Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 7 January 2003
Accepted 28 March 2003

Local squaring functions for non-spherical templates

Jeffrey Roach* and Charles Carter Jr

Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599, USA. Correspondence e-mail: roachjm@k.ro

The local squaring functions representing the likelihood that a particular atom type occupies a particular location in the unit cell are generalized so that the likelihood that a multiatom fragment occurs at a given position can be modeled at less than atomic resolution. The orientation space of the fragment is parameterized according to the *special unitary* 2-group in order to facilitate both interpretation and interpolation. The representation theory of this group is related to the harmonic analysis for functions of rotation. As an example of computer-aided model construction, local squaring functions for several protein backbone fragments are applied to a 2.0 Å model electron density. The resulting functions are sufficient to reconstruct the atomic structure from the probable template placements.

© 2003 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

Local squaring functions (Roach & Carter, 2002) are a formulation of atomic shape constraints more closely resembling solvent-flattening formulations (Wang, 1985; Xiang *et. al.* 1993) than more traditional convolutional structure-factor equations (Sayre, 1952; Woolfson, 1958; Rothbauer, 2000). A *local squaring function* for a particular atomic form measures the likelihood that an atom of the given atomic shape occupies a particular location in the unit cell. In analogy to the Sayre *squaring method equations*, these functions can be constructed from experimentally derived structure factors, amplitude or phase, and structure factors biased by some initial model. Furthermore, their closed-form expression, in terms of convolution, can be calculated efficiently using the fast Fourier transform.

With a combination of several local squaring functions, the entire unit cell is subjected to a probabilistic filter that dampens density in regions unlikely to contain atoms of any of the given forms, while enforcing atomic shapes at locations likely to contain atoms, thereby lifting the requirement for equal atom types. By enabling effective density modification within the molecular envelope, the method extends the solvent-flattening approach throughout the unit cell. Additionally, an atomic model can be extracted from the likely atomic locations expressed in the local squaring functions. Refinements using spherical templates have been generally successful with small protein structures with a high-resolution set of either experimental or model-biased phases (Roach & Carter, 2002).

Effective phase refinement based on local squaring functions suggests that atomic form constraints possess significant potential. Unlocking this potential, however, has been an elusive goal. The impact of atomic form constraints has never been as impressive as other techniques, *e.g.* solvent flattening and more general density modification (Terwilliger, 2000). Furthermore, atomic form constraints for general atom types have never figured prominently in an established phase-refinement procedure. Although this has been due, in part, to limited phase-refinement effectiveness, the requirement of high-resolution data has also severely restricted their scope.

Even at high resolution (0.9 Å), some atoms are not sufficiently separated to be identified individually by the local squaring functions of spherical templates. This was first observed in the carbonyl group of the protein backbone. In the 0.9 Å resolution initial electron density of the C-terminal Kunitz-type domain of the $\alpha 3$ chain of human type VI collagen (Arnoux *et al.*, 2002), our atomic local squaring function investigations could identify only one atom of the carbonyl pair in 55 of 58 residues with significant density. In no residue could both carbon and oxygen be separated and in many cases the atomic placement suggested by the minimum of the local squaring function fell between the two atomic centers.

To illustrate how this is possible, consider the model-calculated electron density of the neuronal acetylcholine receptor antagonist, α-conotoxin PnIA. Containing only 16 residues, this very small protein is a potent toxin derived from cone snails, *Conus pennaceus*. The X-ray structure determination revealed a compact and intricate arrangement of a β-turn followed by two α-helical turns all stabilized by two disulfide bridges (Hu *et al.*, 1996).

Fig. 1 shows a 1.0 Å electron density map at two different contour levels: 3σ (*a*) and 1σ (*b*). Note that, even at relatively high contour levels, 3σ, the individual atoms of the carbonyl group remain joined. Furthermore, comparison of the 1σ level to the 3σ level shows that a significant portion of the electron density clearly takes a value lower than that required to separate most atoms.

It is important to note that this example shows a highly sharpened electron density calculated from a model whose atoms have average temperature factor less than 8 Å^2. Thus, even in ideal conditions, there is not sufficient separation in the carbonyl group to treat each individual atom as a completely independent entity.

Related problems are shared by all direct-methods approaches that are predicated on the assumption that atoms are placed independently (Bricogne, 1997*b*). In reality, of course, the atomic locations are highly dependent upon one another regardless of resolution; however, at high resolution this independence approximation is less drastic. Coordinate-refinement techniques are also compromised by limited resolution. For example, the real-space refinement method of Chapman (1995) attempts to compensate for the absence of high-resolution data with specially chosen atomic scattering factors.

Needless to say, electron-density maps typically encountered in macromolecular crystallography are not this sharply defined. In most cases, even with noise and other imperfections suppressed, taking the atoms to be completely independent entities remains a crude approximation. Therefore, stereochemical interactions must be taken into account. In the case of local squaring functions, overcoming the lack of separation at lower resolution requires the implementation of functions that recognize multiple atom fragments as well as individual atoms.

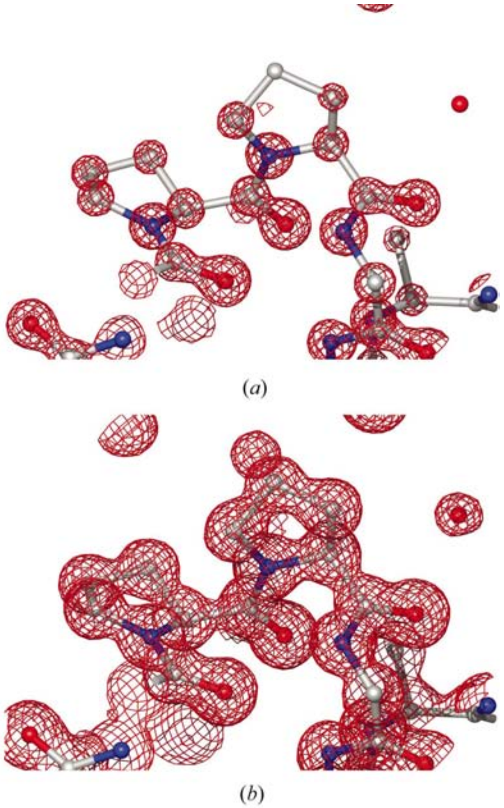

Figure 1
Electron density $\{|F_c|, \varphi_c\}$ at 1.0 Å resolution calculated from published atomic coordinates for α-conotoxin PnIA. (*a*) Contoured at the 3σ level. Carbonyl pairs remain unseparated. (*b*) Contoured at the 1σ level. A significant portion of the electron density occurs at levels where most atoms are unresolved.

2. Local squaring functions: translation

Recall the definition of the local squaring function (Roach & Carter, 2002). For any particular template density ρ_0,

$$O(\rho, y) = \int_V |\rho(x)\rho_0(x-y) - \rho_0^2(x-y)|^2\,\mathrm{d}x^3,$$

where V denotes the unit cell. Although it has been previously assumed that ρ_0 represents a spherical atomic form, in the above definition this is not strictly required. Let $\rho_0(x, \vartheta)$ denote a non-spherical density with some fixed orientation ϑ. At this point, the representation of this orientation is not necessary; however, it will be discussed in detail shortly. Given this template density,

$$O(\rho, \vartheta, y) = \int_V |\rho(x)\rho_0(x-y, \vartheta) - \rho_0^2(x-y, \vartheta)|^2\,\mathrm{d}x^3$$

models the likelihood that a particular template ρ_0 occurs with orientation ϑ at location y. For each template, the local squaring function corresponding to that template is a function of both location and orientation. Therefore it will have a translational Fourier expansion in terms of location for any fixed orientation and a rotational Fourier expansion in terms of orientation for a fixed location.

Determining the translational Fourier coefficients, *i.e.* structure factors, proceeds analogously to the determination for spherical templates. To simplify the notation, let ρ_0 denote some template density with fixed orientation ϑ. Assuming the electron density ρ takes only real values, consider the expansion of $O(\rho, \vartheta, y)$:

$$O(\rho, \vartheta, y) = \int_V [\rho^2(x)\rho_0^2(x-y) - 2\rho(x)\rho_0^3(x-y)]\,\mathrm{d}x^3 + K,$$

where the constant K is given by

$$K = \int_V \rho_0^4(x-y, \vartheta)\,\mathrm{d}x^3.$$

With the asterisk signifying complex conjugation and $|V|$ denoting the volume of the unit cell, the structure factors of the local squaring function are given by

$$\hat{O}_h = [(f_h^{(2)})^*/|V|]\sum_k F_k F_{h-k} - 2F_h(f_h^{(3)})^* + \delta(h)K, \qquad (1)$$

where $f_h^{(n)}$ denotes the scattering factor of ρ_0^n and $\delta(h) = 0$ for all $h \neq 0$ and $\delta(0) = 1$.

The methods necessary to construct the rotational Fourier expansion are somewhat less well known. In the sequel, the problem of generating a Fourier series with respect to orientation for a local squaring function of fixed location is discussed in detail.

3. Functions of rotations

Consider the rotation function familiar from molecular replacement. In order to determine the correct orientation of a molecular fragment known to be in the unknown structure, the Patterson function of the known fragment P_s with orientation ϑ is compared to the observed Patterson P_t. The rotation function R is given by

$$R(\vartheta) = (1/|V|) \int\limits_{\Omega} P_t(r) P_s(\vartheta^{-1} r)\, \mathrm{d}r^3$$

(Rossmann & Blow, 1962; Navaza, 2001). The interpretation of the rotation function is that rotations ϑ with large values of R correspond to orientations of the known fragment that coincide with orientations of similar fragments in the unknown structure. Therefore, the rotation function compares Patterson functions to identify likely orientations of the known fragment within an unknown structure.

For a fixed location, the local squaring function compares the electron density of a known fragment to an electron density derived from initial phase estimates. By considering all possible orientations of the fragment, the local squaring function indicates orientations that the fragment is likely to take at a given location. Note, however, that large values of the rotation function correspond to likely orientations in contrast to the local squaring function where values near zero denote probable orientations at a given location. The need in both cases to consider the set of all possible rotations suggests that an approach to the orientation component of the local squaring function could proceed analogously to the construction of the rotation function.

Calculating a rotation function requires parameterizing and sampling of the space of rotations. It is necessary to sample the parameterization appropriately to produce a uniform sampling of the rotation space. For example, given the Euler-angle parameterization of the rotation group, sampling the set of all possible rotations uniformly is not the same as sampling the set of all possible Euler angles uniformly. This approach, the fast rotation function formulation of Crowther (1972), does however have the advantage of being highly suitable for calculation: reducing to a standard two-dimensional Fourier transform.

Unfortunately, the resulting function is distorted because rotations that have similar parameters need not correspond to similar rotations. In fact, the metric on the rotation group cannot be reduced to a Cartesian metric (Navaza, 2001). As a result, the rotation function can be difficult to interpret. Some modification can be made to reduce these difficulties, for example: reformulation in polar coordinates (Tanaka, 1977) or considering distortion-free two-dimensional sections (Burdina, 1971; Lattman, 1972). Nonetheless, the distortion is due to topological properties of the rotation group and can never be completely eliminated. Details of the relationship between the fast rotation function and the topology of the rotation group are given by Navaza (2001); however, the important point that needs to be emphasized is the development of the spherical harmonics and the role of the irreducible representations of the rotation group in the fast rotation function.

For the orientation component of the local squaring function, a different parameterization of the orientation space will be considered. This parameterization has the advantage of being topologically equivalent to a sphere in four dimensions. Therefore, sampling, interpretation and interpolation follow from a familiar three-dimensional analog: a globe.

3.1. Classical linear groups

Before the orientation parameterization can be described in detail, it is necessary to relate some of the theory of classical linear groups (Zhelobenko, 1973), *i.e.* groups of linear transformations. For example, the rotation group of the previous section corresponds to the group of orthogonal linear transformations fixing the origin that preserve chirality. In three Euclidean dimensions, this group is referred to as the *special orthogonal group* of dimension three over the real numbers (SO3). Another example would be the group of linear transformations preserving volume, the *special linear group*. The term special refers to the fact that the matrix representation of elements of these groups have determinant one. Often it is convenient to identify the matrix representation with the group itself. For example, typically it is sufficient to refer to the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\vartheta) & \sin(\vartheta) \\ 0 & -\sin(\vartheta) & \cos(\vartheta) \end{bmatrix}$$

as a rotation of ϑ about the x axis. Therefore, in terms of matrix representations, the rotation group or SO3 corresponds to 3×3 real orthogonal matrices of determinant one.

The group of interest in connection with the spherical parameterization of the orientation space is the *special unitary group* of two complex dimensions (SU2). The matrix representation of SU2 consists of 2×2 complex unitary matrices of determinant one. Every element of SU2 is given by two complex numbers, a and b, as:

$$\begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix},$$

where $|a|^2 + |b|^2 = 1$. If we let $a = x + iy$ and $b = z + it$, this last condition implies that $x^2 + y^2 + z^2 + t^2 = 1$. Consequently, SU2 is topologically equivalent to the surface of a sphere in four (real) dimensions.

The special unitary group (SU2) is pertinent to the local squaring function because each rotation, or element of SO3, corresponds to two diametrically opposite elements of SU2. Explicitly, the elements given by a, b and $-a$, $-b$ both correspond to:

$$\begin{bmatrix} |a|^2 - |b|^2 & 2\Re(a^* b) & 2\Im(a^* b) \\ -2\Re(ab) & \Re(a^2 - b^2) & -\Im(a^2 + b^2) \\ -2\Im(ab) & \Im(a^2 - b^2) & \Re(a^2 + b^2) \end{bmatrix}. \qquad (2)$$

It can be easily checked that (2) is orthogonal and has determinant one when $|a|^2 + |b|^2 = 1$. Mathematically, this situation is described as SU2 being *locally isomorphic* to SO3 and *covering* it twice (Zhelobenko, 1973, p. 35). The advantage of this parameterization is that the topology of the space is reasonably familiar: a sphere in four (real) dimensions. Therefore, the difference between two rotations corresponds to the great circle distance on the sphere, facilitating the interpretation and interpolation of a function of rotations.

These advantages have made use of the special unitary group common in computer animation (Shoemake, 1985), molecular modeling (Leach, 1996) and Bayesian molecular replacement (Bricogne, 1997*a*; Morris *et al.*, 2001). In some cases, elements of SU2 are called unit *quaternions*, referring to elements of the *quaternion algebra* with magnitude one (Zhelobenko, 1973, p. 33).

3.2. Fourier expansions

Periodicity is a familiar concept in crystallography: periodicity allows the electron density to be expressed in terms of a Fourier series, *i.e.* using structure factors. This illustrates a fundamental result in the theory of group representations: the use of group symmetry or periodicity with respect to some group in developing a linear basis for functions exhibiting that symmetry. The familiar case uses the group symmetry of the translation group to express the electron density as a series of linearly independent functions. We will see that the particular symmetry of the rotation group leads to the spherical harmonic expression of the fast rotation function. In order to exploit this idea for the local squaring function, it will be necessary to describe the abstract Fourier transform on the special unitary group SU2.

Most of the technical details are given in the references (Zhelobenko, 1973; Kirillov, 1994) and the theory of Fourier transforms on linear groups more or less parallels the theory of Fourier series. It is first necessary to determine the irreducible representations of SU2 then an appropriate Fourier series can be assembled from them.

A *representation* of a group corresponds to the action of the group on some linear space. For example, consider the homogeneous polynomials of degree N in indeterminants u, v with complex coefficients. As a vector space V^N, this set has a basis given by $\{u^n v^{(N-n)} | n = 0, \ldots, N\}$. To be concrete, choose degree 2. The vector space of homogeneous degree 2 polynomials, V^2, has basis u^2, uv and v^2. That is, each homogeneous degree 2 polynomial can be expressed as

$$P = \alpha u^2 + \beta uv + \gamma v^2 \quad \alpha, \beta, \gamma \in \mathbf{C}.$$

Now let g be some element of SU2, say

$$g = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix},$$

where a and b are complex with $|a|^2 + |b|^2 = 1$. Consider the following action of g on the vector space V^N:

$$\begin{aligned}[T_N(g)P](u, v) &= P\left(\begin{bmatrix} u & v \end{bmatrix}\begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}\right) \\ &= P(au - b^*v, bu + a^*v). \qquad (3)\end{aligned}$$

For the vector space V^2, $T_2(g)$ is a linear transformation and therefore has a matrix representation. In this case, with respect to the basis u^2, uv and v^2, $T_2(g)$ takes the form

$$\begin{bmatrix} a^2 & ab & b^2 \\ -2ab^* & |a|^2 - |b|^2 & 2a^*b \\ (b^*)^2 & -a^*b^* & (a^*)^2 \end{bmatrix}$$

for g an element of SU2. An *irreducible representation* V preserves only the trivial subspaces $\{0\}$ and V. Equation (3) defines the complete set of irreducible representations for SU2 (Zhelobenko, 1973, p. 93). That is, there is one irreducible representation of degree $n+1$ for each non-negative integer n.

For a wide class of functions, specifically those that are square-integrable under some appropriate measure, the irreducible representations define a series expansion analogous to the Fourier series. Within this class, a function $f(g)$ on SU2 can be expressed as

$$f(g) = \sum_{n=0}^{\infty}(n+1)\, Tr\big(\hat{f}(n)T_n(g)\big), \qquad (4)$$

where Tr denotes the matrix trace. The 'Fourier coefficient' $\hat{f}(n)$ is an $(n+1) \times (n+1)$ complex matrix given by

$$\hat{f}(n) = \int f(g)T_n(g^{-1})\,\mathrm{d}g, \qquad (5)$$

where the integral is taken over the entire group SU2 (Zhelobenko, 1973, p. 74; Sugiura, 1975, p. 76).

4. Local squaring functions: rotation of a general fragment

For a fragment possessing no symmetry, every possible rotation corresponds to a unique orientation of the fragment. Therefore, parameterizing the orientation space corresponds exactly to parameterizing the space of rotations. Parameterizing the rotation space for local squaring functions does not require the full generality of (4) and (5). In particular, since the local squaring functions take only real values, an analogous result to Friedel's law reduces the number of independent entries in the Fourier coefficient matrices of (5). Furthermore, the local squaring functions are functions of rotations and not of general elements of SU2. Since a rotation corresponds to two elements of SU2, the Fourier expansion must take the same value on both elements of SU2 corresponding to the same rotation.

The double covering of rotations is apparent in the Fourier series for functions of SU2. The Fourier coefficients for even values of n can be expressed using only real numbers. This fact can be deduced either abstractly (Kirillov, 1994, p. 68) or by choosing an appropriate basis. For example, using the bases uv, $-i(u^2 + v^2)/2$ and $(v^2 - u^2)/2$, the irreducible representation T_2 takes only real values and is equal to the matrix (2) converting elements of SU2 to rotations. The Fourier coefficients for odd n are quaternion representations and are nonzero for functions that take different values for elements of SU2 corresponding to the same rotation. This observation, reflecting the relationship between SU2 and the rotation group SO3, leads to a derivation of the spherical harmonics from the irreducible representations of SU2 (Zhelobenko, 1973, p. 100). Therefore, given that the local squaring functions take only real values and that they take the same value for diametrically opposite elements of SU2, the SU2-Fourier expansion of a local squaring function contains only the even terms, $2n$, and each of the even Fourier coefficients is a real $(2n+1) \times (2n+1)$ matrix.

Even with these reductions, constructing and minimizing an extensive rotational Fourier expansion of the local squaring function for a particular fragment at each point in the unit cell is too costly to be an option. The simplest approach would be to calculate the local squaring function using equation (1) for a large number of orientations sampled uniformly throughout the orientation space. Choosing the orientation taking the minimal value gives an approximation to the optimal orientation. The quality of this approximation, of course, depends on the sampling density; therefore, this involves a trade-off between accuracy and efficiency.

4.1. Global interpolation

A more sophisticated approach is analogous to the use of Hendrickson–Lattman coefficients. Hendrickson–Lattman coefficients are the first four Fourier coefficients of the phase probability distribution function. Thus, a phase estimate can be generated by maximizing a short Fourier series determined by the Hendrickson–Lattman coefficients. Analogously, an estimate of the best orientation can be determined by minimizing a short rotational Fourier expansion of the local squaring function. In this case, the accuracy of the estimate will be determined by how well the true function is globally approximated by the short Fourier series.

Consider a two-term Fourier expansion:

$$f(g) = \hat{f}(0) + Tr\big(\hat{f}(2)T_2(g)\big),$$

where g is an element of SU2 corresponding to the orientation of the fragment and the Fourier coefficients, $\hat{f}(0)$ and $\hat{f}(2)$, are 1×1 and 3×3 real matrices, respectively. For g given by $x + iy$ and $z + it$, the series takes the form of a quadratic polynomial in x, y, z and t, where $x^2 + y^2 + z^2 + t^2 = 1$:

$$\begin{aligned} f(x, y, z, t) = {} & A_0 + A_1(x^2 + y^2) + A_2(x^2 + z^2) + A_3(x^2 + t^2) \\ & + A_4xy + A_5xz + A_6xt + A_7yz + A_8yt + A_9zt. \end{aligned}$$

Once the ten real parameters have been determined, either by least-squares interpolation or by some other means, the function can be maximized. Since the solution is constrained to the sphere of unit radius, the optimization requires the method of Lagrange multipliers. That is:

$$f(x, y, z, t) - \lambda(x^2 + y^2 + z^2 + t^2 - 1)$$

must be taken to a maximum. In this particular case, where the objective function f is quadratic with no linear terms and the constraint is homogeneous of degree two, maximizing the Lagrangian reduces to determining the eigenvectors and eigenvalues of the Hessian H of f. Each eigenvalue corresponds to a value of the Lagrange multiplier, λ, and the normalized eigenvector corresponding to it represents an orientation where f takes an extreme value. Therefore, the best estimate can be obtained by testing the value of f for each eigenvector. Since the Hessian is only 4×4 and consequently there are at most four possible extrema, this calculation can be performed at each point of the unit cell within a reasonable time frame.

4.2. Local polynomial interpolation

A second approach combining the sampling technique of the simple approach with the interpolation aspects of the short Fourier-series approach would be to construct a polynomial function that locally approximates the true function in the region surrounding the minimal sampled value.

In order to construct a local polynomial approximation, it is necessary to establish a set of local coordinates for the 4-sphere. These local coordinates are analogous to using polar coordinates to describe a point on the surface of a unit sphere in three dimensions or a phase to describe a point on the unit circle in two dimensions. Given that

$$x^2 + y^2 + z^2 + t^2 = 1,$$

there exists some value of ϑ_1 such that $x = \cos\vartheta_1$. Thus we have:

$$y^2 + z^2 + t^2 = \sin^2\vartheta_1.$$

Consequently, except for the circle where $z^2 + t^2 = 0$, the SU2 can be represented by the local coordinates ϑ_1, ϑ_2, ϑ_3 such that:

$$\begin{aligned} x &= \cos\vartheta_1 & y &= \sin\vartheta_1\cos\vartheta_2 \\ z &= \sin\vartheta_1\sin\vartheta_2\cos\vartheta_3 & t &= \sin\vartheta_1\sin\vartheta_2\sin\vartheta_3 \end{aligned}$$

for $0 \le \vartheta_1 \le 2\pi$, $0 \le \vartheta_2 \le \pi$ and $0 \le \vartheta_3 \le 2\pi$ (Sugiura, 1975, p. 56). It is interesting to note that another choice of local coordinates, the Euler angles, leads to the spherical harmonic expansion for functions of the sphere in three dimensions (Zhelobenko, 1973, p. 100).

Once a set of local coordinates is chosen, a quadratic polynomial in the three variables can be interpolated from the sample points surrounding the purported optimum. A quadratic polynomial is particularly suitable because its optimal value, if it exists, can be easily calculated. The difficulty in this technique lies in the construction of an initial orientation sample that covers enough of the orientation space to identify the general region where an optimum lies yet is still fine enough to construct a local function that accurately represents the function in that region.

5. Local squaring functions: rotation of a symmetric fragment

When the template density possesses additional symmetry, it is no longer the case that the set of all possible orientations is parameterized by the set of rotations given by SU2 or SO3. Consider a molecular fragment consisting of spherical atoms representing the carbon–oxygen pair in a carbonyl group. The orientation of this fragment is fixed by any rotation about the axis formed by the center of the two atoms. Centering the carbonyl fragment on the carbon atom, it is easy to see that the complete set of orientations is given by the location of the oxygen atom. That is, the space of orientations is parameterized by a sphere in three dimensions as opposed to a sphere in four dimensions.

Consider a molecular fragment that is fixed by any rotation of the x axis. Mathematically, this corresponds to a subgroup K of SU2 given by elements of the form

$$g = \begin{bmatrix} a & 0 \\ 0 & a^* \end{bmatrix}.$$

If $a = \exp(i\vartheta)$, the above matrix represents a rotation of 2ϑ about the x axis. Unfortunately, the subgroup K is not normal in SU2; it is, nonetheless, possible to describe an orthonormal basis for the square-integral functions on the cosets G/K, *i.e.* those functions that remain fixed by K.

It is possible to determine equations that the Fourier coefficients of a function on G/K must satisfy, *viz*:

$$\begin{aligned} f(g) &= \sum_{n=0}^{\infty}(n+1)\,Tr\big(\hat{f}(n)T_n(g)\big) \\ &= \sum_{n=0}^{\infty}(n+1)\,Tr\big(\hat{f}(n)T_n(kg)\big) = f(kg) \qquad (6) \end{aligned}$$

for all $k \in K$. This method is laborious but intuitive and it can be shown that the Fourier coefficient $\hat{f}(n)$ is zero for all columns except the first.

An alternative approach is available using the theory of spherical functions (Takeuchi, 1994) or homogeneous spaces (Zhelobenko, 1973). Let B_1 be some basis element of the irreducible representation V^{2N} that is fixed by the subgroup K. In this particular case, the basis element $u^N v^N$ is invariant with respect to K. The set of functions R_i given by

$$R_i(gK) = \langle T_{2N}(g)B_1, B_i\rangle,$$

where $\langle , \rangle$ is a SU2-invariant Hermitian form (*c.f.* Zhelobenko, 1973; Kirillov, 1994), is an orthogonal basis for the square-integrable functions that are invariant under the action of K. This basis corresponds to the first row of the matrix for the irreducible representation T^{2N}. Therefore, the same class of functions is specified by the intuitive approach (6). For T^2,

$$R_1 = x^2 + y^2 - z^2 - t^2, \quad R_2 = 2(xz + yt), \quad R_3 = 2(xt - yz).$$

Note that $R_1^2 + R_2^2 + R_3^2 = 1$, thus R_1, R_2 and R_3 specify a sphere in three dimensions.

For a template with rotational symmetry about any particular axis, the local squaring function can be expanded as a series with basis functions given by (7). Unlike the general case, construction and minimization of a short Fourier series does not reduce conveniently to an eigenvector decomposition. The local approximation, however, proceeds analogously to the general case requiring only two local parameters to describe the orientation space. Furthermore, because one less dimension need be sampled, fewer sample points are necessary to achieve reasonable accuracy.

6. An example

As a concluding example, we return to α-conotoxin to show how the local squaring functions can be used to construct an initial atomic model. We considered the local squaring functions associated with three different backbone fragments: a carbonyl group, an α-carbon-to-α-carbon *trans*-peptide link, and a nitrogen-to-carbon backbone segment. Coordinates for these molecular fragments were taken from Prince *et al.* (1992) and centered on the carbon in the carbonyl pair.

A sampling of the orientation space for each fragment was constructed in the following fashion. A set of 800 points was built by combining several uniform samplings of the sphere (Sloane *et al.*, 2000). Then, for each point in a set of 40 points chosen to maximize the minimal distance between points (Sloane *et al.*, 2000), the 20 closest neighbors from the 800 element set were added. The resulting sample set consisted of 400 unique sample points on the sphere in four dimensions and 383 sample points on the sphere in three dimensions.

Local squaring functions for each fragment and each orientation were constructed using equation (1) and applied to

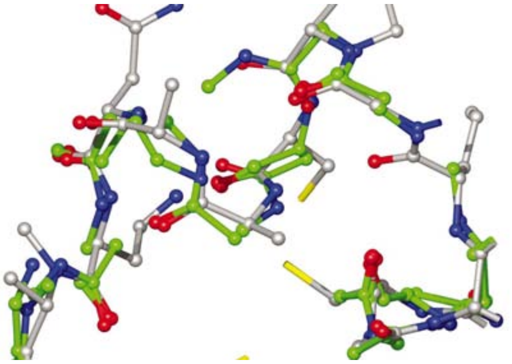

Figure 2
Model determined by the sampled local squaring function method (green) superimposed on the correct model (grey). Template orientations are not perfect; however, most of the backbone is reconstructed accurately.

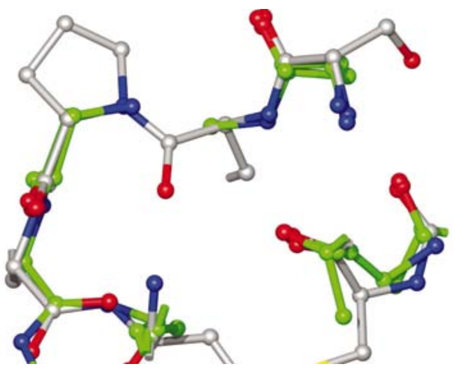

Figure 3
Each template is centered on the carbonyl carbon atom. In this case, no template is placed by the local squaring function at the carbon. However, templates correctly placed on the adjoining residues cover most of this residue.

a 2.0 Å electron density calculated from the published conotoxin model. At each point in the unit cell, a local polynomial approximation was interpolated from the 15 orientations nearest to the orientation with the minimum local squaring function value. In the case of the two larger fragments, a short Fourier approximation was interpolated.

The short Fourier-series approximation gave some impression of the rotational component of the local squaring function at a given point. However, as a method of determining optimal orientation, it was largely unreliable. The local polynomial approximation was more effective, resulting in optimal orientations for several fragments. In the cases where the local polynomial did not specify a minimum, the optimal orientation was chosen from the sampled orientations.

Inspection of the rotational component of the local squaring function at locations corresponding to correctly placed fragments revealed a multimodal function whose local minima were relatively isolated. The exact nature of the function, however, appeared to be highly dependent on the type of fragment as well as the location in the unit cell.

A list of optimal orientations for each fragment was sorted to obtain the locations of the 50 fragments of each type with the lowest local squaring function value. Considering overlap between the fragments, multiple specification due to symmetry and the given electron density, we constructed an initial model (Fig. 2). Although some variation in orientation is apparent both between overlapping components of modeled fragments and between the proposed structure and the true model, it can be seen that a large portion of the complete backbone structure is obtained from the local-squaring-function-based initial model. In fact, backbone fragments for nine of sixteen residues are correctly oriented by two or more templates and four backbone fragments are specified by a single template. The remaining three residues may not have a template placed correctly at the carbon in the carbonyl pair; however, templates from the neighboring residues overlap correctly (Fig. 3) allowing the backbone to be correctly reconstructed.

It is interesting to note that two (Asn 11 and Asp 14) of three side-chain carbonyl groups were located. The missing side-chain carbonyl group mostly was not identified because of its high temperature factor ($T > 30$). Furthermore, some degree of resonance was indicated in Asp 14 by placing two carbonyl templates of imperfect orientation.

Even in this small case, the necessary computation is considerable. Calculating the local squaring function for a single oriented template requires time proportional to a fast Fourier transform. However, structure and scattering factors of the oriented templates constitute a library that can be accessed for calculations with other proteins and calculation for larger crystal structures scales approximately with the crystallographic Fourier transform. Furthermore, the procedure is highly suited to parallel implementation. Each of the local squaring functions, for each template, and each orientation, can be computed independently. Once these have been computed, optimal orientations for each template can be calculated at each point in the unit cell independently. Therefore, provided there is a sufficient distributed computing environment, much of the analysis can proceed simultaneously.

In explicit terms, the initial work required to construct the fragment templates and orient them appropriately required about 15 min. This initialization procedure need only be used once. The calculation of all local squaring functions required about 4 min. The generating maps from the local squaring functions took about 23 min. Finally, determining optimal orientation, interpolation and constructing the initial atomic model needed slightly less than 12 min. All these calculations were completed on a single 500 MHz DEC Alpha.

This example serves as a concrete demonstration of one practical aspect of the theory. Obviously, with a structure this small, model building by hand is not particularly laborious. However, the extension to much larger structures is evident. Obstacles remain to be overcome in applying this method to larger systems. Also, developing data analytic methods to automate initial model construction from the set of likely fragments will be necessary for phase-refinement applications. Nonetheless, local squaring functions for non-spherical templates provide an effective way to combine phase refinement and map interpretations with other direct methods.

This work has been supported by the National Science Foundation under grant CCR-0086013. We would like to thank Professor Leonard Scott for helpful discussions on the representation theory of homogeneous spaces.

References

Arnoux, B., Ducruix, A. & Prangé, T. (2002). *Acta Cryst.* D**58**, 1252–1254.

Bricogne, G. (1997*a*). *Methods Enzymol.* **276**, 424–448.

Bricogne, G. (1997*b*). *Methods Enzymol.* **277**, 14–18.

Burdina, V. I. (1971). *Sov. Phy. Crystallogr.* **15**, 545–550.

Chapman, M. S. (1995). *Acta Cryst.* A**51**, 69–80.

Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossman, pp. 173–178. New York: Gordon and Breach.

Hu, S. H., Gehrmann, J., Guddat, L. W., Alewood, P. F., Craik, D. J. & Martin, J. L. (1996). *Structure*, **4**, 417–423.

Kirillov, A. A. (1994). *Representation Theory and Non-commutative Harmonic Analysis I*, edited by A. A. Kirillov, pp. 1–156. New York: Springer-Verlag.

Lattman, E. E. (1972). *Acta Cryst.* B**28**, 1065–1068.

Leach, A. R. (1996). *Molecular Modeling. Principles and Applications.* Essex: Addison Wesley Longman.

Morris, R. J., Blanc, E., Roversi, P. & Bricogne, G. (2001). American Crystallographic Association Annual Meeting, Los Angeles, USA.

Navaza, J. (2001). *International Tables for Crystallography*, Vol. F. *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, pp. 269–274. Dordrecht: Kluwer.

Prince, E., Finger, L. W. & Konnert, J. H. (1992). *International Tables for Crystallography*, Vol. C. *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson, pp. 613–617. Dordrecht: Kluwer.

Roach, J. & Carter, C. W. Jr (2002). *Acta Cryst.*, A**58**, 215–220.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Rothbauer, R. (2000). *Z. Kristallogr.* **215**, 157–168.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Shoemake, K. (1985). *Comput. Graphics*, **19**, 245–254.

Sloane, N. J. A., Hardin, R. H. & Smith, W. D. (2000). *Tables of Spherical Codes*, http://www.research.att.com/njas/packings/.

Sugiura, M. (1975). *Unitary Representations and Harmonic Analysis: an Introduction.* New York: Wiley.

Takeuchi, M. (1994). *Modern Spherical Functions.* Providence: American Mathematical Society.

Tanaka, N. (1977). *Acta Cryst.* A**33**, 191–193.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Woolfson, M. (1958). *Acta Cryst.* **11**, 277–283.

Xiang, S., Carter, C. W. Jr, Bricogne, G. & Gilmore, C. J. (1993). *Acta Cryst.* D**49**, 193–212.

Zhelobenko, D. P. (1973). *Compact Lie Groups and their Representations.* Providence: American Mathematical Society.