

## 13. MOLECULAR REPLACEMENT

using noncrystallographic symmetry constraints, but this is not known to have been done in practice [types (3) and (4)].

### 13.1.5. The power of noncrystallographic symmetry in structure analysis

#### 13.1.5.1. Relevant parameters: standard case

A poorly determined structure, if known at sufficient resolution and accuracy, can be improved by structural refinement of an atomic model to fit the observations. These methods often use existing structural knowledge (of bond lengths and angles, for example) to improve the convergence of the refinement process.

The most important contributions of noncrystallographic symmetry arise before this point of structure determination is reached. In this stage of structural analysis, the distribution of scattering density may be constrained by the requirements of noncrystallographic symmetry. The density may be improved by imposing noncrystallographic symmetry on poorly defined scattering density, and, in favourable cases, cyclical improvement leads to a unique corrected structure. To avoid any confusion with refinement of the atomic structure, this process will be referred to as ‘symmetry correction’ or ‘correction’ (Hoppe & Gassmann, 1968).

Noncrystallographic symmetry can also be used to improve the accuracy and convergence of atomic structural refinement by increasing the number of observations to a given resolution (Section 13.1.5.5).

The power of correction methods in improving an unknown structure in the standard case (Section 13.1.2.1) depends on:

- (1) the resolution of the analysis,  $d$ ;
- (2) the number,  $N$ , of subunits per asymmetric unit;
- (3) the volume fraction,  $NU/V_a$ , of the asymmetric unit over which the noncrystallographic symmetry operation applies;
- (4) whether the density between subunit volumes is constant;
- (5) the degree of similarity of the subunits being matched; and
- (6) the extent to which the noncrystallographic symmetry operations differ from the crystal symmetry operations.

The first three parameters are expressed in quantitative terms; parameter (4) might be true or false, but more often lies between these; and parameters (5) and (6) are not easily expressed in measurable form.

The resolution  $d$  should ideally be matched to the level of similarity of the subunits. The root-mean-square displacement between an atom in one subunit and the rotated and translated position of the corresponding atom from another subunit (or model subunit) provides an order of magnitude for the resolution  $d$  which can be used effectively. In many cases, the resolution is worse than this for practical reasons of crystal disorder and data collection.

This limit was encountered by Huber *et al.* (1974) working at 1.9 Å resolution. They found that a model structure with a mean coordinate difference of 1.9 Å was not usable for molecular replacement, while another model agreeing to 0.75 Å gave results which allowed the structure to be refined. This suggests that agreement significantly better than the resolution is required.

#### 13.1.5.2. Information gain from ideal noncrystallographic symmetry

Rossmann & Blow (1962) wrote, “The effect of noncrystallographic symmetry . . . results in decreasing the size of the structure to be determined, while the number of observable intensities remains the same. This ‘redundancy’ in information might be used to help solve a structure.” This idea is developed below.

First, it will be shown that (in the absence of noncrystallographic symmetry) there is a constant ratio between the number of

independent measurements required to specify the scattering density at a chosen resolution and the volume of the asymmetric unit. Then the effect of noncrystallographic symmetry on this ratio is discussed. The importance of the ratio (volume of symmetry-constrained unit/volume of asymmetric unit) ( $= U/V_a$ ) is stressed. Another ratio is developed – available no. of measurements/ideally required no. of measurements – and this is referred to as the overdetermination ratio.

Consider a noncentrosymmetric crystal whose asymmetric unit volume is  $V_a$  and whose diffraction data have been measured to a resolution  $d$ . If the multiplicity of the space group (number of asymmetric units in the primitive unit cell) is  $Z$ , the volume of reciprocal space,  $V^*$ , per point of the primitive reciprocal lattice is given by

$$V^* = 1/V = 1/ZV_a,$$

where  $V$  is the volume of the primitive unit cell.

The number of independent orders of diffraction (the number of independent intensities) within the resolution sphere of radius  $1/d$  is given by

$$N_{\text{ref}} = \left(\frac{4\pi}{3d^3}\right) \left(\frac{1}{V^*}\right) \left(\frac{1}{2Z}\right) = \frac{2\pi V_a}{3d^3}. \quad (13.1.5.1)$$

In this formula, Friedel’s law is supposed to apply. A set of  $2Z$  reflections have identical intensity due to the combined effects of Friedel’s law and crystal symmetry. When  $N_{\text{ref}}$  is expressed in terms of  $V_a$ , the multiplicity factor disappears.

To calculate the scattering density over the volume  $V_a$  at resolution  $d$ ,  $2N_{\text{ref}}$  independent quantities need to be specified (say, the real and imaginary parts of each of the  $N_{\text{ref}}$  independent structure factors). The required number of measurements is

$$R = 2N_{\text{ref}} = (4\pi/3d^3)V_a. \quad (13.1.5.2)$$

If only the diffracted intensities can be measured, they provide exactly half the  $2N_{\text{ref}}$  measurements required to calculate the density at resolution  $d$ .

In what follows, it is assumed that the required number of measurements,  $R$ , to specify the scattering density at the chosen resolution is proportional to the volume over which the density must be specified. This is true when the volume is a crystallographic asymmetric unit [equation (13.1.5.2)], and it agrees with another analysis discussed below. Following this argument, the overdetermination ratio

$$\frac{\text{available no. of measurements}}{\text{ideally required no. of measurements}} = \frac{N_{\text{ref}}}{R} = \frac{V_a}{2X}, \quad (13.1.5.3)$$

where  $X$  is the volume whose density is unknown.

Next, consider that ideal noncrystallographic symmetry applies. The crystal asymmetric unit contains  $N$  identical subunits and no other scattering matter. Since the symmetry is noncrystallographic, it is never possible to fit the subunit volumes together so as to fill the unit cell exactly. The volume assigned to each subunit,  $U$ , has to be less than  $V_a/N$ , leaving some parts of the unit cell not assigned to any subunit. In the case of ideal noncrystallographic symmetry, these regions are necessarily empty. In this case  $X = U$ , which is less than  $V_a/N$ , so from equation (13.1.5.3)

$$\text{overdetermination ratio} = (V_a/2U) > N/2.$$

Even where  $N$  is only 2, more intensity data are available than the number of measurements ideally required to specify the electron density at resolution  $d$ .

A more sophisticated analysis of the number of variables required to define a structure with noncrystallographic symmetry has been made in terms of sets of orthogonal ‘eigendensity

### 13.1. NONCRYSTALLOGRAPHIC SYMMETRY

functions', which satisfy the noncrystallographic symmetry (Crowther, 1967). Any structure satisfying the symmetry requirements can be constructed from the appropriate set of eigendensities. Crowther (1969) demonstrated that the number of eigendensities  $m$  is approximately  $(2N_{\text{ref}}U/V_a)$ .

The structure is specified by  $m$  weights, which are applied to the  $m$  allowed eigendensities (which depend only on the symmetry constraints), so the overdetermination ratio

$$\frac{\text{available no. of measurements}}{\text{ideally required no. of measurements}} = \frac{N_{\text{ref}}}{m} = \frac{V_a}{2U},$$

the same result as before, showing that the two methods of analysis approximately agree.

#### 13.1.5.3. Information gain in the non-ideal case

In the non-ideal case, the definition of volume  $U$  assigned to each subunit assumes an important role. It is particularly important that the volumes should not overlap, since this may set up a chain of unrealizable constraints. In imposing noncrystallographic symmetry, the volumes between subunits are often unconstrained, allowing for differences in solvent structure and surface side chains following from their different environments. In addition to the volume  $U$  of one subunit, whose structure is to be defined, an additional volume,  $V_a - NU$ , is left unconstrained. The volume  $X$  of unknown electron density is  $U + (V_a - NU)$ , and, using equation (13.1.5.3), the overdetermination ratio

$$\frac{\text{available no. of measurements}}{\text{ideally required no. of measurements}} = \frac{V_a}{2[V_a - (N - 1)U]} \quad (13.1.5.4)$$

If  $N = 2$ ,  $U$  must be less than  $V_a/2$ , and the overdetermination ratio in equation (13.1.5.4) must be less than 1, so in the non-ideal case there is no chance of convergent correction. This confirms the practical observation that although averaging electron density with  $N = 2$  can improve the structure (Matthews *et al.*, 1967), it does not lead to convergent correction (B. W. Matthews, unpublished results). Slowly convergent *ab initio* structure correction was reported at 6.3 Å resolution for  $N = 4$  (Argos *et al.*, 1975). In this case, the volume  $4U$  of the constrained tetramer was reported to be only about  $V_a/2$ . Substituting  $N = 4$ ,  $U = V_a/8$  in the above expression gives an overdetermination ratio of only 1.6, which was sufficient to allow convergent correction.

An alternative possibility is to constrain the density between subunits to a constant value, even when this may not be precisely correct, in order to improve the convergence of symmetry correction. There is a close analogy to solvent-flattening techniques used in density modification and atomic structural refinement (Schevitz *et al.*, 1981; Wang, 1985). The volume constrained to a constant value is now  $V_a - NU$ . The volume whose structure is to be determined is only  $U$ , and in place of equation (13.1.5.4),

$$\frac{\text{available no. of measurements}}{\text{ideally required no. of measurements}} = \frac{V_a}{2U},$$

as in the ideal case. Such a constraint, while only approximately valid, may allow structure correction to proceed convergently, as found by Rossmann *et al.* (1992). The constraint may be released at a later stage.

This analysis also emphasizes the importance of specifying the size and shape of the subunit volume  $U$  as closely as possible (Wilson *et al.*, 1981). Methods of automatic refinement of the chosen volume are available (Rossmann *et al.*, 1992; Abrahams & Leslie, 1996).

#### 13.1.5.4. Relevant parameters: generalized case

In the generalized case it is obvious that noncrystallographic symmetry makes more measurements available, since the data from more than one type of crystal are being used. The volume of the subunit  $U$  must be less than  $V_a$  in each type of crystal. Making a simple assumption of two crystals, each with one subunit in each asymmetric unit, the available number of measurements per volume of the subunit is, in the ideal case,

$$\frac{N_{\text{ref1}} + N_{\text{ref2}}}{U} = \left(\frac{2\pi}{3d^3}\right) \left(\frac{V_{a1} + V_{a2}}{U}\right) > 2 \left(\frac{2\pi}{3d^3}\right).$$

Thus, the overdetermination ratio  $= (N_{\text{ref1}} + N_{\text{ref2}})/(4\pi U/3d^3) > 1$ , so even in this case the structure is theoretically overdetermined.

This type of reasoning can be applied to analysis of a crystal which includes a unit of known structure and also another unit whose structure is unknown [type (6), Table 13.1.4.1]. A complex between an enzyme of known structure with an unknown inhibitor provides a familiar example. Note first that the envelope defining the volume  $U$  of known structure must be tightly defined, since otherwise unwanted features will be taken over into the unknown structure.

If this known structure of volume  $U$  appears once in the asymmetric unit  $V_a$  of the partly unknown structure, can noncrystallographic symmetry correction be used to define a unique structure at resolution  $d$ ?

The unconstrained volume is  $V_a - U$ . The number of measurements required to define this density at the given resolution is  $4\pi(V_a - U)/3d^3$ . The partially unknown structure provides  $2\pi V_a/3d^3$  measured intensities to this resolution, and specification of the contents of  $U$  at resolution  $d$  is equivalent to  $4\pi U/3d^3$  measurements. Thus the overdetermination ratio is

$$\frac{(2\pi V_a/3d^3) + (4\pi U/3d^3)}{4\pi(V_a - U)/3d^3} = \frac{V_a + 2U}{2(V_a - U)}.$$

The overdetermination ratio is greater than 1 if  $U > V_a/4$ , that is, if only a quarter of the asymmetric unit represents known structure. Although this relationship applies to an 'ideal' case and is therefore certainly too optimistic, it indicates the remarkable power of the molecular-replacement method. If, for example, half the unit cell is devoted to unknown structure ( $U = V_a/2$ ), the overdetermination ratio is ideally 2.

#### 13.1.5.5. Noncrystallographic symmetry in atomic coordinate refinement

In atomic coordinate refinement, noncrystallographic symmetry again provides a useful increase in the ratio of the number of observed quantities to the number of atomic parameters to be refined. As is discussed by Cruickshank in Chapter 18.5, the application of restraints in refinement (on quantities like bond lengths, bond angles and the elimination of short contacts) is formally equivalent to an increase in the number of observational equations. However, if these restraints are tightly applied, they act more like constraints, and their effect is more like a reduction in the number of parameters to be determined. Meaningful refinement is not possible unless the number of observations exceeds the number of parameters, and in practice it usually needs to do so by a factor of 2 or so. If noncrystallographic symmetry is imposed, the number of observations required to define the structure is reduced, because the volume of unknown structure is reduced.

Noncrystallographic symmetry can thus provide a crucial advantage in leading to unambiguous interpretation of structure at relatively poor resolution (say, 3.0 to 3.8 Å), where the ratio of

### 13. MOLECULAR REPLACEMENT

refined parameters to the number of observations is marginal. Consider two crystals of the same material, one of which has one subunit per asymmetric unit and the asymmetric volume is  $V_1$ . The other has  $N$  subunits in an asymmetric unit of volume  $V_N$ . To the same resolution, the available number of observed reflections is increased in the ratio  $V_N/V_1$ , in order to obtain the same number of parameters if noncrystallographic symmetry is imposed.

What effect do these  $N$  subunits have on the precision of the final coordinates? The crystal allows the determination of  $N$  sets of atomic coordinates. If the errors were independent of each other, the precision of the mean value of each coordinate could be improved in the ratio  $N^{-1/2}$  (compared to a well refined  $V_1$  structure).

This improvement will be lost when constraints are applied to the mean coordinates (to make them conform to given bond lengths and angles, for example). If this is done, the errors are no longer independent, and the increase of precision will be less.

Cruickshank (1999 and Chapter 18.5) shows that at high resolution (examples at 0.94 and 1.0 Å) and for atoms of low  $B$  factor (less than say  $10 \text{ \AA}^2$ ), restraints make little difference to the precision of refinement. Under these conditions,  $N$  independent subunits in the asymmetric unit might improve the precision of the mean coordinates by a factor approaching  $N^{-1/2}$ . But at such good resolution, it is very possible that the differences between the calculated subunit conformations are not due to error, but reflect real structural differences. If so, the precision of the mean coordinates is less significant.

At less high resolution (example given at 1.7 Å), Cruickshank has shown that the precision of unrestrained refinement is significantly worse than the precision of the restraints. In this case, imposing noncrystallographic symmetry on the structure should provide some improvement. But because the coordinate errors then cease to be independent, the improvement in the mean coordinates would be less.