RESEARCH PAPERS

Acta Cryst. (1996). D**52**, 425–434

MAGICSQUASH: More Versatile Non-crystallographic Averaging with Multiple Constraints

DAVID J. SCHULLER

Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92717-3900, USA. E-mail: schuller@uci.edu

(Received 18 April 1995; accepted 26 September 1995)

Abstract

The simultaneous application of multiple constraints such as non-crystallographic symmetry averaging, solvent levelling, histogram matching and Sayre's equation has proven to be very effective for phase refinement and extension. An existing program for this purpose, *SQUASH*, has been altered to increase the flexibility of its non-crystallographic symmetry averaging. The modified program, *MAGICSQUASH*, can handle multiple-domain averaging and multiple space group averaging and has aided in the solution of several structures. Examples are described which include the first simultaneous application of solvent levelling, histogram matching and Sayre's equation in two crystal forms while averaging between them.

1. Introduction

Many macromolecules crystallize with several identical monomers in the asymmetric unit. It has long been recognized that taking advantage of this non-crystallographic symmetry (NCS) by averaging the electron density of the identical monomers can reduce the noise in the electron-density map (Rossmann & Blow, 1962; Bricogne, 1974, 1976). When applied iteratively, NCS averaging can be used to refine and extend phases.

It also has been shown that when NCS averaging is applied simultaneously with other real-space and reciprocal-space constraints such as solvent levelling, histogram matching and the use of Sayre's equation, the improvement in phase refinement and extension can be greater than for any of the techniques applied individually (Cowtan & Main, 1993; Zhang, 1993). A program, *SQUASH*, is available for this purpose.

During the development of the theory and algorithms for NCS averaging, it was recognized that the concept also could be applied to identical monomers from more than one crystal form (Rossmann & Blow, 1962; Bricogne, 1974, 1976). Multiple space group NCS averaging in various implementations has indeed been utilized in structural determinations. A few exotic examples involving more than two crystal forms are listed here: Brown *et al.* (1993); Baumann, Wu, Flaherty & McKay (1993).

Some occurrences of NCS are not amenable to treatment with a single averaging procedure and a mask covering an entire monomer. This may be the case when the NCS relationships which apply to part of a monomer do not apply to all of the monomer. For example, Schuller, Grant & Banaszak (1995) and Jones *et al.* (1995) have recently reported hinged multidomain structures in which the hinge orientation differs in the two monomers within a single asymmetric unit. For such cases, benefit may be derived from carrying out several independent averaging procedures, each with a mask covering a separate domain and NCS operators refined for that domain. This shall be referred to here as multiple-domain NCS averaging.

A growing number of program suites are now available which can handle most NCS-related procedures, including multiple-domain and multiple space group averaging (see *Limitations* below). However, none of these suites incorporates the use of histogram matching or Sayre's equation as *SQUASH* does. In addition, a monolithic phase refinement and extension program such as *SQUASH* has clear advantages of speed and ease of use over suites in which the application of several separate programs may be required for each cycle of phase extension.

This report describes the incorporation of multiple-domain and multiple space group NCS averaging into the context of phase refinement and extension with multiple simultaneous constraints. To accomplish this, the NCS-averaging portion of *SQUASH* was altered. The modified program, *MAGICSQUASH*, has proven useful in the determination of a number of protein structures. Examples are described which make use of the multiple-domain and multiple space group averaging.

2. Methods

2.1. *Conventions*

An NCS operator M relating a point in one domain to the equivalent point in a duplicate domain consists of

a rotation followed by a translation, such that,

$$x' = M(x) = |R|(x) + T.$$

Since the treatment of averaging operators is well established (Bricogne, 1974, 1976), in this paper the notation for figures will be simplified. The rotation matrix and translation vector of an operator will not be listed separately; symbols will instead refer to the composite operator. More explicit formulations may be derived by applying the following expansions.

The inverse operation which restores the original position x from a transformed position x' can be expressed as,

$$M^{-1}(x) = |R|^{-1}(x) + |R|^{-1}(-T),$$

the combination of successively applied operators may be expressed as,

$$M_a[M_b(x)] = |R_a||R_b|(x) + |R_a|(T_b) + T_a.$$

Where,

$$M_a(x) = |R_a|(x) + T_a,$$

and

$$M_b(x) = |R_b|(x) + T_b,$$

are two NCS operations.

In the following discussion, the application of an NCS operator which transforms a reference domain onto one of the various domain copies will be referred to as a forward transformation (Fig. 1). These are the operators supplied by the user. The application of the inverse of a user-supplied NCS operator, which transforms a domain copy onto the reference domain will be referred to as a back transformation.

During program execution, coordinates must be converted from map array indices to fractional coordinates and then to orthogonal coordinates. After application of the NCS operators in orthogonal space, this process must be reversed. These steps may be accomplished either by serially applying the operators, or by combining the operators and applying the result. These details will not be explicitly stated in the following discussion for the sake of clarity.

2.2. *Existing SQUASH functionality*

The version of *SQUASH* upon which *MAGICSQUASH* is based uses the *CCP*4 *MTZ* reflection file format (Cowtan & Main, 1993; Collaborative Computational Project, Number 4, 1994). Another implementation of *SQUASH* with NCS averaging uses the older LCF reflection file format (Zhang, 1993). Neither version includes capability for multiple-domain or multiple space group averaging and both descend from an earlier version of *SQUASH* which did not incorporate NCS averaging at all (Zhang & Main, 1990).

Minimal changes were made in the operation of *SQUASH* other than to the NCS averaging code. Even some details of the original NCS code were kept, such as the interpolation algorithm. The map used for NCS interpolation may be calculated on a grid finer than that selected for other procedures in order to reduce interpolation error.

SQUASH generates the solvent-levelling mask separately from the averaging mask, by traditional methods (Wang, 1985; Leslie, 1987). This feature was maintained and is essential for averaging with multiple domains. Parts of the structure which fall between domains or in domains not being averaged would otherwise be lost during solvent levelling. For unusual cases, the ability to override solvent-mask calculation with a mask file has been added.

2.3. *Design goals*

An attempt was made to keep the new NCS routines as general as possible, so as not to introduce any artificial restrictions. In this way the program might be used for a number of structural solutions without coding changes. Transformations are treated in the most general way; *e.g.* for multiple space group averaging, the set of operators relating the reference mask to the various domain copies

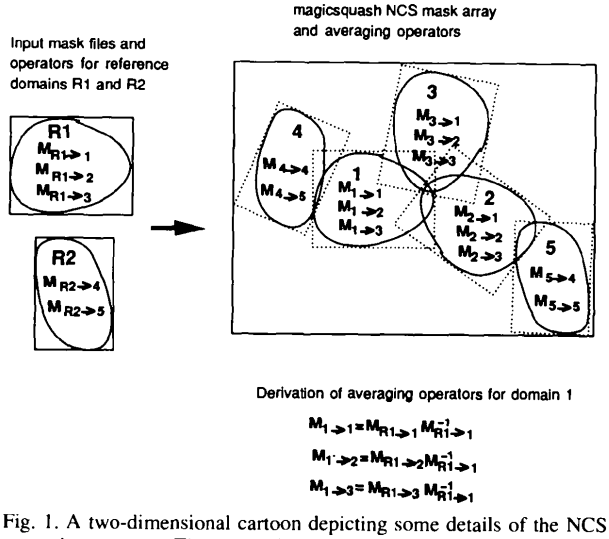


Fig. 1. A two-dimensional cartoon depicting some details of the NCS mask treatment. The array size needed to contain the NCS mask with the grid spacing chosen for map calculation can be determined by applying the NCS operators of each averaging procedure to the corners of the relevant mask file parallelipiped. This can be carried out with information obtained from the command script and the mask file header, before the array is needed. The placement of the individual domain masks into the NCS mask array is described in the text. Also shown is the derivation of the averaging operators for NCS procedure 1; they can be obtained by combining the inverse operator pointing from domain 1 back to the reference domain *R*1 with the operators pointing from *R*1 forward to the various domain copies. $M_{R1\to1}$ refers to the NCS operator which would transform a point in reference domain *R*1 to the equivalent point in task domain 1 of the NCS mask array. Operators for the other domains are derived in the same manner.

is likely to contain the identity transformation (identity matrix and zero length vector) in only one of the space groups, so the identity transform is not assumed and must be specified where applicable. Key array limits, such as the maximum number of domains, operators and space groups, are parameterized in an include file so that they may be easily increased for extraordinary applications.

Ease of use was another prime consideration. The amount of information required from the user was kept to a minimum, with as many routine functions as possible being carried out automatically. An example of this is the interpolation of the averaging and solvent-mask files into the grid spacing chosen for map calculation, so that the user need not be concerned with matching grid sampling during mask preparation.

2.4. *Multiple averaging tasks*

Multiple-domain averaging is incorporated by allowing the specification of several independent NCS procedures in the command script; each with a mask file covering a reference domain and a set of NCS operators relating that reference domain to each copy in the NCS multimer. Each input NCS procedure with N operators is then broken down into N dependent averaging tasks. Each dependent task consists of a mask covering one of the domain copies and a set of N operators relating that copy to each related domain copy, including itself. The operators of a task can be calculated by combining the NCS operators and inverse operators of the parent procedure (Fig. 1). Averaging tasks resulting from separate input procedures are independent. Dependent and independent tasks may be treated differently during mask overlap removal (see below).

2.5. *NCS mask array*

While *MAGICSQUASH* keeps the solvent-levelling mask and various electron-density maps in unit-cell-sized arrays, doing so with the NCS mask array would cause complications. For a mask point falling outside the basic unit cell ($0 \le x < 1, 0 \le y < 1, 0 \le z < 1$ in fractional coordinates), applying unit-cell translations to restore the point to the basic unit cell would require keeping account of the translation applied and reversing it before the application of the NCS operators. (The application of an NCS operator to two points differing only by unit-cell translations does not give equivalent results; the operator is after all non-crystallographic.) To avoid these complications, *MAGICSQUASH* instead keeps the NCS mask in an array large enough to include all of the domains of the NCS multimer without the application of crystallographic operators. The array dimensions required for this can be determined as described in Fig. 1.

The NCS mask array is integer rather than Boolean, which is necessary for the handling of improper symmetry (Bricogne, 1976). This also allows for multiple-domain averaging.

2.6. *NCS mask array creation*

In order to assign the individual domain mask belonging to an NCS task into the NCS mask array, the mask file covering the reference domain of an NCS procedure is read into a temporary array. The NCS mask array is scanned and each grid point is back transformed by each individual inverse NCS operator of that procedure. A value is interpolated from the temporary reference mask array to determine if the back-transformed point falls within that reference domain. If so, the original grid point of the NCS mask array is assigned an integer value unique to the task specified by the applied inverse NCS operator. These steps are repeated for each of the NCS procedures and in this way the domain masks for all averaging tasks are generated in the NCS mask array. Since interpolation is used, the grid spacing and extent of the reference mask file(s) need not agree with that of the NCS mask array, which uses the same grid spacing chosen by the program for map calculation. This last point is useful for multiple space group averaging, since the same reference mask may be used for all of the crystal forms involved.

2.7. *Overlap removal*

The various domains of the NCS mask are not allowed to overlap. This serves two purposes: (1) the exclusion of overlap allows all of the individual domain masks to be kept in a single array. (2) Since improper symmetry or multiple-domain symmetry will apply different operators to each separate domain, it is unclear which set of transformations should be applied to overlap regions.

Direct overlap between domains of independent averaging tasks (generated by separate NCS procedures) is disallowed in the NCS mask array. For dependent tasks (generated by a single NCS procedure), the treatment of direct overlap between domains depends on whether the NCS for those tasks is proper or improper (see below and Fig. 2). After the exclusion of direct overlap, the crystallographic symmetry operators are applied to all assigned grid points of the NCS mask array and the resulting points are translated into the basic unit cell. If any overlap in the unit cell results from the crystallographic operations, the original grid points in the NCS mask array are disallowed, whether the overlap occurs between dependent or independent task domains. Since the solvent-levelling mask is generated independently of the NCS mask, the disallowal of overlap points from averaging does not cause serious errors. It will diminish the fraction of the asymmetric unit being averaged, so the experimenter should take care to generate masks with minimal overlap beforehand.

2.8. *Proper versus improper symmetry*

Proper symmetry consists of a closed group of operators; *i.e.* the same NCS transformations apply to each

dependent task of a procedure with proper symmetry. Improper symmetry is any case for which the symmetry relationships of an NCS procedure are not proper; *e.g.* a dimer with a rotation of other than 180° is improper. For improper symmetry, the NCS operators which apply to one task are not identical to those of other dependent tasks. *MAGICSQUASH* checks for symmetry propriety by comparing the averaging operators generated for dependent tasks of an NCS averaging procedure.

For improper symmetry, separate domain masks for the individual dependent tasks must be generated in the NCS mask array. With proper symmetry a single mask could be generated covering all dependent task domains (Bricogne, 1976). However, in *MAGICSQUASH* a separate domain mask is maintained for each dependent task in order to allow the calculation of correlation statistics. Proper and improper cases are treated differently only with respect to overlap removal in the NCS mask (Fig. 2). This distinct treatment of proper and improper symmetry by *MAGICSQUASH* allows a user to simplify his/her mask creation in cases of proper symmetry. For example, a sphere containing a complete NCS multimer could be used as the reference mask for a case of single domain, proper symmetry. The averaging would be carried out correctly and only the statistical reporting would be compromised.

2.9. *Averaging*

Averaging may be carried out in a two-step procedure by back transforming all of the separate domain copies into a reference cell and summing the density, followed by restoring the averaged reference density to each position with a forward transformation (Bricogne, 1976; '*h*-cell' of Rossmann *et al.*, 1992). Alternatively, the inverse operator pointing from a task domain back to the reference domain may be combined with the NCS operators pointing from the reference domain forward to the various dependent task domains. The density of each transformed position of a grid point may thus be obtained in a single step by application of one of the combined operators. The latter approach is used in *MAGICSQUASH* as it requires only one interpolation step rather than two, which should result in higher accuracy (Bricogne, 1976). The former method has the benefit of allowing a mask to be constructed from the summed reference domain, but this would be very difficult to integrate with the use of multiple domains.

When it is time for *MAGICSQUASH* to average the map, the NCS mask array is scanned. For each grid point belonging to an averaging task domain, each averaging operator associated with that task is applied. The resulting coordinates are translated into the basic unit-cell and density values are calculated from the interpolation map and summed. The grid point itself is then translated into the basic unit cell (remember that the NCS mask array is not limited by unit-cell boundaries) and the corresponding value in the working map is replaced with the sum. Once all interpolation is finished, the density covered by each averaging task domain is normalized to its original mean and variance. Density modification then continues with solvent levelling and histogram matching (Fig. 3). Bricogne's double sorting algorithm (1976) is not implemented.

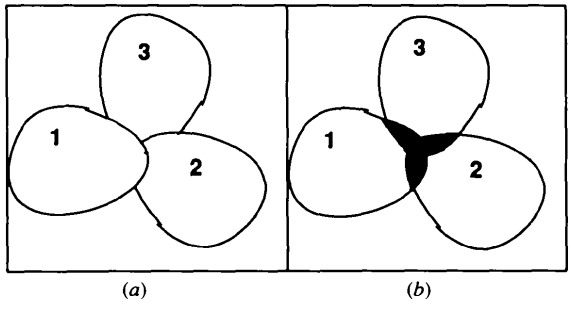


Fig. 2. The scheme for treatment of direct overlap between domains of dependent tasks in the NCS mask array depends on whether the symmetry relating the tasks is proper or improper. (*a*) If the symmetry is proper, overlap grid points are arbitrarily assigned to the domain with the lower index number. (*b*) If the symmetry is improper, overlap grid points are excluded from all averaging task domains. See the text for a description of proper *versus* improper symmetry. Overlap between domains of independent tasks and overlap due to crystallographic operations is excluded as in (*b*).

2.10. *Multiple space group averaging*

When multiple space group averaging is desired, a separate *MAGICSQUASH* job is started for each crystal form in which phases will be refined or extended. Averaging procedures which will be applied across crystal forms are specified in the command script. All jobs involved should use the same reference mask file(s), since all the involved NCS operators should apply to the same reference domain. When it is time for averaging, each job writes its unit-cell map to a file. After scanning the NCS mask array and carrying out any local averaging tasks, each job reads the map file written by each of the other jobs, scans the NCS mask array again and interpolates density from the foreign unit cell as required, adding the interpolated values to the summed density (Fig. 3).

To make sure the writing of a unit-cell file is completed before it is read, each job writes a 'flag file' for each shared NCS procedure after writing its map file. A job will communicate its NCS operators for an averaging procedure, along with synchronization information, in the flag file it writes for that procedure. A job will not read a map file until it finds the appropriate flag files. If an expected file is not found within a certain period of time, a job will terminate. After all averaging steps are carried out, each job proceeds independently with normalizing the averaged density and then continues with solvent levelling, *etc.*, just as it would in single space group applications.

3. Examples

MAGICSQUASH has already been used in the solution of a number of macromolecular structures such as phosphoglycerate dehydrogenase (Schuller, Grant & Banaszak, 1995), toxic shock syndrome toxin I (Prasad *et al.*, 1993), cytochrome $P450_{BM\text{-}3}$ (Li & Poulos, 1995), ascorbate peroxidase (Patterson & Poulos, 1995) and chloroperoxidase (Sundaramoorthy, Mauro, Sullivan, Terner & Poulos, 1996). Two examples will be detailed here, one involving multiple-domain averaging and the other multiple space group averaging. The discussion will center on the phase refinement; for description of the structures themselves see the original papers.

3.1. *Phosphoglycerate dehydrogenase*

In the case of phosphoglycerate dehydrogenase (PGDH) (Schuller Grant & Banaszak, 1995) the heavy-atom multiple isomorphous replacement (MIR) phases were of poor quality and limited to 3.5 Å resolution. Native data were available to 2.75 Å. Solvent levelling and histogram matching improved the map so that some β-strands and α-helices could be seen, but the connectivity between strands was poor. The crystals were of space group $P2_12_12$ and two monomers inhabited the asymmetric unit. Examination of an unaveraged map and the heavy-atom sites led to the solution of the non-crystallographic symmetry relating

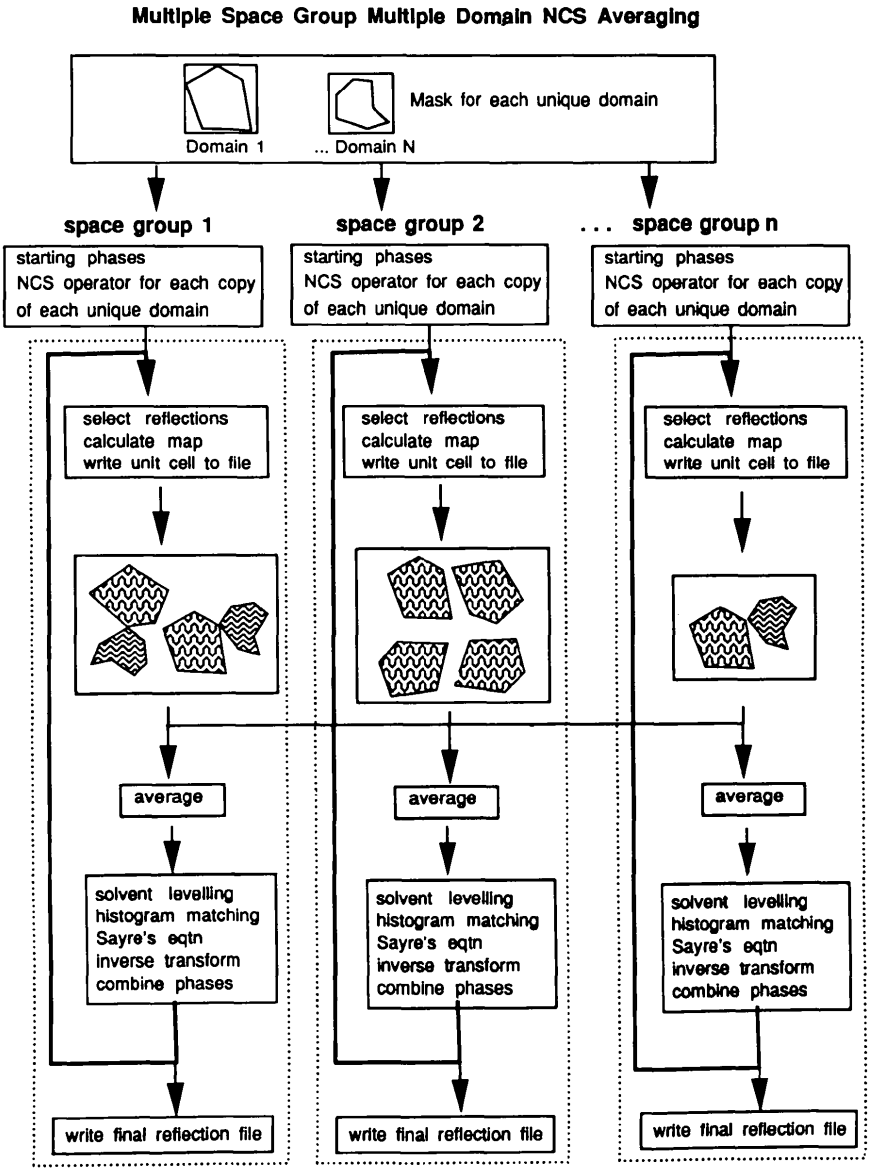


Fig. 3. Flow diagram for iterative phase refinement and extension. For multiple space group averaging, separate jobs for each space group exchange their unit-cell maps through files. The steps within a single *MAGICSQUASH* job are bounded by a dotted rectangle.

the nucleotide-binding domains. Averaging of the nucleotide domains (approximately one half of the polypeptide) improved that portion of the map so that the chain tracing throughout the nucleotide-binding domains could be completed. The coordinates of a homologous enzyme, glycerate dehydrogenase, were also quite helpful at that stage (Goldberg, Yoshida & Brick, 1994).

The connectivity of the unaveraged portion of the molecule could still not be resolved and including it in the averaging mask did not help. This was due to different conformations of an interdomain hinge in the two monomers (Fig. 4 of Schuller, Grant & Banaszak, 1995). The NCS operator for the other half of the protein was refined from a map which was solvent levelled and histogram matched with the nucleotide domains averaged. Phase refinement and extension was then carried out with each half of the monomer being averaged independently.

The first mask of each domain consisted of a sphere. Better masks were created from PDB coordinates as the model was improved. Mask creation and NCS operator refinement was accomplished with the *RAVE* programs *MAMA* and *IMP*, respectively (Kleywegt & Jones, 1994). The masks were converted to *CCP*4 format with *TRANSMASK*, a utility supplied by Ohlendorf (1993). Statistics for phase-extension trials with varying constraints are detailed in Table 1. Portions of the electron-density maps of selected trials are compared in Fig. 4.

3.2. *Chloroperoxidase*

In the case of chloroperoxidase (CPO) (Sundaramoorthy, Mauro, Sullivan, Terner & Poulos, 1996), two crystal forms diffracted to better than 2.0 Å resolution, $C222_1$ and $P2_12_12_1$, with MIR phases available to 2.7 Å for each. Each crystal form contained only one molecule in the asymmetric unit. Unaveraged maps in each crystal form were insufficient for complete map fitting. The NCS relationship between the two monomers was discovered by comparison of the heavy-atom sites for like derivatives. A mask for one of the monomers was constructed with *RAVE* program *MAMA* and the NCS operator was refined with *RAVE* program *MAVE* (Kleywegt & Jones, 1994). The mask was converted to *CCP*4 format with *TRANSMASK* (Ohlendorf, 1993). Statistics comparing phase-extension trials are displayed in Table 2. Portions of the electron-density maps of selected trials are compared in Fig. 5.

Table 1. *Phase extension for PGDH*

Extending phases of PGDH from 3.5 to 2.75 Å in 40 steps. CC is the correlation coefficient between a $2F_o - F_c$ map with model phases and a map created with F_o and the phase and figure of merit output by *MAGICSQUASH*, as reported by program *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994). Maps were calculated from 10.0 to 2.8 Å with $F/\sigma F > 2.0$ data. The initial NCS correlation coefficients for the NAD domain and other domains were 0.31 and 0.22, respectively, with masks based on the final model. When the second half of the protein was misaveraged by the NAD domain NCS operator, the initial correlation was 0.10. CPU times reported are for a Silicon Graphics Indigo with 100 MHz R4000 CPU (SPECint92 = 59, SPECfp92 = 61) and 64 MB RAM. SOLV is solvent levelling. HIST is histogram matching. SAYR is Sayre's equation, full solution mode. NCS is non-crystallographic symmetry averaging with an interpolation grid factor of 2.0.

	CC	CPU (min)
SOLV	0.40	20
SOLV + HIST	0.44	20
SOLV + HIST + SAYR	0.54	92
SOLV + HIST + NCS(bad)*	0.61	56
SOLV + HIST + NCS(nad)†	0.61	49
SOLV + HIST + SAYR + NCS(nad)	0.79	129
SOLV + HIST + NCS(both)	0.85	56
SOLV + HIST + SAYR + NCS(both)‡	0.87	138

* (bad) indicates that the NAD-binding domain is averaged correctly, but the other half of the polypeptide is misaveraged using the operator refined for the NAD-binding domain. † (nad) indicates averaging of only the NAD-binding domain. ‡ (both) indicates the NAD-binding domain and the other half of the polypeptide have been averaged separately and correctly, each with its own mask and refined NCS operators.

Table 2. *Phase extension for CPO*

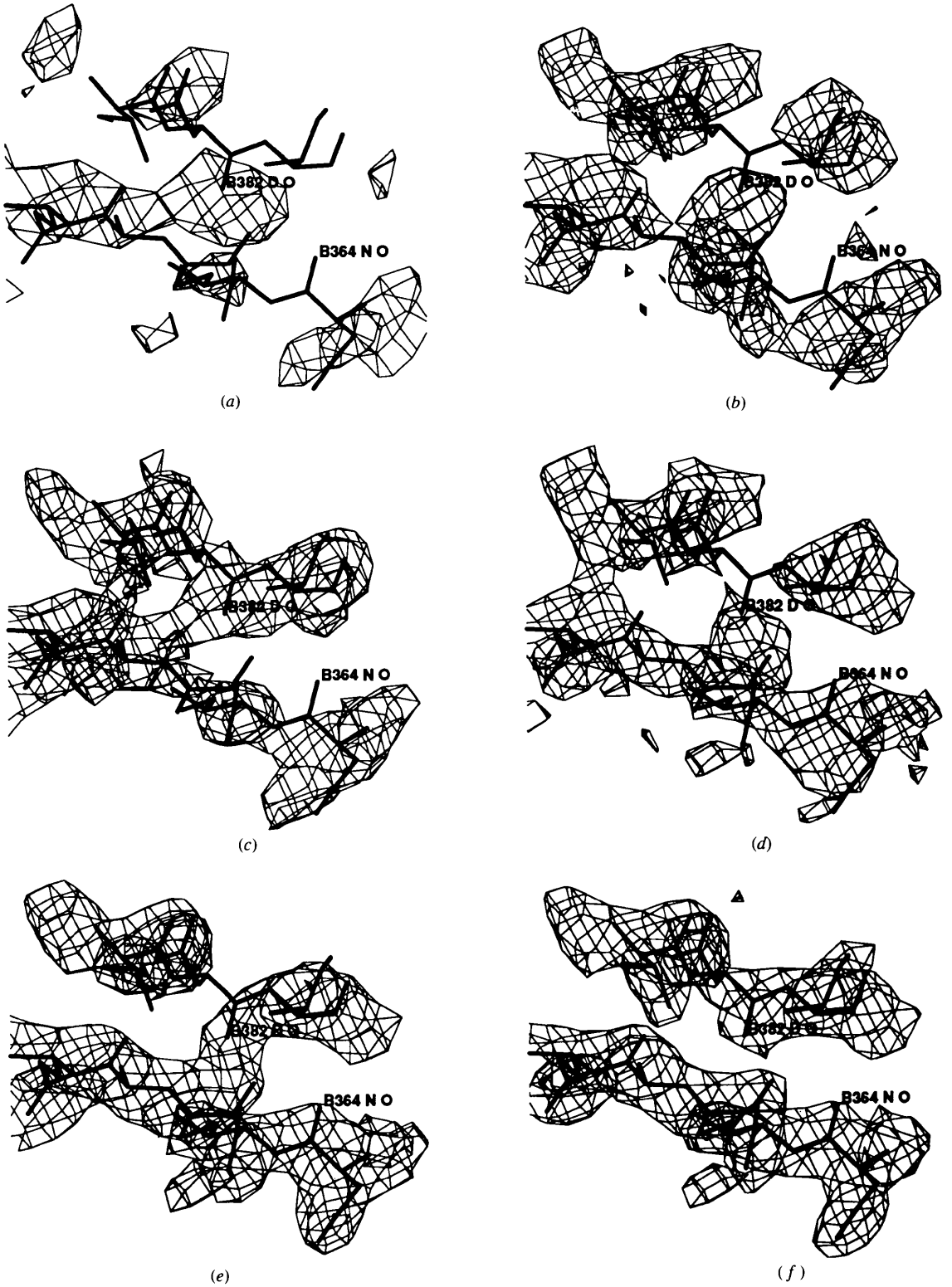
Extending phases for CPO from 2.7 to 2.1 Å in 25 cycles. Map correlations are between the $2F_o - F_c$, model phases map and one calculated with F_o, with phase and figure of merit output by *MAGICSQUASH*. Correlation coefficient as reported by *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994). Maps were calculated with $F/\sigma F > 2.0$ data between 10.0 and 2.1 Å. The averaging mask used for these trials was based on the final polypeptide and heme model, excluding the carbohydrates. The initial NCS correlation was 0.39. Elapsed computing time for the SOLV + HIST + SAYR + NCS jobs on a pair of networked computers like that listed in Table 1 was 217 min. SOLV is solvent levelling. HIST is histogram matching. SAYR is Sayre's equation, full solution mode. NCS is non-crystallographic symmetry averaging with an interpolation grid factor of 2.0.

	CC($C222_1$)	CC($P2_12_12_1$)
SOLV	0.49	0.53
SOLV + HIST	0.59	0.63
SOLV + HIST + SAYR	0.65	0.71
SOLV + HIST + NCS	0.84	0.84
SOLV + HIST + SAYR(P*) + NCS	0.87	0.88
SOLV + HIST + SAYR + NCS	0.88	0.89

* Sayre's equation applied only to the $P2_12_12_1$ space group in this trial.

4. Discussion

Structures with complex NCS requirements, such as PGDH and CPO, could not have been averaged fully using *SQUASH* without the *MAGICSQUASH* modifications. The necessary averaging might have been carried out with a general purpose averaging suite, such as those listed below under *Limitations*. However, these suites generally do not incorporate histogram matching or Sayre's equation, whose effectiveness can be judged from examination of Tables 1 and 2. At best, cycles



(*a*) (*b*) (*c*) (*d*) (*e*) (*f*)

Fig. 4. Electron density of PGDH at 50.0–2.8 Å resolution, except as noted. The location is in the regulatory domain. The final model is displayed as thick lines, with the electron density appearing as thin lines. Notation is the same as for Table 1. All maps are contoured at 1.5σ. Figs. 4 and 5 were generated with *FRODO* (Jones, 1982). (*a*) MIR map at 50.0–3.5 Å resolution. (*b*) SOLV + HIST + SAYR. (*c*) SOLV + HIST + NCS(bad). (*d*) SOLV + HIST + NCS(nad). (*e*) SOLV + HIST + SAYR + NCS(nad). (*f*) SOLV + HIST + SAYR + NCS(both).

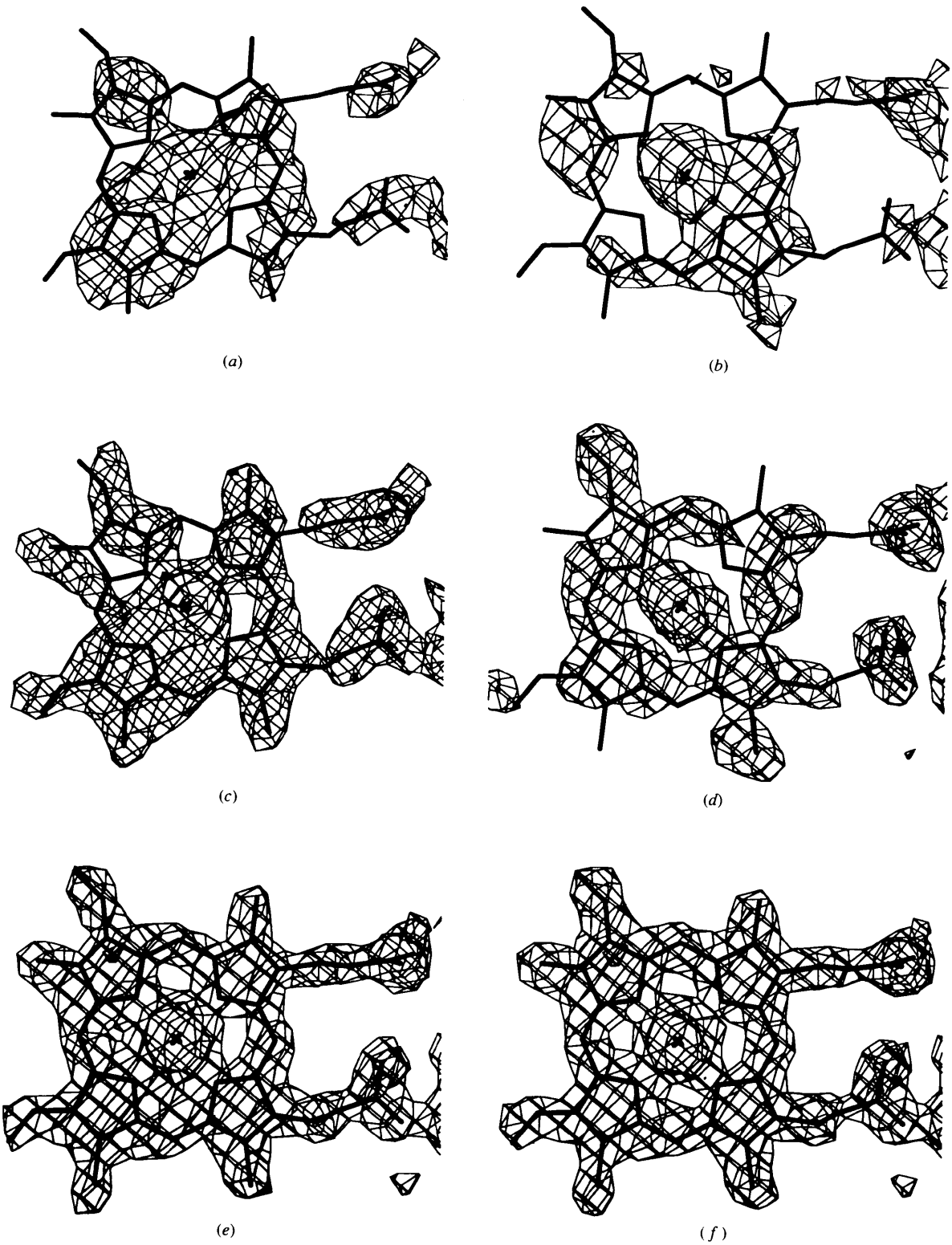


Fig. 5. Electron density of CPO at 50.0–2.1 Å resolution, except as otherwise noted. Notation is the same as for Table 2. All maps are contoured at 1.5σ. (*a*) MIR map at 50.0–2.7 Å resolution ($C222_1$). (*b*) MIR map at 50.0–2.7 Å resolution ($P2_12_12_1$). (*c*) SOLV + HIST + SAYR ($C222_1$). (*d*) SOLV + HIST + SAYR ($P2_12_12_1$). (*e*) SOLV + HIST + NCS ($P2_12_12_1$). (*f*) SOLV + HIST + SAYR + NCS ($P2_12_12_1$).

of averaging could have been interspersed with rounds of solvent levelling, histogram matching and Sayre's equation. This would be a compromise, since ideally all constraints should be applied simultaneously and should be maintained throughout phase extension.

Aside from the expanded averaging capability, *MAGICSQUASH* maintains much of the speed and ease of use of *SQUASH*. Computing times are listed in Tables 1 and 2 for comparison with other averaging packages. The *SQUASH* innovation of incorporating the phase-extension loop inside a single program (Fig. 3) also keeps the command scripts relatively short.

In the example of PGDH (Fig. 4), it is impossible to find one location that is truly representative of the overall quality when comparing so many maps which are treated differently in different regions. The location for Fig. 4 was chosen because it shows a break in the peptide chain with only one domain averaged and Sayre's equation applied (Fig. 4*e*) which is not found in the maps with both domains averaged (Fig. 4*f*). In truth the (Fig. 4*e*) map is elsewhere generally quite good and the number of such breaks is not great, as may be deduced from the values in Table 1. This is impressive considering the resolution is marginal for the application of Sayre's equation. The map with the substrate binding and regulatory domains mis-averaged by the NCS operator for the NAD binding domain (Fig. 4*c*) varies regionally in quality from good to terrible, depending on the superposition error introduced by the hinge shift.

The solution of CPO is the first case in which solvent levelling, histogram matching, Sayre's equation and multiple space group averaging have been simultaneously applied. In Fig. 5, the averaged maps for CPO without (Fig. 5*e*) and with (Fig. 5*f*) Sayre's equation are both very good and either would be adequate for complete model fitting. The map with Sayre's equation is slightly better, as can be seen in the increased distinction between the porphyrin ring and the Fe atom. Elsewhere in the map, the non-Sayre's map contains a few breaks in the peptide chain not found in the Sayre's map, these breaks tend to occur at surface loops. For more marginal applications, the improvement due to the use of Sayre's equation might be enough to make a difference in the ability to fit a model to the map.

Aside from the benefit of simultaneous application, the other constraints can mesh with NCS averaging in other ways. Solvent levelling, histogram matching and Sayre's equation can all be applied when the NCS relationships are not known. They can, therefore, be used to obtain maps for use in NCS mask creation and for the discovery and refinement of operators. With PGDH, perhaps the clearest example of this can be seen when only the NAD-binding domains are averaged. Adding Sayre's equation to the other constraints improves the map quality dramatically, even in map regions which are not averaged (Table 1, Fig. 4*e*). The resulting map is more than adequate for resolving the NCS for the remaining half of the molecule, it would in fact have been adequate for fitting a large fraction of the model, although historically this possibility did not arise during the PGDH structure solution. A coding error, since corrected, prevented the use of Sayre's equation at the time. Notably, this example is also a situation where using the NCS mask for solvent levelling, instead of calculating the solvent-levelling mask separately, would have been disastrous.

In practice, early averaging masks may consist of spheres covering most of a domain while not intruding into neighboring domains. The mask may be improved as the model improves. Even as the model becomes fairly complete, it is useful to update the phase extension with the best available mask, since a map from *MAGICSQUASH* is biased by the applied constraints but not by the model.

4.1. *Limitations*

To make use of the described procedure, starting phases must exist for each space group to be treated.

MAGICSQUASH has no provisions for creation of averaging masks or refinement of NCS operators. These are not viewed as part of the phase-extension loop, but as a prerequisite. A number of general-purpose NCS suites are now available which possess these capabilities, such as *SKEWPLANES* (Bricogne, 1976), *PHASES* (Furey & Swaminathan, 1990), *AVGSYS* (Bolin, Smith & Muchmore, 1993), *RAVE* (Kleywegt & Jones, 1994), *DEMON* (Vellieux, Hunt, Roy & Read, 1995) and *GAP* (Grimes & Stuart, 1995), *RAVE*, *GAP* and *DEMON* in particular are fairly recent full-featured NCS-averaging suites which are claimed to perform both multiple-domain and multiple space-group averaging. The *CCP*4 suite also contains some tools for mask creation (Collaborative Computational Project, Number 4, 1994; Cowtan, 1995).

MAGICSQUASH takes a substantial amount of computer memory. Workstations with 64 megabytes or more of random access memory (RAM) are suggested. Both *SQUASH* and *MAGICSQUASH* carry out all real-space procedures on the unit cell. It would be possible to use the asymmetric unit instead with considerable savings in array sizes for most space groups, although this would require an increase in coding complexity.

Setting up and synchronizing multiple space group averaging applications takes some care. For multiple space group applications in which N space groups are treated, N workstations are recommended, with disk space shared between them for the exchange of the unit-cell map files. If multiple space group applications are carried out on a single computer, it should have extensive RAM and preferably multiple processors. The implementation of multiple space group averaging as separate jobs communicating through files may not be ideal; but it seems to work well enough on computing

configurations which are common today and the implementation was possible with minimal changes to the existing *SQUASH* code.

MAGICSQUASH must be compiled with the *CCP*4 subroutine libraries (Collaborative Computational Project, Number 4, 1994). *MAGICSQUASH* is written mostly in FORTRAN 77, with a few extensions. The file handling required for the multiple space group averaging necessitated the use of some Unix operating system specific code. The program compiles and runs on versions of Unix from Silicon Graphics and Digital and may run on other modern Unix systems.

4.2. *Availability*

Further information about *MAGICSQUASH* version 4.0 may be obtained from the author. Because the described modifications were made to a licensed program, *MAGICSQUASH* can only be distributed to researchers already licensed for *SQUASH* through Molecular Structure Corp. (1995).

Another phase-refinement and extension program, *DM*, has multiple-domain averaging capability based partly on that of *MAGICSQUASH* (Cowtan, 1994). At the time of this writing, multiple space group averaging in *DM* is still under development (Cowtan, 1995). *DM* is available as part of the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

I would like to thank L. J. Banaszak for funding and facilities (NSF grant MCB 9318699), T. L. Poulos for funding and facilities (NIH grant GM33688 & NSF grant MCB9405218), D. H. Ohlendorf for asking helpful questions during program development, M. Sundaramoorthy for providing the chloroperoxidase test case, P. Dunten and K. D. Cowtan for useful discussions and members of the Poulos laboratory for proofreading the manuscript.

References

Baumann, U., Wu, S., Flaherty, K. M. & McKay, D. B. (1993). *EMBO J.* **12**, 3357–3364.

Bolin, J. T., Smith, J. L. & Muchmore, S. W. (1993). Proc. Am. Crystallogr. Assoc. Meet., Albequerque, NM, Vol. 21, p. 51.

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

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

Brown, J. H., Jardetsky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993). *Nature (London)*, **364**, 33–39.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. (1994). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38.

Cowtan, K. D. (1995). Personal communication.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Furey, W. & Swaminathan, S. (1990). 14th Am. Crystallogr. Assoc. Meet., New Orleans, Abstr. PA33.

Goldberg, J. D., Yoshida, T. & Brick, P. (1994). *J. Mol. Biol.* **236**, 1123–1140.

Grimes, J. & Stuart, D. I. (1995). Unpublished results.

Jones, E. Y., Harlos, K., Bottomley, M. J., Robinson, R. C., Driscoll, P. C., Edwards, R. M., Clements, J. M., Dudgeon, T. J. & Stuart, D. I. (1995). *Nature (London)*, **373**, 539–544.

Jones, T. A. (1982). *Computational Crystallography*, edited by D. Sayre, pp. 303-317. Oxford University Press.

Kleywegt, G. J. & Jones, T. A. (1994). *From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59-66. Warrington, England: Daresbury Laboratory.

Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.

Li, H. & Poulos, T. L. (1995). *Acta Cryst.* D**51**, 21–32.

Molecular Structure Corporation (1995). *SQUASH*, Molecular Structure Corporation, 3200A Research Forest Drive, The Woodlands, TX, 77381 USA.

Ohlendorf, D. H. (1993). Personal communication.

Patterson, W. R. & Poulos, T. L. (1995). *Biochemistry*, **34**, 4331–4341.

Prasad, G. S., Earhart, C. A., Murray, D. L., Novick, R. P., Schlievert, P. M. & Ohlendorf, D. H. (1993). *Biochemistry*, **32**, 13761–13766.

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

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J.-B., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.

Schuller, D. J., Grant, G. A. & Banaszak, L. J. (1995). *Struct. Biol.* **2**, 69–76.

Sundaramoorthy, M. (1995). *Structure*, **3**, 1367–1377.

Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.

Wang, B. C. (1985). *Methods Enzymol.* **115**. 90–112.

Zhang, K. Y. J. (1993). *Acta Cryst.* D**49**, 213–222.

Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 377–381.