BRÜNGER, A. T. (1988*a*). *X-PLOR Manual, Version* 1.5. Yale Univ., New Haven, USA.
BRÜNGER, A. T. (1988*b*). *J. Mol. Biol.* **203**, 803-816.
BRÜNGER, A. T. (1989). *Acta Cryst.* A**45**, 42-50.
BRÜNGER, A. T., CAMPBELL, R. L., CLORE, G. M., GRONENBORN, A. M., KARPLUS, M., PETSKO, G. A. & TEETER, M. M. (1987). *Science*, **235**, 1049-1053.
BRÜNGER, A. T., CLORE, G. M., GRONENBORN, A. M. & KARPLUS, M. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 3801-3805.
BRÜNGER, A. T., KARPLUS, M. & PETSKO, G A. (1989). *Acta Cryst.* A**45**, 50-61.
CHOTHIA, C. & LESK, A. M. (1986). *EMBO J.* **5**, 823-826.
CLORE, G. M., BRÜNGER, A. T., KARPLUS, M. & GRONENBORN, A. M. (1986). *J. Mol. Biol.* **191**, 523-551.
FEHLHAMMER, H. & BODE, W. (1975). *J. Mol. Biol.* **98**, 683-692.
FRAUENFELDER, H., PETSKO, G. A. & TSERNOGLOU, D. (1979). *Nature (London)*, **280**, 558-563.
FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.
HAUPTMAN, H. (1982). *Acta Cryst.* A**38**, 289-294.
HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107-112.
HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* A**32**, 778-780.
HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.
HUBER, R. (1985). In *Molecular Replacement.* Proc. of the Daresbury Study Weekend, Daresbury, February 1985, pp. 58-61. SERC Daresbury Laboratory, Daresbury, Warrington, England.
KARPLUS, M. & MCCAMMON, J. A. (1983). *Annu. Rev. Biochem.* **52**, 263-300.
LATTMAN, E. E. (1972). *Acta Cryst.* B**28**, 1065-1068.
LATTMAN E. E. (1985). *Methods Enzymol.* **115**, 55-77.
NORDMAN, C. E. (1972). *Acta Cryst.* A**28**, 134-143.
NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747-754.
POWELL, M. J. D. (1977). *Math. Programming*, **12**, 241-254.
RABINOVICH, D. & SHAKKED, Z. (1984). *Acta Cryst.* A**40**, 195-200.
RAO, S. N., JIH, J.-H. & HARTSUCK, J. A. (1980). *Acta Cryst.* A**36**, 878-884.
ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
SCHIRMER, T., HUBER, R., SCHNEIDER, M., BODE, W., MILLER, M. & HACKERT, M. L. (1986). *J. Mol. Biol.* **188**, 651-676.
SHAKKED, Z., RABINOVICH, D., CRUSE, W. B. T., EGERT, E., KENNARD, O., SALA, G., SALISBURY, S. A. & VISWAMITRA, M. A. (1981). *Proc. R. Soc. London Ser. B*, **213**, 479-487.
STEIGEMANN, W. (1974). PhD Thesis. Technische Univ. München, Federal Republic of Germany.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800-804.
TAYLOR, W. R. (1988). *Protein Eng.* **2**, 77-86.
WANG, D., BODE, W. & HUBER, R. (1985). *J. Mol. Biol.* **185**, 595-624.
WÜTHRICH, K. (1986). *NMR of Proteins and Nucleic Acids.* New York: Wiley.

Acta Cryst. (1990). A**46**, 57-68

Direct Phase Determination for the Molecular Envelope of Tryptophanyl-tRNA Synthetase from *Bacillus stearothermophilus* by X-ray Contrast Variation

BY CHARLES W. CARTER JR, KATHERINE V. CRUMLEY, DAVID E. COLEMAN AND FRANK HAGE

Department of Biochemistry, CB No. 7260, *University of North Carolina at Chapel Hill, Chapel Hill, NC* 27599-7260, *USA*

AND GERARD BRICOGNE

LURE, Bâtiment 209*D*, 91405 *Orsay, France*

(*Received* 7 *April* 1989; *accepted* 21 *August* 1989)

Abstract

Monoclinic crystals of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase grown in the presence of substrate trytophan (space group $P2_1$) display evidence of a low-resolution trigonal space group ($P321$). The origin and averaging transformations for the local 32 point group of this unusually clear sixfold non-crystallographic symmetry may be inferred without prior estimation of the electron density. This local symmetry was exploited in conjunction with solvent density contrast variation to determine the shape of the molecular envelope. X-ray intensities measured from crystals equilibrated in mother liquors of three different electron densities were used to estimate three parameters for each reflection: the modulus of the envelope transform, $|G_\mathbf{h}|$; and components, $X_\mathbf{h}$ and $Y_\mathbf{h}$, relative to $G_\mathbf{h}$, of the structure-factor vector for the transform of intramolecular density fluctuations. The moduli $\{|G_\mathbf{h}|\}$ behave somewhat like structure-factor amplitudes from small-molecule crystals, and estimation of their unknown phases was successfully carried out by statistical direct methods. Reflections to 18 Å resolution, which obey rather well the symmetry of space group $P321$, were merged to produce an asymmetric unit in that space group. $|G_\mathbf{h}|$ values for the 34 strongest of these were phased using the small-molecule direct-methods package *MITHRIL* [Gilmore (1984). *J. Appl Cryst.* **17**, 42-46]. The best phase set was expanded back to the $P2_1$ lattice and negative density was truncated to generate initial phases for all reflections to 18 Å

0108-7673/90/010057-12$03.00

resolution. Phase refinement by iterative imposition of the local 32 symmetry produced an envelope with convincing features consistent with known properties of the enzyme. The envelope implies that the tryptophanyl-tRNA synthetase dimer is an elongated structure with an axial ratio of about 4:1, in which the monomers have two distinct domains of unequal size. The smaller of these occurs at the dimer interface, and resembles the nucleotide binding portion of the tyrosyl-tRNA synthetase. It may therefore contain the amino-terminal one hundred or so residues, including all three cysteines, previously suggested to comprise a nucleotide-binding domain in the tryptophanyl enzyme. A purely crystallographic test of the overall features of this envelope was carried out by transporting it to a tetragonal crystal form of the same protein in which the asymmetric unit is a monomer. The small domain fits snugly inside three mercury and one gold heavy-atom binding sites for this crystal form; and symmetry-related molecules provide excellent, but very different, lattice contacts in nearly all directions.

Introduction

Bacillus stearothermophilus tryptophanyl-tRNA synthetase is a dimer of identical 37 000 dalton subunits (Fersht, Ashford, Bruton, Jakes, Koch & Hartley, 1975). It crystallizes in several different space groups, depending only on the bound ligands (Carter & Carter, 1979; Coleman & Carter, 1984; Carter & Coleman, 1984). This crystal polymorphism and complementary fluorescence experiments (Andrews *et al.*, 1985; Merle *et al.*, 1986) suggest that the different crystal forms represent important conformational differences related to catalysis (Carter & Coleman, 1984). This very conformational flexibility has been disadvantageous because heavy-atom binding tends to destroy the isomorphism between parent and derivatized crystals, and it has not yet been possible to phase structure factors for any of the crystal forms by isomorphous replacement. Crystal form type II*, grown in the presence of tryptophan and belonging to space group $P2_1$, has three enzyme dimers in the asymmetric unit. Evidence for trigonal space group $P321$ at low resolution suggests that these three dimers are arranged about a local threefold axis nearly parallel to the crystallographic c axis, with molecular dyads along local twofold directions such that the local 32 point group symmetry of the asymmetric unit is very nearly crystallographic at low resolution. Thus, both origin and averaging transformations for the local 32 point group can be deduced from the diffraction pattern and employed from the outset. This additional source of phase information makes crystal form II* an unusually favorable candidate for direct approaches to phase determination.

Numerous recent articles have put forward the possible role of a maximum-entropy criterion as an aid to direct phase determination (Bricogne, 1982*a*, 1984; Collins, 1982; Wilkins, Varghese & Lehmann, 1983; Piro, 1983; Narayan & Nityananda, 1982; Britten & Collins, 1982). Bricogne has shown in particular that entropy maximization provides a better procedure for calculating joint and conditional probability distributions of structure factors than those used in traditional direct methods (Bricogne, 1984). This is especially so in macromolecular crystallography, where the existence of solvent channels gives rise to strongly non-uniform distributions of atoms (Bricogne, 1988). The latter circumstance was also shown to lead to a strengthening of phase relations which had not been anticipated by direct methods, and which explains in statistical terms the power of solvent flattening at improving phases. These arguments have led us to seek to determine the molecular envelope by experimental means, so as to provide a firm starting point for subsequent direct phase extension. As the initial step in applying this approach to a new structure determination, we have combined the non-crystallographic symmetry in type II* crystals with X-ray density contrast variation (§ 1) to determine phases for the transform of the molecular envelope directly from its experimentally determined amplitudes (§ 2).

1. The contrast-variation method

Low-resolution X-ray intensities are not readily related to the transform of the molecular envelope because the mother liquor in crystals often nearly matches the mean protein density, significantly reducing the contrast at low resolution. Moreover, most macromolecular data-collection procedures are not designed for measuring accurate low-resolution data. As a result, agreement between observed and calculated structure factors for well refined macromolecular structures is nearly always poor for low-resolution reflections (below 10–15 Å), and these are therefore normally omitted from structure refinements and from published structure-factor lists.

Nevertheless, it is possible to enhance the envelope transform signal in low-resolution reflection intensities by modifying the solvent electron density in macromolecular crystals, and to exploit it in the structure solution (Bragg & Perutz, 1952; Harrison, 1969; Jack, Harrison & Crowther, 1975; Harrison & Jack, 1975; Ibel & Stuhrmann, 1975; Roth, Lewit-Bentley & Bentley, 1984; Bentley, Lewit-Bentley, Finch, Podjarny & Roth, 1984; Moras *et al.*, 1983; Podjarny, Rees *et al.*, 1987). The present study is motivated by three precedents:

(1) Bragg & Perutz (1952) showed that diffracted intensities could be changed by altering the salt concentration and hence the electron density of the crystal mother liquor. They noted that changes in centrosymmetric structure-factor amplitudes are

directly proportional to the transform of solvent-accessible volumes in the crystal. Diffraction effects arising from intramolecular texture are not affected by changes in solvent electron density, and occur mainly at higher resolution. Contrast-variation intensity changes therefore provide information only about the macromolecular boundary. The resulting structure-factor differences were interpreted as amplitudes for the transform of an ellipsoidal envelope, and phases were estimated for centric reflections using a minimum-wavelength principle and intensity measurements from shrinkage states.

(2) Harrison and his co-workers supplemented low-resolution contrast-variation data with a three-dimensional reconstruction of the tomato bushy-stunt virus particle obtained by electron microscopy (Harrison, 1969; Jack *et al.*, 1975). The reconstruction represented stain-accessible regions of the particle and hence the envelope itself. Appropriately scaled structure factors calculated by Fourier inversion of the reconstruction provided structure-factor *vector* differences between observed structure factors, whose amplitudes had been measured at different salt concentrations. These were used as heavy-atom structure-factor vectors, and phases were determined for both centric and acentric reflections by analogy with a single-isomorphous-replacement calculation. Phases obtained in this manner were sufficiently good that even minor heavy-atom positions could be correctly deduced from difference Fourier maps (Harrison & Jack, 1975). Electron density maps calculated with these phases and the X-ray amplitudes also revealed several minor artefacts in the three-dimensional reconstructions.

(3) Ibel & Stuhrmann (1975) explicitly related the contrast dependence of solution scattering to basic scattering functions arising separately from the solute shape and its internal fluctuations and showed that these functions could be recovered from a contrast-variation series. Since the neutron scattering cross section of water can be changed dramatically by mixing deuterated water in defined proportions, their analysis has since made it possible to use contrast-variation neutron diffraction measurements in solving macromolecular crystal structures at low resolution (Roth *et al.*, 1984; Bentley, Lewit-Bentley, Finch, Podjarny & Roth, 1984; Moras *et al.*, 1983; Podjarny, Rees *et al.*, 1987). The linear dependence of the complex structure factor on the contrast can be used to scale contrast-variation data sets together and extract relative phase information for the series (Roth *et al.*, 1984). These applications have used phase differences at different contrasts to image separately protein and nucleic-acid components of nucleo-protein complexes.

Our use of contrast variation is a natural extension of these ideas. We have explicitly separated diffraction effects due to the envelope from those due to internal density fluctuations. As will be described in the following section, analysis of the contrast sensitivity of individual reflections gives rise to an extrapolated set of structure-factor amplitudes. Enhanced contrast in these new amplitudes, and elimination of effects due to internal density fluctuations, mean that statistical correlations between them are much stronger than those between observed amplitudes at any contrast. One consequence of using amplitudes from the envelope transform is therefore to increase the effectiveness of direct phasing methods which depend on these correlations. We have determined these amplitudes experimentally using X-ray diffraction, and used direct methods to determine their unknown phases and hence the molecular envelope.

1.1. *The envelope indicator function, χ_U, and its structure factors, $\{G_\mathbf{h}\}$*

The indicator function, χ_U (Bricogne, 1974), of the solvent, excluding volume U is defined to have a value of one within and a value of zero outside this region. A complementary indicator, χ_{V-U}, is similarly defined for the solvent-accessible region, $V-U$, where V is the unit-cell volume. This construction demarcates between a volume containing atoms of the protein or macromolecule, and one containing only rapidly rearranging solvent molecules. In contrast to the well defined positions and low temperature factors which characterize atoms in the macromolecule, solvent atoms have very large temperature factors, and thus generate an approximately uniform electron density. The total electron density can be written as a sum of terms involving convolutions of the two indicator functions with the corresponding Gaussian distributions of the mean protein and solvent-atom excursions. If we let the protein-atom excursions be contained implicitly in the definition of ρ_p, the protein density, we have

$$\rho(\mathbf{x}) = \{\rho_s * [(4\pi/B_s)^{3/2} \exp(-4\pi^2\|\mathbf{x}\|^2/B_s)]\} \times \chi_{V-U} + [\bar{\rho}_p + (\rho_p - \bar{\rho}_p)] \times \chi_U \quad (1)$$

where ρ_s = solvent electron density, ρ_p = electron density within the protein volume, $\bar{\rho}_p$ = mean electron density within the protein volume, B_s = isotropic temperature factor for solvent atoms, and $*$ denotes convolution and $\times$ denotes multiplication.

Complementarity of the two indicator functions implies that $\chi_{V-U} = 1 - \chi_U$, so this expression can be simplified. Inverse Fourier transformation ($\bar{\mathcal{F}}$) of the result gives for the structure factors

$$F_\mathbf{h} = [\rho_s \times V \times \delta_{0,\mathbf{h}}] + [\bar{\rho}_p - \rho_s \times \exp(-B_s S^2)] \times \bar{\mathcal{F}}[\chi_U]_\mathbf{h} + \bar{\mathcal{F}}[(\rho_p - \bar{\rho}_p) \times \chi_U] \quad (2)$$

where $S = (\sin\theta)/\lambda$ and $\delta_{0,\mathbf{h}}$ is the Kronecker delta symbol.

Of the three resulting terms, the first contributes only to the F_{000} term; the third contributes only to the remaining structure factors, and may be simplified to $\bar{\mathcal{F}}(\rho_p - \bar{\rho}_p)$ since the protein density has U as an envelope. Thus, structure-factor vectors become.

$$\begin{aligned} F_{000} &= \rho_s \times V + (\bar{\rho}_p - \rho_s) \times U \\ &= \bar{\rho}_p \times U + \rho_s \times (V - U) \qquad (3) \\ F_{\mathbf{h}\neq 0} &= [\bar{\rho}_p - \exp(-B_s S^2)\rho_s] \times G_{\mathbf{h}} + \Delta_{\mathbf{h}} \end{aligned}$$

where $G_{\mathbf{h}}$ = the Fourier transform of the envelope at $\mathbf{h}$, and $\Delta_{\mathbf{h}}$ = the Fourier transform of the internal density fluctuation $\{\rho_p - \bar{\rho}_p\}$.

Expression (3) shows how experimental measurements of $F^i_{\text{obs},\mathbf{h}}$ at different contrasts (i) can be used to separate the transform of the envelope from that of the internal fluctuations. The coefficient of $G_{\mathbf{h}}$ is the mean density difference between solvent and protein, and is called the contrast (Stuhrmann & Kirste, 1965; Roth *et al.*, 1984; Jacrot, 1976). Simulations of the behavior of quantities related by (3) (Carter, Dumas & Bricogne, unpublished results; Dumas, 1988), described briefly in § 2.3 and to be described in more detail separately, support the following comments:

(i) The transform, $G_{\mathbf{h}}$, of the indicator function determines the magnitude of the solvent-contrast dependence of $F^i_{\text{obs},\mathbf{h}}$. For reflections at which $G_{\mathbf{h}}$ is zero, $F^i_{\text{obs},\mathbf{h}}$ will consist entirely of the fluctuation term, and these will be insensitive to contrast variation. The power spectrum of $|G_{\mathbf{h}}|$ in simulated data sets declines sharply to approximately the scale of the fluctuation term at moderate resolution (5–7 Å).

(ii) The $\{|G_{\mathbf{h}}|\}$ values span a greatly extended dynamic range, compared with the $\{|F^i_{\text{obs},\mathbf{h}}|\}$, as a result of enhanced contrast. Solvent electron densities in a crystal will generally be close to that of the macromolecular component. For example, the mean electron density of a protein is approximately 0·425 e Å^{-3}; that of 4 M ammonium sulfate, often used to stabilize protein crystals for data collection, is 0·408 e Å^{-3}. Under these conditions contributions to $F^i_{\text{obs},\mathbf{h}}$ from diffraction by the envelope component, χ_U, are only about 0·02/0·425 = 5% of the actual $G_{\mathbf{h}}$, whatever that value may be. Large intensity changes for low-order reflections which accompany changes in solvent density indicate that these $|G_{\mathbf{h}}|$ values are very much larger than the corresponding $|F^i_{\text{obs},\mathbf{h}}|$.

(iii) The resolution to which $|G_{\mathbf{h}}|$ can be estimated reliably by experimentally changing the electron density of the mother liquor in equilibrium with the crystal is limited by the fact that the highly thermalized solvent atoms tend to increase the contribution of the envelope transform to $|F^i_{\text{obs},\mathbf{h}}|$ thereby reducing the masking effect of the solvent at higher resolution. From (3), assuming the same temperature factor for all solvent atoms, $\|F^i_{\text{obs},\mathbf{h}}| - |F^j_{\text{obs},\mathbf{h}}\| = \exp(-B_s S^2)(\rho^j_s - \rho^i_s) \times |G_{\mathbf{h}}|$. Thus, contrast-variation differences represent a smaller fraction of $|G_{\mathbf{h}}|$ at higher resolution.

A semi-quantitative assessment of the strength of the contrast-variation signal can be obtained from the resolution dependence of the correlation coefficients between $|F^i_{\text{obs},\mathbf{h}}|$ for different contrasts. This statistic is quite variable at very low resolution, but increases asymptotically to about 1·0 for simulated and to >0·85 for our experimental contrast-variation data sets. For the former, which did not include the effects of high thermal parameters for solvent atoms, the asymptote occurs at about 12·5 Å, whereas for the experimental data it occurs at about 16 Å. In practice, the contrast-variation signal will depend on the accuracy of the intensity data.

(iv) The structure-factor expression (3) provides natural parameters for the extrapolation to zero solvent density. Since components sum vectorially in the complex plane, the observed structure-factor vector, $F^i_{\text{obs},\mathbf{h}}$, can be written as the resultant of three vectors as shown in the Harker construction in Fig. 1(a), where all ignorance of phase is absorbed into the G component by assuming an arbitrary phase of 0° for $G_{\mathbf{h}}$. Two of these are always antiparallel and proportional to $|G_{\mathbf{h}}|$ and separately to the temperature-factor-weighted mean electron densities of protein and solvent compartments. If the components of Δ are X and Y, then the calculated structure-factor amplitude for contrast i becomes

$$\begin{aligned} |F^i_{\text{calc}}| = K'(\{X + [\bar{\rho}_p - \exp(-B_s S^2)\rho^i_s] \\ \times |G|\}^2 + Y^2)^{1/2}. \qquad (4) \end{aligned}$$

The contrast-variation series provides sets $\{|F^i_{\text{obs}}|\}$, from which one can extrapolate to maximum contrast and estimate separately the structure-factor amplitudes for the molecular envelope and the density fluctuations within it. Least-squares refinement against the closure error residual, $\sum_i \sum_{\mathbf{h}} \sigma^{-2}_{\mathbf{h}} \{|F^i_{\text{obs}}| - |F^i_{\text{calc}}|\}^2$, provides a way to estimate both $|G|$ and Δ, to within a sign ambiguity involving the imaginary component Y. If the solvent for any contrast contains anomalous scatterers, the sign of Y can also be estimated from the resulting Bijvoet differences (Fig. 1b).* In practice, it is necessary also to refine scale constants, K', for all but one contrast, which is kept constant.

(v) The parameters $X_{\mathbf{h}}$ and $Y_{\mathbf{h}}$ represent both the magnitude of $\Delta_{\mathbf{h}}$ and the phase difference between $G_{\mathbf{h}}$ and $\Delta_{\mathbf{h}}$. Thus, once phases are found for $|G_{\mathbf{h}}|$, estimates are available for $\Delta_{\mathbf{h}}$. This choice of parameters differs from that given by Worcester & Franks (1976), in which the independent variable is the scattering density of the solvent, rather than the contrast, and for which the 'slope' parameter is $G_{\mathbf{h}}$, and the

* We are grateful to Hal Wyckoff for suggesting the use of anomalous scatterers in the solvent.

'intercept' parameter is the structure factor for uncontrasted protein, F_{prot}:

$$F^i_{obs,\mathbf{h}} = F_{prot,\mathbf{h}} - \rho^i_s \times G_{\mathbf{h}}.$$

With this formulation the fluctuation term, $\Delta_{\mathbf{h}}$, is still embedded in the $F_{prot,\mathbf{h}}$ term; there is no compelling reason to associate the ignorance of phase with either of the two components. Roth *et al.* (1984) obtained relative phase differences between contrasts using this formulation.

1.2. *Phase determination for* $\{|G_{\mathbf{h}}|\}$, *structure-factor amplitudes for the molecular envelope*

Although unphased, the estimated structure-factor amplitudes, $\{|G_{\mathbf{h}}|\}$, represent a very useful transformation of the amplitude data for several reasons. First, they have been resolved from the transform of density fluctuations within the envelope, so they represent a simpler structure. In fact, the complexity of the envelope does not increase dramatically with the molecular weight of the unit-cell contents, because large macromolecules tend to remain globular. Second, because contributions from the protein and solvent regions are antiparallel, reflections with a large sensitivity to contrast have very large $|G_{\mathbf{h}}|$ values, in some cases an order of magnitude larger than the corresponding $|F^i_{obs,\mathbf{h}}|$. Simulations of the behavior of $|G_{\mathbf{h}}|$ with resolution using coordinates from known crystal structures confirm the expectation that the mean power of $|G_{\mathbf{h}}|$, weighted by the square of the resolution, is very much stronger than that of $\Delta_{\mathbf{h}}$ at low resolution. The steep decline in the mean power of $|G_{\mathbf{h}}|$ with increasing resolution implies that the data set $\{|G_{\mathbf{h}}|\}$ is, in some sense, complete at very moderate resolution limits. Finally, since the indicator function is by definition equal to its square, its sampled transform satisfies a Sayre equation (Sayre, 1952).

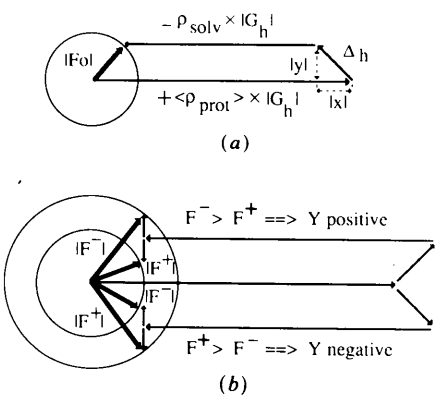


Fig. 1. Structure-factor vector relationships in the contrast-variation method. (*a*) The important 'masking' effect of the solvent regions, and the relationship between parameters *G*, *X* and *Y*, estimated by program *GFROMF*. (*b*) Anomalously scattering solvents can help determine the sign of the imaginary component, *Y*, of the fluctuation term for acentric reflections.

The set of structure factors $\{|G_{\mathbf{h}}|\}$ therefore should exhibit similar properties - simplicity, large normalized moduli, and connectedness through Sayre's equation - to those that make small-molecule data sets amenable to solution by traditional direct methods. These parallels between the set $\{|G_{\mathbf{h}}|\}$ and structure-factor data to atomic resolution for a small molecule underlie our motivation to use contrast variation in this way. Our plan for phase determination separates into two distinct tasks:

(i) Phasing the $\{|G_{\mathbf{h}}|\}$ to determine the molecular envelope, using the fact that they behave similarly to structure factors for a small molecule.

(ii) Phasing the observed structure-factor amplitudes using estimated phases available from the $\{\Delta_{\mathbf{h}}\}$ (Fig. 1*a*); restrictions imposed by the envelope; and, in the present case, phase restrictions due to the sixfold noncrystallographic symmetry.

2. Results

2.1. *Determination of the origin and expression of averaging transformations for the local* 32 *point group*

Type II* crystals have unit-cell parameters $a = 157{\cdot}7$, $b = 88{\cdot}4$, $c = 90{\cdot}8$ Å, $\beta = 91{\cdot}7$ °. The crystallographic asymmetric unit contains three enzyme dimers, arranged with 32 point-group symmetry so as to produce an approximately trigonal lattice with *P*321 space-group symmetry at low resolution, on a unit cell similar to that originally described for type II crystals (Carter & Carter, 1979). These noncrystallographic symmetry elements were verified by examining the three-dimensional self-rotational autocorrelation function, and the native Patterson map. The self-rotation function shows a local threefold axis nearly parallel to **c** as well as local twofolds near the *ab* plane (Figs. 2*a*,*b*). The native Patterson shows a pseudo origin peak at (0·5, 0·5, 0) relating local 32 symmetry groups (Fig. 2*c*). This peak is elongated toward +**u** and +**w**, indicating that the vector between local origins is either slightly shorter or slightly longer than (0·5, 0·5). The two different vectors represent alternate choices for the origin. Both origin and orientation can therefore be specified using only this information. Averaging transformations were obtained by specifying an origin (0·245, 0·5, −0·012) and orientation for the 32 point group, and using symmetry operations from the space group *P*321.

The solvent-accessible volume, estimated from the molecular weight, is about 55% of the asymmetric unit. The noncrystallographic symmetry and large solvent-accessible volume of these crystals offer substantial phasing power. Phasing power has been estimated several different ways. It is certainly inversely related to the fractional volume of the chemical asymmetric unit, and has been estimated as $(NV/U)^{1/2}$,

where N is the redundancy due to noncrystallographic symmetry (Bricogne, 1974, 1976). For these crystals this value is about 3·6, representing the attenuation of root-mean-square noise expected from each cycle of averaging refinement. Arnold & Rossmann (1986) have suggested the criterion $P = [(Nf)/R \times V/U]^{1/2}$ as a measure of potential map quality, where f is the fraction of the data collected and R is an index of the accuracy of the structure-factor amplitudes. With this relationship, and the assumption of an R factor of 2·3% (see Table 1), this P criterion can be estimated to be about 240 for this crystal form, about twice the value for the most favorable case cited by Arnold & Rossmann.

2.2. *Contrast-variation data sets - equilibration and data collection* (*Crumley*, 1989)

Equilibration of type II* crystals with alternate solutions is complicated by their high equilibrium solubility (2·8 mg ml^{-1}) in their native mother liquor (1·8M KH_2PO_4, prepared from a 2:1 mixture of mono- and dibasic salts to achieve a pH of 7·5; 2 mM tryptophan, 5 mM MgOH, 0·1 mM phenylmethane sulfonyl fluoride, 14 mM β-mercaptoethanol). We therefore prepared protein-saturated mother liquors of different electron densities for several solvents

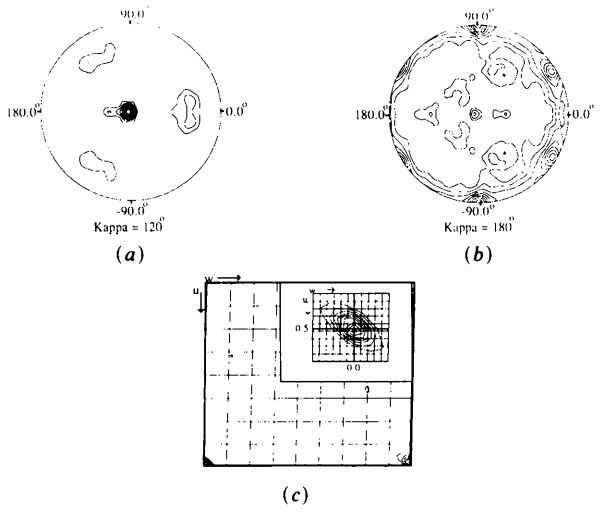


Fig. 2. Origin and placement of the local 32 point group in type II* crystals. (*a*) Self-rotation function polar angle map, section $\kappa = 120°$, showing the orientation of the local threefold axis approximately 5° away from $\mathbf{c}^*$, which is at the pole. (*b*) Self-rotation polar angle map, section $\kappa = 180°$, showing the orientation of the three local twofold axes, also rotated about 5° from the horizontal. The direction $\varphi = 90°$ corresponds to the crystallographic b axis. (*c*) The native Patterson map, section $v = 0·5$, showing the pseudo origin peak that arises from vectors between atoms at different local ($P321$) origins. The insert shows the region around the peak itself, and demonstrates that the peak is elongated along the uw diagonal, consistent with the fact that the a axis in type II* crystals is slightly longer than the long diagonal of the $P321$ unit cell.

before transferring crystals. Crystal stabilization time and prior history of solvent transfers are also important variables. Successful equilibration of crystals with alternate mother liquors required the following steps. Crystals grown in 1·8 M KH_2PO_4 were first stabilized by a 3-4 week incubation in 2·4 M KH_2PO_4. Then they were transferred individually into sitting drops of the new mother liquor. Several such transfers were made before mounting crystals for data collection. We have used X-ray data for three solvents: 2·25 M Li_2SO_4, 0·5 M $NH_4CH_3CO_2$ ($\rho_s = 0·387$), 2·4 M KH_2PO_4 ($\rho_s = 0·412$), and 4·1 M $(NH_4)_2SeO_4$ ($\rho_s = 0·455$). Transfer to the selenate solvent required prior equilibration to phosphate-buffered 3·5 M $(NH_4)_2SO_4$ lacking β-mercaptoethanol.

To avoid corrupting low-resolution reflections by shadowing, a beam stop was constructed from stainless steel tubing 1·6 mm in diameter, filled with a concave lead insert. This beam stop was placed 4·0 cm from the crystal and translated to a slightly eccentric position, in order that reflections with resolution greater than 95 Å would be free of its penumbra. A single multiwire area detector (San Diego Multiwire Systems, San Diego, CA) was used to record images. Alignment of the crystal and detector were determined using software developed by L. Weissmann, and images were integrated using *PROFL*, a program with three-dimensional graphic presentation of selected reflection profiles written by F. Hage. Virtual films within a data set were scaled using the Fourier surface scaling program of Weissmann (1982). Statistics for these data sets are assembled in Table 1.

2.3. *Scaling and estimation of* $\{|G_\mathbf{h}|\}$: *program GFROMF*

Contrast-variation data sets were simultaneously placed on a common scale and used to estimate $|G|$, $|X|$ and $|Y|$ using a program called *GFROMF* because it produces $\{|G_\mathbf{h}|\}$ from the $\{|F^i_{\mathrm{obs},\mathbf{h}}|\}$ (Carter & Bricogne, 1987). *GFROMF* incorporates two levels of non-linear least-squares refinement. The scheme is similar to that used for multiple-isomorphous-replacement phasing. An inner loop over reflections carries out refinement of local parameters G, X and Y from scaled data by least-squares minimization of the closure-error residual defined using (4), summed over the contrasts. The algorithm used for global scale constant refinement in the outer loop involves modifying the partial derivatives of calculated structure factors to reflect entrainment of these values by changes in the scale constants, as put forward for MIR parameter refinement by Bricogne (1982*b*).

To verify its functionality, procedures were developed to generate simulated data sets from published coordinates for solved structures. Electron density maps were generated from protein coordin-

Table 1. *Summary of contrast-variation data sets*

Data set name	Solvent electron density (e Å^{-3})	Bijvoet differences	% asymmetric unit	Resolution limits (Å)	R_c* (% on intensities)	Observations	Unique reflections
22Y87	0·387		95	95–5·3	4·50	28197	9143
9D86	0·412		100+	95–3·8	4·18	60656	25092
18S87	0·455	+	95	95–5·3	14·30	5636	1428

* $R_c = \sum_i \sum_h |\langle I(h)\rangle - I_i(h)| / \sum_i \sum_h \langle I(h)\rangle$.

ates in the absence of contrasting solvent (the F_{prot} map) (CCP4, 1986). These maps were processed to generate maps of the indicator, χ_U (by replacing values above a threshold by 1 and those below it by 0 while effecting a smoothing of the boundary to preserve the relationship $\chi_{V-U} = 1 - \chi_U$), and the fluctuation density {by replacing values inside the envelope by $[\rho(x) - \langle\rho_{prot}\rangle]$} after accumulating the necessary statistics from the maps (Carter, Dumas & Bricogne, unpublished; Dumas, 1988). Fourier inversion of these maps gave simulated values for $|G_h|$, $|X_h|$ and $|Y_h|$. These simulations verified quantitatively the relationships illustrated in Fig. 1(*a*). The F_{prot} map was then modified by replacing solvent pixels with a constant electron density and applying the same smoothing algorithm. Maps were generated for solvent electron densities of 0·36, 0·38, 0·40, 0·42 and 0·44 e Å^{-3} and inverted to give simulated $\{|F^i_{obs}|\}$. Program *GFROMF* correctly rescaled misscaled contrast-variation data sets, and faithfully recovered $|G_h|$, $|X_h|$ and $|Y_h|$ from the simulated $\{|F^i_{obs}|\}$ to within the precision of the simulations.

With the type II* experimental X-ray contrast-variation data, the scaling algorithm proved to be stable and convergent. Repeated use of the program with various combinations of the measured contrast-variation data in Table 1 and other sets measured subsequently have given $\{|G_h|\}$ sets which scale to each other with R factors of about 20%. S^2-weighted radial intensity distributions of the scaled data sets are compared as a function of the resolution in Fig. 3.

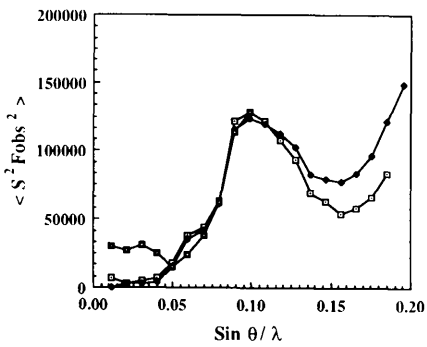


Fig. 3. Spherically averaged intensity distributions of the scaled data sets used to estimate $\{|G_h|\}$ values. Values in each bin were weighted by the square of the mean reciprocal-lattice vector magnitude, $(d^*)^2$, and thus represent the mean scattering power. □ 22Y87; ◆ 9D86; ■ 18S87. Data sets have been scaled with scale constants determined by *GFROMF*.

As expected, many $|G_h|$ values estimated by *GFROMF* are very much larger than any of the corresponding observed F^i_{obs} values. Two examples showing the effect of contrast variation together with estimated $|G_h|$ and Δ_h values are shown in Fig. 4. In both cases the $|G_h|$ values are about an order of magnitude greater than the largest $|F^i_{obs}|$.

2.4. *Phase determination*

Results described below represent a successful application of this approach to the task of determining the molecular envelope to 18 Å resolution. Initial phase determination for the $\{|G_h|\}$ proceeded as follows:

(i) Reflections with integral trigonal indices were reduced from the native $P2_1$ lattice and merged to a single asymmetric unit in the low-resolution space group $P321$. $|G_h|$ values were estimated from the

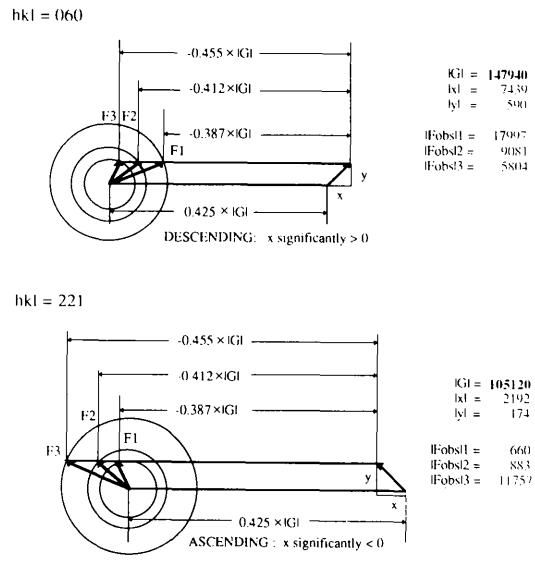


Fig. 4. Examples of the behavior of two acentric reflections from the type II* contrast-variation data sets. Harker diagrams are shown for the 060 and 221 reflections. Both reflections are intense and both have strong sensitivity to contrast. The two reflections show opposite dependences on contrast, because the real component of the fluctuation term is large and of opposite sign in the two cases.

trigonally averaged asymmetric units of the contrast-variation series to a resolution of 18 Å, at which the monoclinic indexing reduced to trigonal indexing with merging R_{sym} values less than 0·25. Beyond this resolution, the merging R_{sym} rose steeply to more than 0·45.

(ii) Normalization of these amplitudes by conventional methods based on assuming a random distribution of atoms is inappropriate, due to unique characteristics of $|G_h|$ values, which were summarized in a previous section. Nevertheless, the simulated data sets provided the basis for an empirical calibration for the expected magnitudes of the largest unitary structure factors, and an absolute G_{000} for the experimental data set was estimated from the volume occupied by protein in the type II* unit cell and the mean electron density of protein. Together, these values were used to estimate the absolute scale of the $\{|G_h|\}$. Unitary structure factors obtained using these estimates, without exponential correction, were then converted to normalized structure factors using an effective number of atoms determined by the condition that it should produce a reasonable number of $|G_{norm}|$ values greater than 1·0, and a r.m.s. $|G_{norm}|$ value as close as possible to 1·0. Strict adherence to the latter criterion failed to yield enough large normalized magnitudes to permit phase determination, but a modest relaxation (r.m.s. $|G_{norm}|$ of about 1·5) succeeded.

The effective number of atoms corresponds roughly to the number of group scatterers evident at low resolution, analogous to the 'globs' considered by Harker (1953). The pronounced peak in the radial intensity distributions shown in Fig. 3 provides an independent gauge for the volume of these scatterers, and hence for their number. It is consistent with an average diameter of about 10·5 Å, or a group volume of about 600 Å^3. Roughly 75 such scatterers would be necessary to fill the expected volume of the protein subunit, leading to an $N^{1/2}$ value of 8·7. In fact, we used a value of $N^{1/2} = 6·3$ in the normalization procedure. This surprisingly good agreement confirms the earlier suggestion (Bricogne, 1984) that structural hierarchy in proteins will lead to a strengthening of phase relations at low resolution; it also places the effective number of scatterers in the range (<80 atoms in the $P321$ asymmetric unit) for which traditional direct methods have proven effective.

(iii) 34 reflections with $|G_{norm}| > 1·0$ were phased by convergence mapping and tangent-formula refinement of 32 different phase sets using the direct-methods package *MITHRIL* (Gilmore, 1984) with the default options. Of these phase sets, seven had combined figures of merit > 2·8. The corresponding electron density maps were examined for chemically reasonable features. The most important of these features included a globular density distribution removed from the local threefold axis, but close

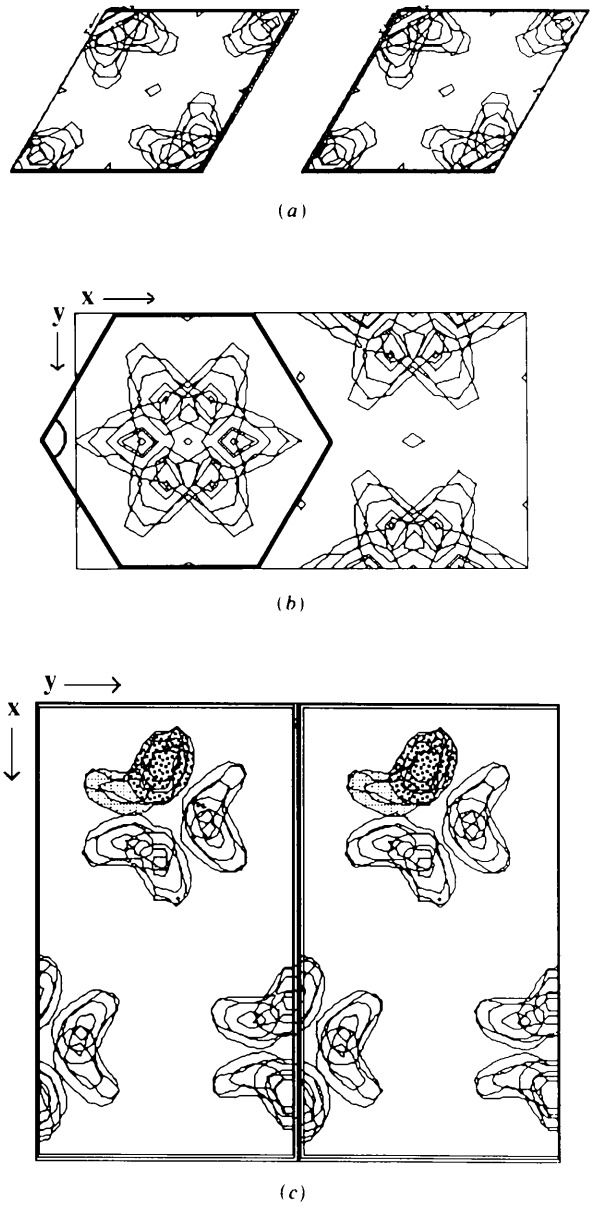


Fig. 5. Evolution of the type II* envelope electron density. All three maps show sections $z = -1/12$ to $1/12$ in $1/12$ths. Contours shown range from $0·7\sigma$ to 10σ in increments of 2σ. (*a*) The map based on 34 reflections, using averaged structure factors in space group $P321$, and phased by convergence mapping and tangent-formula refinement. (*b*) The same map after expanding reflections to the native monoclinic ($P2_1$) unit cell. Note the nearly absolute symmetry in interactions of each 'subunit' with both its neighbors, *i.e.* the map is completely ambiguous with respect to the enzyme molecular dimer interaction. The hexagon represents the envelope boundary, within which averaging was carried out. The circular wedge was omitted from each vertex of the included volume for the first averaging cycle. (*c*) The noncrystallographic-symmetry-averaged map, based on 206 reflections to 18 Å resolution. Note that the dimer interaction, which was ambiguous, is clearly indicated by stacking of density from the upper darkly-shaded monomer directly on the lower lightly-shaded monomer, and by the absence of contact between dimers in the asymmetric unit.

Table 2. *Comparison between structure factors before and after phase refinement by sixfold noncrystallographic symmetry averaging*

Number of reflections	*R* factors		Correlation coefficients		Mean phase changes(°)	
	Initial	Final	Initial	Final	Initial	Final
65*	0·45	0·19	0·44	0·95	20	44
141†	0·60	0·42	0·57	0·81	66	81
206‡	0·51	0·28	0·63	0·95	52	69

* Reflections expanded to $P2_1$ from the 34 strongest reflections phased initially by *MITHRIL* in space group *P*321.
† Reflections phased initially by truncating negative densities.
‡ Overall totals for all reflections.

enough to it to form convincing dimer contacts between peaks related by the crystallographic twofold directions. This requirement is imposed by the fact that the enzyme is a dimer while the asymmetric unit (in *P*321) contains a single monomer. The seven phase sets included three pairs which gave identical electron densities, further limiting the selection. Of the four remaining maps, two were rejected out of hand because density was concentrated on a threefold axis. The other two had substantially similar density distributions, differing only in the relative density at different locations. One of these showed an arrangement of electron density near the crystallographic twofold axis in *P*321 that could be interpreted as a dimer interaction. This electron density distribution had the highest combined figure of merit and was most consistent with the known chemical properties of the enzyme, including the expected dimer interaction. It was therefore selected for subsequent phase development by rotational averaging in space group $P2_1$ (Fig. 5*a*).

(iv) Phased reflections at this stage were expanded to the monoclinic lattice (Fig. 5*b*).

(v) All negative density values were reset to zero using the absolute scale and G_{000} values determined in (ii), and the resulting map was inverted to provide phases for all 206 reflections on the monoclinic lattice to 18 Å resolution.

(vi) Calculated structure-factor amplitudes from this map were replaced by the corresponding observed $|G_h|$ values. An envelope corresponding to the entire volume of the crystallographic asymmetric unit was constructed by sectioning a prism with a hexagonal cross section bounded by the X axis and running from $Z = -0{\cdot}5$ to 0·5 (Fig. 5*b*). In the first averaging cycle, a fraction of a small cylinder was omitted from the edges of this prism. Seven cycles of rotational averaging within this envelope (Bricogne, 1976) were applied to produce a map with an *R* factor of 0·28 for agreement between observed and averaged amplitudes (Fig. 5*c* and Table 2).

2.5. *Features of the synthetase envelope in type* II* *crystals*

A balsa-wood model constructed using the lowest significant contours of the 18 Å map is shown in Fig. 6. Although no information beyond the local-symmetry-averaging transformations was introduced during phase determination, this envelope map is strikingly informative. The following factors argue that the envelope we have determined has many correct features:

(i) The six monomers in the asymmetric unit are arranged around the local origin in such a way that they can form dimers across the local twofold axes. 32 symmetry was imposed, but no constraint was placed on where density should appear. This feature was inherent in the choice of the 34-reflection trigonal phase set with the best combined figure of merit, and, although less clearly apparent, is also characteristic of the second-best phase set.

(ii) Monomers are paired unambiguously such that there are three continuous regions of electron density, corresponding to dimers, about the local threefold axis. This feature was not apparent in the initial 34-reflection map; it emerged decisively after noncrystallographic averaging (Fig. 5*c*). At the current resolution, the actual boundary of each monomer cannot be resolved, either with respect to the putative dimer interface or with respect to the packing contacts between molecules along the c axis.

(iii) The dimer has an overall length of about 150 Å and a girth of about 35 Å for an axial ratio of about 4·3, consistent with its limiting diffusion coefficient in dilute solution (4–5:1; unpublished measurements). The volume of the model accounts for about 45% of the unit-cell volume, in keeping with expectation that six dimers with a hydrated volume of $1{\cdot}0\ \mathrm{cm^3\ g^{-1}}$ will occupy 0·48 of the unit cell.

(iv) Each end of the dimer associates with the tips of five other monomers from other local 32 point groups. Together with the dimer interaction, these lattice contacts provide 'periodic bond chains' accounting for crystal growth in all principal directions. Although the dimer contacts are unambiguous, those between the 'tips', while extensive, are not as well delineated in the map. This property is consistent with the fact that these interactions connect unit cells along the c axis, which assumes at least two slightly different β angles in type II* and type II crystals (Carter & Carter, 1979).

(v) Monomers appear to contain two 'domains' of unequal size. The dimer interface occurs between the

smaller domains, which account for about 30% of its mass. The larger domains at the tips account for about 70% of the mass. The enzyme is susceptible to limited proteolysis, which produces two unequal fragments of approximate $M_r = 25\,000$ and 12 000 (Winter, Hartley, McLachlan, Lee & Muench, 1977; Coleman, 1988). The smaller fragment contains the amino terminus, the larger fragment contains the carboxy terminus of the polypeptide (Winter *et al.*, 1977).

The model (Figs. 5*c* and 6) is distinctly noncentrosymmetric, owing to the disposition of the 'neck' connecting the two domains. The initial 34-reflection map was essentially centrosymmetric, so this property of the electron density was induced during the sixfold averaging in space group $P2_1$, probably by the tilt of the local threefold axis with respect to the envelope. We cannot, of course, determine the correct molecular enantiomorph from the available data.

2.6. *Transport of the envelope to another crystal form of the same protein* (*Coleman*, 1988)

A proper analysis of the correctness of this direct-phase determination will be presented once the answer is known with certainty from a higher-resolution structure. However, the unusual crystal polymorphism of the tryptophanyl-tRNA synthetase already provides an exacting and purely crystallographic test for the validity of this envelope. The same envelope, probably with some conformational changes, should be compatible with the different crystal forms we observe for differently ligated forms of the protein. We have transported the envelope to the unit cell of type IV crystals, a tetragonal form [space group $P4_12_12$ or its enantiomorph; $a = b = 60{\cdot}9$, $c = 234{\cdot}4$ Å (Carter & Coleman, 1984)] in which the asymmetric unit is a monomer. Molecular and crystallographic symmetry restrictions in this cell limited the degrees of freedom for the search to a rotation about the crystallographic dyad, and a translation along it. The orientation of the envelope in the type IV cell was fixed by the use of cross-rotation function maps. The translation along the dyad was fixed by fitting the envelope to one gold and three mercury heavy-atom sites which have recently been determined for type IV crystals using anomalous scattering for the gold and successive derivatization of the non-isomorphous gold derivative with mersalyl to form an isomorphous pair (Coleman, 1988). As shown in Fig. 7, the smaller domain of the envelope fits nicely to the four heavy-atom sites. Three of the four sites lie precisely on the surface of the envelope, the fourth lies within 3 Å of it.

Despite a profound difference in unit-cell size and shape, space-group symmetry and lattice contacts, the envelope affords a convincing packing arrangement that fills the unit cell with few packing incompatibilities. In contrast to the cylindrical interwound packing in type II* crystals, packing in type IV crystals occurs in layers of parallel dimer molecules, elongated in a direction close to the *ab* plane, each of which rotates by 90° along the *c* axis (Coleman, 1988). The sole packing conflict occurs between the distal tip of one molecule and the domain at the dimer interface of a neighboring molecule along **a** and **b**. This contact could be resolved by a slight hinge bending motion between the two domains. Conformational differences are expected between the two crystal forms (Andrews *et al.*, 1985; Merle *et al.*, 1986).

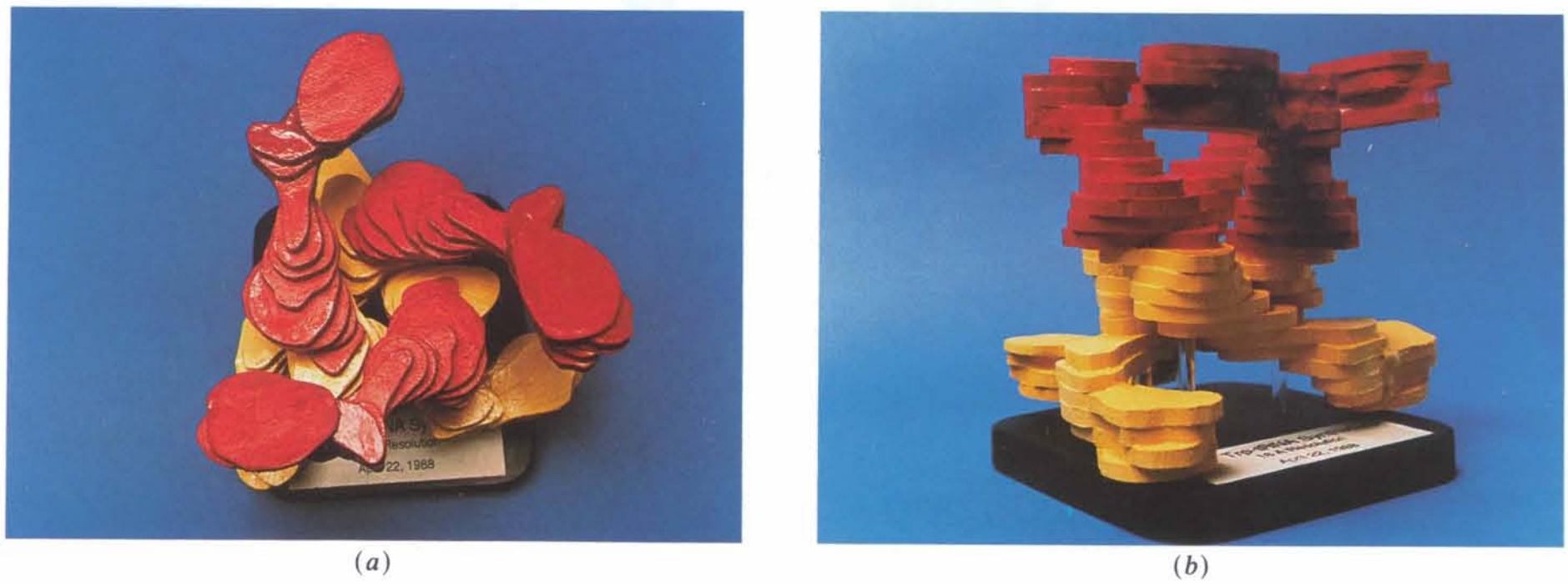
3. Discussion

3.1. *Comparison with other synthetases*

The mass distribution in the enzyme is similar to that in the tyrosyl-tRNA synthetase (YRS), which also has a large domain close to its dimer interface, and for which the dimer has an overall length of about 130 Å (Irwin, Nyborg, Reid & Blow, 1976). The actual length of the tyrosyl enzyme is not currently known because disorder in the crystal prevents visualization of about 100 amino acids at the C terminus, which would presumably increase its length. In YRS a large domain near the interface contains a nucleotide binding site with numerous connecting loops (Blow *et al.*, 1983). These connecting loops account for most of the additional mass of YRS, compared with that of tryptophenyl-tRNA synthetase (WTS). The overall organization of WTS, judged from primary sequence alignments (Webster, Lathrop & Smith, 1987) suggests that it also contains a mononucleotide binding fold at its amino terminus, as does YRS. The fit of the envelope to the type IV crystal mercury sites also suggests that the small domain contains the only three cysteines in the sequence and hence that it represents the segment thought to contain the nucleotide binding fold (Winter *et al.*, 1977; Webster *et al.*, 1987). Thus, the envelope map at this resolution is entirely consistent with what is known of its primary structure and those of other synthetases. Just how extensively this apparent similarity at low resolution persists when the path of the main chain emerges for the tryptophanyl enzyme will be an important question.

3.2. *Experimental determination of a macromolecular envelope using direct methods*

This work constitutes a genuine extension of the use of amplitude information directly in *ab initio* protein crystal structure determination. The approach used here differs substantially from direct phasing methods used previously for macromolecules (Bentley *et al.*, 1984; Moras *et al.*, 1983; Podjarny, Rees *et al.*, 1987), which used refinement of



(*a*) (*b*)

Fig. 6. Photographs of a balsa-wood model constructed from contours of the 18 Å electron density map of the molecular envelope from type II* crystals. Views are approximately down the symmetry directions of the non-crystallographic 32 point group, and the model represents an asymmetric unit containing three enzyme dimers. Each dimer contains one red and one yellow subunit. (*a*) View approximately down the local threefold axis. The extended tips of the six monomers in the asymmetric unit lie about 60 Å from this axis, whereas the putative dimer interface lies close to it, giving rise to a pronounced six-pointed 'star of David' in the *hk*0 projection. (*b*) View approximately down a local twofold axis. The dimer is elongated by a factor of about 4:1 and is somewhat curved, giving rise to the appearance of coiling of the dimers to form the crystallographic asymmetric unit. The trimeric association of enzyme dimers has not been observed in solution, and probably occurs only under conditions of crystallization.

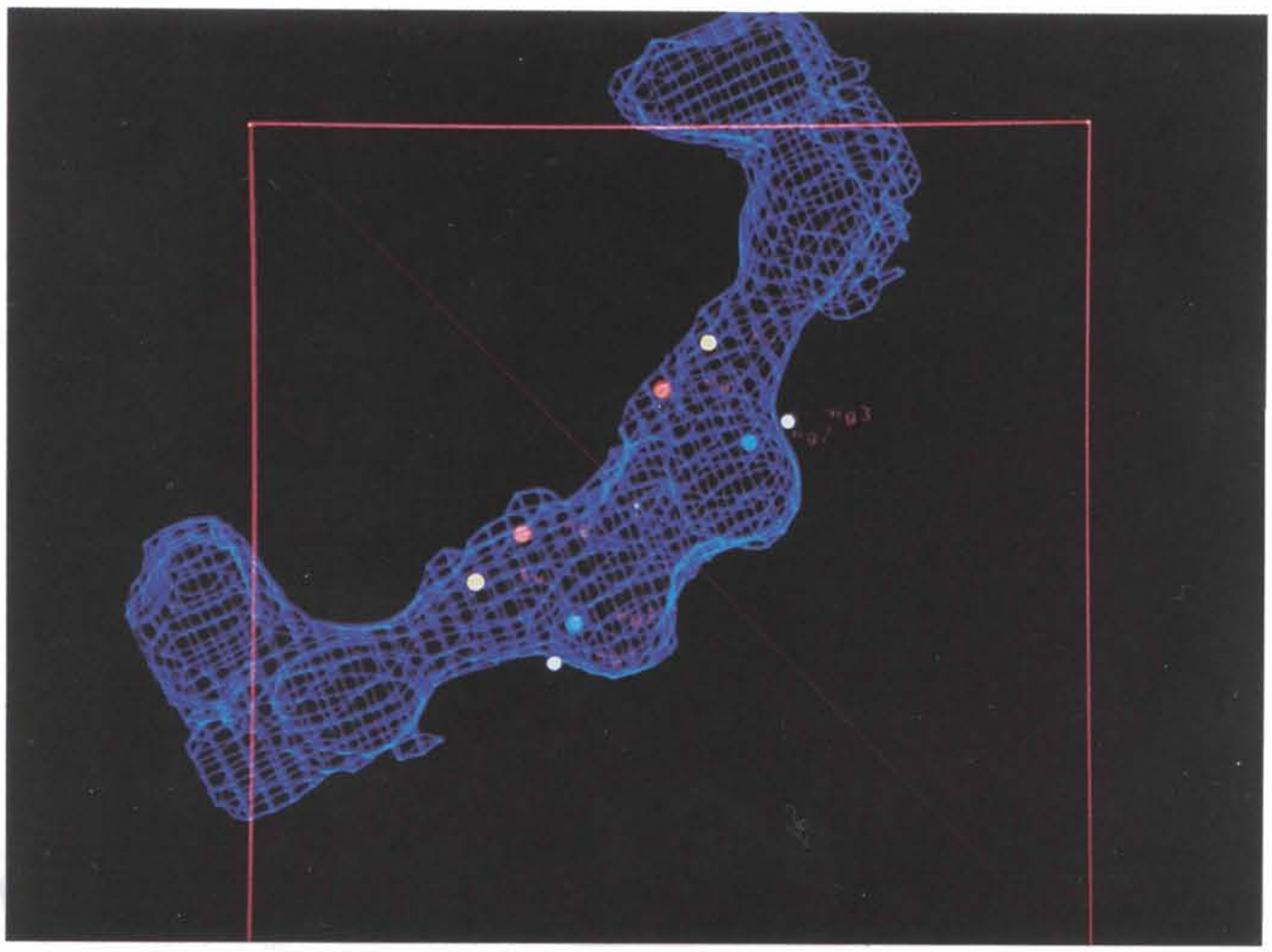


Fig. 7. The molecular envelope of the tryptophanyl-tRNA synthetase dimer obtained for type II* crystals fitted to four heavy-atom sites in the type IV crystal unit cell. The square box outlines a region in the *ab* plane containing four 60·9 Å type IV unit cells. The diagonal indicates a crystallographic and molecular twofold axis at $Z = 1/2$. The positions of three mercury atoms (red, blue, and white) and a single gold (yellow) are indicated on each subunit.

[*to face page* 66

pseudoatoms against the native X-ray or neutron structure factors. They also used a contrast which masked one of two components in the unit cell, which permitted molecular replacement with a known model. The following features are unique to our approach:

(i) We have used experimental X-ray intensity measurements to determine envelope amplitudes, $\{|G_{\mathbf{h}}|\}$. This approach differs markedly from other methods for envelope determination, all of which require some prior estimation of the electron density and hence of phases for the observed structure factors. Although multiple data sets were measured, their purpose was not to generate phase information by triangulation, but to transform the amplitudes into a set representing a simpler structure at higher contrast, and which therefore is considerably more robust with respect to the behavior of statistical direct methods.

(ii) Phase determination was model-independent and hence freer of built-in biases. The only sources of phase information used in the procedure were the statistical correlations between the amplitudes themselves, and the non-crystallographic symmetry. No assumptions were made regarding the shape of the chemical asymmetric unit, *e.g.* that it was ellipsoidal or composed of ellipsoidal 'pseudoatoms'. The envelope within which density was averaged coincided approximately with the crystallographic asymmetric unit, and thus introduced minimal *a priori* information regarding where the protein was likely to be located. The resulting map therefore resembles a protein, not an aggregate of spheres. It is distinctly *not* ellipsoidal, nor is it easy to envisage a comparable map based on any combination of pseudoatoms.

(iii) Although the resolution of the map represented in Fig. 5 is modest by comparison with the majority of new macromolecular structures, we are now in a position to apply the phase-determination algorithms proposed by Bricogne (1982*a*, 1984, 1988).

3.3. *Extension of this approach to direct-phase determination*

Finally, two comments should be made with regard to extending this approach to higher resolution:

(i) We have phased reflections thus far more or less in order of their importance: the 34 phased reflections from tangent formula extrapolation are indeed the most important in defining the molecular shape. They are not necessarily the strongest low-resolution reflections, but are experimentally the most sensitive to contrast variation. Their phases will exert a strong influence on all phases yet to be determined by constraining them to introduce new structural detail only within the envelope. Knowledge of the envelope and the averaging transformations should provide considerable phasing power for phase extension (Arnold & Rossmann, 1986; Podjarny, Bhat & Zwick, 1987).

(ii) Although the number of reflections in a shell increases with the cube of the resolution, this increase is compensated by the increasing detail of the molecular envelope, which permits increasingly large steps for phase extension (Arnold & Rossmann, 1986). Entropy maximization has been shown to be very powerful once a structure is known to moderately high resolution. One such refinement for the crambin structure at 3 Å resolution extrapolated correct phases and amplitudes to 1·5 Å resolution (Bricogne, 1984). Similar successes have been reported recently in the resolution range 5-3 Å (Prince, Sjölin & Svensson, 1988). Experiments are now under way using entropy maximization to extend phases for $\{|G_{\mathbf{h}}|\}$ beyond 18 Å.

This work was supported by NIGMS RO1-26203 and by a Fogarty Senior International Fellowship (TW0-1961). CWC Jr is grateful to the Laboratoire pour Utilisation du Rayonnement Electromagnetique, Orsay, France for their hospitality during tenure of this fellowship. The authors acknowledge Dr C. J. Gilmore for helpful discussions throughout this work, and especially for assistance using *MITHRIL*. The X-ray data collection for this work formed the Master's thesis research project of KVC, and the fitting of the envelope to the type IV crystal form was part of the Doctoral thesis research of DEC.

References

ANDREWS, D., TREZEGUET, V., MERLE, M., GRAVES, P.-V., MUENCH, K. H. & LABOUESSE, B. (1985). *Eur J. Biochem.* **146**, 201-209.

ARNOLD, E. & ROSSMANN, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489-5493.

BENTLEY, G., LEWIT-BENTLEY, A., FINCH, J. T., PODJARNY, A. D. & ROTH, M. (1984). *J. Mol. Biol.* **176**, 55-75.

BLOW, D. M., BHAT, T. N., METCALF, A., RISLER, J.-L., BRUNIE, S. & ZELWER, C. (1983). *J. Mol. Biol.* **171**, 571-576.

BRAGG, W. L. & PERUTZ, M. F. (1952). *Acta Cryst.* **5**, 277-289.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395-405.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.

BRICOGNE, G. (1982*a*). In *Computational Crystallography*, edited by D. SAYRE, pp. 258-264. Oxford: Clarendon Press.

BRICOGNE, G. (1982*b*). In *Computational Crystallography*, edited by D. SAYRE, pp. 223-230. Oxford: Clarendon Press.

BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410-445.

BRICOGNE, G. (1988). *Acta Cryst.* A**44**, 517-545.

BRITTEN, P. L. & COLLINS, D. M. (1982). *Acta Cryst.* A**38**, 129-132.

CARTER, C. W. & CARTER, C. W. JR (1979). *J. Biol. Chem.* **254**, 12219-12223.

CARTER, C. W. & COLEMAN, D. E. (1984). *Fed. Proc.* **43**, 2981-2983.

CARTER, C. W. JR & BRICOGNE, G. (1987). *GFROMF: a Computer Program for Scaling and Estimating Envelope Structure Factors from Contrast Variation Data.* Department of Biochemistry, CB No. 7260, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA.

CCP4 (1986). The Cooperative Computing Project in Crystallography. SERC Daresbury Laboratory, Warrington, England.

COLEMAN, D. E. (1988). PhD thesis. Univ. of North Carolina at Chapel Hill, USA.
COLEMAN, D. E. & CARTER, C. W. JR (1984). *Biochemistry*, **23**, 381-385.
COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.
CRUMLEY, K. V. (1989). MS thesis. Univ. of North Carolina at Chapel Hill, USA.
DUMAS, C. (1988). Thèse DSN. Univ. de Paris-Sud, Centre D'Orsay, France.
FERSHT, A. R., ASHFORD, J. S., BRUTON, C. J., JAKES, R., KOCH, G. L. E. & HARTLEY, B. S. (1975). *Biochemistry*, **14**, 1-4.
GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42-46.
HAGE, F. (1986). *PROFL: a Computer Program for Integration and Display of Virtual Area Detector Films.* Department of Biochemistry, CB No. 7260, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA.
HARKER, D. (1953). *Acta Cryst.* **6**, 731-736.
HARRISON, S. C. (1969). *J. Mol. Biol.* **42**, 457-483.
HARRISON, S. C. & JACK, A. (1975). *J. Mol. Biol.* **97**, 173-191.
IBEL, K. & STUHRMANN, H. B. (1975). *J. Mol. Biol.* **93**, 255-265.
IRWIN, M. J., NYBORG, J., REID, B. R. & BLOW, D. M. (1976). *J. Mol. Biol.* **105**, 577-586.
JACK, A., HARRISON, S. C. & CROWTHER, R. A. (1975). *J. Mol. Biol.* **97**, 163-172.
JACROT, B. (1976). *Rep. Prog. Phys.* **39**, 911-953.
MERLE, M., TREZEGUET, V., GRAVES, P. V., ANDREWS, D., MUENCH, K. H. & LABOUESSE, B. (1986). *Biochemistry*, **25**, 1115-1123.
MORAS, D., LORBER, B., ROMBY, P., EBEL, J.-P., GIEGE, R., LEWIT-BENTLY, A & ROTH, M. (1983). *J. Biomol. Struct. Dynam.* **1**, 209-223.
NARAYAN, R. & NITYANANDA, R. (1982). *Acta Cryst.* A**38**, 122-128.
PIRO, O. E. (1983). *Acta Cryst.* A**39**, 61-83.
PODJARNY, A., BHAT, T. N. & ZWICK, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351-374, and references contained therein.
PODJARNY, A., REES, B., THIERRY, J. C., CAVARELLI, J., JESIOR, J. C., ROTH, M., LEWIT-BENTLEY, A., KAHN, R., LORBER, B., EBEL, J. P., GIEGE, R. & MORAS, D. (1987). *J. Biomol. Struct. Dynam.* **5**, 187-198.
PRINCE, E., SJOLIN, L. & SVENSSON, L. A. (1988). *Acta Cryst.* A**44**, 218-222.
ROTH, M., LEWIT-BENTLEY, A. & BENTLEY, G. A. (1984). *J. Appl. Cryst.* **17**, 77-84.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.
STUHRMANN, H. B. & KIRSTE, R. G. (1965). *Z. Phys. Chem.* **46**, 247-270.
WEBSTER, T. A., LATHROP, R. H. & SMITH, T. F. (1987). *Biochemistry*, **26**, 6950-6957.
WEISSMANN, L. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 56-63. Oxford: Clarendon Press.
WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* A**39**, 47-60.
WINTER, G. P., HARTLEY, B. S., MCLACHLAN, A. D., LEE, M. & MUENCH, K. H. (1977). *FEBS Lett.* **82**, 348-350.
WORCESTER, D. L. & FRANKS, N. P. (1976). *J. Mol. Biol.* **100**, 359-368.

Acta Cryst. (1990). A**46**, 68-72

R Factors in X-ray Fiber Diffraction. III. Asymptotic Approximations to Largest Likely *R* Factors

BY R. P. MILLANE

The Whistler Center for Carbohydrate Research, Smith Hall, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 13 *June* 1989; *accepted* 21 *August* 1989)

Abstract

The largest likely R factor is useful for evaluating the significance of R factors obtained in structure determinations, and is smaller in fiber diffraction than in traditional crystallography. Very simple approximations to functions used to calculate the largest likely R factor in fiber diffraction are derived. For example, the largest R factor (R_m) for m overlapping terms is very well approximated by $R_m \simeq (2/\pi)^{1/2} m^{-1/2}$. These are a useful alternative to the exact, but quite complicated, expressions derived previously. More significantly, they provide insight into the behavior of R factors in fiber diffraction and may be useful in further analysis.

1. Introduction

The largest likely R factor (that for a structure uncorrelated with the correct structure) is a useful yardstick for evaluating the significance of R factors obtained in structure determinations. The largest likely R factor for single crystals was determined by Wilson (1950) and has recently been determined for fiber diffraction (Stubbs, 1989; Millane, 1989*a*, *b*). R factors in fiber diffraction are generally smaller than in single-crystal analyses because the diffraction pattern is cylindrically averaged. The R factor depends on the number of overlapping complex Fourier-Bessel structure factors at different positions in reciprocal space, and therefore on the diameter and symmetry of the diffracting particle and the maximum resolution of the diffraction data. The largest likely R factor in fiber diffraction, while easily calculated, is a rather complicated expression involving special functions (Millane, 1989*b*), making its interpretation obscure. Approximations to largest likely R factors in fiber diffraction are derived here by developing asymptotic approximations to components of this expression.

0108-7673/90/010068-05$03.00