The Fourier transforms of curves and filaments and their application to low-resolution protein crystallography

Andrew C. Hausrath and Alain Goriely


Many research topics in condensed matter research, materials science and the life sciences make use of crystallographic methods to study crystalline and non-crystalline matter with neutrons, X-rays and electrons. Articles published in the *Journal of Applied Crystallography* focus on these methods and their use in identifying structural and diffusion-controlled phase transformations, structure–property relationships, structural changes of defects, interfaces and surfaces, etc. Developments of instrumentation and crystallographic apparatus, theory and interpretation, numerical analysis and other related subjects are also covered. The journal is the primary place where crystallographic computer program information is published.

**Crystallography Journals Online** is available from [journals.iucr.org](http://journals.iucr.org)
The Fourier transforms of curves and filaments and their application to low-resolution protein crystallography

Andrew C. Hausrath* and Alain Gorielyb

*Department of Biochemistry and Molecular Biophysics, University of Arizona, USA, and bProgram in Applied Math, Department of Mathematics and Bio5 Institute, University of Arizona, USA. Correspondence e-mail: hausrath@email.arizona.edu

A numerical method for computing the Fourier transform of an arbitrary space curve is described. The method is applicable to all sufficiently smooth curves and relies on the local geometric parameters describing a curve. The numerical results for a helical curve are compared with the exact analytical theory for the transform of a helix. It is shown that the transform of a filamentary density distribution radially symmetric around a curve is equivalent to the transform of that curve scaled by an appropriately defined weight function. These filamentary density distributions in conjunction with the numerical transform evaluation method can be used for simulating low-resolution diffraction data for protein crystals. Crystallographic structure factors obtained from a filament model representing a simple three-helix-bundle protein are compared with those calculated from conventional coordinate models. At low resolution, the filamentary representation provides an excellent approximation of the structure factors obtained from the standard coordinate model, but requires far fewer independent parameters.

1. Introduction

The determination of protein structures from well ordered and strongly diffracting crystals is now routine. While great progress has been made in structure determination from high-quality data, many interesting proteins yield only poor-quality crystals which diffract to limited resolution. In some of these difficult cases, crystal quality can be improved by better crystallization conditions, improved sample preparation techniques, limited proteolysis, judicious recolonization of expression constructs or chemical modification (Heras & Martin, 2005; Newman, 2006; Makabe et al., 2006; Walter et al., 2006; Neau et al., 2007; Abergel, 2004; Dale et al., 2003; Longenecker et al., 2001; Samygina et al., 2000; D’Arcy et al., 1999). These improvements push the diffraction limit of crystals and permit high-resolution structure determination. However, there are many cases where crystals will not yield diffraction to a resolution where conventional crystallographic methods for structure determination can be applied. This difficulty arises typically with membrane proteins, higher-order complexes or molecules that are intrinsically conformationally heterogeneous. Therefore, even when such molecules can be induced to crystallize, their conformational heterogeneity may prevent the long-range order necessary for strong diffraction.

Crystalllographic analyses are rarely attempted at resolutions below 3.5 Å (Brünger, 2005; DeLaBarre & Brünger, 2006). Below this resolution, electron-density maps do not contain sufficient detail for the construction of atomic coordinate models. This is due to the fact that atomic coordinate models involve a large number of parameters which cannot be constrained from low-resolution experimental observations. To circumvent this problem, various refinement protocols have been developed to reduce the number of degrees of freedom, such as the torsion angle refinement (Rice & Brünger, 1994) and the normal mode refinement (Kidéra & Go, 1990; Poon et al., 2007). Despite these optimization techniques, atomic models will always be underdetermined at low resolution.

X-ray crystallography is fundamentally an image-formation technique, and the electron-density maps it produces can be of very high quality even in cases where the diffraction limit is low. It is important to draw a distinction between the quality of a map and the accuracy of a model. Map quality is limited by how accurately the crystallographic phases can be determined and is independent of resolution. The accuracy of a model that can be constructed from a map is, however, limited by resolution. Although this distinction is clear, in practice the two notions are linked, since, in general, the only method to optimize the phases apart from experimental estimates is through model refinement. Indeed, in current practice, phase improvement is usually regarded as a by-product of model creation rather than as an objective in itself.

An important aspect of X-ray crystallography is the existence of quantitative methods for assessing the degree of confidence in its conclusions (Brünger, 1992; Laskowski et al., 1993). However, these methods make reference to the
agreement of a model with the electron density from which it is derived, or the consistency of the model with independently determined characteristics of other proteins. Hence, these methods cannot be applied at low resolution. However, this does not mean the images formed by the technique at low resolution are intrinsically unreliable, but at present few appropriate methods exist for the assessment (or optimization) of these images.

As an illustration of the issues related to low-resolution imaging, we compare briefly the problem of creating images from either single-particle reconstruction from electron microscopy (EM) or from X-ray crystallographic data. Crystallographic data can, in many cases, be obtained to low resolution (e.g. 6 Å resolution) even when higher-resolution data remain unattainable. However, the central problem in interpreting crystallographic data is the missing phases and a model must be used to define the phases. For EM, individual images are obtained directly but obtaining sufficient data for a 6 Å resolution reconstruction is a significant technical challenge. The more easily obtainable X-ray data at comparable resolution are not currently used, despite the fact that they may contain equally valuable structural information. Our purpose here is to develop a modeling method capable of supplying the phases which will permit use of such data.

To do so, we use a continuous model of protein structure based on the differential geometry of space curves, which can be tuned to different levels of detail. The number of free parameters in the representation can be adapted to the information available from a given experimental data set. The main advantage of this approach is that, in principle, a model can be well determined at different resolutions and that larger-scale features of protein structure such as the overall fold or conformational rearrangements may be apparent at low resolution.

2. Theory

2.1. Geometric background

Classical differential geometry teaches us that sufficiently smooth space curves in three dimensions can be described in terms of two local quantities: the curvature \( \kappa \) which describes the bending of the curve and the torsion \( \tau \) which describes twisting of the curve out of the plane defined by its local bending. More precisely, we consider a curve \( \mathbf{r}(s) = (x(s), y(s), z(s)) \) parameterized by its arc length \( s \) (Gray, 1998), that is, the length along the curve between two points on the curves \( \mathbf{r}(s_1) \) and \( \mathbf{r}(s_2) \) is \( |s_2 - s_1| \). The distance between the same two points is \( |\mathbf{r}(s_2) - \mathbf{r}(s_1)| \) where \( |\mathbf{a}| \) represents the length of the vector \( \mathbf{a} \). We assume that the curve is regular and three times differentiable, that is, the first, second and third derivatives \( \mathbf{r}'(s), \mathbf{r}''(s) \) and \( \mathbf{r}'''(s) \) exist, and \( \mathbf{r}'(s) \) never vanishes. Then, the quantities \( \kappa \) and \( \tau \) are defined as

\[
\kappa = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3} \quad \text{and} \quad \tau = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{|\mathbf{r}' \times \mathbf{r}''|^2}. \tag{1}
\]

We refer to \( \kappa(s) \) and \( \tau(s) \) collectively as the curvatures. For a given curve, the curvatures can be obtained from the curve \( \mathbf{r}(s) \) by differentiation. Conversely, the curve \( \mathbf{r}(s) \) can be obtained, up to a rigid-body motion, from the curvatures by integration (Gray, 1998). The representation in terms of curvatures and the representation in terms of coordinates are equivalent. One advantage of a curvature-based representation is that it contains the intrinsic geometry of the curve rather than its orientation in space. For instance, a helix in space is simply defined by a constant curvature and torsion. To construct a curve \( \mathbf{r}(s) \) from its curvatures, we first introduce a local orthogonal coordinate system, the Frenet frame, attached at every point on the curve and defined in terms of derivatives of the curve. Following standard differential geometry, the Frenet frame consists of the tangent \( \mathbf{t}(s) \), the normal \( \mathbf{n}(s) \) and the binormal \( \mathbf{b}(s) \) vectors

\[
\mathbf{t}(s) = \frac{\mathbf{r}'(s)}{|\mathbf{r}'(s)|}, \quad \mathbf{n}(s) = \frac{\mathbf{t}'(s)}{|\mathbf{t}'(s)|}, \quad \mathbf{b}(s) = \mathbf{t} \times \mathbf{n}. \tag{2}
\]

The curvature and torsion specify the changes in the orientation of the Frenet frame as it moves along the curve. The vectors \( \mathbf{t}, \mathbf{n} \) and \( \mathbf{b} \) satisfy the Frenet equations:

\[
\begin{align*}
\mathbf{t}'(s) &= \mathbf{t}, \\
\mathbf{n}'(s) &= \kappa \mathbf{n}, \\
\mathbf{b}'(s) &= -\tau \mathbf{n}. \tag{3}
\end{align*}
\]

The curve \( \mathbf{r}(s) \) and its Frenet frame are obtained by solving this system of 12 differential equations for a given initial condition \( \mathbf{t}(0), \mathbf{n}(0), \mathbf{b}(0), \mathbf{r}(0) \) which fixes the rigid-body motion.

2.2. The Frenet–Fourier transform of a weighted curve

In crystallography, the structure factor of a body in space is defined as the Fourier transform of the body over the entire space. Explicitly, let \( \mu(x, y, z) \) be a density associated with a body at a point in the physical space \( x = (x, y, z) \). The structure factor \( F_k \) at a point \( \mathbf{k} = (k_x, k_y, k_z) \) in the Fourier space is defined as

\[
F_k = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y, z) \exp[i(\mathbf{k} \cdot \mathbf{x})] \, dx \, dy \, dz \tag{4}
\]

\[
\equiv F_k \mu = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mu(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) \, d\mathbf{x}, \tag{5}
\]

where we have introduced \( \mathcal{F}[\mu](\mathbf{k}) \) to denote the standard Fourier transform of a density \( \mu \). Now, assume that the object of interest is a weighted curve, that is, it is described by a curve in space \( \mathbf{r}(s) \) on which a line density \( \lambda(s) \) characterizing the variation in scattering weight along the curve is defined.\(^1\) This weighted curve is a body in space whose density is zero at any point \( \mathbf{x} \) not on the curve and equal to \( \lambda(s) \) for a point \( \mathbf{x} \) coinciding with the point \( \mathbf{r}(s) \). That is, the density of the body is

\(^1\) Formally, the spatial density \( \mu(\mathbf{x}) \) and the line density \( \lambda(s) \) are defined as tempered distributions.
with the initial condition 

$$\mu(x) = \int_0^L \lambda(s) \delta[x - r(s)] \, ds, \quad (6)$$

where $\delta(x) = \delta(x)\delta(y)\delta(z)$ is Dirac’s distribution in three dimensions. After substituting the density $\mu(x)$ in equation (4) and using the standard rules of integration for Dirac’s delta, the structure factor becomes

$$F(k) = (1/2\pi) \int_0^L \lambda(s) \exp[ik \cdot r(s)] \, ds. \quad (7)$$

The structure factor is now defined in terms of a curve $r(s)$ and a line density $\lambda(s)$. This is an interesting and central object in the theory of diffraction. Mathematically, we define $F(k)$ as the Frenet–Fourier transform of a weighted curve. This Frenet–Fourier transform takes two arguments, the line density $\lambda(s)$ and the curve $r(s)$, and yields a complex number, the value of the transform at $k$. It is denoted $F_\lambda[r](k)$ so that we can write

$$F[\mu] = F_\lambda[r](k). \quad (8)$$

Physically, the factor $\exp[ik \cdot r(s)]$ associated with the point $r(s)$ represents the phase of scattering relative to the origin. The transform is the accumulated scattering from all the points on the curve, taking the phase into account.

Further theoretical and numerical progress can be achieved if the curve is prescribed in terms of its curvatures. Then, the Frenet–Fourier transform can be coupled to the Frenet system of a curve at a point $k$ and can be evaluated by integrating the system

$$F_\lambda = \lambda(s) \exp[ik \cdot r],$$

$$r' = t,$n

$$t' = \kappa(s)n,$$n

$$n' = -\kappa(s)t + \tau(s)b,$$n

$$b' = -\tau(s)n,$$n

with the initial condition $t(0), n(0), b(0), r(0)$ and $F_\lambda(s = 0) = 0$ for all $k$.

The Frenet–Fourier system above provides an alternative method to obtain the Fourier transform in equation (4) which is appropriate for filamentary objects. The two formulations are mathematically equivalent. Equation (4) gives the transform by integrating the density distribution $\mu(x)$ over the entire space. This amounts to adding up the scattering from all parts of the object $\mu$. Equation (9) accomplishes the same result for filamentary objects by taking advantage of the fact that the density is zero everywhere except on the curve. The structure factor can then be obtained by starting at one end of $r(s)$ and adding up the scattering along the length of $r(s)$ until the other end is reached. Because the coordinate representation and curvature representations are equivalent, the Fourier transform defined through integration over coordinates may also be obtained through integration over curvatures.

![Figure 1](image-url)

Disc integration. Coordinate system for the angular and radial integrals in equation (12).

2.3. The Frenet–Fourier transform of a filamentary density distribution

In the previous section we assumed that a line density $\lambda(s)$ was defined on a curve. When considering the scattering of a filamentary object, we can model the density distribution of the body in terms of a curve and a distribution around this curve. Here we compute the value of the line density $\lambda(s)$ for a filament whose density distribution $\rho(s, R)$ is radially symmetric about a curve $r(s)$. Explicitly, we assume that the filamentary density distribution may be written as

$$\mu(x) = \int_0^L ds \int_0^{\infty} R dR \int_0^{2\pi} \rho(s, R) \times \delta[x - [r(s) + Ra(s)\cos \theta + Rb(s)\sin \theta]] \, d\theta, \quad (10)$$

where $(R, \theta)$ are the polar coordinates in the plane normal to the curve at a point $r(s)$ spanned by the normal and binormal vectors $a$ and $b$ defined in equation (3) (see Fig. 1). Inserting this expression for the density into equation (4) and integrating over the $(x, y, z)$ variables, we obtain

$$F(k) = (1/2\pi) \int_0^L ds \int_0^{\infty} R dR \int_0^{2\pi} \rho(s, R) \times \exp[ik \cdot r(s) + Ra(s)\cos \theta + Rb(s)\sin \theta]] \, d\theta \quad (11)$$

$$= (1/2\pi) \int_0^L ds \int_0^{\infty} R dR \int_0^{2\pi} \exp[ik \cdot (Ra(s)\cos \theta + Rb(s)\sin \theta)] \, d\theta. \quad (12)$$

The last integral can be written

$$I = \int_0^{2\pi} \exp[iR[C_a(s)\cos \theta + C_b(s)\sin \theta]] \, d\theta \quad (13)$$

with $C_a(s) = k \cdot n$ and $C_b(s) = k \cdot b$. It can be further simplified by the variable substitution $\theta = \varphi + A$ and standard trigonometric identities to obtain
Table 1

<table>
<thead>
<tr>
<th>( \rho(R) )</th>
<th>( \lambda(s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha R \leq R_{\text{max}} )</td>
<td>((\alpha R_{\text{max}}/D)J_1(RD))</td>
</tr>
<tr>
<td>( 0 \ &lt; \ R \leq R_{\text{max}} )</td>
<td>(\alpha/[\beta(1 + D^2/\beta^2)^{1/2}])</td>
</tr>
<tr>
<td>( \alpha \exp(-\beta R) )</td>
<td>((\alpha/\beta)\exp[-(D^2/4\beta)])</td>
</tr>
<tr>
<td>( \alpha/R )</td>
<td>(\alpha \text{ sign}(R/D))</td>
</tr>
</tbody>
</table>

\[
I = \int_{A}^{2\pi + A} \exp[iR \cos \varphi(C_1(s) \cos A + C_2(s) \sin A)] + iR \sin \varphi[C_2(s) \cos A - C_1(s) \sin A] \, d\varphi. \quad (14)
\]

We can now choose \( A \) such that \( \tan A = -C_1/C_2 \), so that

\[
I = \int_{A}^{2\pi + A} \exp(iRC \sin \varphi) \, d\varphi, \quad (15)
\]

where \( C = (C_1^2 + C_2^2)^{1/2} \). Since we are integrating a periodic function over its entire period, the integration bounds can be shifted arbitrarily and we have

\[
I = \int_{-\pi}^{\pi} \exp(iRC \sin \varphi) \, d\varphi = 2\pi I_0(RC). \quad (16)
\]

where we have identified the integral as the Bessel function \( I_0 \) of order 0. Substituting \( I \) back into equation (12), the structure factor now reads

\[
F(k) = \int_{0}^{L} \exp[i \cdot k \cdot r(s)] \, ds \int_{0}^{\infty} \rho(s, R) J_0(\rho R) \, dR. \quad (17)
\]

We recognize the last integral as the Hankel transform of \( \rho \),

\[
H_0[\rho] = \int_{0}^{\infty} \rho(s, R) J_0(\rho R) \, dR, \quad (18)
\]

and the structure factor becomes

\[
F(k) = \int_{0}^{L} H_0[\rho] \exp[i \cdot k \cdot r(s)] \, ds. \quad (19)
\]

We can now compare this last relation with the line density defined in equation (7) to obtain

\[
\lambda(s) = 2\pi H_0[\rho]. \quad (20)
\]

This establishes the general result that the Fourier transform of a filamentary density distribution \( \mu \) consisting of a radial density distribution \( \rho \) around a curve \( r \) is the Frenet–Fourier transform of the Hankel transform of \( \rho \) along the curve \( r \), that is

\[
\mathcal{F}[\mu] = \mathcal{F}_r[2\pi H_0[\rho], r]. \quad (21)
\]

Further progress can be accomplished by considering specific density models of filamentary structure, that is, by specifying \( \rho(s, R) \). As a first example, we assume a Gaussian cross section for the filament density, \( \rho(s, R) = \alpha(s) \exp[-(\beta(s)R)^2], \) and compute its Hankel transform

\[
\lambda(s) = \int_{0}^{\infty} \alpha \exp(-\beta R^2) J_0(\alpha R) R \, dR = \left[ \alpha(s)/2\beta(s) \right] \exp[-C^2(s)/4\beta(s)]. \quad (22)
\]

Now that \( \lambda(s) \) is known for each \( s \), the evaluation of the transform can be accomplished by solving equation (9). Note that the constant \( C(s) \) depends explicitly on the normal and binormal vectors and equation (9) becomes a system of nonlinear differential equations.

Some additional radial density distributions for which the weight function \( \lambda(s) \) can be explicitly evaluated (the so-called ‘Hankel pairs’) are given in Table 1.

3. Methods

3.1. The Frenet–Fourier transform of a helix

For any given curve, the Frenet–Fourier transform can be computed numerically by direct numerical integration of the Frenet–Fourier system. For a helical curve, it is well known that the Fourier transform can be obtained analytically (Cochran et al., 1952; Klug et al., 1958). This exact result serves as a convenient benchmark for comparison with numerical evaluation using equation (9) and we briefly summarize the derivation of the transform here. The classical result for the transform of a helix is obtained by using a parametric representation of the helix and evaluating equation (7) explicitly. Briefly, let \( h(z) \) be a helix of radius \( r \) and pitch \( P \) oriented with its axis along \( z \) parametrized as

\[
h(z) = (r \cos (2\pi z/P), r \sin (2\pi z/P), z). \quad (23)
\]

Then, the transform of this infinite curve is given by

\[
F_{\text{helix}}(R, \psi, k_z) = (1/2\pi) \sum_{n=-\infty}^{\infty} \delta(k_z - n/P) J_n(2\pi Rr) \times \exp[\imath n(\psi + \pi/2)], \quad (24)
\]

where \( (R, \psi, k_z) \) are the cylindrical coordinates of a point \( k \) in Fourier space. Owing to the periodicity along the axial direction, the transform of a helix is concentrated on layer lines at intervals proportional to \( 1/P \) in \( k_z \). The angular dependence on \( \psi \) contributes only in the phase factor \( \exp[\imath n(\psi + \pi/2)] \). The radial dependence on the \( n \)th layer line is given by the Bessel function \( J_n(2\pi Rr) \). In higher-order Bessel functions, the value of the argument at which the first maximum occurs becomes increasingly large, and so the magnitude of the transform is small near the axis, resulting in the characteristic ‘X’ shape in the diffraction pattern of a helix (Cochran et al., 1952). Therefore, from the exact theory, the radial intensity on layer line \( n \) is

\[
|F_{\text{helix}}(R, \psi, n/P)|^2 = \frac{1}{4\pi^2} |J_n(2\pi Rr)|^2 = \frac{1}{4\pi^2} J_n^2(2\pi Rr) \quad (25)
\]

because the phase contributes a factor of unity to the intensity.

To test the numerical approach, we compared the intensity \( |F_{\text{helix}}(k)|^2 \) on different layer lines with the value of \( |\mathcal{F}_r(k)|^2 \) obtained by evaluation of equation (9). For numerical calculations we employed a helix with \( \kappa = 0.38 \) and \( \tau = 0.15 \) which
values closely approximate the helical curve which passes through $C_\alpha$ positions in an $\alpha$-helix.

3.2. Simulation of low-resolution crystallographic data

As an application of the Frenet–Fourier method, we consider the problem of calculating crystallographic structure factors at low resolution using a filament with a Gaussian density cross section given by equation (22), with constant $a$ and $b$. For the purpose of illustration, we use a curve-defined density distribution representing a small three-helix-bundle protein structure: residues 5–49 of the E domain of protein A (Protein Data Bank code 1edi; Starovasnik et al., 1996). The three-helix-bundle model is shown in Fig. 2(a) and the corresponding curve model is shown in Fig. 2(b). This model was represented with the piecewise-constant curvature profiles shown in Fig. 3 following the method described by Hausrath and Goriely (2006, 2007).² A piecewise-constant curvature profile consists of a series of segments of variable lengths $L_i$ within which the curvature $\kappa_i$ and torsion $\tau_i$ are constant and is therefore characterized by a list $P = \{ (\kappa_i, \tau_i, L_i), i = 1, \ldots, N \}$. Note that, while the curvatures vary abruptly at the junction between segments, the curve itself is continuous. The Frenet frame at the end of one segment is inherited by the subsequent segment, ensuring continuity of the curve and the Frenet frame vectors. With this specification, the number of parameters needed to specify the curve is $3N$. In the example, we employ a list with seven segments, and so a total of 21 curvature parameters specifies this curve model (listed in Table 2). The corresponding coordinate model contains 351 atoms, and so requires $3 \times 351 = 1053$ $x$, $y$, and $z$ parameters.

Crystallographic structure factors were obtained by calculating the Fourier transform at reciprocal-lattice points using equation (9). Numerical integration was performed with a standard fourth-order explicit Runge–Kutta method. The particular values $k$ are determined by the unit-cell constants $a$, $b$ and $c$. For the simplest type of unit cell (space group $P1$, all cell angles $= 90^\circ$) the structure factors $F_{hkl}$ are obtained from equation (9) using

$$F_{hkl} = F \left[ \frac{2\pi h}{a} \wedge \frac{2\pi k}{b} \wedge \frac{2\pi l}{c} \right]. \quad (26)$$

² In previous work, we have used this piecewise-constant specification of curvature profiles to represent proteins and have developed fitting methods for determining piecewise-constant curvature profiles from coordinates which we have used in this case. However, the numerical approach is applicable to general curvature profiles and is not restricted to piecewise-constant profiles.
Table 2

<table>
<thead>
<tr>
<th>Segment</th>
<th>$\kappa$</th>
<th>$\tau$</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3800</td>
<td>0.1500</td>
<td>43.0487</td>
</tr>
<tr>
<td>2</td>
<td>0.2145</td>
<td>0.1180</td>
<td>15.6619</td>
</tr>
<tr>
<td>3</td>
<td>0.3755</td>
<td>-0.7419</td>
<td>11.2732</td>
</tr>
<tr>
<td>4</td>
<td>0.3800</td>
<td>0.1500</td>
<td>54.7950</td>
</tr>
<tr>
<td>5</td>
<td>0.0144</td>
<td>0.6796</td>
<td>4.0778</td>
</tr>
<tr>
<td>6</td>
<td>0.4880</td>
<td>0.0716</td>
<td>7.1421</td>
</tr>
<tr>
<td>7</td>
<td>0.3800</td>
<td>0.1500</td>
<td>46.0000</td>
</tr>
</tbody>
</table>

For comparison between the Gaussian filament- and coordinate-derived structure factors, models were placed in a simulated P1 unit cell with $a = 50.0$, $b = 60.0$, $c = 70.0$ Å, $\alpha = \beta = \gamma = 90.0^\circ$. Coordinate-model-derived structure factors $F_i$ were calculated using the CCP4 suite (Collaborative Computational Project, Number 4, 1994).

In order to use the Gaussian filament representation (Fig. 2c), the parameters $\alpha$ and $\beta$ must be given. Initial estimates for the best values to approximate a protein model at low resolution for these parameters were chosen by calculating structure factors with Miller indices from $-6$ to $6$ for a range of values of $\beta$ in increments of 0.001 up to 0.05, and in increments of 0.005 from 0.05 to 0.3. The agreement between the curve- and coordinate-model-derived structure factor sets was quantified using correlation coefficients. Since both amplitude and phase are obtained from equation (9), they can be correlated separately. The parameter $\alpha$ is a multiplicative constant and does not affect the correlation, but the choice of $\beta$, the radial width parameter of the density filament, has a significant effect.

Side chains vary in the number of atoms of which they are composed, so the scattering mass per residue varies along the chain. However, the backbone atoms contribute a constant scattering mass per residue. In contrast, the Gaussian filament density is constant along the direction of the curve. To investigate the importance of matching the linear density along the structure, we used four different coordinate models with different linear densities: a C$\alpha$ model, the backbone model and the all-atom model, comprising residues 5–49 of 1ed1. We also considered a C$\alpha$ model obtained by taking points exactly on the curve spaced 3.8 Å apart to assess the importance of centering the strongest density from the filament precisely on the atoms. The amplitude and phase of structure factors calculated from these four atomic models were compared with those obtained from filamentary models over a range of $\beta$ values.

At higher values of resolution, an alternative and more sensitive measure of the agreement between maps derived from the continuous and discrete representations was employed. Structure factor sets at variable resolution cut-offs up to 2 Å resolution were calculated for $\beta$ up to 0.1 in increments of 0.005 and then to 0.5 in increments of 0.05. The agreement between $F_i$ maps calculated from these structure factor sets and $F_i$ maps calculated from the coordinate models at resolutions of between 30 and 2 Å was quantified using the map correlation coefficient (calculated with MAPMAN; Kleywegt & Jones, 1996). Figures were made with Chimera (Pettersen et al., 2004), Molscript/Raster3D (Kraulis, 1991; Merritt & Bacon, 1997) and Mathematica (Wolfram Research, 2004).

4. Results

4.1. Comparison between the exact theory and numerical calculations for the transform of a helix

As a preliminary test of the Frenet–Fourier method, we compared numerical values of $|F_{helix}(k)|^2$ obtained with equation (9) with the exact values for $|F_{helix}(k)|^2$ obtained from equation (25).

Fig. 4 shows the radial intensity profile on selected layer lines from $n = 0$ to $n = 50$. The numerical approach quantitatively matches the radial profiles expected from helical diffraction theory. Overall, the agreement is quite good, although the accuracy of the numerical approach declines somewhat with increasing resolution. For example, on the

![Figure 4](image)

**Figure 4**

Radial layer line intensity for transform of a helix. For each layer line, the exact values of $|F_{helix}(k)|^2$ from the analytical theory (Cochran et al., 1952) are shown in black, and gray dots indicate the values of $|F_{helix}(k)|^2$ obtained by the numerical procedure [equation (9)]. Residual error $|F_{helix}(k)|^2 - |F_i|$ is shown in the panels to the right of each layer line. The panels (a) to (g) correspond to layer lines 0, 1, 2, 3, 10, 20 and 50.
layer line corresponding to \( n = 0 \), the maximum error is of the order of 0.01%. At the highest resolution tested, on the layer line corresponding to \( n = 50 \), the maximum error is about 0.3%. Note that the finite extent of the helix is not taken into account in the analytical formula and, therefore, perfect agreement is not expected. Nevertheless, this excellent quantitative agreement indicates that the numerical approach reproduces the exact theory for the known case of a single helix. We can now use the general Frenet–Fourier method on complex structures where exact solutions cannot be derived.

4.2. Comparison of curve-derived and coordinate-derived crystallographic structure factors

The conventional method for calculation of crystallographic structure factors requires construction of an electron-density map from a coordinate model, followed by Fourier transformation of the map to obtain the structure factors (Ten Eyck, 1977; Navaza, 2002). We compared the structure factors obtained from coordinate models in this manner with those obtained from solution of the Frenet–Fourier system of differential equations.

For direct comparison of structure factors at low resolution, we calculated the correlation between coordinate-derived and curve-derived structure factor sets. Because of the greater importance of the phase in producing a high-quality map, structure factor amplitudes and phases were correlated separately. As an example, scatter plots showing the correlations between the structure factors from the coordinate models and the Gaussian filament model with the particular value \( \beta = 0.15 \) are shown in Fig. 5. In this example, the amplitude correlations are similar but the more important phase correlations show greater variability. The general trend is that the phases obtained from equation (9) are most accurate for the less detailed models. The Co model derived from the curve shows a very strong correlation and so use of the continuous density of the filament results in very good phase estimates for this coordinate model. However, the majority of the phases are well approximated in each of these four cases, but with an increasing proportion of outliers as more detail is included in the models.

To determine the best value of \( \beta \) to use, such correlations were calculated over a range of \( \beta \) values. Overall, at up to 12 Å resolution the structure factor amplitudes (Fig. 6) using the Gaussian filament approximation are in excellent agreement with the correct values for all four models as assessed by correlation, achieving values in excess of 0.9 in all cases. Both amplitude and phase correlations rose rapidly from low values near \( \beta = 0 \) to a plateau above \( \beta \approx 0.05 \), beyond which there was little improvement in the range of \( \beta \) values tested. The accuracy of the phases was more strongly dependent on the resolution (see Fig. 6). However, the phase correlation is not as good for the detailed models. The Gaussian filament representation achieves phase correlations of about 0.7 for the main-chain model and about 0.55 for the all-atom model. An examination of the scatter plots in Fig. 5 suggests that the problem stems from a minority of points far from the diagonal in these graphs, but that the majority of points lie near the diagonal. In other words, a large number of phases are well approximated over the whole resolution range tested. An
alternative comparison less sensitive to such outliers (see below) is to compare the maps directly rather than to compare the structure factors.

The four different coordinate models have scattering mass with different radial distributions along the chain. Thus, one might expect that different values of $\beta$ would be appropriate for each, and that models would perform less well for values of $\beta$ away from the value best matching the radial density distribution of the particular model. However, comparison of the amplitudes and phases between structure factors shows that, at a given resolution beyond the threshold value of about 0.05, there is no strong dependence on the parameter $\beta$. The reason that the correlation holds up over this range is the unanticipated property of the Gaussian filament representation that it matches the coordinate model density by different mechanisms at different resolutions. There is some interplay between the width of the Gaussian distribution applied to the curve (the parameter $\beta$) and the resolution of the map. At low resolution, maps cannot display features of the original object below a certain size. Therefore, models that lack such fine features may agree well with low-resolution maps, but show poorer agreement when compared with maps at higher resolution. Fig. 7 illustrates the phenomenon using a series of maps calculated at increasing resolution with a tight Gaussian fila-

Figure 6
Structure factor amplitude and phase correlations. Correlation of structure factor amplitudes (left) and phases (right) from Gaussian curve model as a function of $\beta$, with $\sigma$ fixed at 1.0. (One individual comparison for $\beta = 0.15$ is shown in Fig. 5.) The traces correspond to the curve-derived $C_\alpha$ model (dotted line), the true $C_\alpha$ model (short-dashed line), the main-chain model (long-dashed line) and the all-atom model (solid line). From top, the graphs use data to 12, 15, 20, 25 and 30 Å resolution.
ment ($\beta = 0.4$) compared with a broad one ($\beta = 0.1$). Below 5 Å resolution the maps are very similar, but the differences become very apparent above this value.

This interplay is illustrated in Figs. 8 and 9. For example, at 6 Å, maps do not have sufficient resolution to delineate the chain location precisely, but the strongest features visible in both the coordinate- and the curve-derived maps are elongated cigar-shaped densities of the three $\alpha$-helices. At this resolution the density corresponding to these strong features arises from the ‘tails’ of the Gaussian of the filament model. By contrast, at 4 Å, the path of the chain is beginning to come into focus and the repeating density along the axis of the $\alpha$-helices is strongly apparent. In this range, the density of the curve model which is contributing to the agreement arises from the center of the Gaussian.

While visual comparison of the maps suggests how the agreement between coordinate and curve models is achieved, it is necessarily somewhat subjective and dependent on the choice of contour level. A more objective and quantitative criterion for the map agreement, which considers the map as a whole rather than the particular features, is the map correlation coefficient. Fig. 10 shows a series of graphs of the correlation coefficient between maps derived from the different coordinate models and Gaussian filament models as a function of $\beta$ and for a range of resolutions. The most important for eventual practical applications is the correlation with the all-atom model, as the others do not resemble experimental maps. While our three-helix-bundle test case is idealized by intent, the filament model can achieve excellent correlations with the atomic model (of the order of 70%) even at 4 Å resolution.

5. Discussion

We have introduced a filamentary representation of proteins based on space curves. This approach is directed towards studies at the level of the fold rather than seeking the finer details of side-chain interactions, which are not accessible at low resolution. The curve representation is a very parsimonious one: in favorable cases, the number of parameters necessary to specify a curve model can be a fraction of that needed for a conventional coordinate model. Our motivation is that such a curve model can be well determined from a data set with a fixed number of independent observations, where a coordinate model would be underdetermined.

The test case employed here is a three-helix-bundle curve model which requires 21 curvature parameters for its specification. The structure factors calculated from this model using equation (9) are compared with structure factors of four coordinate models with variable levels of detail. In general, the filamentary model can achieve good correlations between calculated structure factor amplitudes in all cases, but with a wider distribution of phase errors for the more detailed models (Fig. 5). However, the best metric for assessment of the effectiveness of the filament models is the real-space map correlation. The filament model achieves correlations of 0.896 at 6 Å resolution and 0.694 at 4 Å resolution with maps calculated from an all-atom model.

The use of an intrinsic geometric description represents a reduction in dimensionality. In general, an arbitrary three-dimensional density distribution requires a three-dimensional integration. However, the Fourier transform of these filamentary density distributions is evaluated in one dimension, along the arc length. The essential idea is that specification of

![Figure 7](image1.png)

**Figure 7**

Effect of resolution and $\beta$ on structural detail present in maps. Surface representations of maps calculated at 15, 12, 10, 9, 8, 7, 6, 5, 4, 3 and 2 Å resolution (from left to right) for $\beta = 0.4$ (top) and $\beta = 0.1$ (below).

![Figure 8](image2.png)

**Figure 8**

Basis for map agreement at 6 Å resolution. Maps calculated at 6 Å resolution from (a) the all-atom model and (b) the Gaussian filament model (with parameter values $\alpha = 1.0, \beta = 0.3$). The map correlation coefficient between these two maps is 0.946. At this resolution the cigar-shaped density features representing the $\alpha$-helices arise from the overlap of the tails of the radial density from nearby parts of the curve, as illustrated schematically in (c). The curve is shown in blue, and the true Ca trace is in red.

![Figure 9](image3.png)

**Figure 9**

Basis for map agreement at 4 Å resolution. Maps calculated at 4 Å resolution from (a) the all-atom model (blue) and (b) the main-chain model (red) have a correlation coefficient of 0.751, because a significant portion of the density arises from the main-chain atoms which are common to both models. [In (c) the two maps from (a) and (b) are shown together.] The green map in (d) calculated from the Gaussian filament model (using parameter values $\alpha = 1.0, \beta = 0.3$) visually resembles the map from the main-chain model at this resolution. The two maps have a correlation coefficient of 0.804 and this agreement is largely due to the central portion of the Gaussian (rather than the tails as in Fig. 8). Therefore, despite its simplicity, the filament model quantitatively represents the main-chain density with good accuracy and thus accurately describes a large part of the density from the all-atom model. The correlation between the map from the model and the map from the all-atom model is 0.694.
the defining curve at one initial point, in conjunction with knowledge of its local geometry along its length, suffices to reconstruct it in its entirety. Therefore, quantities derived from the curve (including the Fourier transform) are also accessible from this local description.

Approximation of the true electron density of a molecule by simpler representations has a long history in structural biology. The current work builds on the classical theory of diffraction by helices (Cochran et al., 1952; Crick, 1953; Klug et al., 1958; Waser, 1955; Cormack, 1957; Benham, 1981; Metoz & Wade, 1997) (see Fig. 4) to derive a more general theory for the diffraction by curves. Our general result in equation (21) indicates that the scattering from a density distribution radially symmetric about a curve is equivalent to the scattering from the curve itself, appropriately weighted. Therefore, the transforms of filamentary distributions may be evaluated using this theory of diffraction by a curve. Employing such a filamentary approximation to the electron density may also be considered an extrapolation of the ‘globbic approximation’ (Harker, 1953; Guo et al., 1999) to a continuum.

The evaluation of the transform using equation (9) requires consideration of the efficiency and precision of the numerical solution technique that is employed. In particular, at high resolution the rapid oscillations in the phase factor \( \exp(ik \cdot r) \) may require special treatment (Iserles, 2004). Note also that the use of a constant radial density distribution as used here may be generalized further: the numerical method can accommodate more complicated weights \( \lambda(s) \) derived from radial density distributions that vary along the curve. The method may also have applications in other crystallographic problems such as low-resolution refinement, phase extension or de novo phasing. A similar geometric approach may be developed for other techniques such as small-angle scattering and fiber diffraction.

A curvature-based description is capable of expressing very complex three-dimensional forms, and we have made use of this formalism for a continuous curve representation of protein structure (Hausrath & Goriely, 2006, 2007). The current work provides the foundation for use of the continuous representation to solve a practical structural biology problem. As a curvature description is fundamentally equivalent to a coordinate-based description, we suggest that it may be useful to regard the relationship between coordinate and curvature representations as analogous to the relationship between real- and reciprocal-space representations. Just as some problems in diffraction are more naturally addressed in reciprocal space rather than in real space, it may also prove that other problems are more conveniently solved in curvature space (Fig. 11).

This material is based in part upon work supported by the National Science Foundation under grant Nos. DMS-0604704 and DMS-IGMS-0623989 (AG).

References