Acta Cryst. (1997). D53, 274–278

Infinite Non-crystallographic Symmetries in Crystals of a Globular Protein

CHRISTIAN WIESMANN AND GEORG E. SCHULZ*

Institut für Organische Chemie und Biochemie, Albert-Ludwigs-Universität, Albertstrasse 21, 79104 Freiburg im Breisgau, Germany. E-mail: schulz@bio5.chemie.uni-freiburg.de

(Received 1 November 1996; accepted 2 December 1996)

Abstract

The monomeric enzyme 6-phospho-β-galactosidase crystallized in three forms, all of which can be considered as fiber associations when judged from the packing contacts. In one case the internal fiber symmetry is a row of parallel twofold axes that are displaced by $\mathbf{a}'$ along the fiber axis. In the crystal, these twofold axes are non-crystallographic, but they define the x axis by $\mathbf{a} = 2\mathbf{a}'$. Antiparallel and asymmetric fiber association along y and z, respectively, results in space group $P2_1$. The two other crystal forms contain fibers with internal 2_1 axes (translation $\mathbf{c}'$) defining the z axis by $\mathbf{c} = 2\mathbf{c}'$. Antiparallel fiber association results in sheets with internal twofold and 2_1 axes defining the yz plane. These sheets associate along x in two ways; in one case the internal sheet symmetry becomes crystallographic, while it remains non-crystallographic in the other. All non-crystallographic fiber symmetries run infinitely through the crystal. In one case they cause exact systematic absences that mimic space group $P2_12_12_1$ instead of the actual $P2_12_12$. The relationship between contact areas and local symmetries is discussed.

1. Introduction

On crystallization, proteins form numerous packing contacts between different patches of their large and diversified surfaces. An analysis of six crystal forms of pancreatic ribonuclease, for instance, showed that these patches scatter over the whole molecular surface, indicating that they are rather unspecific (Crosio, Janin & Jullien, 1992). In general, crystal contacts are smaller than oligomer interfaces which generally extend over more than 1000 Å^2 buried solvent-accessible surface per subunit (Janin, Miller & Chothia, 1988; Finkelstein & Janin, 1989; Jones & Thornton, 1996). There are cases like the tetrameric NADH peroxidase, however, where the packing contacts are larger than the oligomer contacts (Stehle, Ahmed, Claiborne & Schulz, 1993). Here it seems likely that the enzyme is crystalline *in vivo*.

Many proteins that are considered to be monomeric in solution crystallize with more than one molecule in the asymmetric unit. Often these molecules are related by rather exact local symmetry axes, exploiting mutual interactions optimally; *e.g.* twice at a twofold axis. If these contacts extend over large areas and/or contain several hydrogen bonds, it becomes unlikely that they are broken apart by a more favorable packing arrangement. As a consequence, local symmetries indicate potential oligomers in solution.

Here, we report on a protein which crystallizes under formation of fibers with internal symmetries that do or do not become crystallographic. In one case, the fibers form a sheet with internal symmetries that failed to become crystallographic on sheet association. The internal symmetries are discussed in conjunction with the contact strengths as derived from areas and hydrogen bonds. Analysis of these contacts will promote our knowledge on weak interactions between proteins, which become of more and more interest, *e.g.* because of their role in signalling pathways (Pawson, 1995; Bray, 1995).

2. Materials and methods

The recombinant enzyme 6-phospho-β-galactosidase from *Lactococcus lactis* was produced in *E. coli* strain ΔH1ΔTrp (De Vos & Gasson, 1989) using the expression plasmid pNZ316 kindly provided by W. De Vos (Wageningen, Netherlands). The protein was mutated (Taylor, Ott & Eckstein, 1985) and overexpressed (Wiesmann, Beste, Hengstenberg & Schulz, 1995). For crystallization we used the hanging-drop method throughout. X-ray diffraction data were collected using an area detector (Siemens, model X1000) on a rotating anode (Rigaku, model RU-200B). All solvent-accessible surfaces and, thus also the buried surfaces, were calculated with program *X-PLOR* (Brünger, Kuriyan & Karplus, 1987).

3. Results

3.1. *Crystallization and structures*

The wild-type enzyme is monomeric in solution. It crystallized under two different conditions yielding crystal forms *A* and *C* (Table 1). The inactive mutant Glu375→Cys cocrystallized with a substrate in form *C** at conditions where the wild type grew as form *A* (Table 1). Mutant Ser256→Cys crystallized in form *B* containing dimers connected *via* a 256—256′ disulfide bridge that crosses a crystallographic twofold axis. Only in the first days after the protein preparation, which had

Table 1. *Crystallization and crystal forms*

Crystal form	*A*	*B*	*C*	*C**
Protein†	Wild type, Ser256→Cys‡ (reduced)	Ser256→Cys‡ (oxidized)	Wild type	Glu375→Cys 40 m*M* *o*-NPG-6P§
Reservoir buffer	28% PEG 4000 0.2 *M* Li_2SO_4	Same as in form *A*	1.9 *M* $(NH_4)_2SO_4$ 2% PEG 600	Same as in form *A*
Space group	$P2_1$	$C222_1$	$P2_12_12$	$P2_12_12$
Cell parameters *a* (Å)	60.5	104.0	107.1	104.5
b (Å)	83.2	179.6	174.5	174.4
c (Å)	105.7	60.6	61.0	60.6
β (°)	93.7			
Resolution (Å)	2.3	2.5	2.7	2.5
V_M (Å^3 Da^{-1})	2.5	2.6	2.6	2.6
Asymmetric molecules	2	1	2	2

† All protein solutions were 12 mg ml^{-1} in 0.1 *M* KH_2PO_4 pH 7.5 and 0.02% NaN_3. All reservoir buffers contained 0.1 *M* Tris pH 7.5 and 0.02% NaN_3. Droplets were mixed from equal volumes of protein solution and reservoir buffer. ‡ The protein was probably dimerized by disulfide formation before crystallization in form *B*. Form *A* crystals grew before oxidation. § *o*-Nitrophenyl-β-D-galactopyranoside-6-phosphate.

Table 2. *Packing contacts in crystal form A*

Contact*	Buried surface (Å^2)	Hydrogen bonds	Contacting patches†
I_1-II_1	900	6	*a–a*
I_1-I_2	410	4	*b–c*
I_1-I_3	410	4	*c–b*
II_1-II_2	560	6	*d–e*
II_1-II_3	560	6	*e–d*
I_1-II_4	730	6	*f–f*
I_1-II_5	320	2	*g–h*
I_1-II_6	250	5	*i–j*
I_1-II_2	150	1	*k–l*
I_1-II_7	50	1	*m–n*

* Molecules I_1 and II_1 were taken as the reference at $[x, y, z]$, the others are at I_2 $[-x, y-\frac{1}{2}, -z+1]$, I_3 $[-x, y+\frac{1}{2}, -z+1]$, II_2 $[-x+1, y+\frac{1}{2}, -z+2]$, II_3 $[-x+1, y-\frac{1}{2}, -z+2]$, II_4 $[x-1, y, z]$, II_5 $[-x, y+\frac{1}{2}, -z+1]$, II_6 $[x-1, y+1, z]$, II_7 $[-x, y+\frac{1}{2}, -z+2]$. † Contacting residues contain atoms at distances below 4.5 Å. The residues in patches *a* to *n* are *a* = 1-**2**, 222-**223**-224, 226, 303, 337-338-**339**-340-341, 350, 380, 390-391-**392**-393, **453**, **468**; *b* = 224-225, **228**, **231**, 305-**306**-307-**308**; *c* = 278, 281-**282**, **285**-286; *d* = **61**-62, **65**-**66**, 386-**387**, 390, **448**-**449**-450-451, 456, 459; *e* = 147-148, 150, 152-**153**, **205**-**206**-207-208-**209**-210; *f* = 43-44, 379-380-381, 384, 431, 434-435-436, **441**, **443**-444-**445**-446, 450; *g* = 273-274-**275**-**276**-277-278; *h* = 225, **228**-229-230, 233; *i* = **87**-**88**-89-**90**; *j* = 276-**277**-**278**-**279**; *k* = 459, **463**-464; *l* = **91**, 94, 98; *m* = **105**-106; *n* = **7**; where hydrogen-bonded residues are in bold.

Table 3. *Packing contacts in crystal form C*

Contact*	Buried surface (Å^2)	Hydrogen bonds	Contacting patches†
I_1-II_1	700	2	*q′–q′*
I_1-II_2	570	4	*r′–s′*
I_1-II_3	570	4	*s″–r′*
I_1-II_4	320	1	*w–x*
I_1-I_3	60	0	*t–u′*
I_1-I_4	60	0	*u′–t*
II_1-II_5	60	0	*t–u′*
II_1-II_6	60	0	*u′–t*
I_1-I_2	70	0	*y–y*

* Taking molecules I_1 and II_1 as the reference at $[x, y, z]$, the others are at: I_2 $[-x, -y+1, z]$, I_3 $[x, y, z+1]$, I_4 $[x, y, z-1]$, II_2 $[-x+\frac{1}{2}, y-\frac{1}{2}, -z]$, II_3 $[-x+\frac{1}{2}, y-\frac{1}{2}, -z+1]$, II_4 $[-x, -y+1, z]$, II_5 $[x, y, z-1]$, II_6 $[x, y, z+1]$. † Contacting residues contain atoms at distances below 4.5 Å. The residues in patches *q* to *y* are *q′* = 97, 101, 104-105, **107**, **148**, 150-151, 153-154, 209-210-211; *r′* = 177-178-179-180, 305-306-307-**308**-309-**310**-311-312, **325**, 327, 329, **333**; *s′* = **321**, 339, 341-342-343-344, 349, 380, **392**-**393**, **396**, 400; *s″* = *s′* plus **390**-391; *t* = 1; *u′* = 271-272; *w* = 384, **387**, 443, 445-446, 448, 450; *x* = 26-27, **54**, 56-57-58, 98; *y* = 26-27; where hydrogen-bonded residues are in bold. Residue lists with primes have closely resembling (unprimed) counterparts in form *B* (Table 4).

been carried out with 1 m*M* dithiothreitol under reducing conditions, did this mutant crystallize in its reduced state in crystal form *A* (Table 1).

The enzyme structure was initially solved by multiple isomorphous replacement in form *A* (Wiesmann, Beste, Hengstenberg & Schulz, 1995). The result was used for establishing the structures of the other crystal forms by molecular replacement (Navaza, 1994). Here, crystal form *C** is always presented synonymously with form C because they are almost identical. Forms *A* and *C* have no packing contact in common, whereas several of the contacts of forms *B* and *C* are in agreement with each other. The packing contacts are documented in Tables 2 to 4. The total solvent-accessible surface of a molecule is 16 200 Å^2.

3.2. *Packing in crystal form A*

Here, the most striking feature is a row of parallel twofold axes displaced from each other by vector **a′**. This row runs vertically in Fig. 1; it is formed by the largest contacts *a–a* and *f–f* (Table 2) indicating an association of molecules in a fiber. The displacement vector **a′** along this fiber defines the *x* axis by **a** = 2**a′**. The lateral association of these fibers in the *y* and *z* directions, however, does not follow the twofold axes within the fiber, such that these axes become non-crystallographic and local in space group $P2_1$. Consequently, this association gives rise to two molecules in the asymmetric unit.

Table 4. *Packing contacts in crystal form B of Ser256→Cys*

Contact*	Buried surface (Å^2)	Hydrogen bonds	Contacting patches†
1–2	420	5	*p–p*
1–3	620	2	*q–q*
1–4	550	7	*r–s*
1–5	550	7	*s–r*
1–6	90	0	*t–u*
1–7	90	0	*u–t*
1–8	50	0	*v–v*

* Taking molecule 1 as the reference at $[x, y, z]$, the others are at: 2 $[-x+1, y, -z+\frac{1}{2}]$, 3 $[x, -y+1, -z+1]$, 4 $[-x+\frac{1}{2}, -y+\frac{1}{2}, z-\frac{1}{2}]$, 5 $[-x+\frac{1}{2}, -y+\frac{1}{2}, z+\frac{1}{2}]$, 6 $[x, y, z+1]$, 7 $[x, y, z-1]$, 8 $[-x, y, -z+\frac{1}{2}]$ † Contacting residues contain atoms at distances below 4.5 Å. Cys256 connects molecule 1 and 2 by a disulfide bridge. The residues in patches *p* to *v* are *p* = 236, **244**, 252, **254**, 256–**257**–**258**, 261, 285, 288; *q* = 94, 97–98, 101, 104–105–106–**107**, **148**, 150–151, 209; *r* = 179–180, 307–**308**–309–**310**–311–312, **325**, 327, **329**, 333; *s* = 339, 341–**342**–343–**344**, 349, **380**, **390**–391–392–393, **396**, 400, 453; *t* = 1; *u* = 273; *v* = 43; where hydrogen-bonded residues are in bold.

The two contacts along *x* are not only the strongest ones, but taken together they amount to 49% of all packing contacts, which is appreciably larger than the expected 33% for one dimension (Table 2). This predominance corresponds to the observed habits; the crystals are rods along *x*.

3.3. *Packing in crystal forms C and B*

The protein crystallizes most readily in form *C*. A striking feature of this form are infinite non-crystallographic 2_1 screw axes along *z* that run horizontally in Fig. 2. The translation $\mathbf{c}'$ of these screw axes defines the *z* axis by $\mathbf{c} = 2\mathbf{c}'$. Here, we can discern a fiber made up by the molecules along this 2_1 axis with two contacts of 570 Å^2 each, amounting to 54% of all packing contacts (Table 3). In contrast to form *A*, however, the largest single contact of 700 Å^2 is not within this fiber but causes antiparallel association of fibers. It extends across a twofold axis shown in Fig. 2. The fiber association creates a sheet of molecules, which is defined by internal twofold and 2_1 axes forming a two-dimensional crystal. In an almost identical arrangement this sheet is also present in crystal form *B*.

In crystal form *C*, the association of sheets in the third dimension does not honor the internal symmetry elements of the sheet such that all of them become non-crystallographic giving rise to two molecules in the asymmetric unit (Table 1). Following common practice the two asymmetric molecules are those related by the local twofold axis *via* contact *q′–q′* (Table 3). Although non-crystallographic, the 2_1 axes within the sheet are infinite and run exactly along *z* (Fig. 3), thus causing the absence of all $00l$ reflections with $l = 2n + 1$. As a consequence, the systematic absences in crystal form *C* point to the very frequent space group $P2_12_12_1$, whereas the actual space group is $P2_12_12$. The resulting, almost exact non-crystallographic relation between two molecules gave rise to a peak with a height of 47% of the origin peak at $(u,v,w) = (0.5, 0.4, 0.5)$ in the native Patterson map.

The sheet character of the packing in crystal form *C* is corroborated by the fact that the sum of all contact areas within the sheet amounts to 85% of the total, indicating a weak connection along the third dimension *x* (Table 3, Figs. 2 and 3). It is further confirmed by crystal form *B* containing essentially the same sheet in its *yz* plane. In contrast to form *C*, however, the protein of form *B* has dimerized by disulfide formation of the newly introduced surfacial Cys256. These disulfides interconnect the sheets as sketched in Fig. 3. Most likely caused by these disulfides, the assembly of sheets along the third dimension, *x*, renders the internal sheet symmetries crystallographic in space group $C222_1$ with only one molecule in the asymmetric unit (Table 4). The relation between the two types of sheet association is illustrated in Fig. 3. In form *B* the contact area between

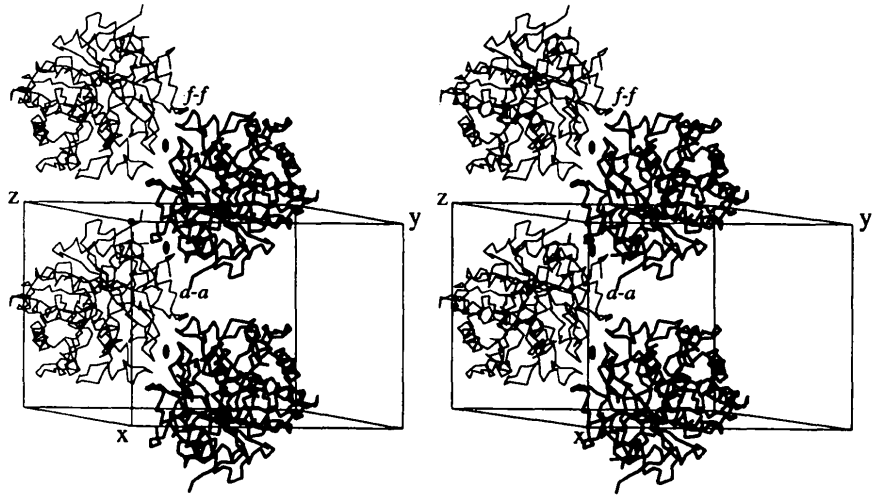

Fig. 1. Stereoview along the local twofold axes in crystal form *A*. The row of twofolds defines a fiber and also the translation **a** in the crystal. Contacts *a–a* and *f–f* within the fiber (Table 2) are the strongest of all, they amount to 49% of the total contact area, which is much more than the expected 33% for one of the three dimensions.

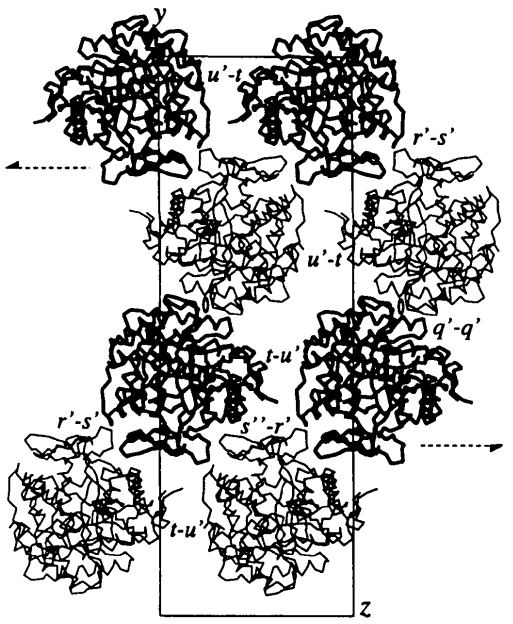

Fig. 2. Molecular packing in the yz plane of crystal form C, which is close to the respective packing in form B. Rows of molecules are connected by non-crystallographic infinite 2_1 axes (broken line symbols) forming horizontal fibers. The contact areas within a fiber amount to 54% of the total. These fibers associate in an antiparallel manner *via* contact q'–q' across local twofold axes (hollow symbols) that are exactly parallel to the x axis, but non-crystallographic in form C. The 2_1 axis defines translation **c**. The antiparallel fiber association defines translation **b**. The contacts within this sheet are labeled, they comprise 85% of the total contact area (Table 3).

the sheets has increased, but it still amounts to only 20% of the total, as compared with 15% in form C.

4. Discussion

The peculiar packings of the globular protein 6-phospho-β-galactosidase show one-dimensional (form A) and one- or two-dimensional (forms B and C) internal building blocks that can be recognized by the sizes and qualities of packing contacts and by non-crystallographic symmetries. In agreement with the observations of Wang & Janin (1993) the local twofold axes in crystal forms A and C are perpendicular to a crystal axis; **a** in form A, and **b** as well as **c** in form C. In form C the local twofold is also parallel to **a**.

All contact areas are smaller than those within stable oligomers (Janin, Miller & Chothia, 1988) rendering fiber and sheet formation in solution an unlikely event. It is conceivable, however, that such contacts could be improved by designed mutations to form oligomers in solution following the experience with protein docking analyses (Jones & Thornton, 1996). In the presented case of 6-phospho-β-galactosidase, the largest contact of 900 Å^2 forming a symmetric dimer (a–a, form A) and the 570 Å^2 contact giving rise to fibers with an internal 2_1 axes (r–s, forms B and C) occlude the active center. An improvement of these contacts should yield oligomers with low enzymatic activity, which could be used for monitoring engineering endeavors.

This work was supported by the Deutsche Forschungsgemeinschaft under Schu406/14-1.

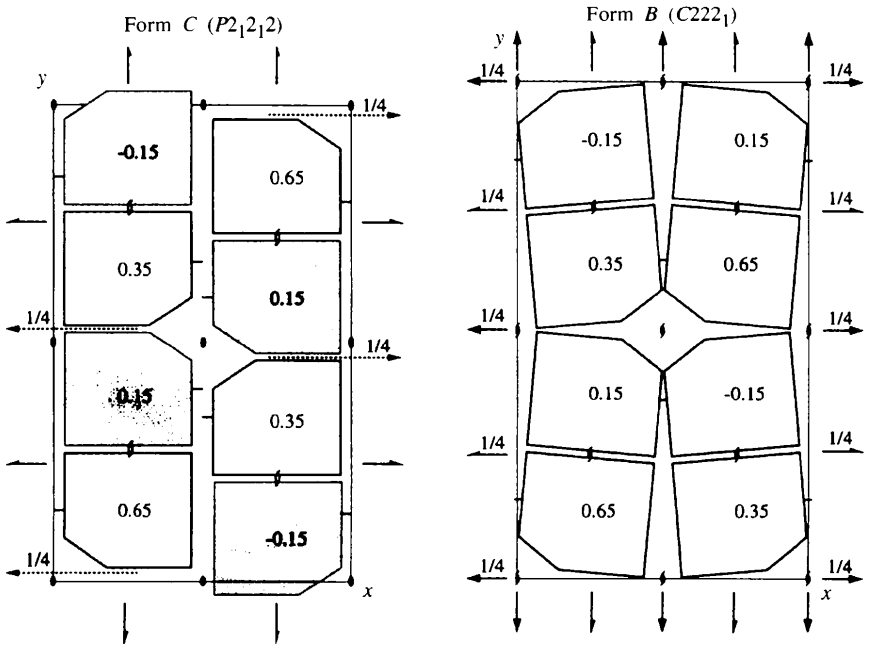

Fig. 3. Sketch of the packing in the xy planes of crystal forms C and B. The local twofold axes of form C are dotted. The infinite non-crystallographic 2_1 axes are along z and denoted by hollow symbols. Form B contains essentially the same (though somewhat pleated) yz sheet as form C (Fig. 2), but differs grossly with respect to sheet association along x. The position of Ser256 in form C is indicated by small bars at the surface. The corresponding disulfide bridge of mutant Ser256→Cys in the crystals of form B is sketched as a connecting bar, it lies in the paper plane across a vertical twofold axis at $z = 0$. The weak non-covalent contacts between the yz sheets gave rise to multiple **a**-axis lengths in form C. Moreover, this axis changed on soaking with heavy-atom compounds as well as during data collection, prohibiting the initial structural analysis of the most readily available crystal form C.

References

Bray, D. (1995). *Nature (London)*, **376**, 307–312.
Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Crosio, M.-P., Janin, J. & Jullien, M. (1992). *J. Mol. Biol.* **228**, 243–251.
De Vos, W. M. & Gasson, M. J. (1989). *J. Gen. Microbiol.* **133**, 563–573.
Finkelstein, A. V. & Janin, J. (1989). *Protein Eng.* **3**, 1–3.
Janin, J., Miller, S. & Chothia, C. (1988). *J. Mol. Biol.* **204**, 155–164.
Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Pawson, T. (1995). *Nature (London)*, **373**, 573–580.
Stehle, T., Ahmed, S. A., Claiborne, A. & Schulz, G. E. (1991). *J. Mol. Biol.* **221**, 1325-1344.
Taylor, J. W., Ott, J. & Eckstein, F. (1985). *Nucleic Acids Res.* **13**, 8765–8785.
Wang, X. & Janin, J. (1993). *Acta Cryst.* D**49**, 505–512.
Wiesmann, C., Beste, G., Hengstenberg, W. & Schulz, G. E. (1995). *Structure*, **3**, 961–968.

Acta Cryst. (1997). D53, 279–289

Molecular Replacement Using Genetic Algorithms

GEOFFREY CHANG AND MITCHELL LEWIS

The Johnson Research Foundation, Department of Biophysics and Biochemistry, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104–6059, USA. E-mail: lewis@crystal.med.upenn.edu

(Received 9 September 1996; accepted 2 December 1996)

Abstract

A new molecular replacement (MR) strategy is introduced which features a continuous transform and a genetic algorithm (GA) for search optimization. This strategy uses a GA to simultaneously search the rotational and translational parameters of a test model while maximizing the correlation coefficient between the observed and calculated diffraction data. This has distinct advantages over conventional MR strategies which require a cross-rotation signal. An important feature of this method is its capability to simultaneously search the overall rotation/translation of the test model in the unit cell while refining the relative orientation/position of internal subdomains. This identifies molecular replacement solutions which would otherwise be completely missed using just a static model, and greatly improve the signal-to-noise contrast.

1. Introduction

The method of molecular replacement (MR) has become a routine tool for determining crystal structures of macromolecules using models that are closely related in structure. The theory of MR and its application to protein crystallography has been described in numerous reviews (see Rossmann, 1990). As an alternative to traditional isomorphous replacement, the molecular replacement method is an elegant computational approach for establishing a preliminary set of phases. The methodology rests on the understanding that structurally similar molecules have closely related molecular transforms and that differences in transforms primarily reflect changes in the orientation and positions of the molecules. Over the years, algorithms for determining the transformations that relate the unknown and known atomic models have evolved both in terms of speed and accuracy. However, as observed by Brünger (1994), approximations made to increase computational efficiency often come at a cost of diminished accuracy.

Conventional molecular replacement methods are, for the most part, based on the properties of the Patterson function. The information that defines the orientation and position of a known structure in an unknown cell is embedded in the intensities of the diffraction data and can be extracted by superposition of the Patterson functions. The Patterson function, or $|F|^2$ synthesis, when calculated using the crystal structure amplitudes, produces a three-dimensional distribution of vectors. There are two types of vectors: self and cross. Self vectors are produced between atoms within a molecule, and cross vectors occur between the atoms of symmetry-related molecules. A Patterson function, calculated from a known atomic model of a related molecule, will produce a constellation of self vectors that are similar to those of the unknown structure in the crystal, but rotated by some as yet undefined set of angles. To ascertain these angles, a rotation function is used to explore the correlation between the observed and model Patterson functions as a function of the rotation space (Rossmann & Blow, 1962; Huber, 1985; Nordman, 1971; Crowther, 1972; Navaza, 1994).

Determining the orientation of a search model with respect to the observed diffraction data can be accomplished in real space (Huber, 1985), reciprocal space (Rossmann & Blow, 1962), or direct space (Brünger, 1994). These three methods are all useful tools and are formally equivalent, but differ in their computational accuracy and efficiency. The real- and reciprocal-space formulations rely entirely on the Patterson function. Both methods compare the observed Patterson, P_c, with that calculated from a model, P_m. The superposition of the two Patterson functions is computed for different orientations of P_m,

$$\mathrm{Rot}(\Omega, r) = \int_{U(r)} P_c(v) P_m(\Omega, v)\, \mathrm{d}v, \qquad (1)$$

and the similarity of the two Patterson functions is assessed as a function of a rotation matrix, Ω, which is specified by a set of angles that samples the angular space. The variables r and $U(r)$ are the radius and the volume of integration, respectively. To obtain the best signal-to-noise ratio, a value of r is chosen that maximizes the number of self vectors while minimizing the number of cross vectors in the volume of integration.

A direct rotation function, introduced by Brünger (1990), differs from the conventional real- and reciprocal-space methods. Instead of rotating the Patterson of the search model P_m, the orientation of the model itself is changed according to $x_i' = \Omega x_i$. Structure amplitudes