The Application of Molecular-Symmetry Constraints to the Constrained–Restrained Parameter Structure-Factor Least-Squares Refinement Program *CORELS*

By A. G. W. Leslie

Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BZ, England

Abstract

A method is described to apply molecular symmetry constraints to the crystallographic refinement of a macromolecule at low resolution using the constrained–restrained parameter structure-factor least-squares refinement program *CORELS* [Sussman, Holbrook, Church & Kim (1977). *Acta Cryst.* A**33**, 800–804]. The parameter shifts calculated by *CORELS* are adjusted to ensure that the refined structure possesses exactly the desired molecular symmetry. This procedure has been used to refine the structure of apo-D-glyceraldehyde 3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 4 Å resolution, imposing 222 molecular symmetry. The starting model was the partially refined structure of the holoenzyme at 2·7 Å resolution. The possible advantages of using *CORELS* refinement when dealing with more general problems involving two distinct molecular conformations which are related by simple rigid-body domain movements are also considered.

Introduction

The structure-factor least-squares refinement program *CORELS* (Sussman *et al.*, 1977) is one of a number of computer programs currently available for the refinement of macromolecular structures. However, *CORELS* and the Gauss–Seidel least-squares procedure of Hoard & Nordman (1979) are the only programs which offer the facility of treating specified atoms as rigid or constrained groups in the refinement. This procedure reduces the number of refined parameters and consequently increases both the rate and the radius of convergence (Scheringer, 1963). It is therefore particularly well suited to the refinement of macromolecules at low resolution (Bragg spacings greater than 4 Å). At this resolution, even the incorporation of stereochemical information in the form of parameter restraints, as implemented in the Hendrickson–Konnert (Hendrickson & Konnert, 1980) or Jack–Levitt (Jack & Levitt, 1978) refinement programs, is not sufficient to provide an acceptable data-to-parameter ratio when individual atomic parameters are allowed to vary.

If the molecule possesses non-crystallographic symmetry, a further improvement in the data-to-parameter ratio can be achieved by constraining or restraining the parameters so that the appropriate non-crystallographic symmetry is maintained. This is a routine feature of the refinement of polynucleotide or polysaccharide structures using the linked-atom least-squares refinement program *LALS* (Smith & Arnott, 1977), where helical symmetry is imposed by means of distance *constraints*. The Hendrickson–Konnert refinement program also allows non-crystallographic symmetry *restraints* to be applied in a very general way.

This paper describes a method of applying molecular symmetry *constraints* to low-resolution refinement using the *CORELS* program. The symmetry of the starting model is preserved by a suitable averaging of the parameter shifts calculated for the individual subunits (or rigid atomic groups) related by the molecular symmetry elements.

The method has been applied to the 4 Å resolution refinement of apo-D-glyceraldehyde 3-phosphate dehydrogenase (apo-GAPDH) from *B. stearothermophilus*, a tetramer of four chemically identical subunits, total molecular weight 145 000. The apoenzyme crystallizes in space group $P2_1$ with the entire tetramer in the crystallographic asymmetric unit. The structure of the holoenzyme (Biesecker, Harris, Thierry, Walker & Wonacott, 1977), with four molecules of nicotinamide adenine dinucleotide (NAD) bound to the tetramer, has been partially refined at 2·7 Å resolution (Wonacott, 1983). This partially refined structure was used as the starting model in the refinement of the apoenzyme.

The method

There are several ways in which non-crystallographic symmetry can be incorporated into a least-squares refinement procedure. One approach, that adopted in the *LALS* refinement program, is to apply molecular symmetry *constraints* through the use of Lagrange undetermined multipliers. As pointed out by Hendrickson & Konnert (1980), when applied to protein structures this method has the disadvantage that any real deviations from molecular symmetry will be completely repressed. In addition, the refined parameters will give no indication of possible errors

in the initial definition of the molecular symmetry operations.

An alternative procedure, as implemented by Hendrickson & Konnert, is to minimize the deviations of each subunit from some 'average' structure. The average structure is obtained by applying the molecular symmetry operations that bring all the subunits into coincidence. The molecular symmetry restraints can then be simply expressed in terms of distance restraints. In those cases where the position of the molecular symmetry axes is not defined by the crystallographic symmetry, the symmetry operations can be refined in successive cycles by finding the transformations which optimally superimpose the subunits. However, it may be necessary to impose additional constraints such as the mutual orthogonality of the three twofold axes in a system with 222 molecular symmetry. The disadvantage of this procedure is that the form of the symmetry restraints is such that they inevitably also restrain the atomic parameters to their current values. This is not usually a serious problem in Hendrickson–Konnert refinement since the atomic shifts are generally rather small. Furthermore, because the data-to-parameter ratio is generally rather unfavourable, it is more important to protect the refinement against excessive parameter shifts than to maximize the rate of convergence. In a *CORELS* refinement, however, the data-to-parameter ratio is much greater, and the refinement is correspondingly more stable. In these circumstances, restraining the model to its current position is likely to hinder seriously the rate of convergence, particularly in the early stages of refinement when large shifts (greater than 3 Å) may be required.

This difficulty can be overcome by applying the molecular symmetry constraints (in this case) to the parameter shifts themselves, rather than including them in the least-squares minimization, and this is the basis of the method described here.

In what follows, 222 molecular symmetry will be assumed whenever an explicit symmetry is required. The extension to other symmetries is quite straightforward.

The following three axial systems are required.

(1) The crystallographic frame. Coordinates in this frame are expressed as fractional crystallographic coordinates and are denoted x.

(2) The intermediate Cartesian frame. This is an orthogonal frame based on the crystallographic axes according to the convention of Rossmann & Blow (1962). Coordinates in this frame, denoted X_g, are expressed in ångströms. For each rigid group in the refinement, the coordinates of all atoms in the group are expressed relative to an origin at the centre of gravity of the group.

(3) The molecular orthogonal frame. This orthogonal set of axes is based on the molecular symmetry axes. In the case of 222 molecular sym-

metry, they simply correspond to the three twofold symmetry axes. Coordinates in this frame, denoted X_M, are also expressed in ångströms.

The relationship between the coordinates in the molecular orthogonal frame, X_M, and the fractional crystallographic coordinates, x, is defined by

$$x = [U][P] X_M + t_M, \qquad (1)$$

where $[U]$ is the matrix converting ångström coordinates in the intermediate Cartesian frame to fractional crystallographic coordinates; $[P]$ is a rotation matrix defining the orientation of the molecular orthogonal frame with respect to the intermediate Cartesian frame and t_M is the vector defining the position of the origin of the molecular orthogonal frame with respect to the origin of the crystallographic frame.

The coordinates of the starting model are expressed in the molecular orthogonal frame and will normally correspond to a refined structure of a different form of the enzyme (*e.g.* in a different state of ligation by substrate or cofactor). This axial frame is convenient because the coordinates are independent of the space group or cell dimensions of the model crystal structure. In addition, in the case of 222 molecular symmetry, there is a trivial relationship between the coordinates of different subunits. Since the molecular symmetry constraints are applied to parameter shifts rather than the actual coordinates, it is essential that the starting coordinates possess exactly the desired molecular symmetry. The coordinates are converted to fractional crystallographic coordinates using (1) prior to refinement with *CORELS*.

In *CORELS*, the refined rigid-body parameters are a translation vector t_g and three rotation angles φ, θ, ρ, which are defined to be successive rotations about the z axis, the new x axis and the new y axis (Sussman *et al.*, 1977).

Fractional   crystallographic   coordinates   are expressed as

$$x = t_g + [U][R]X_g, \qquad (2)$$

where X_g are the starting coordinates of a rigid group of atoms expressed in ångströms in the intermediate Cartesian frame and relative to the centre of gravity of the group. t_g is the vector defining the centre of gravity of the group in fractional crystallographic coordinates, and $[R]$ is the rotation matrix corresponding to the rotations φ, θ, ρ defined earlier. For the starting model, the rotations φ, θ, ρ are all zero, and hence $[R]$ is the unit matrix. However, it should be noted that the t_g will not, in general, be zero.

If the origin and orientation of the molecular symmetry axes with respect to the crystallographic axes are known, then the rigid-body parameters t_g and $[R]$ can be transformed so that the rotations and translations are applied to coordinates expressed in the molecular orthogonal frame.

In this frame we may write:

$$\mathbf{X}_{M'} = \mathbf{T}_M + [S_M]\mathbf{X}_M. \qquad (3)$$

$\mathbf{X}_{M'}$ and $\mathbf{X}_M$ are the coordinates in ångströms of the refined and starting models, respectively, $\mathbf{T}_M$ is the translational shift vector (in ångströms) and $[S_M]$ is the rotation matrix, all with respect to the molecular orthogonal frame. There will be a separate pair of parameters $\mathbf{T}_M$ and $[S_M]$ for each rigid body, but in this description the atomic coordinates are not referred to the centre of gravity of the group. The initial model thus corresponds to zero translation vectors $\mathbf{T}_M$ and unit matrices $[S_M]$.

It can be shown that the relationship between the rotation matrices $[R]$ in (2) and $[S_M]$ in (3) is given by:

$$[S_M] = [P]^{-1}[R][P], \qquad (4)$$

where $[P]$ is defined in (1).

The translational shift vector $\mathbf{T}_M$ in (3) is given by

$$\mathbf{T}_M = [P]^{-1}\{[U]^{-1}(\mathbf{t}_{g'} - \mathbf{t}_M) - [R][U]^{-1}(\mathbf{t}_g - \mathbf{t}_M)\}. \qquad (5)$$

Here, $\mathbf{t}_{g'}$ is the refined centre-of-gravity vector for a given group, and other symbols are as defined above.

Using (4) and (5) the rotation matrices and translation vectors for rigid groups that are related by the molecular symmetry can be expressed in the molecular orthogonal frame. If the molecular symmetry of the starting model is to be maintained, there must be a simple relationship between these different matrices and vectors. This relationship can be demonstrated most clearly by considering a specific example. For atoms in subunit '1' in a 222 symmetric system we may write (3) as

$$\mathbf{X}_{1'} = \mathbf{T}_1 + [S_1]\mathbf{X}_1 \qquad (6)$$

and similarly for atoms in subunit '2':

$$\mathbf{X}_{2'} = \mathbf{T}_2 + [S_2]\mathbf{X}_2. \qquad (7)$$

Since the starting model has 222 symmetry, we may write

$$\mathbf{X}_2 = [M_{12}]\mathbf{X}_1, \qquad (8)$$

where $[M_{12}]$ is the molecular symmetry operation relating subunit two to subunit one. If the refined model is to have the same symmetry we require

$$\mathbf{X}_{2'} = [M_{12}]\mathbf{X}_{1'}. \qquad (9)$$

Substituting (8) and (9) in (7) we get

$$[M_{12}]\mathbf{X}_{1'} = \mathbf{T}_2 + [S_2][M_{12}]\mathbf{X}_1$$

or

$$\mathbf{X}_{1'} = [M_{12}]^{-1}\mathbf{T}_2 + [M_{12}]^{-1}[S_2][M_{12}]\mathbf{X}_1. \qquad (10)$$

Comparing (6) and (10), we find that the conditions for the refined model to possess 222 symmetry are

$$\mathbf{T}_1 = [M_{12}]^{-1}\mathbf{T}_2$$

and

$$[S_1] = [M_{12}]^{-1}[S_2][M_{12}].$$

We may therefore define a reduced rotation matrix and vector as

$$[S_R] = [M]^{-1}[S][M] \qquad (11)$$

and

$$\mathbf{T}_R = [M]^{-1}\mathbf{T} \qquad (12)$$

for each of the rigid groups related by the molecular symmetry where $[M]$ is the molecular symmetry operation relating each group to the corresponding group in some arbitrarily chosen base subunit.

In the case of 222 molecular symmetry, there are four matrices $[M]$, the unit matrix and three corresponding to 180° rotations about the three twofold axes. The matrices are then:

$$[M_{11}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad [M_{12}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$[M_{13}] = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$[M_{14}] = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If the refined model is to obey the molecular symmetry, then these reduced rotation matrices and translation vectors should be identical for all the rigid groups related by the molecular symmetry.

In practice, however, there will be differences in these parameters due to errors in the model and in the data. The most expedient way of imposing the molecular symmetry on the refined model is to force the reduced matrices and vectors to be identical. In the case of the translation vectors, the vector for all groups related by the molecular symmetry can be set equal to the simple average of the individual vectors. The rotation matrices $[S_R]$ can be decomposed into the three rotation angles φ_R, θ_R, ρ_R and these angles can be similarly averaged and the 'best' rotation matrix computed using the averaged values. The averaged parameters can then be transformed back into the *CORELS* frame using (4), (5), (11) and (12), and then applied to the starting model to generate a refined model with perfect molecular symmetry.

Because the dependence of the final atomic coordinates on the rotation angles φ, θ, ρ is non-linear, it is strictly invalid simply to average these angles over different subunits. However, because φ, θ, ρ represent shifts from the starting model, their values will be small (typically less than 5°) and therefore a simple averaging will not introduce serious errors.

In the case of 222 molecular symmetry, the parameter averaging may introduce a systematic error, because it will eliminate any component in the shifts that represent either a translation of the molecular origin or a rotation of the molecular twofold axes with respect to the crystallographic axes. These shifts must therefore be reintroduced *after* the parameter averaging has been performed. This can be achieved using the refinement procedure outlined below.

(1) The model coordinates in the molecular orthogonal frame are converted to fractional crystallographic coordinates using (1). These coordinates are used to generate rigid-group parameters according to (2).

(2) The rigid-group parameters are refined with *CORELS*, and a new set of fractional crystallographic coordinates is calculated. These refined coordinates are used to refine the orientation matrix $[P]$ and vector $\mathbf{t}_M$ so as to minimize, in a least-squares sense, the overall deviation of the tetramer from exact 222 symmetry. The function minimized is

$$\Omega = \sum_{\text{subunits}} \sum_{\text{atoms}} (\mathbf{X}_M^{\text{OBS}} - \mathbf{X}_M^{\text{CALC}})^2,$$

where $\mathbf{X}_M^{\text{OBS}}$ are the coordinates of an atom in the molecular orthogonal frame calculated from the crystallographic coordinates using the current values of $[P]$ and $\mathbf{t}_M$ and $\mathbf{X}_M^{\text{CALC}}$ are the average coordinates of this atom. The average is taken over the four subunits, taking into account the differences in sign of these coordinates in the different subunits. The updated matrix $[P]$ and vector $\mathbf{t}_M$ are then used to average the refined *CORELS* parameters $\mathbf{t}_g$ and $[R]$ imposing 222 molecular symmetry, which will eliminate those components due to translation of the molecular origin or rotation of the molecular symmetry axes.

(3) A new set of fractional coordinates is calculated from the starting model using (1), but now using the updated parameters $[P]$ and $\mathbf{t}_M$. This will re-introduce the components of the shifts eliminated in the averaging. When the averaged *CORELS* parameters are applied to this new set of fractional coordinates, a refined set which includes components arising from a shift or rotation of the molecular axes will be obtained.

The refinement of $[P]$ and $\mathbf{t}_M$ normally converges very rapidly (within two or three cycles), after which it is no longer necessary to recalculate the fractional crystallographic coordinates.

Two additional features of this general approach are worthy of mention. Firstly, because the molecular symmetry constraints are not applied in the calculation of the parameter shifts by *CORELS*, it is possible to examine the calculated shifts for evidence of any significant deviation from the presumed molecular symmetry. This is particularly straightforward when the parameters are expressed in the molecular frame.

In the case of apo-GAPDH, it was of some importance to detect if the tetramer was behaving as a dimer of dimers rather than with strict 222 symmetry. Secondly, because the refined parameters are defined to be a set of rotations and translations to be applied to the starting-model coordinates, it is trivially easy to incorporate a new (and presumably improved) set of starting coordinates should one become available. This feature was employed twice during the refinement of apo-GAPDH as described in the following section.

Application to the 4 Å resolution refinement of apo-D-glyceraldehyde 3-phosphate dehydrogenase

The structure of *B. Stearothermophilus* holo-GAPDH, a tetramer of MW 145 000 made up of four chemically identical subunits has previously been determined at 2·7 Å resolution (Biesecker *et al.*, 1977). A schematic representation of one subunit is shown in Fig. 1. A 6 Å resolution electron density map of the apoenzyme with phases calculated from two isomorphous derivatives (Wonacott & Biesecker, 1977) indicated a significant conformational change with respect to the holoenzyme structure, characterized by a rotation of the entire coenzyme-binding domain away from the active site. In order to investigate the detailed differences in structure between the apo- and holoenzymes
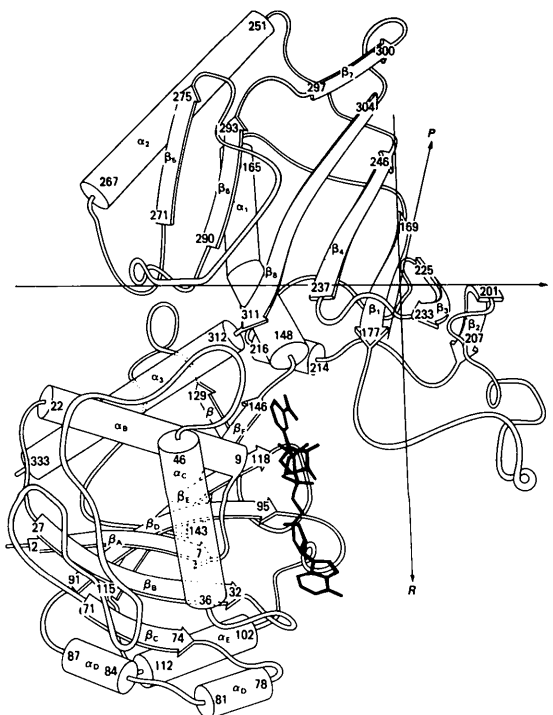

Fig. 1. A schematic representation of the structure of one subunit of *B. stearothermophilus* holo-GAPDH. [Reprinted by permission from *Nature* (*London*), Vol. 266, No. 5600, pp. 328–333. Copyright © 1977 Macmillan Journals Limited.]

Table 1. *A summary of the progress of the refinement of apo-GAPDH using CORELS*

Strict 222 molecular symmetry was imposed up to cycle 23.

Cycles	Resolution of data (Å)	Number of reflections	Number of rigid groups in each subunit	Number of refined parameters	Crystallographic residual (%)		Comments
					Initial	Final	
1–6	10 to 6	3098	3	74	44·5	39·4	Rigid groups are co-enzyme-binding domain, catalytic domain and C-terminal helix
7–8	10 to 6	3098	4	98	39·4	39·3	Co-enzyme-binding domain divided into two parts at residue 77
9–11	10 to 6	3098	5	122	39·3	38·9	Helices α_D and α_E treated as a new rigid group
12–14	10 to 6	3098	5	122	37·6	37·0	Introduce an improved set of holoenzyme coordinates for starting model following a rebuild of this structure
15–16	10 to 5	6056	5	122	39·1	37·4	Only significant changes are in scale and temperature factors
17–20	10 to 4	12 746	10	242	37·5	34·2	Coenzyme-binding domain divided into eight rigid groups (Table 2)
21–23	10 to 4	12 746	10	242	32·5	32·3	Introduce improved holoenzyme coordinates following refinement of rebuilt model at 2·7 Å resolution
23–25	10 to 4	12 746	10	242	32·3	31·6	Remove molecular symmetry constraints

diffraction data to 2·5 Å resolution have been collected from crystals of apo-GAPDH.

In view of the results of the low-resolution study, the problem appeared well suited to a rigid-body-type refinement, using the partially refined holoenzyme structure as a starting model.

Evidence of molecular symmetry for the apoenzyme was derived from an analysis of the positions of heavy-atom binding sites in the tetramer which indicated excellent 222 molecular symmetry (Leslie & Wonacott, 1983).

The orientation and origin of the molecular twofold axes derived from the heavy-atom-site analysis was used to position the holoenzyme structure in the apoenzyme unit cell, and this served as a starting model for refinement. Initially, 3098 reflections between 10 and 6 Å resolution were included, and each subunit was divided into three rigid groups; the coenzyme-binding domain (residues 0–147), the catalytic domain (residues 148–311) and the C-terminal helix (residues 312–333). This gave a total of 74 refined parameters, or a data-to-parameter ratio of better than 40:1.

In spite of the excellent data-to-parameter ratio, the predicted shifts for the coenzyme-binding domain (which shows the largest movement), varied appreciably for different subunits. This was consistent with a 6 Å resolution difference electron density map, calculated with phases derived from the starting model, which displayed large variations in the magnitude of quasi-equivalent features in different subunits,

in addition to many features in the solvent region. On averaging the difference electron density map over the four subunits (Bricogne, 1976), the overall appearance of the electron density improved dramatically. There were many fewer small spurious peaks occurring without corresponding positive or negative mates and the solvent region was now almost featureless. This suggested that the imposition of 222 symmetry on the *CORELS* refinement might improve both the stability of the refinement and the rate of convergence.

Strict 222 symmetry was therefore imposed on the parameter shifts following every cycle of *CORELS* refinement. The orientation matrix $[P]$ and the molecular origin vector t_M were also redetermined after every cycle, although they did not change significantly after the second cycle. As the refinement proceeded, the number of groups in each subunit was increased, and the resolution of the data was extended to include 12 746 reflections between 10 and 4 Å. The progress of the refinement is summarized in Table 1. At cycle 12, the original holoenzyme coordinates were replaced by a new set resulting from a minor rebuild, and at cycle 21 these were in turn replaced by coordinates of the refined rebuilt model, which gave a crystallographic residual of 28·3% at 2·7 Å resolution for the holoenzyme. On both occasions the crystallographic residual for the apoenzyme structure decreased appreciably, demonstrating the considerable similarity in the secondary structures of the two forms of the enzyme.

Table 2. *Definitions of the rigid groups used in the final cycles of the CORELS refinement of apo-GAPDH at 4Å resolution*

CORELS Group	Secondary structural units	Residue numbers
1	$\beta_A, \beta_B, \beta_C$	0–8, 23–34, 71–77
2	α_B	9–22
3	α_C	36–52
4	β-structure	53–70
5	α_D	78–88
6	α_E	101–113
7	$\beta_D, \beta_E, \beta_F$	89–100, 114–120, 142–147
8	$\beta_E-\beta_F$ loop	121–141
9	Catalytic domain	148–311
10	C-terminal helix	312–333

The choice of how to cut each subunit into independent rigid bodies was made using several different criteria. The difference electron density maps were examined for evidence of concerted movement of one or a group of secondary structural elements (α helices or strands of β sheet). The general topology of the subunit was also inspected to see which parts of the structure might be expected to move together as a rigid group. The final choice of groups was then subjected to the limitations that no group was smaller than twenty residues when refining at 6 Å resolution, or ten residues when refining at 4 Å resolution. The rigid-body groups used in the final stages of the refinement are listed in Table 2.

The constrained refinement converged to give a final crystallographic residual of 31·3% for all data to 4 Å resolution. In order to detect any possible asymmetry in the real structure, a further two cycles of refinement were performed without applying the molecular symmetry constraints. This refinement reduced the crystallographic residual to 32·6% at 4Å resolution, and an application of Hamilton's (1965) significance test shows that this reduction is significant at greater than 99·5% confidence. An analysis of the deviations of the coordinates from perfect 222 symmetry and a difference electron density map for the refined asymmetric structure suggests that the asymmetry is due to the presence of a residual amount of the NAD cofactor bound preferentially to one of the four subunits. The NAD content of the crystals used was estimated to be less than 0·2 molecules of NAD per tetramer, but if this residual NAD shows ordered binding to one subunit, as is observed for crystals of GAPDH containing approximately one molecule of NAD per tetramer (Leslie & Wonacott, 1983), the resulting conformational change in that subunit could easily account for the observed asymmetry.

In order to assess the quality of the final symmetric (cycle 23) model for the apoenzyme, structure factors were calculated to 3·5 Å resolution. These structure factors gave a crystallographic residual of 35·0% for 19 557 reflections between 10 and 3·5 Å. A difference

electron density map was calculated at 3·5 Å resolution and averaged (Bricogne, 1976) over the four subunits. The largest features on this map were large negative peaks on arginine, lysine and glutamic acid side chains (presumably the result of thermal or positional disorder) and positive peaks at the known inorganic and substrate phosphate positions, which, in common with the holoenzyme structure, clearly represent ordered sulphate ions. Other features reflected errors in the starting holoenzyme model, since they were also present in a 2·7 Å resolution difference map of that structure. Additional features presumably reflect genuine differences in the internal structures of the rigid groups used in the apoenzyme refinement.

Stereochemical restraints were included in the *CORELS* refinement of the apoenzyme, but with such low weight that they had no significant effect on the least-squares minimization. In spite of this, the final model displayed good stereochemistry, with only six short contacts in error by more than 0·5 Å from the dictionary values. All of these involved atoms at the ends of side chains. The r.m.s. deviation in bond length for the 52 peptide bonds broken in dividing the structure into rigid groups was 0·36 Å, with a maximum deviation of 1·11 Å.

The apoenzyme structure is related to the holoenzyme by an approximate rigid-body rotation of the entire coenzyme-binding domain by 5° away from the catalytic domain. The r.m.s. shift of α-carbons in the coenzyme-binding domain is 1·2 Å, and the largest α-carbon r.m.s. shift for a rigid group was 2·3 Å for the helix α_D. The features of the structure will be described elsewhere (Leslie & Wonacott, 1984).

A comparison of constrained and unconstrained refinement of apo-GAPDH at 6 Å resolution

An assessment of the effectiveness of the molecular symmetry constraints in the *CORELS* refinement of apo-GAPDH is complicated by the discovery that, at 4 Å resolution, the crystallized species investigated is certainly asymmetric, although it would appear that this is due to the presence of residual bound NAD rather than inherent asymmetry of the apoenzyme. However, the degree of asymmetry observed is relatively small [the r.m.s. difference in α-carbon coordinates of the symmetric (cycle 23) and asymmetric (cycle 25) models is only 0·13 Å, with a maximum value of 0.52 Å].

A good indication of the effect of imposing molecular symmetry constraints can therefore be obtained by comparing both the course of the refinement and the final parameter values when refining both with and without the molecular symmetry constraints. This test was carried out using data to 6 Å resolution, and with each subunit divided into three rigid groups as detailed in the previous section. The

original refinement at 6 Å was extended to nine cycles, in order to be certain of convergence, and these nine cycles were then repeated without the application of molecular symmetry constraints.

The variation in some of the rigid-group parameters for the coenzyme binding domain, which shows the largest movement, is shown as a function of cycle number for the constrained and unconstrained refinements in Figs. 2(a) and (b). The angles plotted here are the reduced rotations corresponding to the rotation matrix [S_R] as defined in (11), and therefore these angles will be identical for the four subunits if perfect 222 symmetry is present. The same parameters for the C-terminal helix, the smallest rigid group in the refinement, are presented in Figs. 2(c) and (d).

The spread in parameter values for different subunits in the unconstrained refinement is quite significant, particularly for the C-terminal helix. In addition, the behaviour of the C-terminal helix in the unconstrained refinement is quite erratic, and even the much larger coenzyme-binding domain is more stable in the constrained refinement. The differences between subunits observed in the unconstrained refinement do not resemble the asymmetry of the refined 4 Å asymmetric structure (cycle 25), and do not therefore have any real significance.

A quantitative estimate of the accuracy of the two sets of refined parameters was obtained by using them to calculate two sets of refined atomic coordinates. These were then compared with the 'best' set of coordinates resulting from the 4 Å refinement (cycle 25). The results of this comparison are presented in Table 3, and indicate that the constrained refinement yields a coordinate set closer to the final model than the unconstrained refinement, particularly for the C-terminal helix.

Table 3. *The r.m.s. differences in α-carbon positions between the final set of coordinates from cycle 25 of the CORELS refinement and the two sets of coordinates obtained at 6Å resolution by refining with and without molecular symmetry constraints*

The differences are shown individually for the three rigid groups used in the 6 Å refinement and for the complete molecule.

| | R.m.s. difference in α-carbon position (Å) | |
Group	Constrained refinement	Unconstrained refinement
Coenzyme-binding domain (residues 0–147)	0·43	0·47
Catalytic domain (residues 148–311)	0·26	0·32
C-terminal helix (residues 312–333)	0·27	0·66
Complete molecule (residues 0–333)	0·33	0·39

Discussion

The results of the 6 Å refinement with and without molecular symmetry constraints clearly demonstrate that the application of symmetry constraints to the refinement of apo-GAPDH produces a better molecular model (as judged by comparison with the final 4 Å refined structure) and an improvement in the rate of convergence. In particular, when the size of the rigid group is relatively small, as in the case of the C-terminal helix, the constrained refinement is far less erratic in its behaviour. The application of symmetry constraints thus allows each subunit to be broken down into smaller rigid groups than would otherwise be possible. This in turn enables structural differences between the apo- and holoenzymes to be modelled with greater accuracy. It is interesting to note that the constrained refinement is superior in
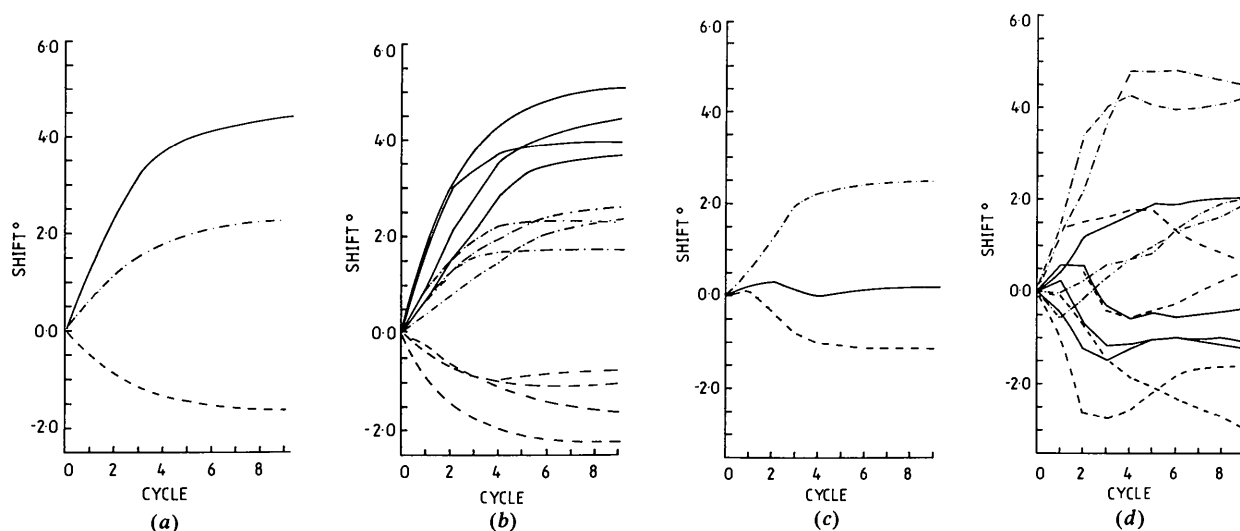

Fig. 2. The variation of the refined rotational parameters φ_R (dashed), θ_R (full), ρ_R (dash–dot) as a function of refinement-cycle number for the coenzyme-binding domain (*a* and *b*) and the C-terminal helix (*c* and *d*). (*a*) and (*c*) represent the results of the constrained refinement, while (*b*) and (*d*) are the results of the unconstrained refinement, both at 6 Å resolution. The four lines for each parameter in (*b*) and (*d*) are the values for the four independent subunits.

spite of the slight asymmetry of the true structure in this instance.

The final model of the apoenzyme has good stereochemistry, a reasonable crystallographic residual (35·0% to 3·5 Å resolution) and gives a very interpretable difference electron density map. It is therefore suitable as a starting model for Hendrickson–Konnert type of refinement at 3·5 Å resolution with very little manual re-building, although it would clearly be preferable to obtain X-ray data from crystals containing significantly less NAD so that more definite statements can be made about the true molecular symmetry of the apoenzyme.

The refinement strategy described here has made possible the high-resolution refinement of the apoenzyme with the absolute minimum of manual model building and without the need for high-resolution isomorphous derivative data. In principle it would have been possible to locate the position and orientation of the molecule in the apo unit cell using molecular replacement techniques with the known holoenzyme structure, rather than by analysis of the heavy-atom binding sites. This would have made possible a structure solution without any requirement for heavy-atom derivatives. In practice, however, the low-resolution apoenzyme electron density map was extremely valuable in determining the nature of the conformational change. A similar approach, *i.e.* the use of *CORELS* with a refined related structure as a starting model, could be applied to other structural problems where large conformational changes induced by cofactor or substrate binding are being investigated. There are now well documented cases where such conformational changes can be approximated by rigid-body movements of one domain relative to another, for example hexokinase (Bennett & Steitz, 1978, 1980), liver alcohol dehydrogenase (LADH) (Ecklund, Samama, Wallen, Branden & Akeson, 1981), and citrate synthase (Remington, Wiegand & Huber, 1982). This approach is applicable irrespective of the presence or absence of non-crystallographic symmetry.

These problems can be divided into two classes depending on the magnitude of the conformational change. If the domain rotation is less than 10°, so that the r.m.s. atomic shift required is of the order of two to three ångströms, then the refinement scheme outlined here can be applied directly (with or without molecular symmetry constraints as appropriate). GAPDH and LADH are examples of this class of problem. When the conformational change is of significantly larger magnitude than this, the required shifts will be outside the radius of convergence of *CORELS* refinement. Hexokinase, with a 12° domain rotation, and citrate synthase, with an 18° domain rotation, are examples of this class. However, if, as is often the case, the structural change involves the rotation of a small domain relative to a large domain,

a modified strategy can be employed. It should be possible in such cases to locate the position and orientation of the large domain with some confidence, using a combination of techniques such as cross-rotation functions, translation functions, Patterson search and molecular packing considerations. (This was the case with the citrate synthase structure.) The resultant model can then be optimized using *CORELS* and very low resolution data (*e.g.* 6 or 8 Å). Once the large domain has been correctly positioned, calculation of a low-resolution (6 Å) electron density map using model or combined model and isomorphous-replacement phases should make it possible to determine the molecular envelope, and possibly recognize some secondary structure in the small (rotated) domain. Using an interactive graphics display system and an α-carbon representation of the structure of the small domain, it should then be possible to determine the position of the small domain which optimally accounts for the observed molecular envelope and secondary structural features (if any). This model could then be used as a starting point for further *CORELS* refinement. The practical feasibility of such an approach requires verification, but its simplicity and the limited amount of computation required make it a very attractive alternative to the more conventional approach of doing many alternating rounds of refinement and manual model building.

The molecular symmetry constraints method as described here is only applicable when refining rigid-body parameters and not the internal dihedral angles of the rigid groups. This is not a serious limitation as dihedral-angle refinement would not normally be appropriate when refining structures at low resolution. In any case, it is possible in principle to extend the parameter averaging to include appropriate constraints on the dihedral angles if desired.

References

BENNETT, W. S. JR & STEITZ, T. A. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 4848–4852.

BENNETT, W. S. JR & STEITZ, T. A. (1980). *J. Mol. Biol.* **140**, 183–209, 211–230.

BIESECKER, G., HARRIS, J. I., THIERRY, J. C., WALKER, J. E. & WONACOTT, A. J. (1977). *Nature (London)*, **266**, 328–333.

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

ECKLUND, H., SAMAMA, J. P., WALLEN, L., BRANDEN, C. I., AKESON, A. & JONES, T. A. (1981). *J. Mol. Biol.* **146**, 561–587.

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502–510.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

HOARD, L. G. & NORDMAN, C. E. (1979). *Acta Cryst.* A35, 1010–1015.

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A34, 931–935.

LESLIE, A. G. W. & WONACOTT, A. J. (1983). *J. Mol. Biol.* 165, 375–391.

LESLIE, A. G. W. & WONACOTT, A. J. (1984). In preparation.

REMINGTON, S., WIEGAND, G. & HUBER, R. (1982). *J. Mol. Biol.* 158, 111–152.

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

SCHERINGER, C. (1963). *Acta Cryst.* 16, 546–550.

SMITH, P. J. C. & ARNOTT, S. (1978). *Acta Cryst.* A34, 3–11.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A33, 800–804.

WONACOTT, A. J. (1983). Personal communication.

WONACOTT, A. J. & BIESECKER, G. (1977). In *Pyridine Nucleotide Dependent Dehydrogenases*, edited by H. SUND. Berlin, New York: Walter de Gruyter.

Investigation of a Silicon Single Crystal by Means of the Diffraction of Mössbauer Radiation

BY K. KREC AND W. STEINER

Institut für Angewandte und Technische Physik, Technische Universität Wien, Austria

Abstract

Diffraction experiments with resonant γ-radiation were performed on a single crystal of silicon at room temperature. The high energy resolution of the Mössbauer effect was used to separate elastically and inelastically scattered radiation. A special experimental setup is described, which in addition allows the determination of the total thermal diffuse scattered intensity. Inelastically scattered radiation was observed in the whole investigated angular range. In the vicinity of Bragg reflections the ratio of the integrated intensities of inelastically to elastically scattered radiation was in very good agreement with the results of calculations using different computer programs based on first-order thermal diffuse scattering theory for a wide range of chosen scan lengths. The measured ratios of the elastically scattered radiation of different Bragg reflections were used to determine the temperature factor B.

I. Introduction

The precise determination of the thermal diffuse scattered (TDS) part of the diffracted radiation from a crystal is of great importance for the structure determination, since this contribution exhibits also a maximum value at the exact Bragg position and an error in the consideration of the integrated intensity can lead to an incorrect temperature factor B which may cause errors of the mean atomic coordinates or of the charge density due to the atomic scattering factors. In ordinary X-ray diffraction experiments an accurate determination of this TDS contribution is usually performed using theoretical models mainly based on the long-wavelength approximation. Several computer programs (Stevens, 1974; Helmholdt & Vos, 1977; Kurittu & Merisalo, 1977) are available which allow the calculation of integrated TDS intensities from the elastic constants taking into account the anisotropy of the elasticity. Experimentally the extraordinary high energy resolution of the Mössbauer effect can be used to separate elastically and inelastically scattered radiation from crystals not containing Mössbauer isotopes (O'Connor & Butt, 1963; Albanese, Ghezzi, Merlini & Pace, 1972; Albanese & Deriu, 1979) offering the possibility to test these theoretical calculations if special care is taken for the intensity measurements. In addition, the temperature factor B can be determined either directly from the elastically scattered intensities or with some assumptions concerning the TDS from the inelastically scattered intensities. Mössbauer diffraction on silicon was used to determine the reflecting power (Ghezzi, Merlini & Pace, 1969; Kashiwase & Minoura, 1983) and the dependence of the TDS intensities on both the reflecting angle and the temperature (Albanese *et al.*, 1972). Whereas the temperature and angle dependence of TDS is in good agreement with the prediction of the lattice wave theory (Warren, 1969), an evaluation of the fraction of the TDS intensities on the total intensities was not possible with reasonable accuracy (Ghezzi *et al.*, 1969) because (i) the quality of the crystals was not known with high accuracy and (ii) an experimental separation of the TDS contribution to the background could not be performed.

It is the aim of the present paper to show that several improvements of the experimental setup allow