Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 18 September 2002
Accepted 14 October 2002

The crystallographic symmetry test for the correctness of a set of phases

P. Tzamalis,[a] K. Bethanis,[a] A. Hountas[a]* and G. Tsoucaris[b]

[a]Physics Laboratory, Agricultural University of Athens, 75 Iera Odos Votanikos, Athens 118-55, Greece, and [b]Laboratoire de recherche de Musèes de France, Paris, France. Correspondence e-mail: hount@aua.gr

One of the most important problems in the application of direct methods for large structures is to establish reliable consistency criteria for the correctness of a phasing trial. The introduction in the twin variables method [Bethanis, Tzamalis, Hountas, Mishnev & Tsoucaris (2000). *Acta Cryst.* A**56**, 105–111] of a new criterion based on the crystallographic symmetry consists of testing the phase extension and refinement algorithm by deliberately sacrificing the space-group-symmetry information in the auxiliary variable set then using its gradual re-appearance as a criterion for correctness. In the present paper, the crystallographic symmetry test has been used in the implementation of the twin algorithm in two different ways: (i) as an overall test throughout the iterations that is likely to reflect the correctness of the phasing procedure for each one of the extension trials in a macromolecular phasing environment; (ii) as a convenient criterion to determine the optimum cycle for freeing the initial phases used by the algorithm for the phase-extension procedure.

1. Notation and definitions

MPE: Mean phase error.

$(\mathrm{SPE})_{\mathbf{H}}$: Symmetry phase error – a measure of the inconsistency among the calculated phases for symmetry-related reflections.

$(\mathrm{S\_R}_{\mathrm{mod}})_{\mathbf{H}}$: Symmetry modulus error – a measure of the inconsistency among the calculated moduli for symmetry-related reflections.

S_MPE: Overall symmetry mean phase error – discrepancy index for the phases of a set of reflections.

$\mathrm{S\_R}_{\mathrm{mod}}$: Overall symmetry mean modulus error – discrepancy index for the moduli of a set of reflections.

Ψ_S_MPE: Overall symmetry mean phase error for the auxiliary Ψ set.

2. Introduction and definitions

The symmetry information plays an important role in direct methods from the simple fact that phasing of one reflection implies transfer of this information to all symmetry-equivalent reflections. When the crystal symmetry is known *a priori*, it seems natural to introduce in a direct-method algorithm the exact structure-factor (phase and moduli) relations implied by symmetry from the very beginning. On the other hand, efficient testing during the process of phase extension is a crucial part of direct methods. The available tests often prove to be insufficient in a macromolecular phasing environment for a strong discrimination between different solutions in a multi-solution algorithm (Gilmore, 1998). The twin-variables method offers the possibility to introduce a new test for successful development of the phase-determining algorithm, based upon symmetry considerations. This possibility stems from the involvement of an auxiliary set of variables in addition to the structure factors as shown below.

To use this test, one must deliberately sacrifice the *ab initio* whole use of space-group-symmetry information. This enables us to use the gradual establishment of space-group symmetry in the course of phase extension as a criterion for the correctness of the calculated phases. This special application of symmetry in the twin algorithm is the object of the present paper. Only a thorough examination of this matter, both theoretically and by computing, is appropriate to bring an answer to the question whether the advantage of a new reliable test compensates for the loss (at least apparent) of the symmetry information, especially in the first steps of the algorithm.

An important feature of the *TWIN* algorithm is the decoupling between the E's, bearing the observed E_H-moduli information, and the auxiliary variables Ψ which alone control the phasing procedure (Hountas & Tsoucaris, 1995; hereafter H-T).

$$E_{\mathbf{H}} = \sum_{\mathbf{K}} \Psi_{\mathbf{K}} (\Psi_{\mathbf{K}-\mathbf{H}})^{*} \quad \overset{\mathrm{FT}}{\longleftrightarrow} \quad \rho(\mathbf{r}) = |\psi(\mathbf{r})|^{2} \tag{1}$$

$$\Psi_{\mathbf{K}} = \sum_{\mathbf{H}} E_{\mathbf{H}} \Psi_{\mathbf{K}-\mathbf{H}}, \quad \text{where } \Psi_{\mathbf{H}} = \mathrm{FT}[\psi(\mathbf{r})]. \tag{2}$$

At this point, it is appropriate to stress the relevance of the above equations to the fundamental quantum-mechanical principles. We wish to find an approximate wave function in direct space $\psi(\mathbf{r})$ such that its squared modulus $\rho(\mathbf{r})$ 'behaves' like a crystallographic electron-density function. It is noteworthy that the definitions expressed by (1) ensure the positivity of $\rho(\mathbf{r})$; correlatively, we note that the positivity of $\rho(\mathbf{r})$ at each step of the algorithm is a difficult question for most direct methods. In a crystallographic context, dominated by the reciprocal-space data, it appears useful to introduce the FT of the wave function $\psi(\mathbf{r})$ denoted here by $\Psi_{\mathbf{H}}$. We wish then that the FT of $\rho(\mathbf{r})$, denoted here by $E_{\mathbf{H}}$, satisfies the observed moduli criterion. This will be achieved by minimizing $M_{\rm mod}$ as a function of all $\Psi_{\mathbf{H}}$:

$$M_{\rm mod} = \sum_{\mathbf{H}} (|E_{\mathbf{H}}| - |E_{\mathbf{H}}^{\rm obs}|)^2 = \sum_{\mathbf{H}} \left(\left| \sum_{\mathbf{K}} \Psi_{\mathbf{K}} (\Psi_{\mathbf{K}-\mathbf{H}})^* \right| - |E_{\mathbf{H}}^{\rm obs}| \right)^2. \tag{3}$$

This FT has a precise physical meaning as the momentum-space wave function: its square modulus represents the probability distribution over the momentum in the same way that $\rho(\mathbf{r})$ represents the probability distribution over the position of a quantum-mechanical particle (Bethanis *et al.*, 2002). In the present context, this physical quantum-mechanical meaning is not directly involved and the set of $\Psi_{\mathbf{H}}$ plays the role of an auxiliary set of variables that determines $E_{\mathbf{H}}$ *via* the left part of (1). In addition, the $E_{\mathbf{H}}$ and $\Psi_{\mathbf{H}}$ sets are linked together *via* (2), the so-called regression equation whose direct-method meaning has been given in §2.3 of the H-T paper. Thus, the twin algorithm aims at determining the phases of the E's through a very large Ψ set, by satisfying a battery of constraints expressed by minimization functions like (3) (Bethanis, Tzamalis, Hountas, Mishnev & Tsoucaris, 2000).

3. The concept of the crystallographic symmetry test

The Ψ's are not restricted by theory to obey the symmetry constraints and, therefore, arbitrary values can be assigned to the Ψ's. Equation (1) allows one to transfer the information from the $\Psi_{\mathbf{H}}$ set to the $E_{\mathbf{H}}$ set. Thus, if we construct an arbitrary $\Psi_{\mathbf{H}}$ set that does not obey the space-group symmetry, equation (1) then will generate $E_{\mathbf{H}}$ values that will not respect the space-group symmetry in both phase and modulus. However, if the further development of phases tends to the correct values, then, naturally, the correct relations between symmetry-related structure factors $E_{\mathbf{H}}$ would also tend to gradually occur. This special feature of the twin algorithm suggests a new test for phase extension which will be called the 'crystallographic symmetry test'.

If we denote a symmetry operation by **R** for the rotational part and **T** for the translation part, the exact structure-factor relations implied by symmetry are:

$$|E_{\mathbf{H}}| = |E_{\mathbf{HR}}| \tag{4}$$

$$\varphi_{\mathbf{HR}} - \varphi_{\mathbf{H}} = -2\pi\mathbf{HT}. \tag{5}$$

In the simple case of space group $P2_1$, with only one pair of symmetry-equivalent reflections, (4) and (5) have the form

$$|E_{\mathbf{H}}| = |E_{\mathbf{HR}}| \tag{4a}$$

$$\varphi_{\mathbf{HR}} - \varphi_{\mathbf{H}} = -k\pi, \tag{5a}$$

where k is the Miller index of the reflection $H = (hkl)$.

For structure factors that do not strictly obey symmetry, we can define discrepancy indices concerning the symmetry-related phases and moduli. In the particular case of $P2_1$ space group, these indices have the form

$$({\rm SPE})_{\mathbf{H}} = |\varphi_{\mathbf{H}} - \varphi_{\mathbf{HR}} \pm k\pi| \tag{6}$$

$$({\rm S\_R}_{\rm mod})_{\mathbf{H}} = [|E_{\mathbf{HR}}| - |E_{\mathbf{H}}|]/|E_{\mathbf{H}}|. \tag{7}$$

These discrepancy indices (6) and (7) may also play the role of a figure of merit for each reflection **H**. From this we obtain general indices for the whole set of reflections, as the mean value of the individual discrepancies, *i.e.*

$${\rm S\_MPE} = \langle ({\rm SPE})_{\mathbf{H}} \rangle \tag{8}$$

$${\rm S\_R}_{\rm mod} = \langle ({\rm S\_R}_{\rm mod})_{\mathbf{H}} \rangle. \tag{9}$$

In the general case, when a space group possesses n symmetry elements, the above indices are defined through the generalized equations

$$({\rm SPE})_{\mathbf{H}} = \sum_{s=1}^{n} |\bar{\varphi}_{\mathbf{H}} - \varphi_{\mathbf{HR}_s} \pm 2\pi\mathbf{HT}_s| \Big/ n \tag{10}$$

$$({\rm S\_R}_{\rm mod})_{\mathbf{H}} = \sum_{s=1}^{n} [|E_{\mathbf{HR}_s}| - |\bar{E}_{\mathbf{H}}|] \Big/ \sum_{s=1}^{n} |E_{\mathbf{HR}_s}|, \tag{11}$$

where by $\bar{\varphi}_{\mathbf{H}}$ and $\bar{E}_{\mathbf{H}}$ are denoted the average values of $\varphi_{\mathbf{HR}_s}$ and $E_{\mathbf{HR}_s}$, $s = 1, \ldots, n$, including the identity.

It is to be noted that an index $R_{\rm sym}$ similar to equation (9) has been calculated and used as the basis of a test for reliability of intensity measurements (Blow, 2002), *i.e.* in a context different from the present paper.

It is worth mentioning that the above definitions are not the only ones for a given set of data. For instance, an alternative index based on the complex normalized structure factors, which combines both phases and moduli, would be:

$$({\rm S\_R}_{\rm complex})_{\mathbf{H}} = \sum_{s=1}^{n} |E_{\mathbf{HR}_s} - \bar{E}_{\mathbf{H}}| \Big/ \sum_{s=1}^{n} |E_{\mathbf{HR}_s}| \tag{12}$$

$${\rm S\_R}_{\rm complex} = \langle ({\rm S\_R}_{\rm complex})_{\mathbf{H}} \rangle. \tag{13}$$

Similar discrepancy indices can be applied to the auxiliary set of Ψ variables. However, in the applications described below, only the Ψ_S_MPE, defined as

$$(\Psi{\rm \_SPE})_{\mathbf{H}} = \sum_{s=1}^{n} |\bar{\omega}_{\mathbf{H}} - \omega_{\mathbf{HR}_s} \pm 2\pi\mathbf{HT}_s| \Big/ n \tag{14}$$

$$\Psi{\rm \_S\_MPE} = \langle (\Psi{\rm \_SPE})_{\mathbf{H}} \rangle, \tag{15}$$

where $\omega_{\mathbf{H}}$ is the phase of $\Psi_{\mathbf{H}}$, *i.e.* $\Psi_{\mathbf{H}} = |\Psi_{\mathbf{H}}| \exp(i\omega_{\mathbf{H}})$, is used.

3.1. Friedel reflections

The above expressions for the space-group symmetry can also be adapted to the Friedel pairs by including the symmetry operation

Table 1
Names and chemical, unit-cell and symmetry data for protein structures.

Structure	Full name	Unit cell (Å, °)	Space group	No. of atoms in asymmetric unit cell	References
Rnase Ap1	Ribonuclease Ap1 of *Aspergillus pallidus*	a = 32.01, b = 49.76, c = 30.67, α = 90.0, β = 115.83, γ = 90.0	$P2_1$	890	Bezborodova *et al.* (1988)
1BKR	Calponin Homology (ch) domain from Human beta-spectrin	a = 31.65, b = 53.95, c = 32.35, α = 90.0, β = 105.48, γ = 90.0	$P2_1$	1095	Banuelos *et al.* (1998)†
1TMY	Chey from *Thermotoga maritima*	a = 32.04, b = 53.95, c = 34.16, α = 90.0, β = 95.56, γ = 90.0	$P2_1$	928	Usher *et al.* (1998)†

† Data were retrieved from the Protein Data Bank.

Table 2
Phase-extension results for three different protein structures.

In all calculations, symmetry indices vary in the same way as MPE does

		Initial set			Extended set			Total (Initial + Extended) set			
	Data set	Resolution (Å)	No. of reflections	MPE	Resolution (Å)	No. of reflections	MPE	No. of reflections	MPE	S_MPE	$S\_R_{\rm mod}$
1	Rnase Ap1	2.5	1000	0°	1.0	16384	15°	17384	14°	1°	1.1%
2		2.5	1000	21°	1.0	17322	60°	18322	58°	63°	39.3%
3	1BKR	2.5	1000	0°	1.0	16828	19°	17828	18°	5°	2.3%
4		2.5	1000	20°	1.0	23030	51°	24030	50°	56°	35.8%
5	1TMY	3.0	1000	0°	1.5	15524	20°	16524	19°	7°	8.5%
6		3.0	1000	20°	1.5	10954	68°	11954	66°	60°	37.8%

$$R = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Note, however, that in the twin algorithm the whole Ψ set may not obey the Friedel pair requirement: $\Psi_{\mathbf{H}} = \Psi_{(-\mathbf{H})}$. Again, the Friedel requirement may be gradually fulfilled in the course of the phase extension and therefore it could complement the space-group symmetry test. However, in the applications described below, the usefulness of the Friedel test does not seem to be significant, and we have imposed the Friedel requirement.

The inner logic of the twin algorithm leads to the use of the whole set of the Friedel independent reflections. Thus, in the following calculations, the number of reflections includes a complete set of Friedel reflections. Clearly, the number of symmetry-independent reflections in $P2_1$ is somewhat larger than half the Friedel independent ones.

4. Practical implementation of the crystallographic symmetry test in the twin variables method

In the present paper, the crystallographic symmetry test has been used in the implementation of the twin algorithm in two different ways described below.

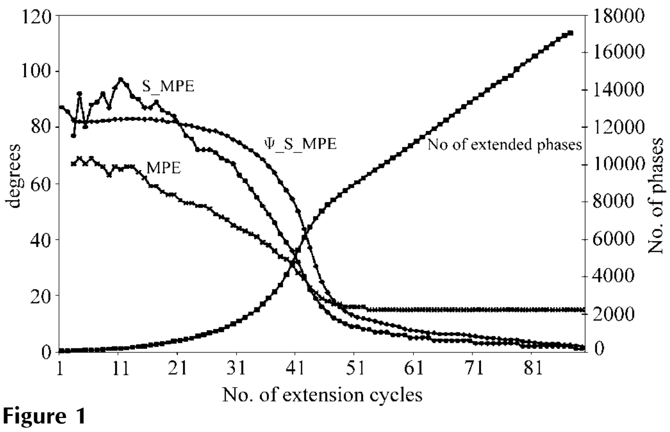

Figure 1
Variation of several symmetry indices during a successful phase extension for protein Rnase Ap1 with a starting set of 1000 known reflections at 2.5 Å resolution. Calculated phases correspond to a resolution range between 2.5 and 1 Å.

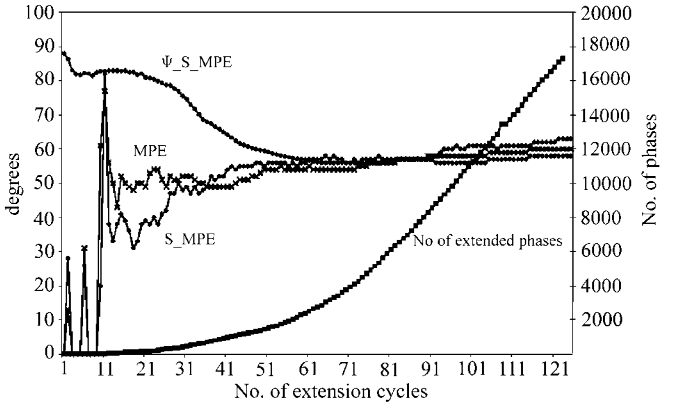

Figure 2
Variation of several symmetry indicators during an unsuccessful phase extension for protein Rnase Ap1 with a starting set of 1000 (non-unique) known reflections at 2.5 Å resolution with a mean phase error 21°. Calculated phases correspond to a resolution range between 2.5 and 1 Å.

Table 3
For protein structure Rnase Ap1, two different phase extensions have been performed.

In the first trial, symmetry was used as a test while in the second trial the symmetry requirements have been imposed from the beginning for all reflections. The final result for the MPE is then improved by only 3°.

Data (protein structure Rnase Ap1)			Phase extension without imposing the symmetry requirements			Phase extension with the symmetry requirements imposed from the beginning		
Resolution (Å)	No. of known phases	Initial MPE	Resolution of the extended set (Å)	No. of calculated phases	MPE	Resolution of the extended set (Å)	No. of calculated phases	MPE
2.5	1000	21°	1.0	17322	60°	1.0	17616	57°

4.1. The crystallographic symmetry test as overall test

S_MPE and S_$R_{\rm mod}$ can be used throughout the iterations as overall indices that are likely to reflect the correctness of the phasing procedure for each one of the extension trials.

4.1.1. Detailed phase-extension calculations on the Rnase Ap1 protein structure. Phase-extension attempts have been made on the protein structure Rnase Ap1 (space group $P2_1$) data (Table 1) using as initial input data a set of 1000 (non-unique) known phases at 2.5 Å resolution. Twin algorithm calculations are successful when error-free initial phases are used (Table 2, row 1). On the contrary, the trial is not successful when an artificial error of 21° is imposed on the values of the initial set of known phases (Table 2, row 2). Fig. 1 provides a detailed description of the variation of S_MPE during the successful phase extension while Fig. 2 represents the unsuccessful trial.

From the above two figures, it appears that in both cases S_MPE varies in the same way as MPE does, and that the symmetry indicators do reflect the correctness of the phasing procedure. This is evidenced below by a detailed examination of the extended reflections.

(*a*) In the successful trial of Fig. 1, there is a considerable variation of S_MPE at the first cycles of extension. But in the unsuccessful trial of Fig. 2, these variations have a larger amplitude and are extended over a larger number of cycles, *i.e.* about 31 cycles in the unsuccessful trial as opposed to 16 cycles of extension for the successful trial.

(*b*) In the successful trial, the progress in the calculation of new phases is quicker than in the unsuccessful one. For example, it can be seen from Figs. 1 and 2 that 2000 new phases have been calculated at the 35th cycle for the successful *versus* the 55th cycle for the unsuccessful. Moreover, in the plot of the number of calculated reflections, the successful trial is characterized by an inflection point that occurs at about the same number of cycles as does the inflection point of the S_MPE variation. Such an inflection point does not exist in the unsuccessful trial (Fig. 3).

(*c*) It is worth noting a difference in the calculation results of S_MPE and Ψ_S_MPE. The second index is calculated for the total number of Ψ's in all cycles, since all Ψ's are used and refined from the beginning of the extension. This is the reason for the small variation of Ψ_S_MPE in the first cycles of extension, in contrast with the large variation of S_MPE calculated only for the accepted (unknown) *E*'s. The latter are very few in the first cycles of the extension process. Thus, the examination of the symmetry of the Ψ set enriches the evaluation of the correctness of the phase extension.

In a previous paper (Bethanis, Tzamalis, Hountas, Mishnev & Tsoucaris, 2000), we have pointed out the extreme dilution of the phase information of the initial set injected into the whole set of the initial random Ψ's. In the present paper, we emphasize the extreme dilution of the symmetry information into the whole set of the initial Ψ's lacking the symmetry

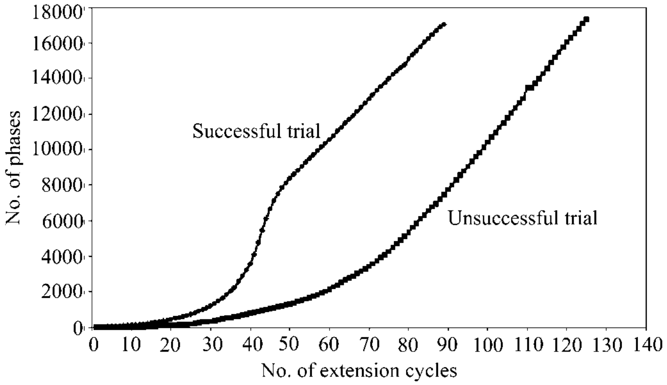

Figure 3
Plotting in the same figure the number of extended phases from Figs. 1 and 2. In the successful trial, the progress in the calculation of new phases is quicker than in the unsuccessful one as can be seen from the inflection point presented in the corresponding line.

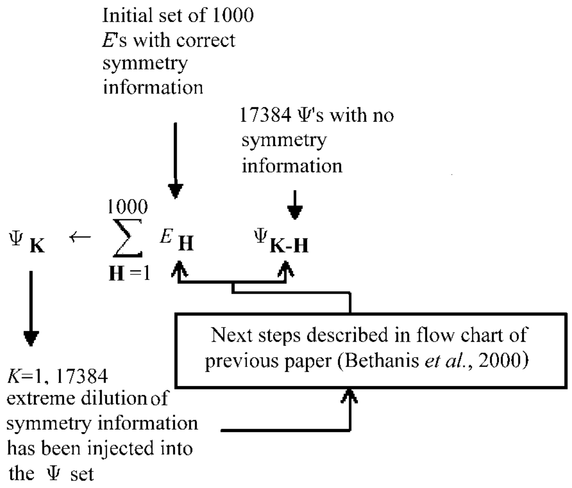

Figure 4
The extreme dilution of the symmetry information presented by an example described in the above flow chart

Table 4
For protein structure 1BKR, several phase-extension calculations have been performed by freeing the initial phases at a different cycle of the procedure for each calculation.

The optimum freeing is observed at the 20th cycle when the S_MPE is minimum.

				Final results	
Trial no.	No. of extension cycle at which initial phases are set free	No. of calculated reflections	S_MPE/S_R$_{\rm mod}$ for the calculated reflections	Number/MPE/S_R$_{\rm mod}$ for the extended reflections	MPE/S_R$_{\rm mod}$ for the initial 6000 reflections
1	10	710	40°/26%	25610/53°/16%	41°/10%
2	20	1616	38°/25.9%	25652/32°/16%	30°/25%
3	40	3938	48°/27%	25138/60°/15%	56°/16%
4	60	7604	40°/19.6%	25318/59°/14%	56°/15%
5	80	11478	41°/16.7%	25488/58°/14%	53°/15%
6	100	18054	42°/14.3%	25620/55°/12%	46°/14%
7	110	-	-	25050/53°/7%	40°/11%

information as schematically shown in Fig. 4. This dilution results in a very weak degree of fulfilment of the space-group-symmetry relations for the Ψ set and can be quantitatively evaluated by computing the Ψ_S_MPE. For instance, in Fig. 1 it is shown that the Ψ_S_MPE at the first iteration step for the 17384 Ψ's is 87°, which is indeed very close to the random value of 90°.

In order to further study the effect of symmetry, we have performed in the case of the unsuccessful trial a new cycle of calculations under exactly the same conditions but with the symmetry requirement imposed on the Ψ set from the beginning for all reflections. The final result for the MPE of the $E_{\bf H}$ values is then improved by only 3° (Table 3). This provides evidence that the symmetry information used from the beginning may not be crucial in the context of the twin algorithm.

4.1.2. Phase-extension calculations for several protein structures. Table 2 shows the final values for S_MPE and S_R$_{\rm mod}$ for trials of phase extension on data corresponding to the three different protein structures presented in Table 1. It can been seen that when the MPE is less than 30° then the values of S_MPE and S_R$_{\rm mod}$ are very low (less than 10° and less than 10%, respectively). On the contrary, when the value of MPE is about 50–60°, the values of S_MPE and S_R$_{\rm mod}$ are over 50° and over 35%, respectively. Here also, the parallel variation of both S_MPE (or S_R$_{\rm mod}$) and MPE permits the use of S_MPE (or S_R$_{\rm mod}$) as a measure of the correctness of the phase-extension procedure.

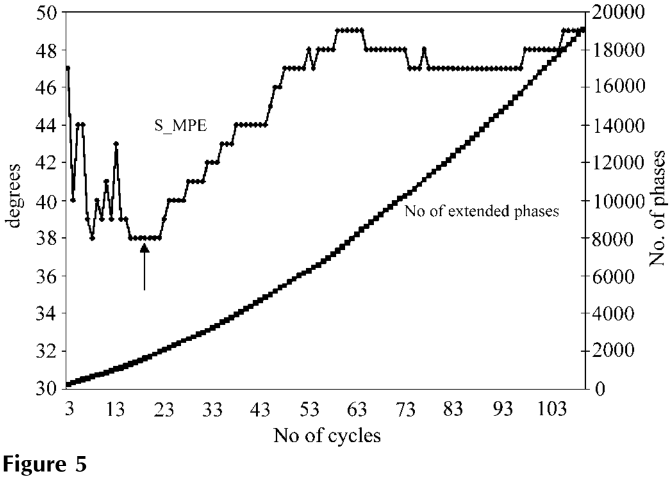

Figure 5
In a preliminary extension calculation, symmetry can monitor the optimum freeing of the initial reflections.

4.2. Optimum freeing of the initial phases by using the crystallographic symmetry test

An important problem in macromolecular crystallography is that of phase extension and refinement when initial phase estimates are available from isomorphous replacement or anomalous scattering or other methods. In most cases, it is necessary to extend the phases either from lower to higher resolution or within the same resolution range. Extensive treatment of this case with the twin variables can be found in previous papers (Bethanis, Tzamalis, Hountas, Mishnev & Tsoucaris, 2000; Bethanis, Tzamalis, Hountas, Tsoucaris *et al.*, 2000). In these calculations, the initial set has been kept constant through the extension procedure. But one expects that the extended reflections can also be used in order to refine the initial approximate phases towards the correct values, provided that a large set of extended phases have already been computed with a good approximation to their correct values. An important question is to determine an appropriate stage at which the variation of the initial set of structure factors is allowed by the program. Precisely, the S_MPE and S_R$_{\rm mod}$ described above prove to be a convenient criterion for this purpose.

Fig. 5 shows the variation of S_MPE and the number of calculated phases in each cycle of extension for the 1BKR protein. In this trial, the starting set consists of 6000 reflections; their phases are obtained by imposing an artificial error of 42° to the correct values. In this figure, there exists a minimum for S_MPE, close to the 20th extension cycle, which is indicated by an arrow. Thus, the next step was to perform several extension calculations by freeing the initial phases at a different cycle of the procedure for each calculation. The results of these calculations are shown in Table 4, from which several observations can be made:

(*a*) The best result, both for the extended and the initial phases, is achieved when the freeing of the initial phases is introduced at cycle 20 where the lowest S_MPE occurs. In this

Table 5
Results of phase extension on different protein structures including S_MPE/S_R$_{\rm mod}$ and R factor of each trial.

In trials 2, 5, 8, both R factor S_MPE/S_R$_{\rm mod}$ and MPE indicate successful phase extension. On the contrary, in trials 1, 3, 4, 6, 7, the R factor is good while as shown by S_MPE/S_R$_{\rm mod}$ and MPE these are unsuccessful phase extensions.

Trial no.	Protein	S_MPE/S_R$_{\rm mod}$	R factor	MPE
1	Rnase Ap1	38°/32.5%	10%	64°
2		6°/5.5%	10%	10°
3	1BKR	56°/35.8%	12%	51°
4		60°/41.6%	17%	58°
5		6°/7%	11%	26°
6	1TMY	60°/37.8%	12%	78°
7		37°/31.9%	13%	61°
8		7°/8.5%	11%	20°

trial, the final MPE for the 6000 initial phases is 30°, compared with the starting MPE of 42°.

(*b*) Freeing the initial phases at a different cycle (10, 40, 60, 80, 100) yields results that are worse compared with the result without freeing the phases (trial no 7). It is clear that freeing the initial phases at the proper stage of extension is of crucial significance for a successful result.

One possible explanation for the improvement of the initial low-resolution phases is that they participate in a large number of triplets. When the extended reflections have been calculated with a small MPE then this new information is conveyed through the convolution equation to the initial poorly phased reflections.

From different trials, we have concluded that an S_MPE of 40° and a number of calculated reflections about 1/6 of the initial number is an acceptable combination of conditions in order to free the initial phases usually with good results. With a time-consuming but more efficient approach, it is possible to perform a preliminary extension in order to determine the exact cycle where S_MPE is minimum and then in the main extension freeing the initial reflections at exactly this cycle. Furthermore, accumulated experience with this kind of trial will enhance our ability to choose the optimum number of cycles for freeing the initial known reflections.

5. Conclusions

The decoupling between the E moduli information and the auxiliary Ψ variables, in the context of the twin method, enabled us to develop a new overall evaluation test based on crystallographic symmetry. This new criterion consists of testing the phase-extension and refinement algorithm by deliberately ignoring the space-group symmetry in the starting set, then using its progressive re-establishment as a criterion for correctness. This provides great flexibility in further developments; as an example, symmetry can monitor the optimum point to free the initial reflections. The calculations of the present paper (Table 2) have allowed us to narrow down the limits of the values of the S_MPE for successful and unsuccessful trials: for S_MPE < 10°, it is probable that phase extension will be successful while for S_MPE > 50° it is probable that phase extension will not succeed.

On the other hand, the most commonly used index in crystallography is the R factor:

$$R = \sum_{\mathbf{H}} ||E_{\mathbf{H}}^{\rm calc}| - |E_{\mathbf{H}}^{\rm obs}|| \Big/ \sum_{\mathbf{H}} |E_{\mathbf{H}}^{\rm obs}|,$$

which measures the inconsistency among the observed and calculated moduli. The R factor is closely related to $M_{\rm mod}$ [equation (3)]. In the context of the twin algorithm, it has been observed that $M_{\rm mod}$ can be readily reduced to practically zero, but this clearly has no physical meaning. This fact, however, greatly emphasizes how easy it is to construct density functions whose Fourier coefficients have the same moduli as a given set of $E_{\rm obs}$. This remark is reflected clearly in Table 5 which summarizes phase-extension procedures for different protein structures.

It should be noted that the idea of the crystallographic symmetry test is essentially the same, at a very fundamental level, as the free R test introduced by Brünger (1993). Brünger has proposed the use of a free R factor in which the intensity data are partitioned into two sets of unique reflections: a working set {W} and a much smaller test set {T}. When a structure is refined, only set {W} is used throughout. At the end of the refinement, an R factor is calculated based only on set {T}, which has also been deliberately sacrificed in order to provide an unbiased criterion of goodness of fit.

$$R_{\rm T}^{\rm free} = \sum_{\mathbf{H}\in{\rm T}} ||F_{\mathbf{H}}^{\rm obs}| - |F_{\mathbf{H}}^{\rm calc}|| \Big/ \sum_{\mathbf{H}\in{\rm T}} |F_{\mathbf{H}}^{\rm obs}|.$$

Further developments include the use of the symmetry tests for the evaluation of the correctness of the phasing process in a multisolution algorithm.

References

Banuelos, S., Saraste, M. & Djinovic Carugo, K. (1998). *Structure*, **6**, 1419.

Bethanis, K., Tzamalis, P., Hountas, A., Mishnev, A. & Tsoucaris, G. (2000). *Acta Cryst.* A**56**, 105–111.

Bethanis, K., Tzamalis, P., Hountas, A. & Tsoucaris, G. (2002). *Acta Cryst.* A**58**, 265–269.

Bethanis, K., Tzamalis, P., Hountas, A., Tsoucaris, G., Kokkinou, A. & Mentzafos, D. (2000). *Acta Cryst.* A**56**, 606–608.

Bezborodova, S. I., Ermekbaeva, L. A., Shlyapnikov, S. V., Polyakov, K. M. & Bezborodov, A. M. (1988). *Biochem. USSR*, **53**, 837–845.

Blow, D. (2002). *Outline of Crystallography for Biologists*, p. 111. Oxford University Press.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Gilmore, C. J. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 159–167. Dordrecht: Kluwer.

Hountas, A. & Tsoucaris, G. (1995). *Acta Cryst.* A**51**, 754–763.

Usher, K. C., De La Cruz, A., Dahlquist, F. N., Swanson, R. V., Simon, M. I. & Remington, S. J. (1998). *Protein Sci.* **7**, 403.