Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Envelope skeletonization as a means to determine monomer masks and non-crystallographic symmetry relationships: application in the solution of the structure of fibrinogen fragment D

Glen Spraggon

Center for Molecular Genetics, Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0634, USA

Correspondence e-mail: gspraggon@ucsd.edu

Received 27 January 1998
Accepted 17 July 1998

An algorithm is described which utilizes the solvent mask generated by the solvent-flattening technique to calculate a monomer molecular envelope. In the case where non-crystallographic symmetry (NCS) is present in the crystal and self-rotation angles are known from a self-rotation function, the resultant monomer envelopes can be used to search for the translation component of the NCS element by a three-dimensional search in real space. In the absence of self-rotation angles, the monomer envelope may be used to derive the NCS operators by reciprocal-space techniques. Thus, an automatic procedure for averaging directly from the solvent-flattening stage can be implemented. The procedure was instrumental in the structure solution of fibrinogen fragment D, which is presented as an example.

1. Introduction

The improvement of experimental phases by solvent flattening (Wang, 1985) is a standard procedure which produces a solvent envelope. Other density-modification techniques have also been proposed to improve initial phases, including histogram matching (Zhang & Main, 1990), incorporation of Sayre's equation (Cowtan & Main, 1993), iterative map skeletonization (Bystroff *et al.*, 1993) and perhaps the most potent technique, molecular averaging (Bricogne, 1974, 1976). Single-crystal averaging techniques may be implemented when the crystal contains more than one identical molecule in the asymmetric unit. The molecules are related by non-crystallographic symmetry (NCS) relationships, the redundancy being exploited to further improve phases. Two kinds of NCS exist in crystals (for a review, see Vellieux & Read, 1997); proper NCS, where all molecules in the crystal form a closed group and the relationships are described by pure rotation operations, and improper NCS, where molecules are related by three rotation angles and three translational components such that the operation to map molecule 1 onto molecule 2 is not identical to the operation to map molecule 2 onto molecule 1. Of course, a mixture of proper and improper NCS elements may be present in the crystal.

Often, the rotational elements of the NCS operators can be determined from the data alone by a self-rotation function (Rossmann & Blow, 1962). In the case of proper NCS, because the operators form a closed group, the rotational symmetry coupled with crystallographic symmetry can be used to derive an envelope for averaging (see, for example, Fry *et al.*, 1993) and the map may be averaged using the envelope of the entire complex. In the case of improper NCS, averaging can only proceed when the monomer envelope and non-crystal-

lographic symmetry matrices are known. Thus, for improper symmetry, the translational component of the NCS operators must also be derived.

Traditionally, the derivation of NCS matrices is carried out by finding identical features in each NCS-related molecule in the unit cell. These features often consist of the positions of heavy atoms bound to identical places in different molecules or identical features of secondary structure fitted into an initial electron-density map. Calculation of NCS matrices is carried out by superposition of such features between the different molecules (Kabsch, 1976, 1978; MacLachlan, 1979). Determination of molecular envelopes needed to average initial maps or refine NCS matrices is performed either by hand with programs such as *O* (Jones *et al.*, 1991), or automatically with a *BONES* skeleton (Jones *et al.*, 1991) or coordinate file derived from the initial map.

Once approximate NCS matrices and molecular envelopes are available the matrices may be refined. This procedure consists of extracting voxels within the defined molecular envelope, applying the NCS operator to the voxels, interpolating the transformed non-integer voxel density and calculating the correlation coefficient between the untransformed and transformed voxels. The six rigid-body parameters are varied in a steepest-descent algorithm which will bring the process to convergence if the initial operators are within the radius of convergence of the procedure. Averaging is then carried out and the molecular envelope iteratively improved.

Recently, correlation mapping (Rees *et al.*, 1990; Stein *et al.*, 1994; Vellieux *et al.*, 1995) has provided a technique for building a molecular envelope automatically with only knowledge of the NCS matrices and an initial electron-density map. The technique is described in Vellieux *et al.* (1995) and is similar to procedures used by Rees *et al.* (1990) to obtain a local correlation map from NCS operators. Briefly, the local correlation map contains at each voxel the linear correlation coefficient between spheres centred on voxels related by NCS operators. An envelope can be constructed by defining its boundaries as a continuous region of the map above a certain correlation-coefficient threshold. The resultant envelope may be modified to exclude regions with overlap due to crystallographic symmetry operations. The problem of defining a monomer envelope thus reduces to the location of accurate NCS operators. Calculation of such operators can prove difficult in the case of heavy-atom positions because heavy atoms often do not bind at identical sites in different molecules, either because they are bound between molecules, or because the chemistry of the environment varies for different NCS-related molecules. Finding the location of features in initial electron-density maps may also be difficult if electron density is poor.

If an approximate envelope and self-rotation angles are known, the translational element may be determined by a steepest-descent three-dimensional search of translation vectors, using points inside the initial envelope as a search model and a correlation coefficient as the target function. If no self-rotation angles are identified, a six-dimensional search can be employed in real space which can be very exhaustive of computer time. Alternatively, one could search in reciprocal space which is relatively fast.

Here, an algorithm is proposed which approximately determines the monomer molecular envelope from a solvent-flattened envelope by a process analogous to map skeletonization (Greer, 1974; Swanson, 1994). The envelope is skeletonized in such a way that, instead of the map peak size acting as a weight, the size of the sphere that can be incorporated in the protein part of the envelope is used and the connectivity of the skeleton is defined as the overlap between such spheres. The translational element of the NCS operators can then be calculated from the centre of mass of the individual envelopes. In the case where self-rotation angles are known, it becomes a simple matter to search around this approximate vector to determine the true vector. A computer program *ENVDER* (envelope derivation) has been written in Fortran77 to implement the algorithm.

2. Algorithm description

2.1. Skeletonization

A simple algorithm is used to skeletonize the solvent-flattened envelope (Fig. 1). The asymmetric unit of the solvent envelope derived from the solvent-flattening technique is used. The envelope adopts a binary format such that a voxel is given a value of 1 for protein and 0 for solvent.

Each voxel is searched for the largest radius sphere which can be incorporated using the voxel coordinates at its centre without penetrating the solvent region. In the case where spheres extend beyond the boundary of the map, voxels are folded back into the asymmetric unit by crystallographic symmetry operators. The result is a set of n points $r_i, \ldots, r_n$

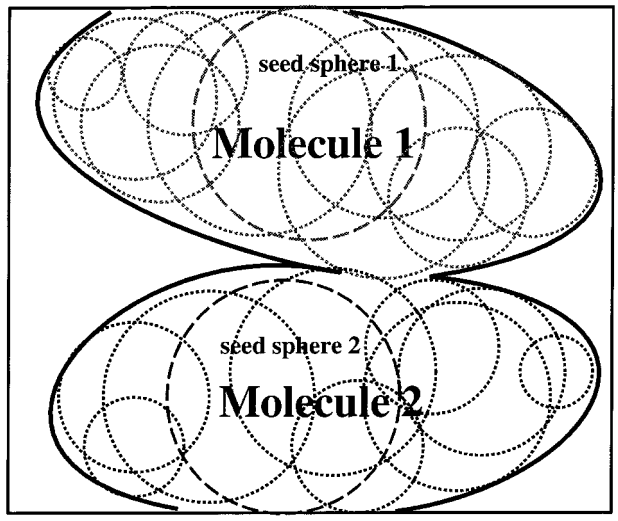

Figure 1
Schematic diagram of the envelope-skeletonization procedure in a hypothetical two-dimensional crystal containing two molecules in the asymmetric unit. The boundary of the solvent envelope is represented by a solid thick line. Circles assigned to molecule one are in light grey, whilst those of molecule two are in black. The seed spheres for both molecules are represented by a dashed line, whilst those grown out of the seed spheres are represented by dotted lines.

($i = 1, n$) each with an associated radius $R_i, \ldots, R_n$, organized in a heap data structure.

The resultant set of points is then searched to eliminate those that are completely contained within other points. Elimination proceeds by calculation of d_{ij} (the minimum distance after all crystallographic symmetry has been applied) between two points r_i and r_j with radii R_i and R_j, respectively. If d_{ij} is less than the modulus of the difference in radii $|R_i - R_j|$, the smaller sphere is completely contained within the larger and the point is removed from the heap.

All of the remaining points in the heap are sorted into descending order of size. Three user-defined criteria are entered: the expected number of molecules in the asymmetric unit (nmol), the overlap criterion to accept that the two spheres are connected (ovlap) and the minimum size of sphere to be considered (minsphere). The overlap criterion is defined as the overlap of the vectors connecting the two spheres r_i and r_j,

$$\begin{aligned} &< 0 \quad d_{ij} > Ri + R_j \\ (R_i + R_j - d_{ij})/(R_i + R_j) &= 0 \quad d_{ij} = R_i + R_j \\ &> 0 \quad d_{ij} < R_i + R_j. \end{aligned} \tag{1}$$

By default, the overlap criterion is set at 0.9 and the minimum sphere radius at half the maximum sphere radius found in the list.

The group of points is then partitioned into nmol sets. This is achieved by selecting a seed point for a particular molecule n, defined as the largest sphere in the list which has no connection to previous seed points. In the case of the first molecule, this is the largest sphere in the list r_1. After the seed point has been selected, the algorithm searches through the remaining unassigned points. Those which have an overlap greater or equal to ovlap and a radius greater than minsphere are assigned to set n. When the desired number of molecules has been realised, the algorithm searches the remaining unassigned spheres for the maximum overlap with that of the assigned spheres. If the maximum overlap of an unassigned sphere with a sphere belonging to set n is greater than ovlap, the unassigned sphere is placed in set n. This procedure is carried out in thin shells of sphere size from the maximum radius to a radius corresponding to minsphere.

What remains is a list of points, stored as the coordinates of the nearest distance to other points in the same set. These points are then expanded into a large map such that if a voxel lies within the radius of a sphere of set n, the voxel is given the value n and the skeletonized envelopes are output. In addition, the centre of mass of each skeletonized envelope is calculated and output.

2.2. Determination of NCS operators

In the general case where n molecules are known and $n - 1$ self-rotation matrices are available, the approximate translation vector $\mathbf{t}_{kjn}$ between molecule k and molecule j can be calculated as

$$\mathbf{t}_{kjn} = \mathbf{v}_j - S_i R_n(\mathbf{v}_k), \tag{2}$$

where R_n is the nth non-crystallographic rotation operator, S_i is the ith crystallographic symmetry operator and $\mathbf{v}_j$ and $\mathbf{v}_k$ are the centres of mass of molecules j and k respectively.

2.2.1. When rotational operators are known. If the rotation operators are known, then determination of the correct $\mathbf{t}_{kjn}$ reduces to a search in a box around $\mathbf{t}_{kjn}$ for all $n - 1$ available self-rotation angles and crystallographic symmetry operators i. The correct $\mathbf{t}_{kjn}$ should give the highest correlation coefficient. If the difference between the true translation vector and that defined is small, programs such as *IMP* from the *RAVE* package (Jones, 1992; Kleywegt & Jones, 1994) or *DM* (Cowtan, 1994) may be used to determine accurate NCS matrices by a steepest-descent search using the correlation coefficient as a target function. Alternatively, a straight three-dimensional search around the calculated translation vector may be conducted over a fine grid, the length of points being sampled at user-defined intervals. The larger the sampling and larger the grid, the more computational time is required.

2.2.2. In the absence of any information on NCS. If no information about the NCS operators is known, rotation operators must be found before a search around the translation vector can be conducted. In real space, the number of sampling points is vast, and the correlation coefficient can be extremely sensitive to the sampling space of the rotational and translational elements. This can be resolved by performing the calculations in reciprocal space. Here, the density contained inside one of the skeletonized envelopes calculated in *ENVDER* can be transferred into a cubic box of side-length equal to roughly twice the maximum dimension of the skeletonized envelope. The resultant density can be fast Fourier transformed to produce a set of structure factors for the masked density. Rotation angles are calculated as in standard molecular-replacement techniques using programs such as *AMoRe* (Navaza, 1994) or *X-PLOR* (Brünger, 1992). The translation vector can be calculated as in §2.2.1 or in reciprocal space using the translation function of the molecular-replacement package.

3. Results and discussion

3.1. Application to fibrinogen fragment D

Fragment D is the principle fragment of proteolysed fibrinogen (~86 kDa). Fibrinogen itself is a large (340 kDa) glycoprotein found in the plasma of all vertebrates and is responsible for providing the scaffold for blood clots by the thrombin-catalysed conversion of fibrinogen to fibrin. Fragment D crystals belong to space group $P2_1$ and contain two molecules in the asymmetric unit with a 50% solvent content (Everse *et al.*, 1995). A self-rotation function performed with the program *POLARRFN* (Collaborative Computational Project, Number 4, 1994) clearly indicated a twofold axis, distinct from the crystallographic twofold, on the $\kappa = 180^\circ$ section, the height of the peak being 84.1% of the crystallographic peak. The structure of fragment D has recently been solved to 2.9 Å resolution and consists of a three-stranded α-helical coiled coil, two of the strands terminating in homo-

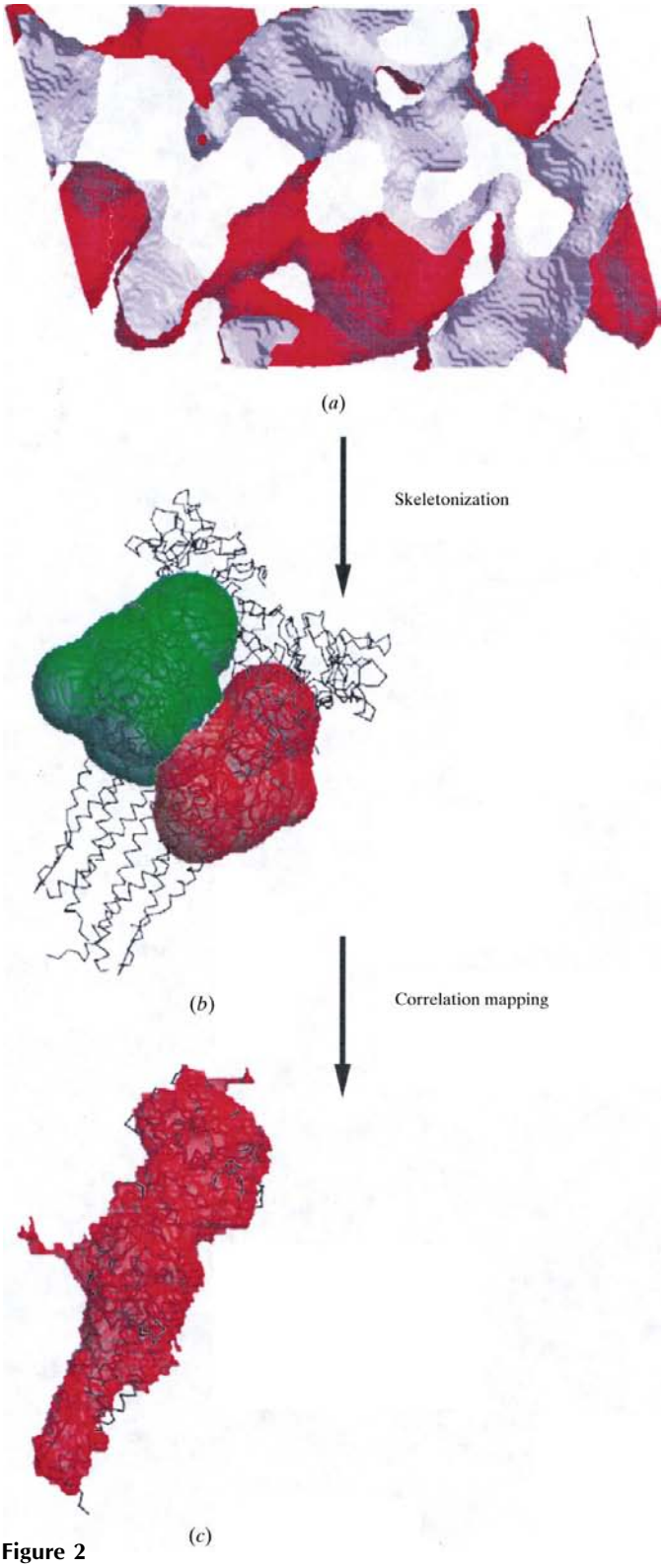

Figure 2
Diagrammatic representation of the envelope-skeletonization procedure. (*a*) Envelope from asymmetric unit of crystal is input into *ENVDER*. (*b*) Envelopes for two molecules of fragment D produced by *ENVDER*; one envelope is coloured green whilst the other is red. The refined coordinates are superimposed on the envelopes. (*c*) The envelope finally used for averaging produced by correlation mapping after NCS operators had been determined. Figure produced by *BOBSCRIPT* (Esnouf, 1997; Kraulis, 1991) and *Raster3D* (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

Table 1
Parameters for the final envelopes and NCS operators produced by *ENVDER*.

Map skeletonization		
Largest sphere (Å)		20.95
Number of spheres in skeletonized envelope		5738
Minimum sphere cut-off (Å)		10.0
Overlap criterion		0.9
Individual envelopes		
	Envelope 1	Envelope 2
Number of spheres	833	639
Volume of envelope (% of unit cell)	9.1	7.9
Centre of mass of envelopes (Å)	71.14, 9.86, 30.00	34.53, 8.54, 48.04
Actual centre of mass (Å)	69.97, 1.07, 31.81	37.77, 15.74, 51.41
Refinement of NCS with *RAVE*		
Rotation angles (initial, Euler)		(0, 90, 180)
Rotation angles (final, Euler)		(−1.89, −89.97, 176.87)
Translation vector (initial, Å)		(4.53, 18.40,−22.99)
Translation vector (final, Å)		(5.83, 18.36, −18.52)
Final correlation (%)		48.5

logous globular domains, whilst the third folds back to form a short helix antiparallel to the three other strands (Spraggon *et al.*, 1997). Initial multiple isomorphous replacement with anomalous scattering (MIRAS) phases were obtained to 4 Å resolution from three derivatives (gold, mercury and uranium) using the *PHASES* package (Furey & Swaminathan, 1997). The initial map was solvent flattened with the program *GAP* (Grimes & Stuart, unpublished work). The resultant map was improved, but visual inspection of the solvent envelope, electron-density map and heavy-atom positions made it difficult to delineate the molecular boundaries or determine initial NCS operators. The asymmetric unit solvent envelope was therefore used as input for the *ENVDER* program. Setting parameters for two molecules with an overlap criteria of 0.9 and largest sphere of 10.0 Å clearly produced two discernible skeletonized envelopes of approximately the same size (with volumes of 9.1 and 7.9% of the unit cell, respectively) and shape (with a volume overlap of 82% when NCS operators are applied, Fig. 2*b*, Table 1).

The translational vector was calculated from equation (2) and the resultant NCS relationship and skeletonized envelope was used with the program *IMP* from the *RAVE* package (Jones, 1992; Kleywegt & Jones, 1994) to refine the NCS relationship, the final correlation coefficient being 48% (Table 1). The refined matrix was then used to improve the mask by means of correlation mapping (Rees *et al.*, 1990; Stein *et al.*, 1994; Vellieux *et al.*, 1995) as implemented in the *CCP*4 programs *MAPROT*, *MAPMASK* and *NCSMASK* (Fig. 2*c*). The correlation envelope was used for averaging and tentative phase extension using the programs *RAVE* and *DM* (Jones, 1992; Kleywegt & Jones, 1994; Cowtan, 1994). This produced an electron-density map of sufficient quality to allow an almost complete chain-trace of the molecule. The mean phase error

Table 2
Results produced by variation of the overlap and minimum sphere criterion.

The correct NCS operator was not derived for the sets of values in italics.

		Minimum sphere radius (Å)																	
		2.0		4.0		6.0		8.0		10.0		12.0		14.0		16.0		18.0	
Overlap		Mol1	Mol2	Mol1	Mol2	Mol1	Mol2	Mol1	Mol2	Mol1	Mol2	Mol1	Mol2	Mol1	Mol2	Mol1	Mol2	Mol1	Mol2
0.1	CC†	*52.5*	*47.3*	*52.7*	*47.6*	*52.*	*48.0*	*52.9*	*47.5*	*53.6*	*49.1*	*52.6*	*51.9*	*55.2*	*51.2*	*55.9*	*25.3*	*65.6*	*9.4*
	Vol ‡	*22.6*	*9.5*	*22.5*	*9.5*	*22.4*	*9.1*	*22.2*	*8.8*	*21.9*	*7.7*	*18.3*	*6.5*	*16.7*	*5.6*	*15.8*	*3.9*	*9.7*	*0.8*
	Dis §	*36.5*		*36.8*		*37.2*		*37.6*		*42.9*		*30.1*		*34.9*		*32.0*		*20.5*	
0.2	CC †	*50.4*	*54.2*	*50.5*	*54.3*	*50.6*	*54.8*	*50.8*	*54.8*	*53.5*	*51.1*	*52.7*	*55.6*	*55.1*	*54.6*	*56.1*	*26.3*	*60.0*	*25.5*
	Vol ‡	*24.2*	*11.2*	*24.1*	*11.1*	*24.0*	*10.8*	*23.9*	*10.6*	*21.5*	*9.6*	*17.8*	*7.4*	*16.1*	*6.4*	*15.1*	*4.1*	*5.4*	*0.9*
	Dis §	*35.7*		*35.9*		*35.9*		*36.0*		*28.5*		*27.3*		*31.5*		*22.9*		*11.7*	
0.3	CC †	*51.8*	*58.1*	*51.9*	*58.2*	*52.0*	*58.7*	*52.8*	*58.8*	*55.9*	*56.4*	*52.2*	*63.3*	*54.8*	*65.2*	*55.6*	*41.9*	63.2	61.5
	Vol ‡	*21.5*	*15.3*	*21.4*	*15.2*	*21.3*	*14.9*	*21.2*	*14.5*	*19.0*	*12.9*	*17.6*	*9.3*	*15.8*	*8.2*	*14.7*	*3.2*	5.1	4.9
	Dis §	*32.7*		*32.8*		*33.0*		*32.9*		*33.0*		*25.2*		*25.2*		*17.8*		2.7	
0.4	CC †	*53.6*	*60.4*	*53.8*	*60.4*	*53.9*	*61.0*	*54.5*	*61.4*	*55.4*	*61.5*	*51.6*	*68.4*	*54.2*	*67.0*	*55.8*	*56.8*	63.2	61.5
	Vol ‡	*19.2*	*17.5*	*19.0*	*17.7*	*19.0*	*17.3*	*19.0*	*16.8*	*18.8*	*13.8*	*17.4*	*10.5*	*15.4*	*9.0*	*14.4*	*5.3*	5.1	4.9
	Dis §	*26.8*		*26.9*		*27.2*		*27.4*		*29.9*		*19.0*		*21.2*		*12.1*		2.7	
0.5	CC †	*52.3*	*59.8*	*53.7*	*61.3*	*53.8*	*61.6*	*54.7*	*62.1*	*55.3*	*63.3*	*51.6*	*71.8*	*56.4*	*59.0*	62.9	61.7	63.2	61.5
	Vol ‡	*19.0*	*19.8*	*18.9*	*19.7*	*18.9*	*19.4*	*18.8*	*19.0*	*18.7*	*14.9*	*17.3*	*11.3*	*15.4*	*10.1*	6.0	5.4	5.1	4.9
	Dis §	*28.8*		*28.4*		*28.5*		*28.5*		*26.9*		*28.6*		*19.5*		2.5		2.7	
0.6	CC †	*53.0*	*60.0*	*53.1*	*60.0*	*53.2*	*60.5*	*54.1*	*60.5*	*55.2*	*65.1*	*51.5*	*71.2*	*55.0*	*58.0*	62.9	61.7	63.2	61.5
	Vol ‡	*19.2*	*21.3*	*19.1*	*21.3*	*19.0*	*21.0*	*18.9*	*21.1*	*18.6*	*14.2*	*17.2*	*11.2*	*14.3*	*5.8*	6.0	5.4	5.1	4.9
	Dis §	*32.8*		*32.8*		*32.2*		*31.3*		*32.8*		*28.6*		*21.2*		2.5		2.7	
0.7	CC †	*54.0*	*58.3*	*54.0*	*58.1*	*54.1*	*58.1*	*54.7*	*60.0*	*56.2*	*70.2*	*52.1*	*69.3*	*56.6*	*58.3*	62.9	61.7	63.2	61.5
	Vol ‡	*18.2*	*24.1*	*18.1*	*24.0*	*18.0*	*23.8*	*18.4*	*21.2*	*18.0*	*12.5*	*16.5*	*9.5*	*13.0*	*5.8*	6.0	5.4	5.1	4.9
	Dis §	*32.6*		*32.9*		*32.9*		*34.3*		*30.2*		*21.9*		*19.6*		2.5		2.7	
0.8	CC †	*49.0*	*53.5*	*54.1*	*56.3*	*54.0*	*58.2*	*53.3*	*67.4*	*57.1*	*72.0*	57.8	65.3	65.1	63.1	63.2	61.3	63.2	61.5
	Vol ‡	*17.3*	*27.1*	*17.7*	*26.3*	*17.6*	*23.7*	*16.7*	*13.9*	*16.3*	*12.0*	13.0	7.8	7.8	5.8	6.0	5.4	5.1	4.9
	Dis §	*32.2*		*32.5*		*37.1*		*25.9*		*16.9*		8.9		6.9		2.5		2.7	
0.9	CC †	*52.1*	*66.3*	*53.2*	*69.1*	*54.2*	*69.0*	65.2	70.0	65.5	71.0	62.7	65.0	62.7	62.2	63.3	60.5	63.2	61.5
	Vol ‡	*13.6*	*15.5*	*13.7*	*14.5*	*13.9*	*11.8*	12.7	9.2	9.1	7.9	7.8	6.3	6.5	5.7	6.1	5.3	5.1	4.9
	Dis §	*22.7*		*22.1*		*16.6*		8.3		5.6		4.9		2.9		2.4		2.7	

† Correlation coefficient = (N1 and N2)/$[(N1)(N2)]^{1/2}$ for two envelopes containing N1 and N2 points, respectively. The correlation coefficient (%) between *ENVDER*-derived envelopes and envelopes derived from final coordinates (using a radius of 4 Å around each atom). ‡ Final volumes of *ENVDER* envelopes expressed as a percentage of the unit cell (the final envelope has a volume of 12.5% of the unit cell). § Euclidian distance between *ENVDER*-derived translation operator and true translation operator.

between averaged phases and phases calculated from the final structure was 62°, whilst that from the unaveraged solvent-flattened structure was 71°.

To demonstrate the technique for a case when the self-rotation angles are undetermined, the skeletonized envelope for molecule 1 was used to mask the density from the initial solvent-flattened map, and the masked density was transferred to a large (120 Å side) cubic box with the program *MAPROT* (Collaborative Computational Project, Number 4, 1994). The map was subsequently fast Fourier transformed with *SFALL* (Collaborative Computational Project, Number 4, 1994) to obtain structure factors. A rotation function in *AMoRe* (Navaza, 1994; Collaborative Computational Project, Number 4, 1994) was carried out using all data between 8.0 and 4.0 Å. This produced two clear solutions at Euler angles (0.43, 0.0, 0.0) (corresponding to the input envelope) and (−1.5, 91.45, 178.73) (corresponding to the envelope's NCS-related partner), with correlation coefficients 20.2 and 19.8%, respectively (the next highest peak was 14.3%). Input of the second solution into the translation function of *AMoRe* produced a clear solution at 28.5% correlation coefficient after rigid-body refinement with all data between 8.0 and 4.0 Å.

3.2. Investigation of skeletonization criteria

Three user-defined criteria are included in the program, the number of molecules in the asymmetric unit (nmol), the overlap criterion (ovlap) and the minimum sphere size to be included (minsphere). The number of molecules in the asymmetric unit is necessary so that the first stage in envelope generation, the seed-sphere generation, will terminate. The procedure will automatically terminate by running out of suitable seed points, however, if minsphere is set to a fairly large value.

To assess the effect of ovlap and minsphere on the skeletonized envelopes, both parameters were varied. The quality of the envelopes was judged by calculating the correlation coefficient between skeletonized envelopes and those calcu-

lated from the final coordinates, the volumes of the skeletonized envelopes, convergence on the correct NCS operator and the inaccuracy of the calculated translation vector (Table 2). As can be seen from Table 2, no derivation of NCS operators is achieved when the overlap criterion is below 0.3 (sets of values in italics). This results from the first seed sphere being connected to an excess of other spheres. Above 0.3, the situation for all calculated quantities improves gradually until at 0.9 the NCS operator can be found for all runs with minimum sphere size above 8 Å (40% of the maximum sphere size). Variation of minsphere allows the envelopes to be modeled first by the largest spheres, then by increasingly smaller spheres. In this way, spheres of decreasing size are slowly incorporating into the envelope. Lowering minsphere creates situations whereby narrow corridors to other molecules can be grown out of the seed sphere by the connection of small spheres. This is indicated by values of the correlation coefficients, which decrease below a minimum sphere size of 10 Å (50% of maximum). Below a minimum sphere size of 6.0 Å, no NCS operator could be found with any ovlap (sets of values in italics); above this, the operator could be found for all minsphere values. When the minimum sphere size is set at 18.0 Å (90% of the maximum) only the two largest spheres are included; this allows derivation of NCS operators in all but two cases and is quick. This could be useful when dealing with globular proteins, but may cause problems in cases of poor envelopes when one molecule contains larger spheres than another. The quality of the envelope is seen to be maximal when the minimum sphere radius is set at 10 Å (50% of the maximum sphere radius) and the overlap criteria is set at 0.9. Here, the correlation coefficients for the two envelopes were 65.5 and 71.0%, respectively.

4. Summary

A method has been described that uses an envelope-skeletonization technique for the automatic determination of monomer envelopes and non-crystallographic symmetry matrices. The envelope-generation procedure is reminiscent of the skeletonization of electron-density maps (Greer, 1974; Swanson, 1994), except the criterion for peak selection is the maximum radii that can be included in the solvent envelope, as opposed to the height of the peak. Connectivity of the skeleton is defined by an overlap criterion, analogous to ridge length in map skeletonization, such that two spheres are connected if the overlap criterion exceeds a certain value.

Here, the technique was instrumental in the solution of the structure of fibrinogen fragment D where the poor quality of a solvent-flattened map and envelope and lack of concrete structure within the map made interpretation and derivation of NCS operators and the molecular boundary difficult.

Although problems may be envisaged for tightly bound molecule complexes, the technique is generally applicable for any solvent envelope to mask density for use in molecular replacement. It can, therefore, provide an automatic procedure for the determination of an averaged map directly from the solvent-flattening stage.

I would like to thank Professor D. Stuart for useful discussion of the manuscript, Professor R. Doolittle for support and careful review of the manuscript and Dr S. Everse for computational support.

References

Bacon, D. J. & Anderson, W. F. (1988). *J. Mol. Graph.* **6**, 219–220.
Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.
Bricogne, G. (1976). *Acta Cryst.* A**32**, 832–847.
Brünger, A. T. (1992). *X-PLOR: Version 3.1: A System for X-ray Crystallography and NMR.* Yale University, Connecticut, USA.
Bystroff, C., Baker, D., Fletterick, R. J. & Agard, D. A. (1993). *Acta Cryst.* D**49**, 429–439.
Cowtan, K. (1994). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Everse, S. J., Pelletier, H. & Doolittle, R. F. (1995). *Protein Sci.* **4**, 1013–1016.
Fry, E. Acharya, R. & Stuart, D. I. (1993). *Acta Cryst.* A**49**, 45–55.
Furey, W. S. & Swaminathan, S. (1997). *Methods Enzymol.* **277**, 590–620.
Greer, J. (1974). *J. Mol. Biol.* **82**, 279–301.
Jones, T. A. (1992). *Molecular Replacement*, edited by E. J. Dodson, S. Glover & W. Wolf, pp. 92–105. Warrington: Daresbury Laboratory.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. A. (1976). *Acta Cryst.* A**32**, 922–923.
Kabsch, W. A. (1978). *Acta Cryst.* A**34**, 827–828.
Kleywegt, G. J. & Jones, T. A. (1994). *From First Map to Final Model*, edited by S. H. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
MacLachlan, A. D. (1979). *J. Mol Biol.* **128**, 49–79.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Navaza, J. (1994). *Acta Cryst.* D**50**, 157–163.
Rees, B., Bilwes, A., Samama, J. P. & Moras, D. (1990). *J. Mol. Biol.* **214**, 281–297.
Rossmann, M. & Blow, D. (1962). *Acta Cryst.* **15**, 24–31.
Spraggon, G., Everse, S. J. & Doolittle, R. F. (1997). *Nature (London)*, **389**, 455–462.
Stein, P. E., Boodhoo, A., Armstrong, G. D., Cockle, S. A., Klein, M. H. & Read, R. J. (1994). *Structure*, **2**, 45–57.
Swanson, S. (1994). *Acta Cryst.* D**50**, 695–708.
Vellieux, F. M. D., Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.
Vellieux, F. M. D. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Zhang, K. Y. & Main, P. (1990). *Acta Cryst.* A**46**, 377–381.