald sphere (S1) of radius equal to the wavenumber $\kappa = 2\pi/\lambda$ (where λ is the wavelength of the x-rays). The view of this Ewald sphere from a direction antiparallel to the X axis is shown in figure 3, and that antiparallel to the Z axis in figure 4.

The measured diffraction pattern samples the 3D reciprocal space of the molecule on this Ewald sphere. In terms of the radial distance q, and polar and azimuthal angles θ and ϕ, specifying points in this reciprocal space, inspection of the geometry of figures 2–4 makes it clear that sets of points on each Ewald sphere, and hence the measured diffracted intensities, may be specified by polar and azimuthal angles in the frame of reference of each diffraction pattern, with the following relation between the polar angle and the radial

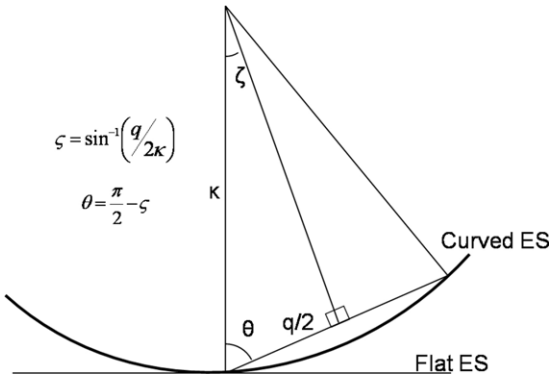

Figure 3. Section through the Ewald sphere S1 (see figure 2), viewed antiparallel to the X axis.

distance q:

$$\theta(q) = \pi/2 - \sin^{-1}(q/2\kappa). \tag{2}$$

This relationship correctly takes account of the curvature of the Ewald sphere for arbitrary x-ray wavenumber κ. Consequently, one may alternatively specify any point on the measured diffraction pattern by a combination of q and ϕ, as illustrated in figure 4. In fact, the measured intensity in a diffraction pattern arising from radiation incident antiparallel to the Z axis is

$$I_Z(q,\phi) = \sum_{lm} I_{lm}(q) Y_{lm}(\theta(q),\phi). \tag{3}$$

In the frame of reference fixed relative to the reciprocal space of the molecule, the intensity $I^{(w)}(q,\phi)$ on a diffraction pattern due to a different molecular orientation specified by an index w may be thought of as sampling a different section (S2) through the same 3D reciprocal space, rotated relative to the above through three Euler angles. Thus, the measured diffracted intensity for the wth molecular orientation may be written as

$$I^{(w)}(q,\phi) = \sum_{lmm'} D^{(w)}_{lmm'} I_{lm'}(q) Y_{lm}(\theta(q),\phi) \tag{4}$$

where $D^{(w)}_{lmm'}$ is a Wigner D-matrix which rotates a spherical harmonic Y_{lm} through the given Euler angles.

Now consider the cross-correlation $J(q,\phi;q',\phi')$ over all diffraction patterns (DPs) of the measured intensities at two pixels specified by (q,ϕ) and (q',ϕ') illustrated in figure 4. If the number of DPs is N,

$$\begin{aligned} J(q,\phi;q',\phi') &= \frac{1}{N}\sum_w I^{(w)}(q,\phi) I^{(w)}(q',\phi') \\ &= \frac{1}{N}\sum_w \sum_{lmm'} D^{(w)*}_{lmm'} I^*_{lm'}(q) Y^*_{lm}(\theta(q),\phi) \\ &\quad \times \sum_{l'm''m'''} D^{(w)}_{l'm''m'''} I_{l'm'''}(q') Y_{l'm''}(\theta'(q'),\phi'). \end{aligned} \tag{5}$$

The Wigner D-functions $D^{(w)}$, which are functions of the three Euler angles specifying full rotations of the molecule, are representations of the full rotation, or $SO(3)$, Lie group. Each group element is specified by a given set of the Euler angles. It is assumed that, for a large number of measured diffraction patterns, this set of angles spans the entire space of group elements uniformly. Note also that the summation over w in (5) involves only the D-functions. Performing the sum over w, which is then effectively a sum over the space of all the elements of the $SO(3)$ group, and applying the great orthogonality theorem (see e.g. Tinkham 2003), we find

$$\frac{1}{N}\sum_w D^{(w)*}_{lmm'} D^{(w)}_{l'm''m'''} = \frac{1}{2l+1}\delta_{ll'}\delta_{mm''}\delta_{m'm'''}. \tag{6}$$

Performing the sum over w first in (5), making use of the great orthogonality relation (6), and then summing over l', m'', and m''' leads to the following simplification of the expression (5) for the intensity cross-correlation function:

$$J(q,\phi;q',\phi') = \sum_l F_l(qq';\phi\phi') B_l(q,q') \tag{7}$$

where

$$\begin{aligned} F_l(qq';\phi\phi') &= \frac{1}{2l+1}\sum_m Y^*_{lm}(\theta(q),\phi) Y_{lm}(\theta'(q'),\phi') \\ &= \frac{1}{4\pi} P_l[\cos\theta(q)\cos\theta'(q') \\ &\quad + \sin\theta(q)\sin\theta'(q')\cos(\phi-\phi')] \end{aligned} \tag{8}$$

where P_l is a Legendre polynomial of order l, and

$$B_l(q,q') = \sum_{m'} I^*_{lm'}(q) I_{lm'}(q'). \tag{9}$$

The indices q and q' are common to the two sides of equation (7). Therefore, for a particular pair of qq' indices, (7) may be written as the matrix equation

$$J_{\phi\phi'} = \sum_l F_{\phi\phi',l} B_l. \tag{10}$$

All elements of the matrix F consist of real-valued Legendre polynomials. Thus, the above equation is purely real, and may be solved for the real coefficients B_l by matrix inversion:

$$B_l = \sum_{\phi\phi'} \{F^{-1}\}_{l,\phi\phi'} J_{\phi\phi'} \tag{11}$$

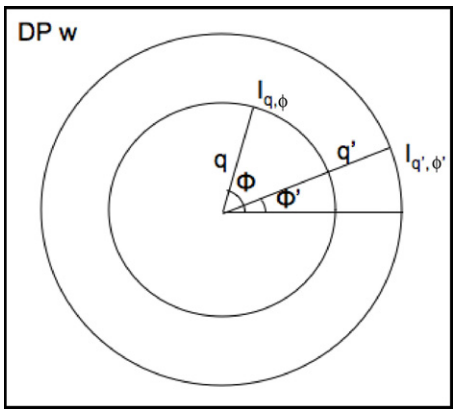

Figure 4. A diffraction pattern pixel may be labeled by the magnitude q of the scattering vector, and an azimuthal angle ϕ in the frame of reference attached to the diffraction pattern. A set of intensity cross-correlations may be constructed by multiplying the intensities $I_{q,\phi}$ and $I_{q',\phi'}$ on each diffraction pattern (w) and summing over all diffraction patterns.

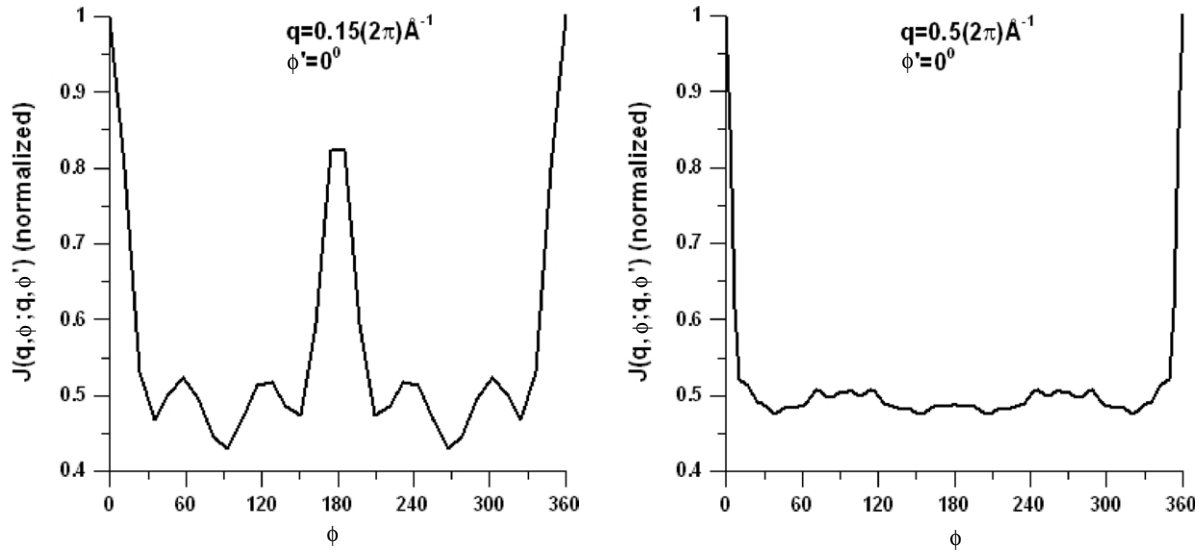

Figure 5. Plots of the intensity correlation coefficient $J(q\phi; q'\phi')$ versus ϕ for $\phi' = 0°$ for the values of $q = q'$ indicated. These quantities were calculated from simulated data for diffraction patterns of a randomly oriented molecule, chignolin (PDB entry: 1UAO) of P1 symmetry. Each of these plots is a linear combination of Legendre polynomials, exactly as predicted by theory. The structural information resides in the magnitudes of the expansion coefficients $B_l(q, q)$ of the Legendre polynomials.

where the cross-correlations J may be evaluated from the measured diffraction data. Putting back the qq' indices, we see that we may find each of the quantities $B_l(q, q')$ (equation (9)) from the measured data.

3. Numerical tests

The equations take a particularly simple form for $q = q'$, i.e. for correlations between intensities on the same resolution ring. Then, equation (8) may be written as

$$F(qq; \phi\phi'; l) = \frac{1}{4\pi} P_l[\cos^2\theta(q) + \sin^2\theta(q)\cos(\phi - \phi')] \tag{12}$$

and hence (7) may be written as

$$J(q\phi; q\phi') = \frac{1}{4\pi}\sum_l P_l[\cos^2\theta(q) + \sin^2\theta(q)\cos(\phi - \phi')] \times B_l(q, q). \tag{13}$$

The quantities $B_l(q, q)$ are real constants, and $J(q\phi; q\phi')$ is a linear combination of Legendre polynomials that, for a given q, is a function of $(\phi - \phi')$ only. This was verified by evaluating $J(q\phi; q\phi')$ from 358 400 simulated diffraction patterns of random orientations of the small protein chignolin (PDB entry 1UAO) via the first equality of (5). The results are shown in figure 5, where $J(q\phi; q\phi')$ is plotted against ϕ, taking $\phi' = 0$ for various values of q.

Furthermore, Friedel's rule

$$I(\mathbf{q}) = I(-\mathbf{q}) \tag{14}$$

imposes limitations on the allowed values of l. Equation (14) may be expanded as

$$\sum_{lm} I_{lm}(q)Y_{lm}(\hat{\mathbf{q}}) = \sum_{lm} I_{lm}Y_{lm}(-\hat{\mathbf{q}}) = \sum_{lm} I_{lm}(-1)^l Y_{lm}(\hat{\mathbf{q}}). \tag{15}$$

The only way the first and third expressions above can be equal is if l is even. Thus, the sum in (13) is over just the Legendre polynomials of even orders, 0, 2, 4, ...

We used (11) to calculate $B_l(q, q)$ from data for the same set of simulated diffraction patterns and have plotted the resulting quantities as functions of l in figure 6 for the particular values of q (=0.15(2π) and 0.5(2π) Å^{-1}) corresponding to the plots in figure 5. The crosses represent values of the same quantity calculated from the spherical harmonic expansion coefficients $I_{lm}(q)$ of the 3D distribution of scattered intensities computed directly from the assumed structure of the model protein (as noted earlier, only even values of l give non-zero contributions). The near perfect agreement between the lines and crosses is an indication of the correctness of the theory, and suggests that the quantities $I_{lm}(q)$ may be found from the intensity cross-correlations of measured diffraction patterns.

If this can be done, the resulting intensity distribution could be evaluated on an 'oversampled' (Miao *et al* 1999) Cartesian grid, from which the molecular electron density may be found using an iterative phasing algorithm.

Finding the coefficients $I_{00}(q)$ from the elements $B_0(q, q)$ is trivial. For $l = 0$, the only allowed value of m is also 0. Then, the summation on the RHS of (9) reduces to a single term, that is

$$B_0(q, q) = I_{00}(q)I_{00}(q), \tag{16}$$

and since $I_{00}(q)$ is real, $I_{00}(q)$ may be found by simply taking the square roots of the quantities $B_0(q, q)$ (the only ambiguity is in the sign of the square root, and this ambiguity is removed by the physical requirement of positive intensities).

$I_{00}(q)$ represents the angular average of the scattered intensity on the (reciprocal-space) resolution shell specified by the magnitude q of the scattering vector. This suggests the following test. One may start with, say, a protein of

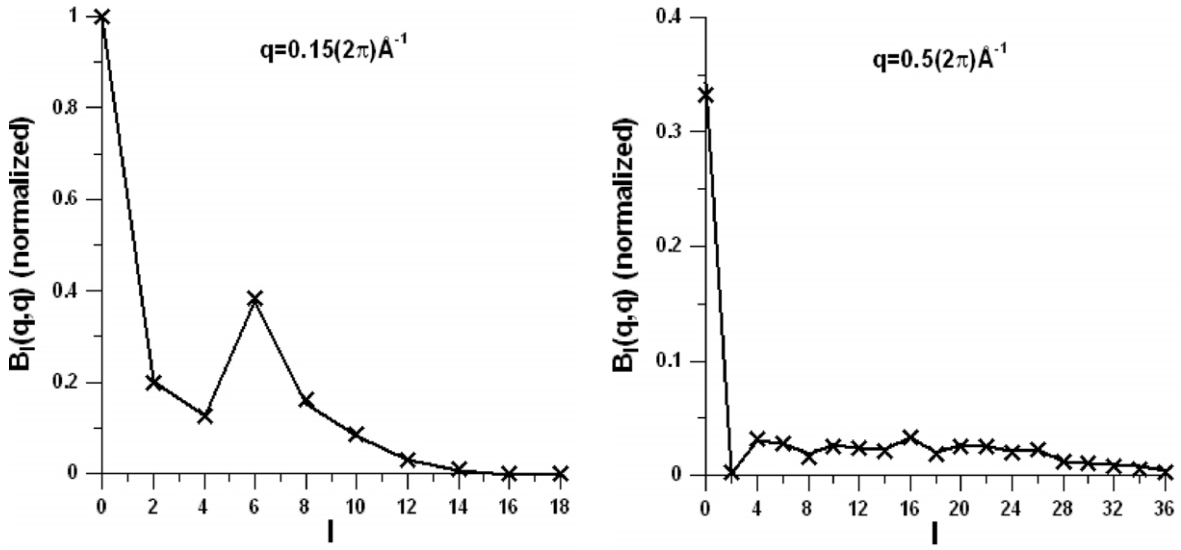

Figure 6. Plots of $B_l(q, q)$ versus l for values of q corresponding to each of the plots in figure 5. These plots give the ratio of the contributions of the angular momenta l to each of the cross-correlations $J(q\phi; q0)$ computed from the simulated diffraction patterns. The crosses represent values of the same quantity calculated from (9) with the spherical harmonic expansion coefficients $I_{lm}(q)$ of the 3D distribution of scattered intensities computed directly from the assumed structure of the model protein (as noted in the text, only even values of l give non-zero contributions). The near perfect agreement between the lines and crosses is an indication of the correctness of the theory, and suggests the possibility that the quantities $I_{lm}(q)$ may be extracted from the intensity cross-correlations of measured diffraction patterns. The two plots shown are on the same scale, with $B_0(0.15(2\pi), 0.15(2\pi))$ taken to be unity, as shown.

known structure from the Protein Data Bank, and for a fixed orientation, calculate directly the distribution of scattered intensities over a set of spherical resolution shells of the 3D molecular reciprocal space. An averaging of these intensities over this set of resolution shells would be expected to yield the quantities $I_{00}(q)$.

One may also simulate a large number of diffraction patterns from random orientations of the same molecule. This set of diffraction data will enable the calculation of the set of cross-correlation coefficients $J(q, \phi; q'\phi')$ described above. From this set of reduced data, one may calculate the coefficients $I_{00}(q)$ from equations (11) and (16) above for the resolution shells characterized by the same set of quantities q.

A comparison of the calculations of the same quantities from a model of the small synthetic protein chignolin (PDB entry: 1UAO) is shown in figure 7. The near perfect agreement is an encouraging indication of the basic correctness of the theory.

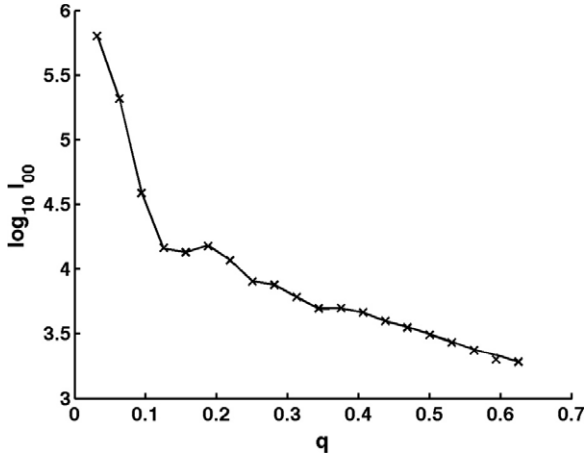

Figure 7. Comparison of $I_{00}(q)$ calculated from the cross-correlations of simulated diffraction pattern intensities (solid curve) and from averages of the intensity (crosses) on spherical shells of varying radius q (in Å^{-1}) of a directly calculated 3D diffraction volume of a fixed molecule of the small synthetic protein chignolin (PDB entry: 1UAO).

4. Independence of angular momentum blocks in the spherical harmonic expansion of the intensity

The quantities $I_{00}(q)$ represent the angularly averaged radial distribution of scattered intensity, which is measured directly from e.g. x-ray scattering from protein molecules in solution, as in small angle x-ray scattering (SAXS). If this were the only quantity found by our method, it would be nothing more than a fancy method of obtaining SAXS data, though to perhaps a higher resolution. However, as with Kam's (1978) analysis of fluctuations of x-ray scattering, a knowledge of the $B_l(q, q')$ coefficients allows the extraction of much more information about a molecule than from SAXS data alone, due to the potential it affords of also finding the higher-order coefficients I_{lm} for $l, m \neq 0$ of the spherical harmonic expansion of the reciprocal space of the molecule. First, we point out why an obvious method for extracting the $I_{lm}(q)$ quantities from $B_l(q, q')$ runs into difficulties.

Returning to equation (9) we note that it should be possible to calculate all elements of the quantity $B_l(q, q')$ from measured diffraction data by matrix inversion and using equation (11). The question is how to calculate the unknown expansion coefficients $I_{lm}(q)$ on the RHS from the known quantities on the LHS.

An initial thought might be to exploit the fact that the angular momentum quantum number l appears on both sides of the equation, by noting that, for each value of l, (9) may be written as

$$B_{qq'} = \sum_m I^\dagger_{qm} I_{mq'}, \tag{17}$$

which is in the form of the matrix equation

$$\mathbf{B} = \mathbf{I}^\dagger \mathbf{I} \tag{18}$$

where † denotes Hermitian conjugation. This suggests that the matrix $\mathbf{I}$ may be found by following the standard method of finding the square root of the Hermitian matrix $\mathbf{B}$: first note that the latter may be written as

$$\begin{aligned}\mathbf{B} &= \mathbf{C}[\{\lambda_m\}]_\mathrm{D}\mathbf{C}^\dagger \\ &= \mathbf{C}[\{\sqrt{\lambda_m}\}]_\mathrm{D}[\{\sqrt{\lambda_m}\}]_\mathrm{D}\mathbf{C}^\dagger \\ &= \mathbf{G}\mathbf{G}^\dagger\end{aligned} \tag{19}$$

where the subscript D after the square brackets indicates that the square brackets represent a diagonal matrix whose diagonal elements are represented by the set of quantities within the brackets. The subscript m specifies the particular diagonal element.

With this notation, $\mathbf{C}$ is the matrix constructed from the column eigenvectors of $\mathbf{B}$, and λ_m is the mth eigenvalue. The matrix

$$\mathbf{G} = \mathbf{C}[\{\sqrt{\lambda_m}\}]_\mathrm{D} \tag{20}$$

is then the 'square root' of $\mathbf{B}$. However, since

$$\mathbf{B} = \mathbf{G}\mathbf{G}^\dagger, \tag{21}$$

then also

$$\mathbf{B} = \mathbf{G}\mathbf{O}\mathbf{O}^\dagger\mathbf{G}^\dagger \tag{22}$$

where $\mathbf{O}$ is a unitary matrix which (by definition) satisfies the equation

$$\mathbf{O}\mathbf{O}^\dagger = \mathbf{U} \tag{23}$$

where $\mathbf{U}$ represents an identity matrix.

Thus, the square root, $\mathbf{G}$, of $\mathbf{B}$ as found by the above method will be ambiguous up to a multiple of an arbitrary $(2l+1) \times (2l+1)$ unitary matrix, $\mathbf{O}$, and thus $\mathbf{G}$ cannot necessarily be identified with $\mathbf{I}^\dagger$. Kam (1978) suggests that this unitary matrix may be identified with a Wigner rotation matrix (or D-function), $D^{(l)}_{mm'}(\alpha, \beta, \gamma)$. If, and only if, the Wigner rotation matrices associated with the different values of l corresponded to the same Euler angles α, β, and γ would this not be a problem, as the 3D diffraction volume, and ultimately the 3D molecular electron density, would merely be subject to an overall rotation through the same Euler angles. The problem is that each angular momentum block is calculated separately, and there is no guarantee that the result of applying the matrix square root operation above will result in rotations of the different angular momentum blocks by the same Euler angles.

5. Determination of a molecular shape function

One approach to obtaining structural information about the molecule from the B matrices calculated as above is via the calculation of the molecular shape function (Stuhrmann 1970a, 1970b; Svergun and Stuhrmann 1991), a target of the analysis of small angle x-ray scattering (SAXS). A brief review of this method is now given, followed by a description of its adaptation to the present problem.

Central to this method is the definition of a molecular shape function $F(\omega)$, where $\omega \equiv (\theta, \phi)$ represents a set of polar angles θ and ϕ in the frame of reference of the molecule. This function is defined by the relations

$$\rho(\mathbf{r}) = \begin{cases} 1 & \text{if } 0 \leqslant r \leqslant F(\omega) \\ 0 & \text{if } r > F(\omega) \end{cases} \tag{24}$$

where

$$F(\omega) = \sum_{lm} f_{lm} Y_{lm}(\omega) \tag{25}$$

where Y_{lm} is a spherical harmonic, and l and m are azimuthal and magnetic quantum numbers.

Thus the shape function $F(\omega)$ is determined if its spherical harmonic coefficients f_{lm} may be found from the experimental data.

This is possible for the following reason. The spherical harmonic coefficients of the electron density

$$\rho_{lm}(r) = \int \rho(\mathbf{r}) Y^*_{lm}(\omega)\, \mathrm{d}\omega \tag{26}$$

are related to those of the scattered amplitude by the Hankel transform

$$A_{lm}(q) = i^l (2/\pi)^{1/2} \int_0^\infty \rho_{lm}(r)\, j_l(qr) r^2\, \mathrm{d}r \tag{27}$$

where j_l is the spherical Bessel function of order l.

From equations (24) to (27), it may be deduced (Shneerson and Saldin 2009) that

$$A_{lm}(q) = i^l (2/\pi)^{1/2} \int R_l[F(\omega)] Y^*_{lm}(\omega)\, \mathrm{d}\omega \tag{28}$$

where

$$R_l[F(\omega)] = \int_0^{F(\omega)} j_l(qr) r^2\, \mathrm{d}r \tag{29}$$

and, from (25), we conclude that $A_{lm}(q)$ is a known function of the expansion coefficients $f_{l'm'}$ of the shape function representation. Symbolically one may write this as

$$A_{lm}(q) = g(q, \{f_{l'm'}\}) \tag{30}$$

where g is a known function of q and of the set $\{f_{l'm'}\}$ of coefficients of the shape function expansion. It should be noted that the full set of coefficients of the shape function expansion contributes to each angular momentum component A_{lm} of the expansion of the amplitudes. Consequently, there is no question of the shape function coefficients being subject to independent rotations in $SO(3)$. It is this feature which

enables the shape function to be determined without ambiguity from an angular momentum decomposition of the 3D intensity distribution, even if the coefficients of that decomposition are subject to arbitrary and independent rotations.

The spherical harmonic coefficients of the scattered amplitudes and intensities are related by (Stuhrmann 1970b)

$$I_{lm}(q) = (-1)^m \sum_{l_1 m_1 l_2 m_2} G(lm, l_1 m_1, l_2 m_2) A_{l_1 m_1 1}(q) A^*_{l_2 m_2}(q) \tag{31}$$

where

$$G(lm, l_1 m_1, l_2 m_2) = (-1)^{m_1} \times \left[\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}\right]^{1/2} \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} l_1 & l_2 & l \\ m_1 & \bar{m}_2 & \bar{m} \end{pmatrix} \tag{32}$$

where the quantities represented by the large parentheses are Wigner $3j$ symbols. From (9) and (31), (32) it can be concluded that

$$B_l^{(\mathrm{th})}(q, q') = h(q, q'; \{f_{l'm'}\}), \tag{33}$$

i.e., $B_l^{(\mathrm{th})}(q, q')$, the theoretical expression for the quantity $B_l^{(\mathrm{exp})}(q, q')$, is a different known function h of the expansion coefficients $\{f_{l'm'}\}$ of the shape function expansion. This suggests that the coefficients $\{f_{l'm'}\}$ may be found from the experimental data by optimization of the agreement with theory by minimizing the cost function

$$\Phi[\{f_{l'm'}\}] = \sum_{qq';l} [B_l^{(\mathrm{exp})}(q, q') - B_l^{(\mathrm{th})}(q, q'; \{f_{l'm'}\})]^2 \tag{34}$$

in a multidimensional space whose coordinates are the shape function coefficients $\{f_{l'm'}\}$.

This is essentially the method (Svergun and Stuhrmann 1991) used in SAXS to find the shape function, except that the experimental data set in SAXS consists only of $I_{00}(q)$. In the terms of our more general theory, the cost function in SAXS contains only the terms for which $q = q'$ and $l = 0$ in (34), above. To put it another way, SAXS data set consists of only a small subset of the much larger data set accessible in the single-molecule diffraction experiment that we describe. Due to the much greater information content of this much larger data set, it must be possible to determine the molecular shape function much more accurately with the data expected to be available in an XFEL experiment. Performing the radial integrals of the Hankel transforms (27) analytically (Shneerson and Saldin 2009) should also allow the inclusion of data from wider range of q (and q') than usual in SAXS.

We have shown earlier how accurately $I_{00}(q)$ (and hence $B_0(q, q)$) may be found from the cross-correlations of simulated diffraction patterns from random orientations of a single molecule. For the purposes of the next test, we simulated the quantities $B_0^{(\mathrm{exp})}(q, q)$ and $B_2^{(\mathrm{exp})}(q, q)$ from the theoretical expressions (9) and (28)–(32) for assumed values of the molecular shape coefficients $\{f_{l'm'}\}$ for lysozyme (PDB entry: 2BPU). We then attempted to recover these quantities in the expression by minimizing the expression (34), starting from random values of $\{f_{l'm'}\}$ in the expressions for $B_l^{(\mathrm{th})}(q, q)$. Of course, for q in the range 0 to about 0.25 Å^{-1} the quantity $B_0(q, q)$ is just the square of the SAXS signal $I_{00}(q)$. Since we were interested in the question of whether the non-SAXS part of the signal from the intensity cross-correlations contains extractable information about the molecular shape, we included only the $B_0^{(\mathrm{exp})}(q, q)$ data for q outside this range in the cost function (34). The optimization was performed using a simulated annealing algorithm (Kirkpatrick *et al* 1983). The resulting optimal fit of the quantities $B_0(q, q)$ and $B_2(q, q)$ between experiment and theory is shown in figure 8.

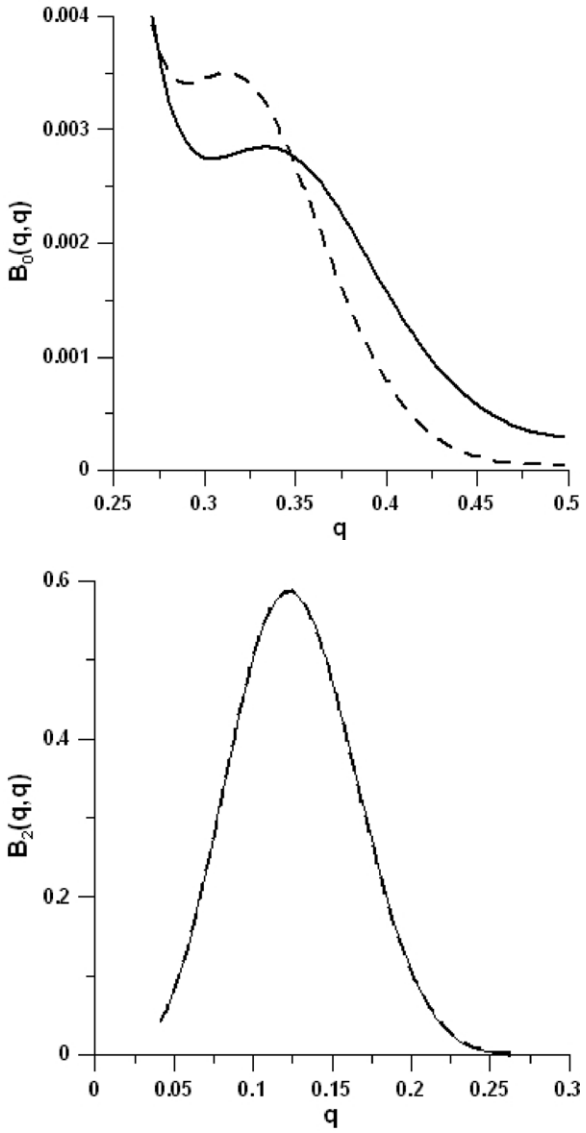

Figure 8. Comparison of variations with q (in Å^{-1}) of $B_0^{(\mathrm{exp})}(q, q)$ and $B_2^{(\mathrm{exp})}(q, q)$ (solid lines) based on assumed values of molecular shape coefficients f_{lm}, and recovery of the same quantities by varying these coefficients by simulated annealing, starting from random values. The best fit curves for $B_0^{(\mathrm{th})}(q, q)$ and $B_2^{(\mathrm{th})}(q, q)$ are displayed as dashed lines. The optimization was performed on only the non-SAXS data, and thus on the $B_0(q, q)$ coefficients in a range of large q (from 0.25 to 0.5 Å^{-1}). Consequently the quantities $B_0(q, q)$ shown are much smaller than $B_2(q, q)$. Hence the apparent disagreement of the $B_0(q, q)$ quantities is relatively insignificant.

From the resulting optimal values of these expansion coefficients, the molecular shape function was evaluated from (25) and the resulting shape plotted in figure 9 using the MASSHA graphics program (Konarev *et al* 2001), with a stick figure of the α-C trace of lysozyme superimposed on it. The excellent agreement is an encouraging indication of the capability of this method of determining the molecular shape from the non-SAXS part of the data.

A major problem with the use of SAXS for the determination of molecular structure is the very limited amount of information in the relatively short range of q of the single $I_{00}(q)$ curve measured. An argument from Shannon suggests that the width of an information element in a SAXS curve from a molecule of diameter D is $\Delta q = \pi/D$. This is about 0.025 Å^{-1} for a typical protein of diameter $\simeq$120 Å, suggesting that a SAXS curve of width $\sim$0.25 Å^{-1} would contain no more than about 10 'Shannon channels' (Svergun *et al* 1996). In turn, this severely limits the number of structural parameters extractable from the data. Consequently, analysis of SAXS data is used mostly to extract just a few leading terms $f_{l'm'}$ of the expansion of the molecular shape function, and this may give no more than a low-resolution shape of the molecule.

The method of analysis proposed here yields not just the SAXS data contained in $B_0(q, q)$, but also many other independent items of data in the independent variations of q and q' and the different values of l in the more general quantities $B_l(q, q')$ extractable from the data, and should thus contain perhaps even orders of magnitude more information about the structure of the molecule.

6. Alignment of different angular momentum blocks by molecular replacement

Details of the internal structure of the molecule may be found if intensity cross-correlations are measured also for a molecule of known structure with substantial structural overlap with the unknown one to be determined, as in the method of molecular replacement in classical protein crystallography (Rossmann and Blow 1962).

Suppose that the above analysis is applied both to the molecule of known structure, and that of the closely related unknown structure. Different B-matrices (17) may be found from the measured intensity correlations of the two molecules. Assuming that the differences between the two structures are significantly smaller than their similarities, we may define

$$\delta B_l(q, q') = B'_l(q, q') - B_l(q, q') \tag{35}$$

where B' refers to the unknown structure, and B to the known one. Taking the variation of equation (9),

$$\delta B_l(q, q') = \sum_m \{I^*_{lm}(q)\delta I_{lm}(q') + \delta I^*_{lm}(q) I_{lm}(q')\} \tag{36}$$

where $\delta I_{lm}(q)$ represents the difference in coefficients of the spherical harmonic expansions of the 2D diffraction intensities of the unknown and known structures. The corresponding expansion coefficients $I_{lm}(q)$ of the known structure are found from e.g. structural data deposited in the Protein Data Bank (PDB), via the expression for the spherical harmonic expansion coefficients of the scattered amplitude (Svergun *et al* 1995):

$$A_{lm}(q) = \sum_j f_j(q) i^l j_l(q r_j) Y^*_{lm}(\hat{\mathbf{r}}_j) \tag{37}$$

where $f_j(q)$ is the scattering factor of an atom at position $\mathbf{r}_j$, and by the use of (31) to relate these coefficients to the corresponding expansion coefficients of the intensity.

The presence of the index l on both sides of (36) suggests once more that the equations may be solved for each angular momentum quantum number in turn. For each value of l, the above equation may be written with subscripts specifying matrix elements as

$$\delta B_{qq'} = \sum_m \{I^*_{qm}\delta I_{mq'} + \delta I^*_{qm} I_{mq'}\}. \tag{38}$$

The quantities on the LHS can be found from the experiment, and the quantities I may be calculated from the structure of the known protein. The only unknowns are the difference expansion coefficients, δI, of the 3D intensity distribution. The fact that the difference intensities $\delta I_{mq'}$ and δI_{qm} appear only in products with corresponding reference quantities for a known structure in which the different angular momentum blocks are correctly aligned ensures that even though the quantities $\delta I_{lm}(q)$ are calculated separately for each value of l, the relative orientations of these quantities across different angular momenta are correctly preserved.

The only question is whether the number of equations, represented by the number of distinct elements of δB, is greater than or equal to the number of unknown elements of δI. Let us denote the number of values of q for which independent intensity values may be measured by (#q) and the number of independent values of m for the given value of l by (#m). Given that δB is a symmetric matrix, the number of

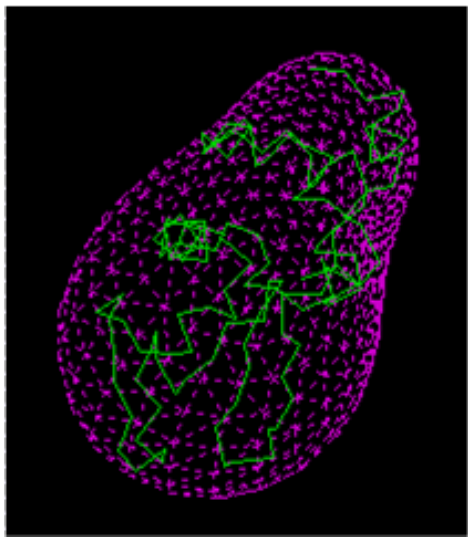

Figure 9. Molecular shape of lysozyme (PDB entry: 2BPU) recovered by optimization of $B_0(q, q)$ and $B_2(q, q)$, as described in the text. The graphics program MASSHA is used to overlap the recovered molecular shape with an α-C trace of the protein.

equations is the number of its independent elements, namely $(\#q)^2/2 + (\#q)/2$. But $(\#m) = (2l + 1)$ (this assumes the elements δI_{mq} are complex but that, due to the reality of $\delta I(\mathbf{q})$, not all elements are unique) and consequently the number of independent elements δI_{mq} for a given l is $(\#q)(2l + 1)$. Thus, it follows that the equations may be solved uniquely for all the elements of δI if

$$(\#q)/2 + 1/2 \geqslant (2l + 1). \tag{39}$$

For a given value of q, the quantity $(\#q)$ may be estimated as $q/(\delta q)$, where $(\delta q) = \pi/D$, where D is the molecular diameter. Thus, $(\#q) = qD/\pi = 2qR/\pi$ (where R is the radius of the molecule). Substituting into (39), we deduce that approximately

$$l \leqslant qR/(2\pi), \tag{40}$$

i.e. that the maximum angular momentum quantum number $l_{\max}$ in the expansion

$$\delta I(\mathbf{q}) = \sum_{lm} \delta I_{lm}(q) Y_{lm}(\hat{\mathbf{q}}) \tag{41}$$

of the 'difference intensity' is

$$l_{\max} = qR/2\pi. \tag{42}$$

The electron density associated with this 'difference intensity' may be found by noting that

$$\delta I(\mathbf{q}) = A^*(\mathbf{q})\delta A(\mathbf{q}) + \text{c.c.} \tag{43}$$

where $A(\mathbf{q})$ is the scattered amplitude due to a known molecule at the point in reciprocal space specified by $\mathbf{q}$, calculable using the formula

$$A(\mathbf{q}) = \sum_j f_j(\mathbf{q}) \exp(\mathrm{i}\mathbf{q} \cdot \mathbf{r}_j) \tag{44}$$

where $f_j(\mathbf{q})$ is a form factor of an atom j at position $\mathbf{r}_j$ in the known structure, and $\delta A(\mathbf{q})$ is the corresponding quantity due to the unknown 'difference electron density' $\delta\rho(\mathbf{r}_l)$, where $\mathbf{r}_l$ is a location in real space (in the same coordinate system as was used in the specification of atom positions $\mathbf{r}_j$), and c.c. denotes complex conjugation. Substituting the discrete Fourier transform

$$\delta A(\mathbf{q}) = \sum_l \delta\rho(\mathbf{r}_l) \exp(\mathrm{i}\mathbf{q} \cdot \mathbf{r}_l) \tag{45}$$

into (43), we deduce

$$\delta I(\mathbf{q}) = \sum_l \delta\rho(\mathbf{r}_l)[A^*(\mathbf{q}) \exp(\mathrm{i}\mathbf{q} \cdot \mathbf{r}_l) + \text{c.c.}], \tag{46}$$

a set of (real) linear equations which may be solved for the difference electron density, as in the holographic method for x-ray crystallography of Szöke (1993), if the number of independent values of $\{\delta I(\mathbf{q})\}$ is greater than or equal to the number of independent values of $\{\delta\rho(\mathbf{r}_l)\}$.

7. Time-resolved structure determination of isolated molecules

The method of time-resolved crystallography (e.g. Schmidt *et al* 2005; Key *et al* 2007) has provided fascinating glimpses of the evolution of protein structures on short timescales after excitation with a pump laser pulse. The technique may be described as a pump–probe experiment, the probe x-ray beam following the pump laser beam at a precisely defined time interval in the range of 100 ps to a second.

The method of analysis of the measured Laue diffraction patterns for different sample tilts assumes a knowledge of the unexcited (or 'dark') structure. The change in the structure upon excitation is determined by a difference Fourier method (Cochran 1951) which operates on the difference between the structure factors before and after the laser excitation.

For single-molecule structure determination, one could envisage the measurement of two sets of diffracted intensity data: one ($\{I\}$) of diffraction patterns of molecules not excited by a pump laser in random orientations, e.g. as proposed by Neutze *et al* (2000), and another ($\{I'\}$) consisting of similar diffraction patterns of molecules excited a short time earlier by a pump laser. Internal correlations amongst these sets of intensities allow the determination of the quantities $B_l(q, q')$ and $B'_l(q, q')$, respectively. Equations (35)–(38) would then allow the determination of quantities $\delta I_{lm}(q)$; as before, assuming a knowledge of the 'dark' structure will enable a calculation of the coefficients $I_{lm}(q)$, as described in section 6. In the present case, the quantities $\delta I_{lm}(q)$ are identified with the spherical harmonic expansion coefficients of the change in the 3D scattered intensity distribution upon excitation with the pump laser. The corresponding change in the electron density may be found from equations (41)–(46).

8. Discussion

The possibility of determining the structures of individual protein molecules without the need for crystallization has been one of the main scientific justifications (Hajdu *et al* 2000, Abela *et al* 2007) for the development of a fourth-generation x-ray source, the x-ray free electron laser (XFEL). The algorithm originally proposed for generating a 3D diffraction volume from measurements of diffraction patterns from random orientations of the molecules in a molecular beam has been the so-called 'common-line' method developed originally for 3D electron microscopy (see e.g. Frank 2006). Our recent work (Shneerson *et al* 2008) has shown that the method may indeed be applied to perform the same task for low-noise diffraction pattern data. However, our experience is that the method is difficult to apply for measured mean photon counts per pixel of less than about 10. A little reflection makes it clear why the common-line method is not optimal for the very low photon counts of about 5 per 100 pixels expected for diffraction of a single XFEL pulse from a typical protein (Shneerson *et al* 2008). A common-line method seeks to find the relative orientations of individual diffraction patterns from just a very small fraction of its data in the vicinity of the common lines.

A method more suited to this problem, which finds the relative reciprocal-space orientations of the measured diffraction patterns from essentially the entire diffracted photon ensemble, has been demonstrated recently (Fung *et al* 2008). This employs a manifold embedding technique to find the latent variables (e.g. the Euler angles specifying the relative orientations of the diffraction patterns).

The present paper proposes the notion that it may not be necessary to find the relative orientations of the individual diffraction patterns in order to extract the quantity of interest, namely the distribution of scattered intensity over the entire 3D reciprocal space of the molecule. It has many similarities to the method of correlation analysis of diffraction data proposed by Kam (1978), except that, in this case, the correlations may be measured directly from the data without the large background isotropic (SAXS-like) contribution. We have indicated how the 3D diffraction volume of the protein whose structure is sought may be found unambiguously if the measured data are combined with data from a similar molecule of known structure. The applicability of this idea to recover the change of the structure in a pump–probe time-resolved experiment is also noted.

9. Conclusions

We have described an algorithm for reconstructing a 3D diffraction volume from diffraction patterns from random orientations of an x-ray scatterer. The method involves the determination of the coefficients of the spherical harmonic expansion of this 3D diffraction volume from the cross-correlations of the intensities of the measured diffraction patterns. The set of these coefficients as a function of the magnitude q of the scattering vector contain more general information than is extractable from small angle x-ray scattering (SAXS), and should allow extraction of much more information about the molecule than the low-resolution molecular envelope which is the usual target of SAXS on biomolecules.

In particular, if a protein whose internal structure is sought is known to be closely related to that of a protein of known structure, we suggest that measurements of intensity cross-correlations of both proteins will allow the determination of differences between the 3D scattered intensities of the two proteins, and thus allow the reconstruction of the *difference in electron density* between the two structures, as in the *molecular replacement* method of traditional protein crystallography. The last step may be implemented even without the invocation of an iterative phasing algorithm.

An adaptation of the method is suggested for the analysis of single-molecule diffraction data from pump–probe experiments for an efficient determination of time-resolved changes of the molecular electron density.

Acknowledgment

We acknowledge financial support from US DOE grants DE-FG02-84ER45076 and DE-FG02-06ER46277.

References

Abela R *et al* 2007 The European x-ray free-electron laser *Technican Design Report* ed M Altarelli *et al* pp 401–20 http://xfel.desy.de/tdr/index_eng.html
Bishop C M 1998 *Neural Comput.* **10** 215
Cochran W 1951 *Acta Crystallogr.* **4** 408
Fenn J B 2002 *J. Biomol. Tech.* **13** 101
Fienup J R 1978 *Opt. Lett.* **3** 27
Fienup J R 1982 *Appl. Opt.* **21** 2758
Frank J 2006 *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Oxford: Oxford University Press)
Franklin R E and Gosling R G 1953 *Nature* **171** 740
Fung R, Shneerson V L, Saldin D K and Ourmazd A 2008 *Nat. Phys.* at press (doi:10.1038/nphys1129)
Gaffney K J and Chapman H N 2007 *Science* **316** 1444
Hajdu J, Hodgson K, Miao J, van der Spoel D, Neutze R, Robinson C V, Faigel G, Jacobsen C, Kirz J, Sayre D, Weckert E, Materlik G and Szöke A 2000 *Structural Studies of Single Particles and Biomolecules, LCLS: The First Experiments* pp 35–62 http://www-ssrl.slac.stanford.edu/lcls/papers/lcls_experiments_2.pdf.
Huldt G, Szöke A and Hajdu J 2003 *J. Struct. Biol.* **144** 218
Kam Z 1978 *Macromolecules* **10** 927
Key J, Srajer V, Pahl R and Moffatt K 2007 *Biochemistry* **46** 4706
Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671
Konarev P V, Petoukov M V and Svergun D I 2001 *J. Appl. Crystallogr.* **34** 527
Miao J, Charalambous P, Kirz J and Sayre D 1999 *Nature* **400** 342
Neutze R, Wouts R, van der Spoel D, Weckert E and Hajdu J 2000 *Nature* **406** 752
Normille D 2006 *Science* **314** 751
Oszlányi G and Süto A 2004 *Acta Crystallogr.* A **60** 134
Oszlányi G and Süto A 2005 *Acta Crystallogr.* A **61** 147
Rossmann M G and Blow D M 1962 *Acta Crystallogr.* **15** 24
Schmidt M, Nienhaus K, Pahl R, Krasselt A, Anderson S, Parak F, Nienhaus G U and Srajer V 2005 *Proc. Natl Acad. Sci.* **102** 11704
Shneerson V L, Ourmazd A and Saldin D K 2008 *Acta Crystallogr.* A **64** 303
Shneerson V L and Saldin D K 2009 *Acta Cryst.* A submitted
Spence J C H, Schmidt K, Wu J S, Hembree G, Weierstall U, Doak B and Fromme P 2005 *Acta Crystallogr.* A **61** 237
Stuhrmann H B 1970a *Z. Phys.* **72** 177
Stuhrmann H B 1970b *Acta Crystallogr.* A **26** 297
Svergun D I, Baberato C and Koch M H J 1995 *J. Appl. Crystallogr.* **28** 768
Svergun D I and Stuhrmann H B 1991 *Acta Crystallogr.* A **47** 736
Svergun D I, Volkov V V, Kozin M B and Stuhrmann H B 1996 *Acta Crystallogr.* A **53** 419
Szöke A 1993 *Phys. Rev.* B **47** 14044
Tinkham M 2003 *Group Theory and Quantum Mechanics* (Dover: Courier)
Watson J D and Crick F H C 1954 *Nature* **171** 737