Deconvolution of Disoriented Fiber Diffraction Data using Iterative Convolution and Local Regression

By Thomas T. Tibbitts

Center for Molecular Biophysics, Boston University Department of Physics, 590 Commonwealth Avenue, Boston, MA 02215, USA

and D. L. D. Caspar

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254-9110, USA

Abstract

Computationally efficient procedures are described for the deconvolution of disoriented fiber diffraction data to the resolution limit of measurable intensity in the patterns. The methods can be applied to diffraction data from imperfectly parallel arrays of one-dimensionally periodic rods or two-dimensionally periodic sheets, randomly rotated about their unique axes, to derive a representation of the intensity distribution corresponding to perfectly parallel orientation. With use of angular convolution and local angular regression, a set of uniform cylindrically averaged squared structure factors are iteratively adjusted, subject to a minimum-wavelength constraint, until they produce a disoriented pattern that fits the observed diffraction data. The results from this deconvolution provide a measure of the properly scaled cylindrically averaged squared structure factors, which can be used with other structural information to construct a physically plausible trial model suitable for further refinement. Sample deconvolutions of simulated X-ray patterns from partially oriented gap junction membranes are presented and the results from point-model deconvolutions are compared to those from constrained deconvolutions that began with the transform of a physically plausible trial model.

Introduction

The field of fiber diffraction encompasses a variety of analytical and computational techniques for structural characterization of oriented polymer chains (Alexander, 1969), rod-shaped protein filaments (Dickerson, 1964) and sheet-like lipid membranes (Makowski & Li, 1983). Since the pioneering X-ray work on the natural fibers hair and wool (Astbury & Street, 1931), a wide variety of biopolymers and biomolecular assemblies have been studied with the increasingly sophisticated techniques of fiber diffraction. Recent papers have concerned, for example, wrinkled DNA (Arnott, Chandrasekaran, Puigjaner, Walker, Hall, Birdsall & Ratcliff, 1983), porin lattices (Büldt, Mischel, Hentschel, Regenass & Rosenbush, 1986), tobacco mosaic virus (Namba, Pattanayek & Stubbs, 1989), bacterial flagella (Namba, Yamashita & Vonderviszt, 1989), *F*-actin (Holmes, Popp, Gebhard & Kabsch, 1990), gap junctions (Tibbitts, Caspar, Phillips & Goodenough, 1990) and pf1 bacteriophage (Nambudripad, Stark, Opella & Makowski, 1991).

X-ray diffraction patterns from oriented rod-like particles or oriented multilayered sheets correspond to the intensity distribution from a single particle or sheet that has been cylindrically averaged. When rods are aligned with their long axes parallel in a fiber, they will generally arrange themselves into an imperfect lattice with the particles not in register and rotationally disordered about their long axes, so that the ensemble is uniformly distributed with respect to the azimuthal angle ψ. Azimuthal disorder also exists in parallel membrane multilayers. The sheets are aligned perpendicular to a unique axis but may be rotated by any random angle about that axis. Diffraction patterns from azimuthally disordered samples of parallel rods or sheets are thus similar to a photograph of a single crystal rotating about a unique axis (see Holmes & Blow, 1966).

Both oriented rod-like biomolecules and membrane sheets that have periodic internal structure give rise to cylindrically averaged diffraction patterns in which the intensity is confined to a set of layer or lattice lines, so that methods for analysis of the two types of patterns are similar. Rods may have a one-dimensional repeating structural motif along their long axis: examples are protein α-helices, nucleic acid double helices and helical arrangements of protein subunits such as *F*-actin (see Dickerson, 1964). One-dimensional periodicity is also observed, for example, in silk fibroin (Bamford, Brown, Elliot, Hanby & Trotter, 1954) and amyloid fibrils (Fraser, Nguyen, Surewicz & Kirschner, 1991). Similarly, in some sheet-like biological membranes a two-dimensional lattice of membrane proteins occurs

naturally or can form during isolation or by reconstitution of isolated membrane components: examples are bacteriorhodopsin, Na^+/K^+-ATPase and gap junctions (see Gennis, 1989).

Patterns from one- and two-dimensionally periodic samples can be useful for the refinement of structural models of fibers and membranes, if they can provide estimates of cylindrically averaged squared structure factors. However, it is difficult to obtain this information from experimental fiber or membrane diffraction data because of imperfect orientation of the particles or sheets. Measurements from these patterns must take into account the effects of azimuthal disorder in the sample combined with imperfect orientation, finite coherence of diffracting domains and finite collimation of the radiation source. Thus, the first and rather difficult step is to estimate a set of positive real numbers from the digitized diffraction pattern to represent the cylindrically averaged structure factors, $\langle F^2 \rangle_\psi$, sampled at the points (R_q, Z_q). These estimates may then be used with other information to formulate a plausible initial structural model.

Numerical angular deconvolution (Makowski, 1978) has been widely used to estimate $\langle F^2 \rangle_\psi$ from mildly disoriented patterns and has enabled successful refinement of the tobacco mosaic virus structure to 2.9 Å resolution (Namba, Pattenayek & Stubbs, 1989). Angular deconvolution uses matrix inversion of a set of equations describing the overlapping components in the digitized pattern and correction factors based on the derived form for the intensity distribution expected in mildly disoriented fiber diffraction patterns (Holmes & Leigh, 1974; Stubbs, 1974).

Generally, deconvolution by matrix inversion has limitations and requires some simplifying assumptions. The correction factors used become progressively less reliable close to the rotation axis of the pattern and thus scaling of the off-meridional to meridional $\langle F^2 \rangle_\psi$ is not straightforward. Perfect beam collimation and regular order within the structural units must often be assumed because correction for the combined effects of disorientation with beam divergence and finite coherence of one- or two-dimensionally periodic structures makes the problem unmanageably complex. At higher scattering angles, overlap of reflections due to disorientation and linewidth broadening eventually limits the resolution to which deconvolution can be applied. Beyond this limit, matrix inversion becomes impossible (Makowski, 1991).

To overcome these difficulties, we have undertaken the development of computationally efficient methods for deconvoluting disoriented cylindrically averaged diffraction patterns, based on repeated simulation of the whole observed pattern (Tibbitts, Caspar, Phillips & Goodenough, 1988). Our angular convolution procedures at least partially overcome the problem of line overlap to allow extraction of the maximum amount of information from poorly oriented samples. In our analysis, a starting set of $\langle F^2 \rangle_\psi$ from equal-magnitude structure factors are iteratively adjusted to fit the disoriented data. Adjustment coefficients are determined by simulation of a disoriented diffraction pattern and performance of local angular regressions of the predicted pattern to the data. Simulation, regression and adjustment are repeated in a loop until the difference between model pattern and data is minimized. By simulating the effects of the various convoluting functions, these procedures can thus provide reliable first estimates of the $\langle F^2 \rangle_\psi$, even from diffraction patterns showing significant disorientation and disorder.

In this paper, formulas for our iterative convolution and local regression procedures are worked out for patterns from one-dimensionally periodic rods and two-dimensionally periodic sheets, and tested with some sample deconvolutions of simulated X-ray diffraction patterns from imperfectly oriented gap junction membranes. In the trials, reasonable estimates of the $\langle F^2 \rangle_\psi$ were obtained that properly scaled meridional and off-meridional data and took into account cylindrical averaging, disorientation, beam divergence and specimen disorder.

Theory and methods

Simulation of azimuthal disorder

Simulation of fiber diffraction patterns from imperfectly oriented rods or sheets requires an efficient method for the calculation of the combined effects of cylindrical averaging and disorientation at each pixel. Our steps were: (1) calculate $I_\parallel$, the cylindrically averaged diffraction pattern from perfectly parallel rods or sheets, with factors that determine the finite width of the lines taken into account; (2) introduce disorientation by smearing each point in the $I_\parallel$ pattern to give a disoriented pattern, $I_\vee$; (3) convolute the disoriented pattern $I_\vee$ with a function to simulate divergence of the incident beam of radiation. The first step, calculating an $I_\parallel$ pattern given a trial set of $\langle F^2 \rangle_\psi$, is slightly different for one- and two-dimensionally periodic samples but all subsequent steps in the simulations are the same.

Calculation of patterns for parallel one-dimensionally periodic rods. For a one-dimensionally periodic particle, intensity in the transform will be negligible except in a series of layer planes perpendicular to the Z axis, which is aligned with the particle axis in real space. The heights of these planes satisfy the relationship $Z = l/c$, where l is the layer-plane index and c is the period along the particle axis. Other symbols used are given in Table 1. In general, particles in a fiber will be rotationally disordered about the fiber axis and intensity within the layer planes will be

Table 1. *Symbols used*

R, ψ, Z	Cylindrical polar coordinates in reciprocal space
v, σ, ψ	Spherical polar coordinates in reciprocal space
F	Structure factor
F^2	Squared structure factor FF^*
$\langle F^2(R_q, Z_q)\rangle_\psi$	Cylindrically averaged squared structure factor
$\langle F^2\rangle_\psi$	The set $\{\langle F^2(R_q, Z_q)\rangle_\psi\}$
R_q, Z_q	Cylindrical coordinate of a sample point $q \in \langle F^2\rangle_\psi$
v_q, σ_q	Spherical coordinate of a sample point $q \in \langle F^2\rangle_\psi$
Re $\{F\}$	Real part of structure factor
Im $\{F\}$	Imaginary part of structure factor
h, k	Miller indices for reciprocal-lattice lines
m	Lattice (or layer) line sample point index
$f(t)$	Lattice (or layer) line lateral profile
p	Coherence length
$I_\parallel$	Intensity from perfectly parallel rods or sheets
$I_\vee$	Intensity from disoriented rods or sheets
I_{exp}	Intensity in expected (or experimental) disoriented pattern
I_{calc}	Intensity in model disoriented pattern
σ_0	The angle σ for a point in $I_\parallel$
$N(\alpha)$	Disorientation distribution
α	Inclination relative to the axis of mean orientation
α_0	Disorientation parameter
$I_{model}(\alpha_0)$	Intensity in Gaussian disoriented pattern
$i_0(x)$	The shape factor, $\exp(-x)\mathcal{I}_0(x)$
$\mathcal{I}_0(x)$	Hyperbolic Bessel function $\mathcal{I}_0(x) = (2\pi)^{-1}\int_{\psi=0}^{2\pi} \exp(x \cos\psi)\,d\psi$
$A(N; \sigma_0, \sigma)$	Angular distribution for a Debye-Scherrer arc
$\kappa(v_q, \sigma_q)$	Adjustment coefficient for a point $q \in \langle F^2\rangle_\psi$
$K\langle F^2\rangle_v$	Adjusted set of structure factors
$\mathcal{R}_I$	Measure of agreement between two disoriented patterns
$\mathcal{R}_{\langle F^2\rangle}$	Measure of agreement between two sets of $\langle F^2\rangle_\psi$
$\rho(r, \theta, z)$	Density distribution in real-space polar coordinates

cylindrically averaged. The intersections of the sphere of reflection with these layer planes are layer lines, which appear as hyperbolas on a flat film (see Guinier, 1963). The intensity distribution from helical assemblies is further limited to a set of layer lines that satisfy a selection rule; at a point at radius R on line l, the intensity can be expressed in terms of Fourier-Bessel structure factors as $\sum_n G_{nl}(R)G_{nl}^*(R)$, where n is the order of the Bessel function (see Klug, Crick & Wyckoff, 1958). To use a general index q for coordinates of the cylindrically averaged squared structure factors $\langle F^2\rangle_\psi$, we define $F_n^2(R, Z) = G_{nl}(R)G_{nl}^*(R)$ and sample along the layer lines at intervals of δ_R,

$$\sum_n G_{nl}(R)G_{nl}^*(R) = \sum_n F_n^2(m\delta_R, l/c) \qquad (1)$$

$$= \sum_n F_n^2(R_q, Z_q) \qquad (2)$$

$$= \langle F^2(R_q, Z_q)\rangle_\psi. \qquad (3)$$

Given a trial set of $\langle F^2\rangle_\psi$, the axial repeat c and a layer-line profile $f(t)$, the general formula for calculation of the diffraction pattern from perfectly parallel rod-shaped helical particles is

$$I_\parallel(R, Z) = \sum_q \langle F^2(R_q, Z_q)\rangle_\psi f(t), \qquad (4)$$

where $t = Z - Z_q$ measures the lateral distance from the layer line centered at Z_q. Stubbs (1974) investi-

gated lateral profiles of layer lines in diffraction patterns from tobacco mosaic virus and paramyosin fibers and suggested the Gaussian form $f_G(p; t) = \exp[-\pi p^2 t^2]$, with standard deviation $t_0 = 1/(2\pi^{1/2}p)$, where p is the coherence length of the particle.

Calculation of patterns for parallel two-dimensionally periodic sheets. Simulation of diffraction patterns from parallel two-dimensional periodic sheets is slightly different from the above since the profile of lattice lines in the pattern depends on the radius R due to cylindrical averaging. For a single planar crystalline array of identical subunits, intensity in the transform is confined to a lattice of parallel lines. In a multi-layered sample with perfect orientation but with the sheets rotationally disordered about their normal direction, the reciprocal-lattice lines are replaced by concentric cylinders. Intersection of the sphere of reflection with these concentric cylinders gives rise to lines in the diffraction pattern, which appear as distorted ovals on a flat film; each line measured from the film represents intensity from one or more reciprocal-lattice lines averaged together (see Guinier, 1963). To use a general index q for the coordinates of the cylindrically averaged structure factors, we sum the contributions from lattice lines (h, k) with the same radius R and sample along each line at intervals of δ_Z,

$$\sum_{h,k} F_{hk}(Z)F_{hk}^*(Z) = \sum_{h,k} F_{hk}^2(R_{hk}, m\delta_Z) \qquad (5)$$

$$= \sum_{h,k} F_{hk}^2(R_q, Z_q) \qquad (6)$$

$$= \langle F^2(R_q, Z_q)\rangle_\psi. \qquad (7)$$

Given a trial set of $\langle F^2\rangle_\psi$ for the lattice constant a and a reciprocal-lattice line profile $f(t)$, from Fig. 1 the general formula for the calculation of a pattern from perfectly parallel two-dimensionally periodic sheets that properly scales the near-meridional to off-meridional data is

$$I_\parallel(R, Z) = \sum_q n\langle F^2(R_q, Z_q)\rangle_\psi \int_{\psi=0}^{2\pi} f(t)\,d\psi. \qquad (8)$$

The n are multiplicity integers that depend upon the rotational symmetry at the origin of the reciprocal lattice. For a hexagonal lattice, $n = 1$ for points on the meridian and $n = 6$ for sample points on off-meriodional lines. From the cosine law, the lateral distance t, measured from the center of the line located at R_q, is given by $t^2 = R^2 + R_q^2 - 2RR_q \cos\psi$. Equation (8) makes no assumptions about the intrinsic shape of the lattice line profile $f(t)$ and therefore $I_\parallel$ can be evaluated by numerical integration [with use of, for example, subroutines provided by Press, Flannery, Teukolsky & Vetterling (1988)] for any empirical function $f(t)$.

However, if we assume that the reciprocal-lattice lines have a Gaussian lateral profile, the integral in (8) can be evaluated analytically. Let $f(t) = f_G(p; t) = \exp(-\pi p^2 t^2)$. With use of the identity $R^2 + R_q^2 = (R - R_q)^2 + 2RR_q$ and the integral form of the hyperbolic Bessel function $\mathcal{I}_0$ [see Table I of Abramowitz & Stegun (1965)], (8) can be rewritten in terms of the shape factor,

$$i_0(2\pi p^2 RR_q) = \exp(-2\pi p^2 RR_q)\mathcal{I}_0(2\pi p^2 RR_q), \quad (9)$$

and the Gaussian profile $f_G(p; t)$ for $\psi = 0$. Thus,

$$I_\parallel(R, Z) = 2\pi \sum_q n\langle F^2(R_q, Z_q)\rangle_\psi$$

$$\times f_G(p; R - R_q)i_0(2\pi p^2 RR_q). \quad (10)$$

The shape-factor term i_0 in (9) is plotted in Fig. 2. It represents the correction to the Gaussian lattice line lateral profile due to cylindrical averaging. Notice
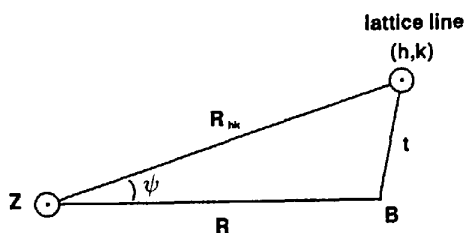
Fig. 1. Calculation of the cylindrically averaged pattern for parallel two-dimensionally periodic sheets. Cross section of the intensity distribution arising from a single two-dimensional lattice with finite coherence at constant Z. Shown emerging from the page are the Z axis and a reciprocal-lattice line parallel to it. In real space, a lattice plane oriented at the azimuthal angle ψ gives rise to a reciprocal-lattice line centered at the radius $R_g = R_{hk}$ with a cylindrically symmetric lateral intensity profile $f(t)$, where t is the distance from the lattice line. To find the intensity contributed at a point B, located at distance R in the cylindrically averaged pattern, consider lattices with azimuthal orientation between ψ and $\psi + d\psi$. These will contribute intensity $F_{hkm}F_{hkm}^* f(t) \, d\psi$ at point B. The total intensity at B is calculated by integration over all possible lattice rotations, with the number of reciprocal-lattice lines that contribute at B taken into account.
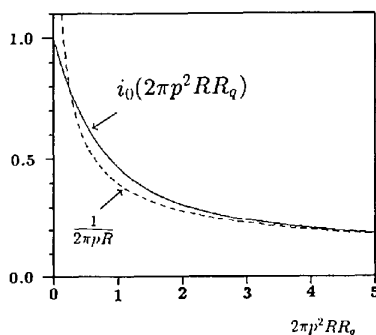

Fig. 2. Cylindrical correction to a Gaussian line profile. The solid curve shows the shape correction factor for the lattice line profile in the cylindrically averaged pattern from perfectly parallel sheets if a Gaussian line profile $f_G(p; t)$ is assumed. The dashed curve represents the correction factor given by (11), which is only valid away from the meridian.

that this function equals unity at the meridian, where $R = 0$.

Equation (10) properly scales near-meridional intensities and reduces to an appropriately simple form for points away from the meridian. To see this, we employ the asymptotic form $i_0(x) = (2\pi x)^{-1/2}$ for large $x = 2\pi p^2 RR_q$, which is also plotted in Fig. 2. Far from the meridian, where $R_q \simeq R \geq 2(2\pi)^{1/2}/p$,

$$i_0(2\pi p^2 RR_q) \simeq (2\pi pR)^{-1}. \quad (11)$$

Substituting (11) into (10), we have

$$I_\parallel(R, Z) \simeq (1/p) \sum_q n\langle F^2(R_q, Z_q)\rangle_\psi$$

$$\times f_G(p; R - R_q)/R_q, \quad (12)$$

which shows that, away from the meridian, reciprocal-lattice lines with Gaussian lateral profile $f_G(p; t)$ give rise to lines with Gaussian profile $f_G(p; R - R_q)$ in the cylindrically averaged pattern and are weighted by $1/R$.

Simulation of disorientation

Proper simulation of diffraction data from samples with arbitrary disorientation requires a general method for the calculation of the intensity distribution in Debye–Scherrer arcs that is valid everywhere in reciprocal space and correctly scales near-meridional and off-meridional data. We describe below how Debye–Scherrer arcs can be calculated from the sample disorientation distribution and used in a lookup table to smear out the $I_\parallel$ pattern.

In a disoriented sample, the probability of finding particles (or sheets) at an angle α to the axis of mean orientation in an element of solid angle $d\Omega$ is $N(\alpha) \, d\Omega/4\pi$. The sample disorientation distribution, $N(\alpha)$, can often be accurately measured or modeled from low-angle diffraction data. The theoretical intensity distribution in a disoriented fiber pattern, given $N(\alpha)$, a layer-line lateral profile and the cylindrically averaged squared transform of a single particle, has been derived by Holmes & Leigh (1974) (referred to here as H&L). To derive the formula for the Debye–Scherrer distributions needed to smear out $I_\parallel$ patterns, we follow the initial steps and notation of H&L.

Consider a point B in a cylindrically averaged pattern from a disoriented one- or two-dimensionally periodic sample and let O be the origin (Fig. 3). The intensity at a point B having the spherical polar coordinates (v, σ) is found by integration over all possible rod or sheet orientations with the density of rod vectors or sheet-normal vectors pointing in any particular direction and the strength of the intensity that such rods or sheets contribute at B taken into account,

$$I = \int_\Omega I_\parallel(v, \gamma)N(\alpha) \, d\Omega/4\pi, \quad (13)$$

where (v, γ) refer to spherical polar coordinates of B in the frame of reference fixed by the rod or sheet, disoriented by the disorientation angle α (Fig. 4). This integration may be effected by consideration of the intensity contributed from an arbitrary cone of vectors lying at an angle $\gamma = \sigma_0$ to the line OB and multiplication by the number of vectors that lie between angles σ_0 and $\sigma_0 + d\sigma_0$. If we then integrate over φ, the azimuthal coordinate around OB, we have
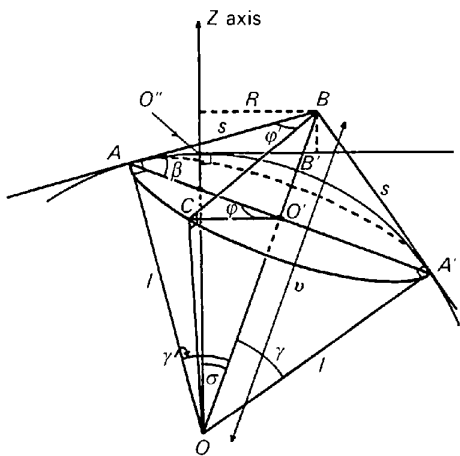
Fig. 3. Calculation of the intensity distribution in Debye–Scherrer arcs. The general point B in the disoriented pattern is described by the spherical polar coordinates (v, σ) or the cylindrical polar coordinates (R, Z). For any given rod or lattice plane, the spherical polar coordinates of B in the rod- or lattice-plane frame of reference are (v, γ) and the cylindrical polar coordinates are (s, l). The axes of the disoriented rods or normal vectors of disoriented lattice planes contributing at the point B may be grouped into cones of semi-angle γ. A rod gives rise to a line of intensity running parallel to the s axis at the coordinate l in the frame of reference fixed in the rod. Similarly, a lattice plane gives rise to a line of intensity running parallel to the l axis at the radial coordinate s in the frame of reference fixed in the sheet. The angle of disorientation of a general rod or lattice plane, defined by the vector $\overline{OC}$, is $O''CO = \alpha$, where $O''O$ is the axis of mean orientation. φ is the azimuthal coordinate of the vector $\overline{OC}$, measured about the line OB. The drawings for Figs. 3 and 4 are from Holmes & Leigh (1974), reproduced with their permission.
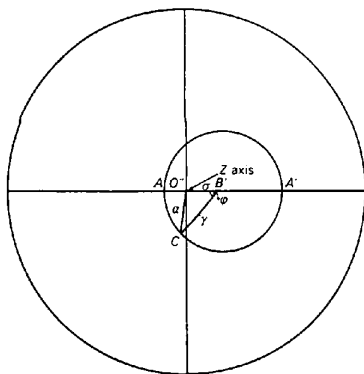

Fig. 4. Angles used in disorientation. Stereogram showing the relationship between angles α, σ, γ and φ, defined in Fig. 3.

the contribution from all such cones,

$$I = (4\pi)^{-1} \int_{\varphi=0}^{2\pi} \int_{\sigma_0=0}^{\pi} I_{\parallel}(v, \sigma_0) N(\alpha) \sin \sigma_0 \, d\sigma_0 \, d\varphi.$$

$$(14)$$

Since $I_{\parallel}(v, \sigma_0)$ is cylindrically symmetric, we rearrange (14) to give

$$I = (4\pi)^{-1} \int_{\sigma_0=0}^{\pi} I_{\parallel}(v, \sigma_0) \left[\sin \sigma_0 \int_{\varphi=0}^{2\pi} N(\alpha) \, d\varphi \right] d\sigma_0$$

$$(15)$$

and we define

$$A(N; \sigma_0, \sigma) = \sin \sigma_0 \int_{\varphi=0}^{2\pi} N(\alpha) \, d\varphi \qquad (16)$$

to be the distribution of intensity in a Debye–Scherrer arc, as a function of the angle σ in the disoriented pattern, which arises from a point in the $I_{\parallel}$ pattern located at the angle σ_0 to the Z axis. Then, from (15), the general formula for the intensity distribution in the whole disoriented pattern, which properly scales the meridional and off-meridional data, is

$$I_{\backslash/}(v, \sigma) = (4\pi)^{-1} \int_{\sigma_0=0}^{\pi} I_{\parallel}(v, \sigma_0) A(N; \sigma_0, \sigma) \, d\sigma_0.$$

$$(17)$$

The Debye–Scherrer distributions $A(N; \sigma_0, \sigma)$ can be evaluated by numerical integration of (16) for any disorientation distribution $N(\alpha)$, with use of the spherical triangle relation $\cos \alpha = \cos \sigma_0 \cos \sigma + \sin \sigma_0 \sin \sigma \cos \varphi$ (refer to Fig. 4 for $\gamma = \sigma_0$). Thus, the shape of the Debye–Scherrer distributions depends on the angle of the unsmeared intensity to the meridian, σ_0, and the disorientation distribution, $N(\alpha)$, but is independent of the spherical radius v. This fact permitted us to construct a lookup table of distributions to smear $I_{\parallel}$ patterns efficiently.

With the assumption of the special case of a Gaussian distribution

$$N(\alpha) = N_G(\alpha_0; \alpha) = (2/\alpha_0^2) \exp(-\alpha^2/2\alpha_0^2) \quad (18)$$

with standard deviation α_0, the integral in (16) may be evaluated analytically [see H&L, equation (11)]. The intensity in a Debye–Scherrer arc under these assumptions is

$$A_G(\alpha_0; \sigma_0, \sigma) = \sin \sigma_0 N_G(\alpha_0; \sigma - \sigma_0)$$

$$\times i_0[(\sin \sigma_0 \sin \sigma)/\alpha_0^2] \quad (19)$$

when expressed as the product of the Gaussian disorientation distribution for $\varphi = 0$ and a shape-factor term i_0.

In Fig. 5, Debye–Scherrer distributions are plotted as a function of the angle of a point in the $I_{\parallel}$ pattern, σ_0, with the assumption of a Gaussian disorientation distribution with standard deviation α_0. Our computer program to calculate these distributions used

the general formula (16) for integrating digital representations of the disorientation distributions. As a test, we compared the shape distributions $A(N; \sigma_0, \sigma)$, calculated for Gaussian $N(\alpha)$, to the corresponding $A_G(\alpha_0; \sigma_0, \sigma)$ calculated using the analytical equation (19) and found no significant differences. With 4° disorientation, the arcs away from the meridian are approximately Gaussian, centered at the angle σ_0 (Fig. 5a). With 15° disorientation, however, the distributions show a much stronger dependence upon σ_0 (Fig. 5b). In this case, the
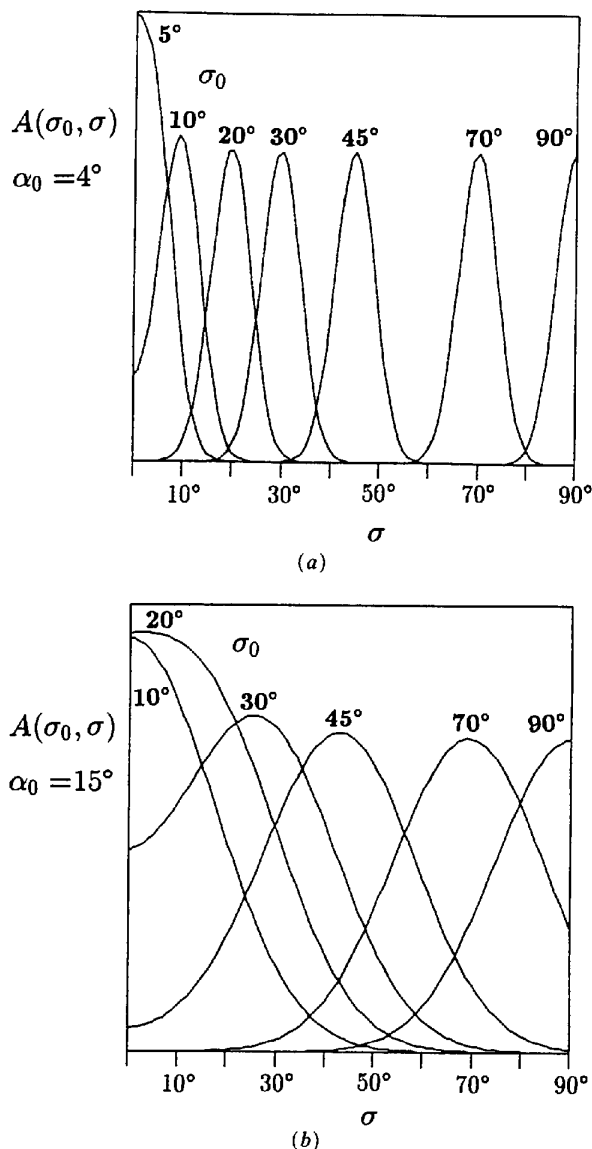
Fig. 5. Plots of Debye-Scherrer arcs for construction of lookup tables. The intensity distribution in a Debye-Scherrer arc arising from a point in an oriented $I_\parallel$ pattern, as a function of angle σ_0, for Gaussian disorientation distributions with standard deviations (a) $\alpha_0 = 4°$ and (b) $\alpha_0 = 15°$. Notice the asymmetry of Debye-Scherrer arcs is most pronounced for points near the meridian, where the shape term i_0 varies most rapidly.

maxima in the Debye-Scherrer arcs do not occur at the corrsponding σ_0 angles, but are shifted towards the meridian; all points at $\sigma_0 \lesssim 20°$ in the $I_\parallel$ pattern correspond to an arc with an maximum on the meridian in the disoriented pattern $I_{\backslash/}$.

Angular convolution lookup table. The completed lookup table was a two-dimensional matrix containing the intensity distributions in Debye-Scherrer arcs, for patterns from samples with disorientation $N(\alpha)$, tabulated in rows for discrete values of the angle σ_0 in the $I_\parallel$ coordinate system. The columns corresponded to discrete values of the angle σ in the coordinate system of the disoriented pattern $I_{\backslash/}$. For computationally efficient smearing of single-quadrant $I_\parallel$ patterns to give single-quadrant $I_{\backslash/}$ patterns, these Debye-Scherrer distributions were first calculated for angles σ_0 over the domain $[0, \pi/2]$ and σ over the domain $[0, \pi]$, both at 1° intervals. Tails of the Debye-Scherrer distributions that extended across the equator at $\sigma = \pi/2$ were then 'folded back' and added to the original distribution,

$$A'(N; \sigma_0, \sigma) = A(N; \sigma_0, \sigma) + A(N; \sigma_0, \pi - \sigma). \quad (20)$$

From the stored lookup table A', single-quadrant disoriented patterns from trial models were calculated rapidly in spherical polar sampling coordinates,

$$I_{\backslash/}(v, \sigma) = \sum_{\sigma_0=0}^{90°} I_\parallel(v, \sigma_0) A'(N; \sigma_0, \sigma). \quad (21)$$

Simulation of X-ray beam divergence

The experimental intensity distribution measured from a diffracting sample will depend upon the collimation of the beam. The effect is that the intensity distribution calculated by the assumption of perfect collimation, $I_{\backslash/}$, is convoluted with the beam shape $I_{beam}(R, Z)$. If the beam, for example, is approximated by a two-dimensional Gaussian distribution with standard deviations b_R and b_Z, then

$$I_{calc}(R, Z) = \sum_{Z'=Z-3b_Z}^{Z+3b_Z} \sum_{R'=R-3b_R}^{R+3b_R} I_{\backslash/}(R', Z')$$
$$\times \exp\left[-(R'-R)^2/2b_R^2 - (Z'-Z)^2/2b_Z^2\right]. \quad (22)$$

Thus, for a trial set of $\langle F^2 \rangle_\psi$, with all the parameters and functions in the simulation accounted for and the assumption of Gaussian distributions for N, f and I_{beam}, the disoriented pattern in spherical polar coordinates is

$$I_{trial}(v, \sigma) = I_{calc}(\langle F^2_{trial} \rangle_\psi, a, \alpha_0, p, b_R, b_Z, \ldots; v, \sigma). \quad (23)$$

Here, a represents the appropriate unit-cell constant(s) and the ellipsis indicates that parameters for additional convoluting functions can be included.

Adjustment of the trial $\langle F^2 \rangle_\psi$ by local angular regression

Once a disoriented diffraction pattern $I_{trial}(v, \sigma)$ had been simulated from the $\langle F^2_{trial} \rangle_\psi$, the goal was to determine a new set of real numbers $\langle F^2_{adj} \rangle_\psi$ that would give a disoriented pattern in better accord with the experimental (or expected) pattern $I_{exp}(v, \sigma)$. In spherical-polar coordinates, adjustment coefficients $\kappa(v_q, \sigma_q)$ were needed to correct each of the $\langle F^2(v_q, \sigma_q) \rangle_\psi$,

$$\langle F^2_{adj}(v_q, \sigma_q) \rangle_\psi = \kappa(v_q, \sigma_q) \langle F^2_{trial}(v_q, \sigma_q) \rangle_\psi. \quad (24)$$

These were estimated from local fits of the trial-model disoriented pattern, I_{trial}, to the I_{exp} pattern with use of robust linear regression (Andrews, 1974) in small angular neighborhoods centered about the (v_q, σ_q) coordinate of each point q in $\langle F^2 \rangle_\psi$. The statistical model for the qth local regression was thus

$$I_{exp}(v_q, \sigma_j) = \kappa(v_q, \sigma_q) I_{trial}(v_q, \sigma_j) \quad (25)$$

for the sample points in the disoriented patterns located on an arc at radius v_q centered at σ_q, where the jth point in the set was located at angle σ_j. The weighted least-squares estimate of κ was given by

$$\hat{\kappa}(v_q, \sigma_q) = \frac{\langle w_j I_{trial}(v_q, \sigma_j) I_{exp}(v_q, \sigma_j) \rangle_j}{\langle w_j I^2_{trial}(v_q, \sigma_j) \rangle_j}. \quad (26)$$

We started each local regression with the $\hat{\kappa}$ obtained with weights $w_j = 1$ for all j and then performed three rounds of iterative weight adjustment. At each iteration, each new weight w_j was based on the normalized absolute residual u_j from the previous fit,

$$u_j = \Delta_j / cM,$$
$$\Delta_j = |I_{exp}(v_q, \sigma_j) - \hat{\kappa}(v_q, \sigma_q) I_{trial}(v_q, \sigma_j)|. \quad (27)$$

For normalization, we chose M to be the median of the Δ_j and $c = 6.0$. If u_j was greater than 1.0, then the sample point j was given the new weight $w_j = 0$. Otherwise,

$$w_j = (1 - u_j^2)^2. \quad (28)$$

This is known as the Tukey biweight function (Mosteller & Tukey, 1977). Use of this iterative weighting function means that, in parts of the patterns that have closely spaced lines, a contaminating itensity coming from a neighboring line on one side will tend to be ignored. Furthermore, the cutoff for $u_j > 1.0$ means that noisy sample points with relatively large Δ_j (*e.g.* outliers in I_{exp} owing to dust on the film) will get zero weight.

After all the local regressions had been performed, each $\langle F^2(v_q, \sigma_q) \rangle_\psi$ was multiplied by the corresponding $\hat{\kappa}(v_q, \sigma_q)$, then a minimum-wavelength constraint was applied in the following way. The lines were individually Fourier transformed and a 'low-pass' filter was applied to remove components corresponding to Patterson vectors larger than the diameter of

the rod or thickness of the sheet. The back-transformed lines were then recombined.

The entire process of model pattern simulation by angular convolution and model adjustment by local regression was repeated in a cycle

$$\langle F^2_i(v_q, \sigma_q) \rangle_\psi = \kappa_{i-1}(v_q, \sigma_q) \langle F^2_{i-1}(v_q, \sigma_q) \rangle_\psi, \quad (29)$$

$$I_i(v, \sigma) = I_{calc}(\langle F^2_i \rangle_\psi; v, \sigma), \quad (30)$$

until the differences between I_i and I_{exp} were minimized. To monitor convergence, the residual

$$\mathcal{R}_I = \frac{\sum_{v,\sigma} v \sin \sigma |I_i(v, \sigma) - I_{exp}(v, \sigma)|}{\sum_{v,\sigma} v \sin \sigma |I_{exp}(v, \sigma)|} \quad (31)$$

was evaluated after each new $I_i(v, \sigma)$ had been simulated. After convergence, the disoriented I_{exp} pattern had been deconvoluted and we obtained the set of cylindrically averaged squared structure factors resulting from the final adjustment,

$$K\langle F^2_{trial} \rangle_\psi = \prod_i \kappa_i(v_q, \sigma_q) \langle F^2_{trial}(v_q, \sigma_q) \rangle_\psi. \quad (32)$$

Sample deconvolutions

Expected model: imperfectly oriented gap junctions

To test our procedures for decolvolution of disoriented diffraction patterns, we chose an expected model that represented imperfectly oriented gap junction membrane pairs containing two-dimensional hexagonal lattices of protein channels. In this model, each protein channel was represented by a dimer of hexamers spanning the membrane pair. Within each identical monomer was an ordered domain containing irregular rod-shaped columns of density. These rods, representing ~30 Å long α-helical segments, were centered within the hydrocarbon cores of the membranes, inclined at small angles to the membrane normals. In the 'gap' between the membrane pairs were additional short rods in each monomer inclined at larger angles to the membrane normals.

In experimental gap junction patterns, the widths of the lattice lines increase with increasing order, which corresponds to a diverging autocorrelation function for the hexagonal array of connexons, *i.e.* the lattice constant fluctuates locally with a standard deviation of ~5%. This effect was simulated by summation of the $I_{\parallel}$ patterns of three component lattices: the mean with lattice constant $a = 79.2$ Å and weight 0.6, and variants with $a = 78.8$ and 79.6 Å and weight 0.2, which approximates the effect of a Gaussian distribution of lattice constants. The electron density profile of the membrane pair in our expected model was based on a published analysis of experimental gap junction X-ray patterns (Makowski, Caspar, Phillips & Goodenough, 1977).

For our tests, we varied the disorientation, α_0, of the expected model diffraction patterns

$$I_{\exp}(\alpha_0) = I_{\text{calc}}(\langle F_{\exp}^2 \rangle_\psi, \alpha_0, p, b_R, b_Z; v, \sigma), \quad (33)$$

which we simulated by smearing the $I_\|$ pattern

$$I_{\|\exp} = I_\|(\langle F_{\exp}^2 \rangle_\psi, p; v, \sigma_0) \quad (34)$$

using lookup tables based on $\alpha_0 = 4$ and $\alpha_0 = 15°$ Gaussian disorientation distributions (Fig. 6). In these expected disoriented patterns, $I_{\exp}(4°)$ and $I_{\exp}(15°)$, neighboring lines are smeared together on the equator near 0.1 Å^{-1}, similar to experimental gap junction patterns (see Tibbits et al., 1990). To deconvolute these test patterns, we started with two different $\langle F_{\text{trial}}^2 \rangle_\psi$: (1) a set $\langle F_{\text{pt}}^2 \rangle_\psi$ corresponding to uniform intensity along each reciprocal-lattice line, which is the transform of a structureless point model; (2) a set $\langle F_{\text{rod}}^2 \rangle_\psi$ predicted by a physically plausible rod model that simulated some features in the expected structure. In our test deconvolutions, only $\langle F_{\text{trial}}^2 \rangle_\psi$ and α_0 varied, the remaining parameters and sampling coordinates were the same; so we have adopted the simple notation $I_{\|\text{trial}} = I_\|(\langle F_{\text{trial}}^2 \rangle_\psi)$, $I_{\text{trial}}(\alpha_0) = I_{\text{calc}}(\langle F_{\text{trial}}^2 \rangle_\psi, \alpha_0)$ and $I_{K\text{trial}}(\alpha_0) = I_{\text{calc}}(K_{\alpha_0} \langle F_{\text{trial}}^2 \rangle_\psi, \alpha)$. Here, K_{α_0} refers to the product of the adjustment coefficients made to the $\langle F_{\text{trial}}^2 \rangle_\psi$ that minimized the difference between $I_{\text{trial}}(\alpha_0)$ and $I_{\exp}(\alpha_0)$.

Starting with a point model

We started with the Fourier transform of a 'point' model, in which $\text{Re}\{F_{hkm}\}$ is a positive constant and $\text{Im}\{F_{hkm}\} = 0$ for all points hkm in the model trans-

Table 2. Fits to the disoriented target patterns

$\alpha_0(°)$	Trial model	$\mathcal{R}_I$
4	Point	0.051
15	Point	0.033
4	Rod	0.092
15	Rod	0.039

form. Since the projected electron density distribution across the membrane was refined using profile fitting to the meridional reflections, the trial meridian was set equal to the expected meridian and not adjusted. The point-model transform was first used to construct an $I_{\|\text{pt}} = I_\|(\langle F_{\text{pt}}^2 \rangle_\psi)$ pattern that was then smeared, using lookup tables of Debye–Scherrer distributions, to give the disoriented patterns $I_{\text{pt}}(4°)$ and $I_{\text{pt}}(15°)$ shown in Fig. 7.

In the two deconvolutions that start from the point model, iterative adjustments were made to the $\langle F_{\text{pt}}^2 \rangle_\psi$ until the disoriented model patterns fitted the corresponding expected patterns. The closeness of the final fits can be judged by comparison of the two $I_{K\text{pt}}(\alpha_0)$ patterns in Fig. 7 with the $I_{\exp}(\alpha_0)$ patterns in Fig. 6; the calculated residuals are given in Table 2. Not surprisingly, a lower residual $\mathcal{R}_I$ was obtained from the comparison of $I_{K\text{pt}}(15°)$ with $I_{\exp}(15°)$ than from the comparison of $I_{K\text{pt}}(4°)$ with $I_{\exp}(4°)$ because the very broad smearing in the patterns with 15° disorientation set a limit to the importance of sharp features in the underlying $\langle F^2 \rangle_\psi$.

Starting with a rod model

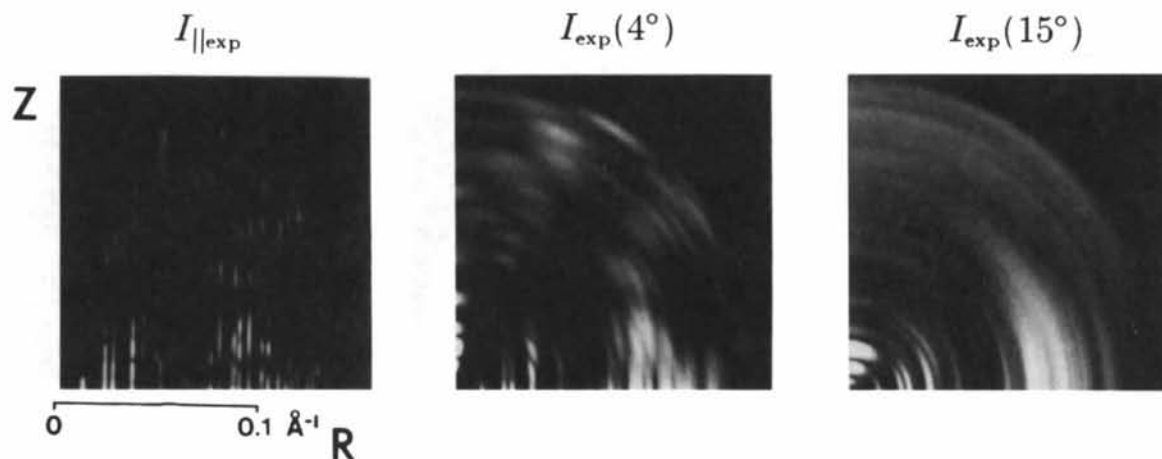The rod trial model represented each protein monomer in the paired connexon hexamers by four

$$I_{\|\exp} \qquad\qquad I_{\exp}(4°) \qquad\qquad I_{\exp}(15°)$$

Fig. 6. Expected patterns. An oriented pattern simulated from the expected model, $I_{\|\exp} = I_\|(\langle F_{\exp}^2 \rangle_\psi, p)$, with a large constant p, which gives narrow lattice lines, and with the meridian omitted for clarity. The $I_\|(\langle F_{\exp}^2 \rangle_\psi)$ pattern we used in our tests was similar but had the meridian included and line widths that increased with R. This pattern was angularly smeared to simulate two disoriented expected patterns, $I_{\exp}(4°) = I_{\text{calc}}(\langle F_{\exp}^2 \rangle_\psi, \alpha_0 = 4°)$ and $I_{\exp}(15°)$. The single quadrant shown corresponds to combined experimental data from untilted and tilted samples in the X-ray beam, after the optical densities scanned from the films are corrected for the curvature of the sphere of reflection (Fraser, MacRae, Miller & Rowlands, 1976), a smooth background is subtracted and noise is reduced by mm-symmetry averaging (Tibbitts et al., 1990). To display the high- and low-angle portion of the disoriented patterns together on a convenient scale, the intensities at each pixel have been mutiplied by the square of their distance from the origin (v^2).

Gaussian untilted transmembrane cylinders. This model was created by building three Gaussian cylinders of length 30 Å and one of length 10 Å into each monomer in a low-resolution map. There were no features in the rod trial model that corresponded to the short rods near the midplane of the connexon pair in the expected map.

Fig. 8 shows the $I_\parallel$ pattern from the rod trial model. Disorientation was added to this pattern using the lookup tables of Debye–Scherrer smearing distributions. Fig. 8 also shows the two $I_{rod}(\alpha_0)$ patterns and the two $I_{Krod}(\alpha_0)$ patterns after model adjustment. To judge the closeness of the fits, compare $I_{Krod}(\alpha_0)$ patterns to the corresponding $I_{exp}(\alpha_0)$ patterns in Fig. 6. By comparing the values of $\mathcal{R}_I$ for the rod-model deconvolutions to those for the point model, we were able to assess the extent to which model adjustment was constrained by introducing node positions based on the rod trial model (Table 2). Incorrect node positions from the rod model caused the final $\mathcal{R}_I$

values to be higher relative to those obtained with the point-model starting point.

Comparison of deconvolution results

Deconvoluted patterns starting from the point model showed that, for the $K_{4^\circ}\langle F_{pt}^2\rangle_\psi$ result, the Z coordinates of two or three near-equatorial nodes in the $\langle F_{exp}^2\rangle_\psi$ were indicated by minima in nearly all the lattice lines (Fig. 9). In the $K_{15^\circ}\langle F_{pt}^2\rangle_\psi$ pattern, a similar number of nodes in the $\langle F_{exp}^2\rangle_\psi$ were indicated by minima, but these were much more shallow. Thus, the uncertainty in the positions of near-equatorial nodes indicated by the results was strongly dependent upon, and increased by, the degree of disorientation.

The results starting with the rod model showed that for the $K_{4^\circ}\langle F_{rod}^2\rangle_\psi$ result most lattice lines between $R = 0.05$ and $R = 0.11$ Å^{-1} had one near-equatorial node, derived from the $\langle F_{rod}^2\rangle_\psi$ that was not present in the $\langle F_{exp}^2\rangle_\psi$ (compare Figs. 8 and 9). The same
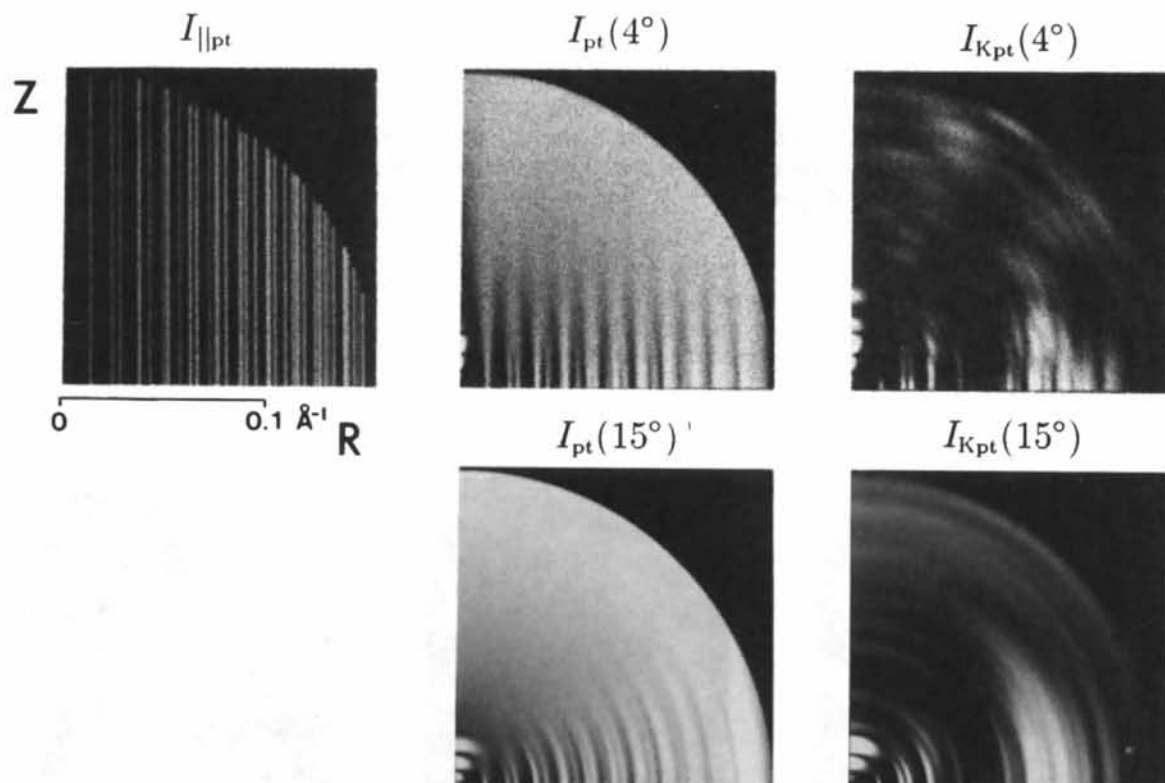

Fig. 7. Adjustment of the point model. The transform of a point model, which has uniform intensity on each lattice line (h, k), was used to calculate the $I_{\parallel pt} = I_\parallel(\langle F_{pt}^2\rangle_\psi)$ pattern. Some lines appear stronger relative to the others because they have multiple overlapping azimuthal components. We generated the $I_{pt}(4^\circ)$ and $I_{pt}(15^\circ)$ patterns by smearing a similar $I_{\parallel pt}$ pattern with the expected model meridian added to it. The following sequence of operations successfully minimized the difference between the model and expected patterns. The once-adjusted $\langle F^2\rangle_\psi$ were used to recalculate a model disoriented pattern, then readjusted to fit the data. This process was repeated for eight cycles. Each line was then individually Fourier transformed and a low-pass filter with a soft edge at 90 Å was applied to impose the minimum-wavelength constraint. A second round of eight simulations/adjustments was then performed. The minimum-wavelength constraint was then applied a second time. With the VMS operating system in a low-priority batch queue and all these operations performed with 385×512 pixel floating-point image arrays (in spherical polar coordinates) required approximately 30 min CPU time on a VAXstation 3100 workstation configured with 32 Mbytes of memory. The adjusted model disoriented patterns are $I_{K_{pt}}(4^\circ) = I_{calc}(K_{4^\circ}\langle F_{pt}^2\rangle_\psi, \alpha_0 = 4^\circ)$ and $I_{K_{pt}}(15^\circ)$.

lattice lines in the $K_{15°}\langle F_{rod}^2\rangle_\psi$ result showed two or more incorrect nodes that retained their positions from the starting model. Thus, as seen before with the deconvolutions starting with uniform intensities, increased disorientation extended the region near the equator where the position of nodes could not be well determined.

To evaluate our deconvolution results quantitatively, we calculated residual curves $\mathcal{R}_{\langle F^2\rangle}(R)$, comparing each $K\langle F_{trial}^2\rangle_\psi$ with the $\langle F_{exp}^2\rangle_\psi$, and plotted the results in Fig. 10. Note the higher residuals of the $K_{15°}\langle F_{pt}^2\rangle_\psi$ and $K_{15°}\langle F_{rod}^2\rangle_\psi$ curves relative to the $K_{4°}\langle F_{pt}^2\rangle_\psi$ and $K_{4°}\langle F_{rod}^2\rangle_\psi$ curves. This indicates that, even though the $I_{Kpt}(15°)$ and $I_{Krod}(15°)$ patterns were better fits to the expected pattern $I_{exp}(15°)$ than the $I_{Kpt}(4°)$ and $I_{Krod}(4°)$ patterns were to $I_{exp}(4°)$, the $\langle F^2\rangle_\psi$ resulting from these deconvolutions were less reliable.

Electron density maps of the adjusted rod model

A three-dimensional electron density map was calculated from the $K_{4°}\langle F_{rod}^2\rangle_\psi$ result and azimuthal separation information from the rod trial model, with use of the fast Fourier transform $\mathcal{F}$,

$$\rho_{Krod}(r, z, \theta) = \mathcal{F}\{K_q^{1/2} F_{hkm,rod}\}. \tag{35}$$

This map was compared to maps of the trial rod and expected models, $\rho_{rod}(r, z, \theta) = \mathcal{F}(F_{hkm,rod})$ and $\rho_{exp}(r, z, \theta) = \mathcal{F}(F_{hkm,exp})$, to see how iterative convolution and regression changed features in the trial map to be more like those in the expected map.

In a transverse section near the center of the membrane bilayer ($z = 39.5$ Å), the expected map showed the connexon represented by high-density protein surrounded by relatively featureless lower-density regions representing lipid hydrocarbon chains (Fig. 11). The corresponding section from the trial map showed a hexagonal arrangement of six monomers, each containing three regular untilted Gaussian rods. In the new ρ_{Krod} map, features appeared that are similar to those in the expected structure: elongated concentrations of density replaced the uniform and circular rod cross sections and regions of low density appeared between the monomers.

Also shown in Fig. 11 are transverse sections near the midplane of the paired hexamers ($z = 5$ Å). For this portion of the structure, there were short segments in ρ_{exp} of rod-like density and ρ_{rod} was simply a low-resolution map. Nevertheless, the ρ_{Krod} map showed that concentrations of density appeared as a result of the adjustments made to the $\langle F_{rod}^2\rangle_\psi$ by our iterative convolution and local regression procedures.
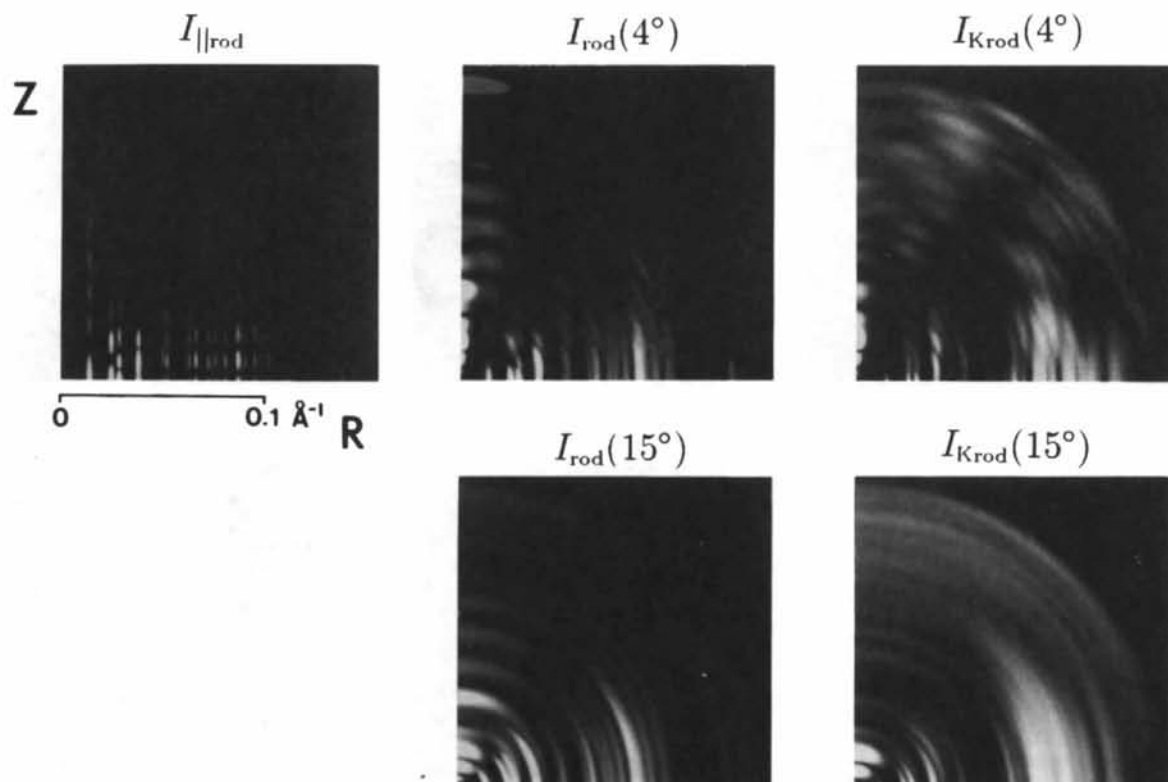

Fig. 8. Adjustment of the rod model. Diffraction patterns from the two deconvolution procedures that used the rod trial model: the disoriented trial model pattern $I_{\|rod}$; the trial disoriented pattern $I_{rod}(4°)$; the adjusted model disoriented pattern $I_{Krod}(4°)$; the trial disoriented pattern $I_{rod}(15°)$; and the adjusted model disoriented pattern $I_{Krod}(15°)$.

Only the relative magnitudes of the cylindrically averaged squared structure factors in the trial model were adjusted but this gave rise to new features in the corresponding map that were a good first approximation to peaks of protein density in the expected structure. The $\rho_{K\,rod}$ sections also show features that have arisen in the space between the hexamers – regions that were featureless in the low-resolution map.

Discussion

Our simulation and local regression procedures, which use a lookup table for the intensity distribution in Debye–Scherrer arcs, provide a simple and efficient way to deconvolute diffraction patterns from imperfectly oriented samples. The simulation procedures we use can take into account disorder within periodic rods or sheets and smearing due to beam divergence and other functions. By using local angular regression, cylindrically averaged squared structure factors based on an arbitrary starting model can be adjusted to be consistent with the diffraction data. Our procedures can therefore provide good initial estimates of the cylindrically averaged squared structure factor from only the degree of disorientation α_0,

a coherence length p and the constant(s) needed to index the lines in the pattern.

In our sample deconvolutions of gap junction diffraction patterns, the choice of the rod trial model enforced some *a priori* constraints on the model adjustment procedure. The rod model, although approximating the expected structure, defined some near-equatorial nodes that were badly placed and these held to their original positions. From a comparison of the final fits, a lower residual $\mathcal{R}_I$ was obtained for $I_{K\,pt}(4°)$, which resulted from the adjustment of uniform intensities to fit the $I_{exp}(4°)$ pattern, than for $I_{K\,rod}(4°)$, which started from the transform of the physically plausible model. Increased smearing due to disorientation, however, diminished the importance of constraints on the trial model; very good fits to the $I_{exp}(15°)$ pattern were found with either trial model as a starting point ($\mathcal{R}_I \simeq 0.03$).

Better fits were found to the $I_{exp}(15°)$ expected pattern than to $I_{exp}(4°)$ but the resulting $K_{15°}\langle F^2_{trial}\rangle_\psi$ values were less reliable than the $K_{4°}\langle F^2_{trial}\rangle_\psi$ deconvolutions, as indicated by their higher residuals $\mathcal{R}_{(F^2)}$. Near the equator in patterns from disoriented two-dimensionally periodic samples, the direction of smearing is parallel to the lattice lines. As the
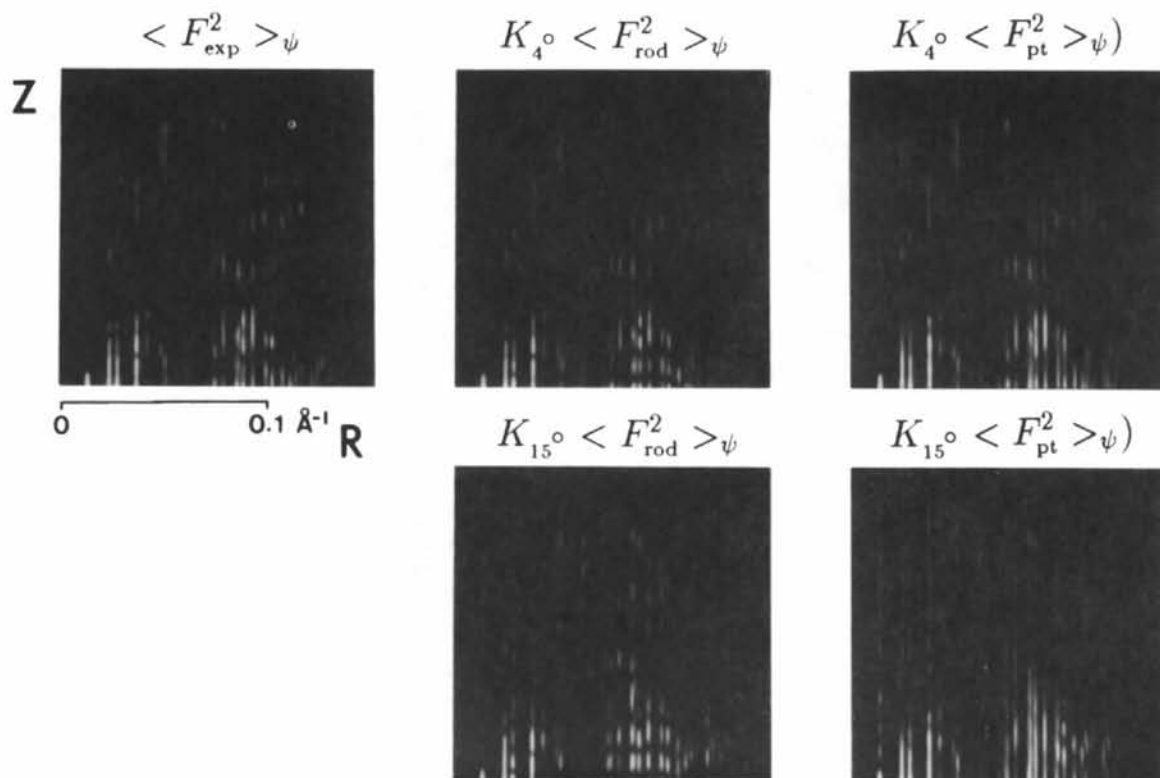

Fig. 9. Deconvolution results. $I_\parallel(\langle F^2_{exp}\rangle_\psi)$, representing the cylindrically averaged squared structure factors from the expected model (which is the same as in Fig. 6), is shown for comparison to the results. $I_\parallel$ patterns are shown for $K_{4°}\langle F^2_{pt}\rangle_\psi$ and $K_{4°}\langle F^2_{rod}\rangle_\psi$, which resulted from deconvolution of the $I_{exp}(4°)$ pattern; and for $K_{15°}\langle F^2_{pt}\rangle_\psi$ and $K_{15°}\langle F^2_{rod}\rangle_\psi$, which resulted from deconvolution of the $I_{exp}(15°)$ pattern. The meridional intensities were included during all the tests but were held constant and are not shown.
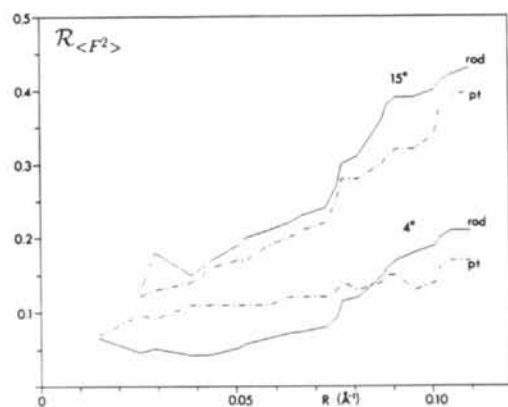
Fig. 10. Reliability of iterative convolution and local regression. To quantitate the agreement between sets of cylindrically averaged squared structure factors $\langle F^2_{model}\rangle_\psi$ and their expected values $\langle F^2_{exp}\rangle_\psi$, we calculated residuals defined as $\mathcal{R}_{\langle F^2\rangle}(R) = \sum_q |\langle F^2_{model}\rangle_\psi - \langle F^2_{exp}\rangle_\psi|/\sum_q \langle F^2_{exp}\rangle_\psi$. The four curves show the dependence of $\mathcal{R}_{\langle F^2\rangle}$ on resolution as lattice lines at increasing lateral distance R from the meridian are included in the comparison. The dashed curves compare $K_{4^\circ}\langle F^2_{pt}\rangle_\psi$ and $K_{15^\circ}\langle F^2_{pt}\rangle_\psi$ to $\langle F^2_{exp}\rangle_\psi$. The solid curves compare $K_{4^\circ}\langle F^2_{rod}\rangle_\psi$ and $K_{15^\circ}\langle F^2_{rod}\rangle_\psi$ to $\langle F^2_{exp}\rangle_\psi$.
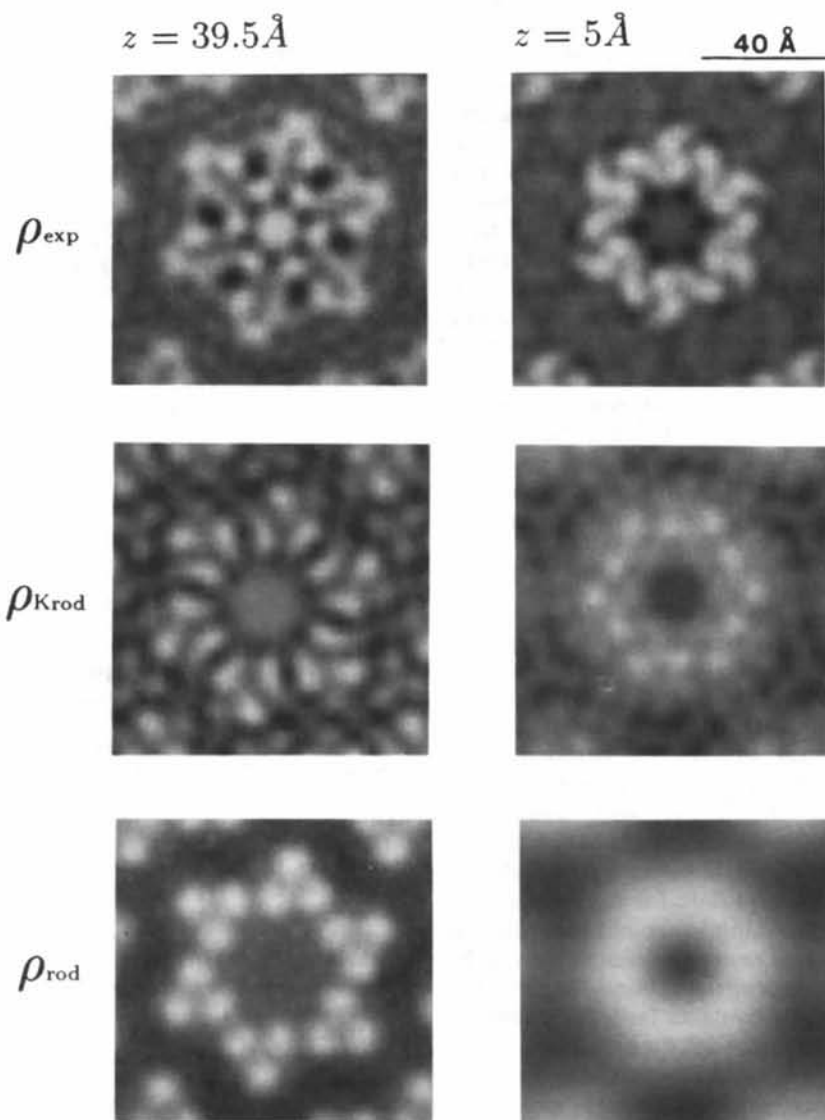

Fig. 11. Comparison of adjusted rod model map to the expected map. Transverse sections of the three-dimensional electron density maps for the expected model, the adjusted rod model and the rod trial model. The sections were taken at $z = 39.5$ and $z = 5.0$ Å, measured from the midplane of the paired connexon hexamers.

disorientation increases, so does the extent to which near-equatorial features of the intensity distribution are blurred. Furthermore, the degree of angular overlap of off-equatorial intensities arising from different lines increases. Thus, for both trial models, the deconvolutions of the $I_{exp}(15°)$ pattern were less reliable than the deconvolutions of $I_{exp}(4°)$, despite the fact that the disoriented patterns $I_{K pt}(15°)$ and $I_{K rod}(15°)$ gave the lowest residuals $\mathscr{R}_I$.

To refine models exactly, iterative convolution adjustment procedures must be combined with independent structural information. The first step is to build a plausible low-resolution model. This can be derived from electron-microscopic data and initial deconvolution of the diffraction pattern using a point model. Each refinement cycle begins with adjustment of the trial model $\langle F^2 \rangle_\psi$ to be consistent with the diffraction pattern. After Fourier transformation of the adjusted model structure factors, modifications are made to the density distribution based on features in electron micrographs and other structural data. The cycle is repeated until the electron density map is no longer changed by the application of the model constraints. Since each deconvolution of the diffraction pattern partially depends upon the trial model used, the progressively better models produced in the refinement will give better initial angular and lateral separation of closely spaced intensities in the diffraction data.

In fiber diffracton patterns from rod-like helical arrangements of subunits such as F-actin, the highest-resolution data usually appear on or near the meridian (see Holmes et al., 1990). Unfortunately, the effects of disorientation are most pronounced near the meridian in fiber patterns, which makes numerical deconvolution of these data difficult. Furthermore, the shape of the Debye–Scherrer arcs near the meridian are asymmetric, so that correction factors based on the assumption of approximately Gaussian arcs cannot scale these data properly to the rest of the pattern. For filaments with well characterized subunits (such as actin and mutant hemoglobins), monomers can be built into a low-resolution map of the polymeric form. Our convolution and local-regression procedures make it possible to refine their positions in the assembled filament against all the measurable fiber diffraction data for consistency. To refine models to near-atomic resolution, stereochemical constraints and energy minimization can be applied to enforce appropriate bond angles and plausible van der Waals contacts (see, for example, Hendrickson & Konnert, 1980a,b). Thus, with sufficient model constraints, our iterative angular convolution and local regression procedures provide a method for the refinement of current models of rod-like and sheet-like biological assemblies to significantly higher resolution using high-angle parts of disoriented diffraction patterns where intensities cannot be reliably deconvoluted by matrix inversion-based numerical procedures.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1965). *Handbook of Mathematical Functions.* New York: Dover.

ALEXANDER, L. E. (1969). *X-ray Diffraction Methods in Polymer Science.* New York: Wiley-Interscience.

ANDREWS, D. F. (1974). *Technometrics,* 16, 523–529.

ARNOTT, S., CHANDRASEKRAN, R., PUIGJANER, L. C., WALKER, J. K., HALL, I. H., BIRDSALL, D. L. & RATCLIFF, R. L. (1983). *Nucleic Acids Res.* 11, 1457–1474.

ASTBURY, W. T. & STREET, A. (1931). *Philos. Trans. R. Soc. London Ser. A,* 230, 75–101.

BAMFORD, C. H., BROWN, L., ELLIOT, A., HANBY, W. E. & TROTTER, I. F. (1954). *Nature (London),* 173, 27.

BÜLDT, G., MISCHEL, M., HENTSCHEL, M. P., REGENASS, M. & ROSENBUSH, J. P. (1986). *FEBS Lett.* 205, 29–31.

DICKERSON, R. E. (1964). *The Proteins,* 2nd ed., pp. 603–778. New York: Academic Press.

FRASER, P. E., NGUYEN, J. T., SUREWICZ, W. K. & KIRSCHNER, D. A. (1991). *Biophys. J.* 60, 1190–1201.

FRASER, R. D. B., MACRAE, T. P., MILLER, A. & ROWLANDS, R. J. (1976). *J. Appl. Cryst.* 9, 81–94.

GENNIS, R. B. (1989). *Biomembranes: Molecular Structure and Function.* New York: Springer-Verlag.

GUINIER, A. (1963). *X-ray Diffraction in Crystals, Imperfect Crystals and Amorphous Bodies.* San Francisco: W. H. Freeman.

HENDRICKSON, W. A. & KONNERT, J. H. (1980a). *Biomol. Struct. Funct. Conform. Evol.* 1, 43–57.

HENDRICKSON, W. A. & KONNERT, J. H. (1980b). *Computing in Crystallography,* edited by R. DIAMOND, S. RAMASESHAN & R. VENKATESAN, pp. 13.01–13.32. Bangalore: Indian Academy of Sciences.

HOLMES, K. C. & BLOW, D. M. (1966). *The Use of X-ray Diffraction in the Study of Protein and Nucleic Acid Structure.* New York: Robert E. Krieger.

HOLMES, K. C. & LEIGH, J. B. (1974). *Acta Cryst.* A30, 635–638.

HOLMES, K. C., POPP, D., GEBHARD, W. & KABSCH, W. (1990). *Nature (London),* 347, 44–49.

KLUG, A., CRICK, H. F. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* 11, 199–213.

MAKOWSKI, L. (1978). *J. Appl. Cryst.* 11, 273–283.

MAKOWSKI, L. (1991). *Acta Cryst.* A47, 562–567.

MAKOWSKI, L., CASPAR, D. L. D., PHILLIPS, W. C. & GOODENOUGH, D. A. (1977). *J. Cell Biol.* 74, 629–645.

MAKOWSKI, L. & LI, J. (1983). *Biomembranes: Structure and Function,* edited by D. CHAPMAN, pp. 43–166. London: Macmillan.

MOSTELLER, F. & TUKEY, J. W. (1977). *Data Analysis and Regression: a Second Course in Statistics,* pp. 353–356. Reading, MA: Addison-Wesley.

NAMBA, K., PATTANAYEK, R. & STUBBS, G. (1989). *J. Mol. Biol.* 208, 307–325.

NAMBA, K., YAMASHITA, I. & VONDERVISZT, F. (1989). *Nature (London),* 342, 648–654.

NAMBUDRIPAD, R., STARK, W., OPELLA, S. J. & MAKOWSKI, L. (1991). *Science,* 252, 1305–1308.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTER-LING, W. T. (1988). *Numerical Recipes in C: the Art of Scientific Computing.* Cambridge Univ. Press.

STUBBS, G. J. (1974). *Acta Cryst.* A30, 639-645.

TIBBITTS, T. T., CASPAR, D. L. D., PHILLIPS, W. C. & GOOD-ENOUGH, D. A. (1988). *Biophys. J.* 53, 634a.

TIBBITTS, T. T., CASPAR, D. L. D., PHILLIPS, W. C. & GOOD-ENOUGH, D. A. (1990). *Biophys. J.* 57, 1025-1036.

Frequency Statistical Method for Evaluating Cosine Invariants of Three-Phase Relationships

By DAVID A. LANGS

Medical Foundation of Buffalo, 73 High Street, Buffalo, NY 14203, USA

Abstract

A new variation on the established procedure to evaluate three-phase structure invariants through quadrupole relationships is described. This method differs from earlier algebraic formulations in that the cosine-invariant estimates are based on a conditional observed frequency distribution of $|E|$ magnitudes for the quadrupole, rather than on the values of the magnitudes themselves. Successful applications of this method to a number of structures that ranged in size from 84 to 317 independent non-hydrogen light atoms are given.

Introduction

The three-phase crystallographic structure invariants play a central role in the determination of crystal structures by direct-phasing methods. Tangent-formula methods for small-molecule determinations have traditionally relied on the 0 (modulo 2π) probability estimate for these 'triples' (Karle & Hauptman, 1956). Efforts to extend these techniques to larger structures have required more precise estimates to be obtained for these phase invariants, though use of algebraic formulae (Karle & Hauptman, 1957; Vaughan, 1958; Hauptman, 1964; Hauptman, Fisher, Hancock & Norton, 1969; Karle, 1970; Duax, Weeks & Hauptman, 1972; Hauptman & Duax, 1972), determinantal joint probability distributions (Tsoucaris, 1970; Messager & Tsoucaris, 1972; Giacovazzo, 1976, 1977a; Karle, 1979, 1980) or probabilistic formulae, as applied to isomorphous-replacement or anomalous-dispersion data (Hauptman, 1982; Giacovazzo, 1983; Fortier, Moore & Frazer, 1985) and to the extended neighborhoods or phasing shells of data that define higher-order relationships into which these triples have been suitably embedded (Hauptman, 1975; Giacovazzo, 1977b; Karle, 1982). This report describes a new method to estimate three-phase

invariants based on examination of the frequency distribution of $|E|$ magnitudes that complete a family of conditionally constructed quadrupoles that are common to the evaluated triple.

Background

One of the earliest strategies in direct-methods research was the development of formulae to evaluate crystal-structure phase invariants and semi-invariants as a means to determine crystal structures. This work was initiated about the same time that the rules for origin and enantiomorph specification and phase-extension techniques were being developed. Formulae to estimate the single-phase structure invariants had an immediate application; they provided a means to reduce the number of algebraic symbols that had to be permuted and tested for a selected starting group of phases. But algebraic formulae that were developed for the determination of the cosine values of the three-phase structure invariants, for example, for $P1$ symmetry (Karle & Hauptman, 1957),

$$|E_h E_{-k} E_{k-h}| \cos (\varphi_h - \varphi_k + \varphi_{k-h})$$

$$\simeq N^{-1/2}(|E_h|^2 + |E_k|^2 + |E_{k-h}|^2 - 2)$$

$$+ \tfrac{1}{2}N^{3/2}\langle(|E_l|^2 - 1)(|E_{l-k}|^2 - 1)(|E_{l-h}|^2 - 1)\rangle_l, \quad (1)$$

however, did not have an immediate impact on phasing practices. Firstly, these formulae were computationally demanding; the average of a product of $|E|^2 - 1$ magnitudes had to be computed over a range of 1 that sampled the whole of reciprocal space and, to ensure that the whole of reciprocal space was explored, the diffraction vectors of the triple, $\mathbf{h}$, $-\mathbf{k}$ and $\mathbf{k} - \mathbf{h}$, had to be cyclicly permuted within the framework of the formula. Secondly, numerical tests indicated that (1) tended to produce unreliable cosine