Acta Cryst. (1995). D**51**, 1020–1024

Use of 'Random-Atom' Phasing Models to Determine Macromolecular Heavy-Atom Replacement Positions

BY DAVID A. LANGS, DONGYAO GUO AND HERBERT A. HAUPTMAN

Hauptman-Woodward Medical Research Institute, 73 High Street, Buffalo NY 14203, USA

(*Received* 31 *January* 1995; *accepted* 9 *May* 1995)

Abstract

A procedure is described by which the phase-invariant translation function may be used to provide starting set phases for the minimal function that are significantly better than those generated from random-atom coordinates. Applications to determine the heavy-atom positions for single isomorphous replacement data from macromolecular structures are very encouraging. In the case of chiral space groups, *e.g.* $P4_12_12$ *versus* $P4_32_12$, unlike Patterson functions, these methods provide the correct enantiomorph coordinates for the heavy-atom sites for whichever space group is chosen.

Introduction

Random-start phasing algorithms (Yao, 1981) provide a significant alternative to traditional basis set phase permutation methods, especially when these latter tangent-formula methods fail because of encountering aberrant phase relationships early in the phase extension of these iterative step-wise procedures. The minimal function (Hauptman, 1988, 1991) is the basis of a newly developed random-start phase-determination procedure which has been referred to as the 'shake-and-bake' or 'SnB' method (Weeks, DeTitta, Hauptman, Thuman & Miller, 1994). The SnB method has been shown to be successful in instances in which tangent-formula direct methods have had difficulties in providing a phase solution (Miller, DeTitta, Jones, Langs, Weeks & Hauptman, 1993). Unlike *RANTAN*-based procedures which require random phase starting sets, the SnB method utilizes phases computed from random-atom starting points. This paper describes how the phase invariant translation function (Langs, 1992) can be used to obtain coordinate-based phases for the SnB process that are significantly better than those supplied by randomly selected atoms.

Background

The minimal function is a minimum variance residual based on the theoretical expected cosine value of the constituent phase invariants (Hauptman, 1988, 1991).

$$R(\varphi) = \sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{hk}} \left\{ \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{h}} - \varphi_{\mathbf{h+k}}) - \tau_{\mathbf{hk}} \right\}^2 \Big/ \sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{hk}}, \tag{1}$$

where $A_{\mathbf{hk}} = 2|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h-k}}|N^{1/2}$, N represents the number of equivalent non-H atoms of the structure in the primitive reduced cell, $\tau_{\mathbf{hk}} = I_1(A_{\mathbf{hk}})/I_o(A_{\mathbf{hk}})$, and I_n is the nth order modified Bessel function of the first kind. Random-start multisolution methods have been used to find solutions to $R(\varphi)$ which minimize the value of the residual under the constraint of atomicity. While phases computed from randomly generated atomic positions have proved effective starting points from which to initiate the SnB refinement procedure, strategies to use the initial value of $R(\varphi)$ to prescreen different 'random-atom' phase sets to identify trials having a greater likelihood of producing a solution have generally not proved useful. The random-atom phase sets which do eventually converge to a solution most often have an intial r.m.s phase error that is not much less than 100°, or very close to that expected for completely random phases, *i.e.* 103.92° ($\pi/3^{1/2}$ radians). Although phase sets that have an r.m.s. phase error in the vicinity of 80° or less almost always converge toward solutions upon SnB refinement, this phase error is much too large for the value of $R(\varphi)$ to be diagnostic of this potential.

An alternative phase invariant function was proposed (Guo & Hauptman, 1990) as an *a priori* figure of merit to help identify those phase sets which have a low initial r.m.s. phase error. This 'maximal' function, $M(\varphi)$, was formulated as the sum of cosines of the triples and quartets phase invariants weighted by their respective A and B values,

$$\begin{aligned} M(\varphi) = &\sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{hk}} \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h+k}}) \\ &+ \sum_{\mathbf{h},\mathbf{k},\mathbf{l}} B_{\mathbf{hkl}}(|E_{\mathbf{h+k}}|^2 + |E_{\mathbf{h+l}}|^2 + |E_{\mathbf{k+l}}|^2 - 2) \\ &\times \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} - \varphi_{\mathbf{h+k+l}}). \end{aligned} \tag{2}$$

$M(\varphi)$ tends to be large and positive for trial sets of random-atom phase which have a lower than average r.m.s. phase error with the true structure. Here $A_{\mathbf{hk}}$ is the triples A value defined as in (1) and $B_{\mathbf{hkl}} = 2|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}E_{\mathbf{h+k+l}}|/N$ for the analogous quartets structure product.

ISSN 0907-4449

Although it may be straightforward to use $M(\varphi)$ to prescreen trial phase sets to identify candidates with lower than average r.m.s. phase errors, it may be computationally prohibitive to exhaustively examine all of the permissible three-dimensional space for positioning a one-atom probe, unless this is done on a fairly course grid interval, say 0.5 Å, or if the search may be restricted to only two dimensions, as in the case of monoclinic $P2_1$ symmetry. It is crucial that a more efficient computation scheme for evaluating (2) would be necessary in three-dimensional applications if the number of points of evaluation were excessive.

Phase-invariant translation function

A phase-invariant translation function was developed that reduced the computational complexity of (2) to an FFT (Langs, 1992).

$$\Phi(\mathbf{r}) + \sum_{\mathbf{h},\mathbf{k}} |E_\mathbf{h} E_\mathbf{k} E_\mathbf{l}|\, |E_\mathbf{h} E_\mathbf{k} E_\mathbf{l}(\mathbf{r})| \cos(\varphi_\mathbf{h} + \varphi_\mathbf{k} + \varphi_\mathbf{l}), \quad (3)$$

$$E_\mathbf{h} E_\mathbf{k} E_\mathbf{l}(\mathbf{r}) = \sum_{j,v,w}^{n} \sum_{p}^{m} E_{\mathbf{h}_j p} E_{\mathbf{k}_v p} E_{\mathbf{l}_w p} \times \exp 2\pi i [(\mathbf{h}_j + \mathbf{k}_v + \mathbf{l}_w)\cdot\mathbf{r}].$$

The Fourier transformable product $E_\mathbf{h} E_\mathbf{k} E_\mathbf{l}(\mathbf{r})$ is composed of molecular transforms computed for the pth molecular fragment at the various symmetry-related reciprocal lattice vectors $\mathbf{h}_j = \mathbf{R}_j^t \cdot \mathbf{h}$, where $\mathbf{R}_j^t$ is the transpose of the jth symmetry matrix operator. Strictly speaking, $\Phi(\mathbf{r})$ is only an approximation of $M(\varphi)$, since the magnitude of the calculated structure product, $|E_\mathbf{h} E_\mathbf{k} E_\mathbf{l}(\mathbf{r})|$, is not constant, but is equal to the product of the computed E values of the one-atom structure at the various positions in the cell. The magnitudes and phases of the molecular transforms of the structure need only be computed once, the brunt of the calculation over the various grid point positions $\mathbf{r}_j$ of the search space is evaluated through the FFT, which is roughly 1000 times faster than computing the phase invariant sums iteratively at each grid point. If in performing these one-atom searches, the initial coordinate position of the search atom is chosen at the origin, all amplitudes $|E\mathbf{h}_j|$ will equal 1.0, the phases $\varphi\mathbf{h}_j$ will equal $\mathbf{h}\cdot\mathbf{t}_j$, the translational components of the jth equivalent position, and the peaks of the map $\Phi(\mathbf{r})$ should correspond to the atomic positions of the structure relative to all the permissible choices of origin and enantiomorph. The peaks of $\Phi(\mathbf{r})$ do not uniquely define the structure. One must examine all permissible shifts of origin and inversion of the enantiomorph to establish how any one peak may be related to another.

Experimental calculations

Three $P2_12_12_1$ test structures of various sizes from 28 to 316 non-H light atoms were used to evaluate whether the three-dimensional one-atom translation synthesis would produce map maxima whose coordinates tended to be in the vicinity of the true coordinate positions of the structure; (I) PR435, $N = 28$, (Weeks, Duax & Wolf, 1976), (II) ILED, $N = 84$ (Pletnev, Galitskii, Ivanov, Smith, Weeks & Duax, 1980), (III) GRAMA, $N = 316$ (Langs, 1988). A total of 3003 triples were generated from 280 phases for (I), 5024 triples from 500 phases for (II) and 17 573 triples from 1500 phases for (III). These correspond to only those triples whose A values exceeding 1.0, 0.7 and 0.5, respectively. $\Phi(\mathbf{r})$ maps were computed for these triples basis sets at a minimal resolution ~ 0.12 Å to ensure that all major peaks would be properly determined using Rollett's 19-point interpolation scheme (Rollett, 1965). A total of 167 non-negative peaks were noted for the PR435 structure, while the top largest 200 and 400 peaks for the respective ILED and GRAM structures were examined. Table 1 notes the total number of peaks which were found within 0.20 and 0.40 Å of a true atom location for each of the three structures. Calculations were also performed using quartet invariants, but were shown to be considerably less effective that those utilizing the more reliable triples lists and will not be discussed further.

Table 1. *Comparison of number of peak positions from the $\Phi(\mathbf{r})$ maps that were within* 0.20 *and* 0.40 Å *of a true atom location for the PR*435, *ILED and GRAM structures*

There are two entries for each structure. The first row of each entry samples only the top N largest peaks, the second row includes peaks that occur further down the list. The percentage of peaks which are expected to fall within these $\delta\mathbf{r}$ limits from randomly generated coordinates are given as the last row of the Table. The 28-atom PR435 structure demonstrates that the structure information from the phase-invariant translation function is concentrated in the top 28 peaks of the map. The larger structures do not mirror these results.

	$\delta\mathbf{r} \le 0.20$ Å		$\delta\mathbf{r} \le 0.40$ Å	
	Number	%	Number	%
PR435	6/28	21.4	13/28	46.4
	13/167	7.8	36/167	21.5
ILED	2/84	2.4	10/84	11.9
	7/200	3.5	35/200	17.5
GRAM	6/200	3.0	43/200	21.5
	13/400	3.2	88/400	22.0
Random	—	2.7	—	21.8

Similar triples-based $\Phi(\mathbf{r})$ maps were computed for a Pt_2I_2(ethylenediamine)$_2$ derivative of the $P4_12_12$ structure of 20β-hydroxy steroid dehydrogenase (β-HSDH) which diffracted to 3.0 Å resolution (Duax & Ghosh, 1995), and three SIR derivatives for the $P2_12_12_1$ structure of cytochrome c_{550} (C550) (Timkovich & Dickerson, 1976). The C550 SIR data were obtained from the Brookhaven Protein Data Bank (entry pdb155c.ent). The native and the three derivatives, $PtCl_4^{2-}$, $Pt(CN)_4^{2-}$ and UO_2^{2+}, all diffract to 2.5 Å resolution, the total number of data deposited were 2994, 2807, 2399 and 2886, respectively. There are 4160 measurable data within the 2.5 Å shell, the majority of

the missing data represent weak reflections that were discarded.

The $PtCl_4^{2-}$ and UO_2^{2+} are essentially full-occupancy, single-site derivatives with the Pt and U ions located at ($x = 0.0110$, $y = 0.0524$, $z = 0.2097$) and ($x = 0.0995$, $y = 0.3012$, $z = 0.7396$). respectively. The $Pt(CN)_4^{2-}$ complex is a four-site derivative with Pt atoms at ($x = 0.0979$, $y = 0.4307$, $z = 0.7934$, oc = 1.00), ($x = 0.4690$, $y = 0.4690$, $z = 0.8628$, oc = 0.62), ($x = 0.3074$, $y = 0.2834$, $z = 0.7649$, oc = 0.45) and ($x = 0.3999$, $y = 0.0976$, $z = 0.7727$, oc = 0.20). The β-HSDH platinum derivative also has four partial occupancy sites ($x = 0.4877$, $y = 0.1705$, $z = 0.3762$, oc = 0.57), ($x = 0.4589$, $y = 0.286$, $z = 0.3482$, oc = 0.21), ($x = 0.4710$, $y = 0.2265$, $z = 0.3610$, oc = 0.17) and ($x = 0.4990$, $y = 0.1472$, $z = 0.3611$, oc = 0.14). In this latter derivative the multiple sites represent local positional or thermal disorder of the $Pt_2I_2(en)_2$ moiety at a single methionine-binding residue.

Normalized structure factors were scaled using an anisotropic temperature-factor model (Blessing & Langs, 1988). Difference structure factors were obtained using the relation,

$$|\Delta E_r| = \left|\sigma_d|E_d| - \sigma_n|E_n|\right|/\sigma_r, \qquad (4)$$

where σ_d, σ_n and σ_r equal to $(\sum Z_j^2)^{1/2}$ for the derivative, native and heavy-atom difference structures, and Z_j is the number of electrons for atoms of each chemical type summed over the contents of the unit cell. These ΔE values represent a lower bound estimate for the heavy-atom scattering for non-centric reflections, and as such the average $|\Delta E|^2$ for the data set will be less than 1.0. Conversely, if the choice is made to process ΔF values resultant from the scaled native and derivative data to obtain ΔE amplitudes, it follows that the average $|\Delta E|^2$ would be scaled to more closely equal 1.0, regardless of the fact that a lower bound estimate cannot be expected to do so. Approximately 185 of the largest $\Delta E_\mathbf{h}$ were selected for each of the single isomorphous replacement (SIR) derivatives, and these in turn were used to generate between 850 and 1000 triples in each case. The two multi-site derivatives were later shown to produce better results if 300 $\Delta E_\mathbf{h}$'s were selected and ~4000 triples were used. The results of these translation-function searches are presented in Table 2.

Initially, two paired sets of minimal-function trials were run, protocol *A* with ΔE data derived from the full data set, and protocol *B* with ΔE data computed only from those E_d and E_n which were larger than 0.70. This latter trial was performed to test whether the large ΔE values obtained using weaker data, which had proportionately larger r.m.s. errors, might provide unreliable estimates for ΔE. The test calculations involved running *ab initio* trials using the minimal-function program to test the rate of success and discriminatory sensitivity as compared to false solutions. The results for the C550 derivatives for the full and truncated SIR data sets are summarized in Table 3.

Table 2. *Summary of triples translation function results for the various SIR derivative data sets*

These results correspond to data obtained when ΔE values were obtained from the full set of paired native and derivative E values. The rank position of the correct solution vectors are given in column five. Column 6 lists the ratio of the top solution peak height as compared to the largest spurious peak in the translation function map. Two row entries are presented for the $Pt(CN)_4$ and $Pt_2I_2en_2$ derivatives; slightly better results were obtained for the $Pt(CN)_4$ data set when the basis set of ΔE values was increased to 300 phases.

Structure	Derivative	No. of ΔE's	No. of triples	Ranks	Ratio
C550	$PtCl_4$	183	897	1	1.72
	UO_2	183	868	1	2.50
	$Pt(CN)_4$	186	833	2*,7	0.60
	$Pt(CN)_4$	300	3712	1*,3,12	1.37
β-HSDH	$Pt_2I_2en_2$	183	1091	2*,7*	0.95
	$Pt_2I_2en_2$	300	4280	2*	0.75

* Peaks resulting from the overlap of two solution vectors.

Further details concerning the actual peak listing resultant from the triples translation function of the $Pt(CN)_4^{2-}$ four-site derivative are given in Table 4(*a*). A listing of the minimal-function peak heights resultant from recycling the top solution peak of the triples translation map into the minimal function refinement are listed in part (*b*) of Table 4. This latter listing removes the origin ambiguity among the solution peaks in the upper part of Table 4.

Discussion

The results listed in Table 1 strongly suggest that the triples translation function is not effective with regard to providing one-atom starting coordinates for structure determination of greater than 80 atoms complexity. Although the results for the 28-atom PR435 structure indicate an eightfold increase above the random probability in finding an atom within 0.20 Å of a true location within the top 28 peaks, this advantage is completely lost in applications to the larger ILED and GRAMA structures. In spite of this discouraging revelation, the method may still prove useful in determining heavy-atom locations from SIR data as is shown in Table 2. The triples translation-function results for four seperate macromolecular data sets are each very encouraging. The solutions for the single-site derivatives both occurred as the largest peak in the map, that being 1.72 and 2.5 times larger than the next spurious peak. For the two four-site derivatives, the correct higher occupancy solution vectors were found within the top three peaks. Only sites representing 20% occupancy or less tended to fall lower down the list.

A typical minimal-function refinement cycle for an *N*-atom structure usually consists of computing phases for the 10*N* largest E values (initially from the coordinates of *N* randomly selected atoms), sequentially optimizing

Table 3. *Comparison of minimal function results from the three C550 SIR data sets*

The two sets of trials correspond to ΔE's derived from (*A*) the full set of data, and (*B*) only those data for which the native and derivative E's both exceeded 0.70. Approximately 185 E's and 1000 triples were employed as was indicated in Table 2. Cosav is the *A*-weighted average of the triples cosine values [Cosav = $\sum A_{h,k} \cos(\varphi_h + \varphi_k - \varphi_{h+k}) / \sum A_{h,k}$]. A low value for $R(\varphi)$ and a high value for Cosav are generally cosistent with a good solution. The solution rates (Sol%) ranged between 0.6 and 9% and were based on 400 random trials unless otherwise noted.

		Protocol *A* $R(\varphi)$/Cosav			Protocol *B* $R(\varphi)$/Cosav	
Derivative	Sol%	Solution	Non-solution	Sol%	Solution	Non-solution
UO_2	9	0.177/0.970	0.475/0.444	3	0.298/0.892	0.348/0.608
$PtCl_4$	6	0.158/0.855	0.475/0.626	0.6	0.318/0.708	0.408/0.593*
$Pt(CN)_4$	4	0.344/0.554	0.463/0.427†			
	1	0.487/0.586	0.546/0.517‡			

* Approximately 5% of the non-solutions gave a pair of spurious peaks that were ±2.0 Å on either side of the y coordinate of the true solution vector location. The 0.6% solution rate represents 12 successful convergences in 2000 random trials.

† The average $|E|^2$ for the data obtained by equation (4) was 0.40. The 4% solution rate is based on 16 out of 400 random trials which gave either the 1.0 or 0.62 occupancy site as the first or second map peak.

‡ The average $|E|^2$ was rescaled to equal 1.0. There were 38 out of 4000 random trials which produced solutions as defined in the footnote above.

Table 4. *Translation-function and minimal-function peaks*

In (*a*) the occupancy of the particular atom associated with each solution peak is indicated to facilitate reference to the coordinate values given in the text. In (*b*) all four substitution sites were correctly located within the top 11 peaks.

(*a*) Translation-function peaks from the four-site $Pt(CN)_4^{2-}$ derivative using the 300 largest ΔE values and 3712 triples

Peak	X/A	Y/B	Z/C	Density	Occupancy
1	0.0896	0.0713	0.2063	16.10*	1.00, 0.20
2	0.2326	0.0800	0.0753	11.76	
3	0.0299	0.0279	0.1278	11.12*	0.62
4	0.2500	0.0703	0.2031	9.95	
11	0.1094	0.0000	0.2031	8.31	
12	0.1861	0.2126	0.2532	8.30*	0.45

(*b*) Minimal-function peaks for $Pt(CN)_4^{2-}$ from the 186 phase refinement given in row three of Table 3

Peak	X/A	Y/B	Z/C	Density	Occupancy
1	0.5922	0.5701	0.2103	0.946†	1.00
2	0.9653	0.5308	0.1318	0.761†	0.62
3	0.6193	0.4316	0.2180	0.380	
4	0.0146	0.0650	0.0949	0.337	
5	0.9547	0.4799	0.0959	0.280	
6	0.9052	0.0310	0.2017	0.278	
7	0.0918	0.4974	0.1579	0.258	
8	0.7985	0.7156	0.2112	0.257†	0.45
9	0.0058	0.1461	0.1476	0.249	
10	0.0519	0.2793	0.1348	0.247	
11	0.9136	0.9007	0.2072	0.235†	0.20

* Solution vectors which are not consistent with the same origin and enantiomorph.

† Solution vectors which are all consistent with the same origin and enantiomorph.

each of these $10N$ phases to minimize $R(\varphi)$ by using a parameter-shift procedure, and computing an E map from the refined E values from which one selects the N largest peaks to reiterate the cycle. In the SIR phasing application we found it necessary to select more than $10N$ phases to insure a triples/phases ratio of 5 or greater, in fact $\sim 200N$ phases were required to achieve this overdeterminacy. In all of the calculations if was best to assume one was searching for N full occupancy sites, plus three others which were given occupancies of 0.75, 0.50 and 0.25. This strategy was required for the single-site derivatives because one would lock onto the intial random-atom position and it would generally not be displaced during ten or more cycles of the phase-refinement process. Searching for three additional atoms, even though they were not anticipated in the heavy-atom structure, allowed the refinement to access peaks below the normal N-atom cutoff and give them an opportunity to alter the direction of the phasing. This alteration of the normal 'SnB' protocol enhanced the solution rate considerably. The three supplementary atoms, could be gradually weighted out of the refinement in the latter cycles, if one wished, say from NCYCLE equal to 6 through 10, by raising the fractional weights (0.75, 0.50, 0.25) to the NCYCLE-5 power. This had a pronounced effect on reducing the solution value of $R(\varphi)$ for single-site derivatives.

The minimal-function results presented in Table 3 establish that it is probably a good strategy to compute ΔE values from the full data set (protocol *A*), and not be concerned with errors expected in the weaker measurements (protocol *B*). Not only are the $R(\varphi)$ and triples average *A*-weighted cosine values (Cosav) more indicative of a correct solution, *i.e.* a low value for $R(\varphi)$ and a high value for Cosav, the difference between $R(\varphi)$ and Cosav for solutions *versus* non-solutions are larger. The lowest values of $R(\varphi)$ for non-solutions for the single-site derivatives are about 0.3 units higher than solutions for protocol *A* (see rows 2 and 3 of columns 3 and 4) as compared for smaller differences of 0.05 and 0.10 in protocol *B* as is shown in columns 6 and 7 of Table 3. Thus, the discrimination between solutions and non-solutions is considerably better for protocol *A* as compared to *B*. Also the propensity for attractive looking false solutions may be reduced. The truncated data set of the $PtCl_4^{2-}$ derivative produced a 'split peak' false solution in about 5% of the random trials (row 2 of column 7). The two false peaks in question had the same x and z coordinate as the true site, but had y coordinates

which differed by ±0.025 (*i.e.* ±2.0 Å) from the correct *y* value. No such false solutions were found in trials using $\Delta E's$ derived from the full data set. It was also noted that the average $|\Delta E|^2$ from the full set of $Pt(CN)_4^{2-}$ data was abnormally low, that being 0.40. Separate minimal-function trials were run on data for which this average was rescaled to equal 1.0 (row four of Table 3). The effect was to worsen the convergence toward solutions, as indicated by the reduced solution rate (4% in row 3 as compared to 1% in row 4), and $R(\varphi)$ values increasing from 0.344 to 0.486. The difference between $R(\varphi)$ and Cosav for distinguishing solutions from non-solutions is also reduced.

The data in Table 4 indicate how the relative peak positions for the four-site $Pt(CN)_4^{2-}$ derivative were obtained from the triples translation function, and how the origin and enantiomorph were resolved by the minimal-function calculation. The minimal-function calculations for these SIR data sets were quite fast when compared to small-molecule *ab initio* calculations. It typically took only ten refinement cycles to produce solutions, such that 100 random trials could easily be performed in 30 min on an SGI Indigo workstation. The triples translation-function calculations required much less time, typically less than 10 s in total. We also note that unlike the minimal-function calculations, the triples translation-function results are insensitive to the relative scaling of the *E* values, and the magnitudes of the peaks of the map will maintain their same relative proportions, regardless of the scale. It also follows that if the top five or ten peaks from the translation-function map are recycled as one-atom fragments in the minimal function, one need not examine 100's of random trials in order to identify a solution. As a word of caution, it may be wise to verify these results by re-inspecting the difference Patterson (Argos & Rossmann, 1976; Terwilliger, Kim & Eisenberg, 1987) or comparing them to other solution methods such as those recently described using genetic algorithms (Chang & Lewis, 1994).

The trials involving SIR data from the $P4_12_12$ structure of β-HSDH help illustrate another advantage that these methods may have over difference-density Patterson analyses to determine the heavy-atom positions. Patterson maps are identical for chiral paired space groups, but the relative positions of multiple heavy-atom sites may be obtained for either chosen space group by vector verification methods. Molecular-replacement translational searches likewise must test both chiral space groups if the search fragment is in general non-centrosymmetric. Single-atom triples translation searches and minimal-function anlayses are different in this regard. Both operate on the space-group specific symmetry information which has been encoded into the algebraic representation of the triples phase-invariant relationships. It does not matter which of the two chiral groups one selects, the solution vectors one obtains will be consistent with that choice. Although the positions (u, v, w) and (v, u, w) will be identical in the $P4/mmm$ symmetry Patterson of the β-HSDH structure, the positions (x, y, z) and (y, x, z) in the one-atom triples translation map will not be equivalent in peak height. If one were to select the opposite space group, the map intensities at thse two locations woud be interchanged. Similarly for the minimal function, the values of $R(\varphi)$ are different for sets of atoms transformed fom (x, y, z) to (y, x, z) in this application, but would be interchanged if the opposite space group had been selected.

Closing note. Jaskolski has recently determined the replacement atom positions for a $P4_32_12$ SIR data set of an 160-residue protein containing four S/Se methionyl amino acids using the techniques described. These positions could not be unambiguously determined using difference Patterson methods. The correctness of the four sites was later reconfirmed by a subsequent MAD phasing experiment in which the structure was determined.

Research support from NIH grants GM-32812 and GM-46733 are gratefully acknowledged. DAL would like to thank Drs Debashis Ghosh and Mariusz Jaskolski for providing SIR and SAS data sets in space groups $P4_12_12$, $P3_121$, $P6_2$ and $P6_322$ upon which to test these procedures. Copies of the programs are available from this author upon request.

References

ARGOS, P. & ROSSMANN, M. G. (1976). *Acta Cryst.* B**22**, 2975–2979.
BLESSING, R. H. & LANGS, D. A. (1988). *Acta Cryst.* A**44**, 729–735.
CHANG, G. & LEWIS, M. (1994). *Acta Cryst.* D**50**, 667–674.
DUAX, W. L. & GHOSH, D. (1995). Unpublished data.
GUO, D. Y. & HAUPTMAN, H. A. (1990). 40th Amer. Crystallogr. Assoc. Meet., New Orleans, LA. Abstract T02.
HAUPTMAN, H. A. (1988). Amer. Crystallogr. Assoc. Meet. Philadelphia, PA. Abstract R4.
HAUPTMAN, H. A. (1991). In *Crystallographic Computing*, Vol. 5, *From Chemistry to Biology*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 324–332. New York: Oxford Univ. Press.
LANGS, D. A. (1988). *Science*, **241**, 188–191.
LANGS, D. A. (1992). *Acta Cryst.* A**48**, 172–174.
MILLER, R., DETITTA, G. T., JONES, R., LANGS, D. A., WEEKS, C. M. & HAUPTMAN, H. A. (1993). *Science*, **259**, 1430–1433.
PLETNEV, V. Z., GALITSKII, N. M., IVANOV. V. T., SMITH, G. D., WEEKS, C. M. & DUAX, W. L. (1980). *Biopolymers*, **19**, 1517–1534.
ROLLET, J. S. (1965). *Computing Methods in Crystallography*, pp. 35–37. Oxford: Pergamon Press.
TIMKOVICH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **231**, 4033–4046.
TERWILLIGER, T. C., KIM, S.-H. & EISENBERG, D. (1987). *Acta Cryst.* A**43**, 1–5.
WEEKS, C. M., DETITTA, G. T., HAUPTMAN, H. A., THUMAN, P. & MILLER, R. (1994). *Acta Cryst.* A**50**, 210–220.
WEEKS, C. M., DUAX, W. L. & WOLF, M. E. (1976). *Acta Cryst.* B**32**, 261–263.
YAO, JIA-XING (1981). *Acta Cryst.* A**37**, 642–644.