Acta Cryst. (1994). A**50**, 210–220

Structure Solution by Minimal-Function Phase Refinement and Fourier Filtering. II. Implementation and Applications

BY CHARLES M. WEEKS, GEORGE T. DETITTA AND HERBERT A. HAUPTMAN

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, *USA*

AND PAMELA THUMAN* AND RUSS MILLER

Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, *USA*

(*Received* 23 *March* 1993; *accepted* 24 *August* 1993)

Abstract

The minimal function, $R(\varphi)$, has been used to provide the basis for a new computer-intensive direct-methods procedure that shows potential for providing fully automatic routine solutions for structures in the 200–400 atom range. This procedure, which has been called *shake-and-bake*, is an iterative process in which real-space filtering is alternated with phase refinement using a technique that reduces the

* Present address: Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

value of $R(\varphi)$. It has been successfully tested using experimental data for a dozen known structures ranging in size from 25 to 317 atoms and crystallizing in a variety of space groups. The details of this procedure, the parameters used and the results of these applications are described.

Introduction

Existing crystallographic methods permit the routine solution of most structures containing fewer than 100 independent non-H atoms, but successful applications to significantly larger molecules are difficult and usually require considerable expertise and painstaking effort. The continual development of ever more powerful computers has motivated the search for a direct-phasing method that is automatically and routinely applicable even to structures containing a few hundred atoms. This search has caused the following question to be asked: with a set of randomly positioned atoms, is it possible to devise a refinement strategy that will produce a correct structure? The strategy proposed here, termed *shake-and-bake*, is an iterative procedure that alternates phase refinement with real-space filtering techniques. The basis for phase refinement is provided by the minimal principle (Hauptman, 1988, 1991; DeTitta, Weeks, Thuman, Miller & Hauptman, 1994).

The minimal principle effectively exploits the information inherent in the conditional probability distributions of the triplet and negative-quartet structure invariants. The triplet, or three-phase, invariants

$$T_{\mathbf{HK}} = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}} \tag{1}$$

and the quartet, or four-phase, invariants

$$Q_{\mathbf{LMN}} = \varphi_{\mathbf{L}} + \varphi_{\mathbf{M}} + \varphi_{\mathbf{N}} + \varphi_{-\mathbf{L}-\mathbf{M}-\mathbf{N}} \tag{2}$$

are generated from a specified basis set of phases $\{\varphi\}$ having the largest corresponding normalized structure-factor magnitudes ($|E_{\mathbf{H}}|$ *etc.*). The parameters $A_{\mathbf{HK}}$ associated with the triplets $T_{\mathbf{HK}}$ and the parameters $B_{\mathbf{LMN}}$ associated with the quartets $Q_{\mathbf{LMN}}$ are defined by

$$A_{\mathbf{HK}} = 2N^{-1/2}|E_{\mathbf{H}}E_{\mathbf{K}}E_{\mathbf{H}+\mathbf{K}}| \tag{3}$$

and

$$B_{\mathbf{LMN}} = 2N^{-1}|E_{\mathbf{L}}E_{\mathbf{M}}E_{\mathbf{N}}E_{\mathbf{L}+\mathbf{M}+\mathbf{N}}|[(|E_{\mathbf{L}+\mathbf{M}}|^2 + |E_{\mathbf{M}+\mathbf{N}}|^2 + |E_{\mathbf{N}+\mathbf{L}}|^2) - 2], \tag{4}$$

where N is the number of atoms, assumed here to be identical, in the unit cell. It should be noted that $B_{\mathbf{LMN}}$ takes on negative values when the cross-term normalized structure-factor magnitudes ($|E_{\mathbf{L}+\mathbf{M}}|$, $|E_{\mathbf{M}+\mathbf{N}}|$, $|E_{\mathbf{N}+\mathbf{L}}|$) are all relatively small, and that the cross-term phases need not be in the set $\{\varphi\}$. It is also convenient to define

$$t_{\mathbf{HK}} = I_1(A_{\mathbf{HK}})/I_0(A_{\mathbf{HK}}), \quad t_{\mathbf{LMN}} = I_1(B_{\mathbf{LMN}})/I_0(B_{\mathbf{LMN}}) \tag{5}$$

and

$$t'_{\mathbf{HK}} = I_2(A_{\mathbf{HK}})/I_0(A_{\mathbf{HK}}), \quad t'_{\mathbf{LMN}} = I_2(B_{\mathbf{LMN}})/I_0(B_{\mathbf{LMN}}), \tag{6}$$

where I_0, I_1 and I_2 are the modified Bessel functions. The *minimal function* can then be expressed as

$$R(\varphi) = \left[\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}}(\cos T_{\mathbf{HK}} - t_{\mathbf{HK}})^2 + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}|(\cos Q_{\mathbf{LMN}} - t_{\mathbf{LMN}})^2\right] \times \left(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}|\right)^{-1}, \tag{7}$$

where the conditional expected value of $\cos(T_{\mathbf{HK}})$ given $A_{\mathbf{HK}}$ is the Bessel-function ratio denoted by $t_{\mathbf{HK}}$ and $t_{\mathbf{LMN}}$ is the expected value of $\cos(Q_{\mathbf{LMN}})$ given $B_{\mathbf{LMN}}$. Then, with a sufficiently large number of phases constrained to values consistent with an atomic structure, $R(\varphi)$ has a global minimum at the point where all phases are equal to their true values for some choice of origin and enantiomorph. Phase refinement can be achieved by altering the values of the phases in such a way that the value of $R(\varphi)$ is reduced.

When the phases are equal to their true values, the value of $R(\varphi)$ is given by

$$R_T = \tfrac{1}{2} + \left(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}|\right)^{-1} \times \left[\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}}(\tfrac{1}{2}t'_{\mathbf{HK}} - t^2_{\mathbf{HK}}) + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}|(\tfrac{1}{2}t'_{\mathbf{LMN}} - t^2_{\mathbf{LMN}})\right] < \tfrac{1}{2}, \tag{8}$$

independent of the choice of origin and enantiomorph. In contrast, the value of the minimal function when the values of the phases are chosen at random is

$$R_R = \tfrac{1}{2} + \left(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}|\right)^{-1} \times \left\{\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}}t^2_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}|t^2_{\mathbf{LMN}}\right\} > \tfrac{1}{2}. \tag{9}$$

With (8) and (9), R_T and R_R can both be calculated *ab initio* without prior knowledge of the phases and R_T can provide an indication of the expected $R(\varphi)$ values for a solution.

Equations (7)–(9) are strictly correct only for $P1$ and the few other space groups having no centrosymmetric phases. In $P\bar{1}$ and other centrosymmetric space groups, the conditional expected values of $\cos(T_{\mathbf{HK}})$ and $\cos(Q_{\mathbf{LMN}})$ are given by $\tanh(A_{\mathbf{HK}}/2)$ and $\tanh(B_{\mathbf{LMN}}/2)$. The centrosymmetric equivalents of (7)–(9) are

$$R(\varphi) = \left\{ \sum_{\mathbf{H,K}} A_{\mathbf{HK}} [\cos T_{\mathbf{HK}} \quad \tanh(A_{\mathbf{HK}}/2)]^2 + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| [\cos Q_{\mathbf{LMN}} - \tanh(B_{\mathbf{LMN}}/2)]^2 \right\} \times \left[\sum_{\mathbf{H,K}} A_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \right]^{-1}, \quad (10)$$

$$R_T = 1 - \left(\sum_{\mathbf{H,K}} A_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \right)^{-1} \times \left[\sum_{\mathbf{H,K}} A_{\mathbf{HK}} \tanh^2(A_{\mathbf{HK}}/2) + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \tanh^2(B_{\mathbf{LMN}}/2) \right] < 1 \quad (11)$$

and

$$R_R = 1 + \left(\sum_{\mathbf{H,K}} A_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \right)^{-1} \times \left[\sum_{\mathbf{H,K}} A_{\mathbf{HK}} \tanh^2(A_{\mathbf{HK}}/2) + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \tanh^2(B_{\mathbf{LMN}}/2) \right] > 1. \quad (12)$$

In non-centrosymmetric space groups, invariants having restricted values should be treated as in the centrosymmetric case.

Phase-determination procedure

The six-part shake-and-bake structure-determination procedure, shown by a flow diagram in Fig. 1, combines minimal-function phase refinement and real-space filtering. It is an iterative process that is repeated until a solution is achieved or a designated number of cycles have been performed (DeTitta, Hauptman, Miller, Pagels, Sabin, Thuman & Weeks, 1991; Weeks, De Titta, Miller & Hauptman, 1993). For the present purposes, 'solution' does not necessarily imply a complete structure but rather a set of atomic positions that can be readily refined and extended by standard least-squares and Fourier techniques. With reference to Fig. 1, the major steps of the algorithm are described next and typical values of the various parameters used in this procedure are given and summarized in Table 1.

Table 1. *Shake-and-bake variables with typical values*

Independent non-H atoms	N'		
Invariant generation:			
Phases per atom	10	10	10
Triplets per atom	100	or 20	or 100
Negative quartets per atom	0	100	500
Initial phasing model	1, 2 or 4 atoms		
Parameter shift			
Step size	16		
Number of steps	±5		
Real space:			
Grid size	0.33 Å		
Peaks selected	$\sim N'$		
Number of cycles	N'	$1.5N'$	

A. Generate invariants

Normalized structure-factor magnitudes ($|E|$'s) are generated by standard scaling methods such as a Wilson plot, and the triplet and negative-quartet (those with $B_{\mathbf{LMN}} < 0$) invariants that involve the largest corresponding $|E|$'s are generated. Parameter choices that must be made at this stage include the numbers of phases, triplets and negative quartets to be used. Successful applications have been made using triplets alone as well as combinations of triplet

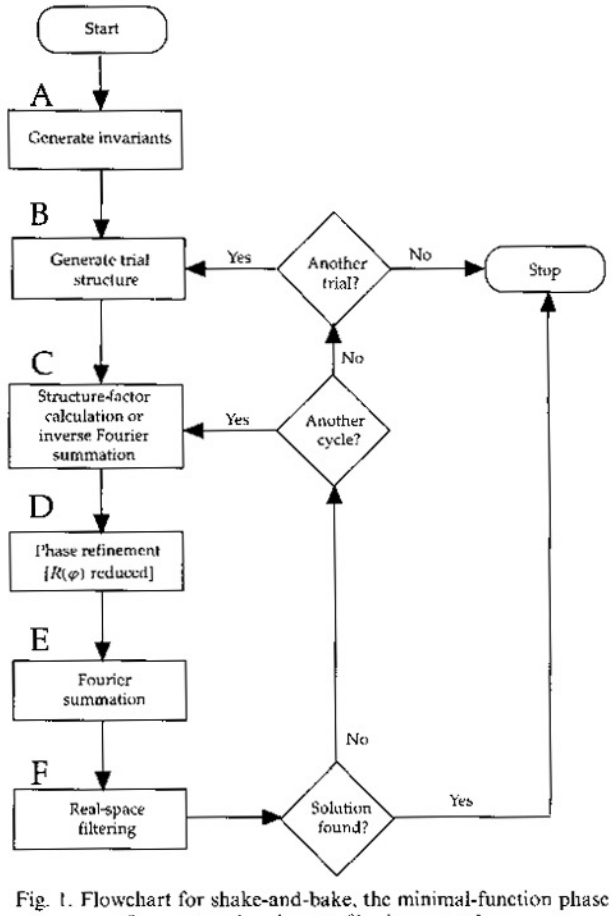

Fig. 1. Flowchart for shake-and-bake, the minimal-function phase refinement and real-space filtering procedure.

and quartet invariants. The total number of invariants is chosen to be at least 100 times the number of atoms. Further, when both triplets and quartets are used, the numbers of the two types of invariants are normally chosen so that $\sum A \simeq \sum |B|$. The ratios of the numbers of phases, triplets and negative quartets to the number of non-H atoms in the asymmetric unit give what is called the 'phase-invariant ratio'. For example, if the phase determination for a structure with 50 independent atoms involved 500 phases, 5000 triplets, and 25000 negative quartets, this ratio would be 10:100:500.

B. Generate trial structure

A trial structure or model is generated that is comprised of a number of randomly positioned atoms and their symmetry-related mates sufficient to specify the origin and enantiomorph for the space group in question. The starting coordinate sets are subject to the restrictions that no two atoms are closer than a specified distance (normally 1.2 Å) and that no atom is within bonding distance of more than four other atoms.

C. Structure-factor calculation

A normalized structure-factor calculation based on the trial coordinates is used to compute initial values for all the desired phases simultaneously. In subsequent cycles, peaks selected from the most recent Fourier series are used as atoms to generate new phase values. In the applications reported here, all non-H atoms were considered to be equal unless stated otherwise.

D. Phase refinement

The values of the phases are perturbed by a *parameter-shift* method in which $R(\varphi)$, which measures the mean-square difference between estimated and calculated structure invariants, is reduced in value. $R(\varphi)$ is initially computed on the basis of the set of phase values obtained from the structure-factor calculation in step *C*. The phase set is ordered in decreasing magnitude of the associated $|E|$'s. The value of the first phase is incremented by a preset amount and $R(\varphi)$ is recalculated. If the new calculated value of $R(\varphi)$ is lower than the previous one, the value of the first phase is incremented again by the preset amount. This is continued until $R(\varphi)$ no longer decreases or until a predetermined number of increments has been applied to the first phase. A completely analogous course is taken if, on the intial incrementation, $R(\varphi)$ increases, except that the value of the first phase is decremented until $R(\varphi)$ no longer decreases or until the predetermined number of decrements has been applied. The remaining phase values are varied in sequence as just described. Note that, when the ith phase value is varied, the new values determined for the previous $i-1$ phases are used immediately in the calculation of $R(\varphi)$. Although this process, when convergent, yields the constrained global minimum of $R(\varphi)$, the procedure described retains a measure of sequential character. The step size and number of steps are variables whose values must be chosen. In some shake-and-bake applications, an alternative method has been used to vary the phase values (Miller, DeTitta, Jones, Langs, Weeks & Hauptman, 1993). In centrosymmetric space groups, each phase takes on the values 0 and 180°, and the value yielding the smaller $R(\varphi)$ is chosen.

E. Fourier summation

Fourier summation is used to transform phase information into an electron-density map. Normalized structure-factor amplitudes, $|E|$'s, have been used at this stage (rather than F's) because phases are available for the largest E's but not for all the largest F's. The grid size must be specified.

F. Real-space filtering

Image enhancement has been accomplished by a discrete electron-density modification consisting of the selection of a specified number of the largest peaks on the Fourier map for use in the next structure-factor calculation. The simple choice, in each cycle, of a number of the largest peaks corresponding to the number of expected atoms has given satisfactory results. No minimum-interpeak-distance criterion was applied at this stage.

Results and discussion

The shake-and-bake procedure has been tested succesfully using the experimentally measured atomic-resolution intensities for the known stuctures listed in Table 2. These structures range in size from 25 to 317 independent non-H atoms in the asymmetric unit and crystallize in seven different space groups. Two structures contain moderately heavy P or Cl atoms. Some of these structures (*e.g.* 9α-methoxycortisol) were easily solved by conventional direct methods, while at least one of them (gramicidin A) required years of painstaking non-routine effort. Several presented some challenge and three (prostaglandin E_2, AZET and APAPA) were included in a suite of difficult structures supplied by the crystallographic group at the University of York, England.

Solutions are trial structures having a close match between peak positions and the known true atomic positions for some choice of origin and enantio-

Table 2. *Test data sets re-solved using the minimal function*

Structure	Atoms	Formula	Space group	Reference
Prostaglandin E_2	25	$C_{20}H_{32}O_5$	$P1$	Edmonds & Duax (1974)
Prostaglandin $F_1\beta$	25	$C_{20}H_{36}O_5$	$C2$	G. T. DeTitta (unpublished)
Alsosterone	27	$C_{21}H_{28}O_5.H_2O$	$P2_1$	Duax & Hauptman (1972)
9α-Methoxycortisol	28	$C_{22}H_{32}O_6$	$P2_12_12_1$	Weeks, Duax & Wolff (1976)
AZET	48	$(C_{21}H_{16}ClNO)_2$	$Pca2_1$	Colens, Declercq, Germain, Putzeys & Van Meerssche (1974)
Tetrahymanol	63	$(C_{30}H_{52}O)_2.H_2O$	$P2_1$	Langs, Duax, Carrell, Berman & Caspi (1977)
APAPA	69	$C_{30}H_{37}N_{15}O_{16}P_2.6H_2O$	$P4_12_12$	Suck, Manor & Saenger (1976)
Antibiotic A204A	71	$C_{49}H_{84}O_{17}.H_2O.C_3H_6O$	$C2$	Smith, Strong & Duax (1978)
Isoleucinomycin	84	$C_{60}H_{102}N_6O_{18}$	$P2_12_12_1$	Pletnev, Galitskii, Smith, Weeks & Duax (1980)
meso-Valinomycin	84	$C_{60}H_{102}N_6O_{18}$	$P\bar{1}$	D. A. Langs (unpublished)
Non-peptidic enkephalin analog	96	$(C_{24}H_{30}N_2O_6)_2$	$P1$	D. A. Langs (unpublished)
Hexaisoleucinomycin	127	$C_{81}H_{110}N_8O_{24}.14H_2O$	$P2_12_12_1$	Pletnov, Ivanov, Langs, Strong & Duax (1992)
Gramicidin A	317	$(C_{99}H_{140}N_{20}O_{17})_2.15C_2H_5OH$	$P2_12_12_1$	Langs (1988)

morph; non-solutions do not have a significant correlation between peaks and atomic positions. In order to evaluate the success or failure of a particular trial during initial experimentation, it was often useful to consider the mean phase error or average absolute value of the deviations of the phases from their values calculated using the final refined coordinates and thermal parameters. Solutions typically have mean phase errors of 30° or less. In space groups $P2_12_12_1$, the mean phase error reported is actually the minimum such error for the 16 possible positions of the structure corresponding to the eight choices of origin and two choices of enantiomorph. In all space groups, including those that have an infinite number of origin positions, solution can be conveniently identified by comparison of the calculated structure invariants for a trial structure to the actual values for the refined structure.

Unless clearly stated otherwise, the results reported here are for the 28-atom $P2_12_12_1$ test structure 9α-methoxycortisol. These results are for 30 shake-and-bake cycles using the phases for the 280 reflections having the largest $|E|$, single randomly positioned atoms to give initial phases, a maximum of five positive or five negative parameter-shift steps of 16° each, and Fourier summations of 0.33 Å resolution from which the largest 28 peaks were selected and used as trial structures in the structure-factor calculations for the subsequent cycle (see Fig. 1 and Table 1). The behavior of the 9α-methoxycortisol data appears to be typical. Test calculations based on other data sets have, so far, confirmed the conclusions and choices of parameter values obtained from the analysis of this small steroidal data set.

Fig. 2 contrasts the behaviors of the minimal function for trial structures that become solutions and for those that do not. Curve III corresponds to a non-solution that consistently has one of the lowest values of $R(\varphi)$ and curve IV is typical of a non-solution with one of the highest $R(\varphi)$ values. In this example, the 280 reflections having the largest $|E|$ values for 9α-methoxycortisol were used to generate the 560 triplets and 2800 negative quartets with the largest A and $B|$ values, respectively (*i.e.* phase-invariant ratio 10:20:100). Under these conditions, a 'solution' corresponds to a mean phase error of approximately 10°. Initial $R(\varphi)$ values were in the range 0.6–0.7 and the values for trials not corresponding to solutions tended to decrease slowly over the first 15 cycles until they were in the range 0.42–0.6. Final $R(\varphi)$ values for solutions were in the range 0.32 0.34 and clearly distinguishable from the values for non-solutions. As shown by the $R(\varphi)$ and mean-phase-error curves for a typical solution, both $R(\varphi)$ and phase-error values fall quickly just before a solution is achieved. Successful trials cannot be distinguished from the unsuccessful on the basis of $R(\varphi)$ until the mean phase is in the range 30–40°, just a few cycles before a successful conclusion. Even partial solutions, with approximately half the atoms in the correct positions, that eventually converge to solutions have $R(\varphi)$ values in the same range as those for non-solutions. However, final $R(\varphi)$ values clearly identify solutions.

The probability of success depends strongly on the initial phase error, as illustrated in Fig. 3. The initial mean phase errors for 10000 randomly positioned

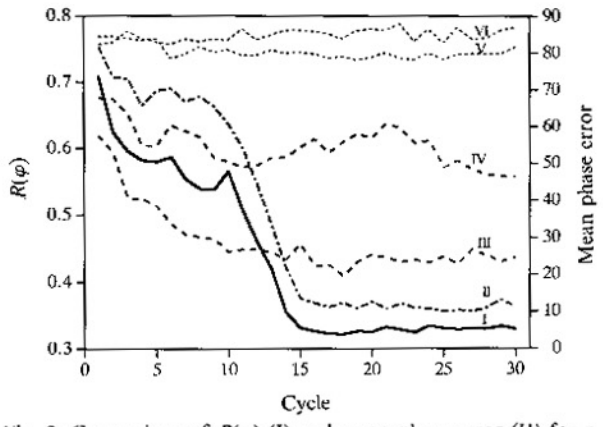

Fig. 2. Comparison of $R(\varphi)$ (I) and mean phase error (II) for a solution trial with $R(\varphi)$ (III, IV) and mean phase error (V, VI) for two non-solution trials.

one-atom 9α-methoxycortisol trial structures were computed, and the number of successful trials in 15 groups of 100 trials having approximately equal initial mean errors was determined after 30 cycles. The success rate is over 80% when the starting error is 65° or less and drops to an approximately constant level of 5% when the error increases to 80° or more. In addition, it has been observed for this structure that, when the mean phase error is reduced to 55° or less during refinement, convergence to a solution is guaranteed.

After 30 shake-and-bake cycles, an overall success rate of about 13% was observed for 500 9α-methoxycortisol trials that were randomly generated. The data presented in Table 3 show that, as the number of atoms included in the initial model is reduced, so is the minimum value of the initial mean phase error that can be obtained for any trial. In contrast, the average initial mean phase error for 10000 trials does not vary much regardless of whether 1, 2 or 28 atoms are included in the initial model. Thus, the success rate does not change significantly regardless of the initial model size even though the few one-atom trials with lowest initial error are virtually certain to succeed. These few best trials may, however, be of critical importance in the case of larger structures that have very few successful trials. Both the minimum and average initial mean phase errors are larger for random phase trials than for any of the atomic models, and this is reflected in the slightly reduced overall success rate of 10%. These observations indicate that suitable random starting phases may be those obtained from a model with the minimum number of atoms required to specify the origin and enantiomorph for the relevant space group.

The relationship between success rate, a single-atom starting point, initial phase error and initial

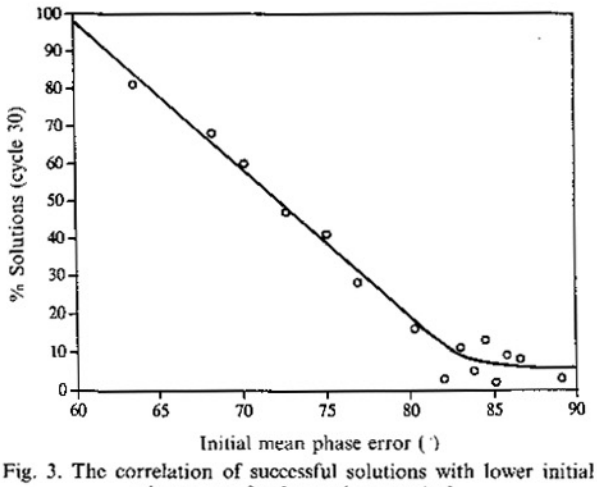

Fig. 3. The correlation of successful solutions with lower initial phase error for 9α-methoxycortisol.

Table 3. *Initial atomic model and success rate for 9α-methoxycortisol*

Initial model: number of atoms	Initial mean phase error (°) (10000 trials) Minimum	Average	Solutions (%) (500 trials, 30 cycles)
1	54.9	82.6	13.2
2	62.0	82.7	12.6
28	70.0	83.1	13.0
0 (random phases)	75.6	87.1	10.2

Table 4. *Correlation of success rate after 30 cycles with initial parameters for 9α-methoxycortisol one-atom starting points 0.25 Å apart*

	Number of trials	Number of solutions	Solutions (%)
	2880	233	8.1
(*a*) Range of distances of initial atom from nearest atomic center (Å)			
0.00–0.25	139	63	45.3
0.25–0.50	766	108	14.1
0.50–0.75	1010	45	4.5
0.75–1.00	624	13	2.1
> 1.00	341	4	1.2
(*b*) Range of initial mean phase error (°)			
54.0–80.5	576	149	25.9
80.5–83.1	576	23	4.0
83.1–84.5	576	27	4.7
84.5–85.8	576	21	3.7
85.8–90.3	576	13	2.3
(*c*) Range of initial $R(\varphi)$			
0.37–0.51	576	55	9.6
0.51–0.54	576	49	8.5
0.54–0.58	576	49	8.5
0.58–0.77	576	44	7.6
0.77–0.96	576	36	6.3

$R(\varphi)$ is further clarified by the data presented in Table 4. Starting atoms were generated by placing a 0.25 Å grid in the asymmetric part of the 9α-methoxycortisol cell (*i.e.* $0 \le x \le 0.25$, $0 \le y \le 0.25$, $0 \le z \le 0.25$ after consideration of all origin and enantiomorph positions). The coordinates of the points so generated were used as starting atoms in 2880 trials, and the phases so obtained were refined for 30 cycles using triplets alone (phase-invariant ratio 10:100:0). The success rate was found to be highly correlated not only with the initial mean phase error but also with the distance of a grid point to the closest atomic position for any choice of origin and enantiomorph. In contrast, the correlation of successful solutions with initial $R(\varphi)$ was relatively weak, in agreement with the results presented in Fig. 2. As would be expected, there is also a correlation between distance from the closest atomic position and initial mean phase error, which is 72° for the 139 points closer than 0.25 Å and greater than 84° for points further than 0.5 Å. The average initial $R(\varphi)$ value is 0.54 for the 139 closest points, and this quantity increases steadily and is 0.69 for the 341

points further than 1.0 Å from the closest atomic position.

The data presented in Tables 5–7 illustrate how variation in a number of parameters affects the success rate for 9α-methoxycortisol. Table 5 shows the results of changing the phase/atom ratio. In this experiment, all possible triplets that could be generated for each set of phases were used, but no negative quartets were included. The success rate appears to reach a maximum at a phase/atom ratio of 9–10 and a ratio in this range appears to be optimum; however, solutions are stable and can be easily distinguished on the basis of $R(\varphi)$ even when this ratio is as low as 3.6. Perhaps the main limiting factor as the number of phases decreases is series termination in the Fourier. This is indicated by the decreasing number of correct atomic positions on solution maps as the number of phases decreases.

Table 6 shows the effects of changing the parameter-shift step size and number of steps. The phase-invariant ratio used for these tests was 10:20:100. These data indicate that it is best to permit a maximum shift of 180° per phase per cycle. Use of smaller step sizes and performance of more evaluations of $R(\varphi)$ actually decreases the success rate while increasing the required computer time. In fact, of the conditions tested, the combination of two 90° steps gives the best results.

Table 7 examines changes in the invariant set used for phase refinement, a decrease in Fourier resolution, and the number of peaks selected as atoms. It also addresses the question of how many cycles should be performed. Reduction of the resolution of the E maps from 0.33 to 0.66 Å significantly reduces the success rate, but use of fewer of the largest peaks as atoms in the structure-factor calculations for the first few cycles does not make a significant difference. It is not cost effective to do more than 30 cycles because the number of additional solutions is less than the number that would be obtained by starting with a comparable set of fresh trials. Tests using other data sets have shown that, if the number of atoms to be found is less than 100, it is appropriate to do about as many cycles as there are atoms. However, until there is more experience with this method, it might be prudent to increase the number of cycles somewhat for larger structures.

Success rates as measured by the number and percentage of random trial structures that converge to solution are presented in Table 8 for a number of the test structures. The corresponding R_T, R_R and $R(\varphi)$ values are given in Table 9. The number of randomly positioned atoms used to compute initial phases was one for the $P2_12_12_1$, $Pca2_1$, $P4_12_12$ and $P\bar{1}$ structures, two for the monoclinic ($C2$ and $P2_1$) structures, and four for the $P1$ structures. In all cases, a maximum of five parameter-shift steps of 16°

Table 5. *Effects of different numbers of phases on success rate after* 30 *cycles for* 500 9α-*methoxycortisol trials with single randomly positioned starting atoms*

Phases, phases per atom	Triplets per atom	Number of solutions	Solutions (%)	Minimum atoms per solution	Maximum $R(\varphi)$ solution	Minimum $R(\varphi)$ non-solution
100, 3.6	9	16	3.2	17	0.17	0.20
150, 5.4	28	27	5.4	19	0.24	0.29
200, 7.1	65	35	7.0	22	0.29	0.35
250, 8.9	124	39	7.8	24	0.31	0.40
300, 10.7	211	38	7.6	26	0.33	0.44

Table 6. *Effects of different parameter-shift variables on success rate after* 30 *cycles for* 500 9α-*methoxycortisol trials with single randomly positioned starting atoms*

Step size (°)	Number of steps	Maximum shift per cycle (°)	Solutions (%)
±4	5	±20	5
±4	10	±40	8
±4	20	±80	11
±4	40	±160	13
±8	5	±40	8
±8	10	±80	11
±8	20	±160	12
±16	3	±48	9
±16	5	±80	13
±16	10	±160	15
±32	3	±96	15
±32	5	±160	18
±45	4	±180	17
±60	3	±180	21
±90	2	±180	26
±180	1	±180	13

Table 7. *Effects of various parameters on success rate for* 500 9α-*methoxycortisol trials with single randomly positioned starting atoms*

Triplets per atom	20	50	100	100	20	20
Quartets per atom	100	250	500	0	100	100
Fourier resolution (Å)	0.33	0.33	0.33	0.33	0.66	0.33
Number of peaks selected	28	28	28	28	28	Variable*
Number of solutions:						
Cycles 1–10	21	22	21	11	11	24
Cycles 11–20	25	19	21	15	13	26
Cycles 21–30	20	12	11	13	8	17
Cycles 31–40	13	18	11	10	5	12
Cycles 41–50	12	12	3	6	3	6
Cycles 51–60	5	7	4	5	2	6
Cycles 61–70	2	3	5	6	0	3
Cycles 71–80	10	4	6	6	0	3
Cycles 81–90	3	4	1	1	0	3
Cycles 91–100	2	2	2	1	0	3
Success rate (cycles 1–100) (%)	23	21	17	15	8	21

* The numbers of peaks selected in successive cycles beginning with cycle 1 were 1, 2, 3, 5, 8, 13, 21, 28, 28, ...

were performed in each cycle, the Fourier resolution was ≃0.33 Å, and the number of peaks selected for the structure-factor calculations was equal to the expected number of atoms. $R(\varphi)$ values for *meso-*

Table 8. *Summary of success rates for the test structures*

Structure	Trials	Cycles	Solutions (%) Triplets only	Solutions (%) Triplets and quartets	Phase-invariant ratios Triplets only	Phase-invariant ratios Triplets and quartets
Prostaglandin E_2	128	30	20	32	10:86:0	10:20:100
		40	29	40		
		100	44	66		
Prostaglandin $F_1\beta$	128	30	8	5	10:100:0	10:100:20*
				11		10:60:60*
				13		10:20:100
Aldosterone	128	30	14	12	10:100:0	10:20:100
		40	16	12		
9α-Methoxycortisol	500	30	8	13	10:100:0	10:20:100
		40	10	16		
AZET (equal atoms)	128	10	0	2	10:100:0	10:100:500
		30	7	9		
		50	13	15		
		70	14	22		
AZET (as Cl_2C_{46})	128	10	5	5	10:100:0	10:100:500
		30	20	18		
		50	21	22		
		70	23	24		
Tetrahymanol	128	70	3	7	10:100:0	10:100:500
		100	4	9		
APAPA (equal atoms)	128	70	0	0	10:100:0	10:100:500
		100	0.8	0		
APAPA (as P_2C_{67})	128	70	1.5	0	10:100:0	10:100:500
		100	1.5	0		
Antibiotic A204A	128	70	4	2	10:100:0	10:100:500
		100	4	3		
Isoleucinomycin	500	100	2	3	7.1:71:0	7.1:71:536
	2048	150	4	4	10:100:0	10:100:550
meso-Valinomycin	1024	150	0.9	1.5	10:100:0	10:100:400
Non-peptidic enkephalin analog	1024	150	21	32	10:100:0	10:100:20*
	2048	150		62		10:100:350
Hexaisoleucinomycin	2000	100		0.2		7.1:71:709

* $\sum A$ was not approximately equal to $\sum\|B\|$.

Table 9. *Summary of R_T, R_R and $R(\varphi)$ for test-structure solutions and non-solutions*

Structure	Triplets only R_T	Triplets only R_R	Triplets only R(solution)	Triplets only R(non-solution)	Triplets and quartets R_T	Triplets and quartets R_R	Triplets and quartets R(solution)	Triplets and quartets R(non-solution)
Prostaglandin E_2	0.11	1.11	0.18 0.22	0.23 0.32	0.32	0.82	0.25–0.30	0.28–0.33
Prostaglandin $F_1\beta$	0.09	1.14	0.31 0.35	0.22 0.52				
10:100:20					0.11	1.11	0.33 0.48	0.29 0.58
10:60:60					0.13	1.09	0.29–0.39	0.32–0.58
10:20:100					0.19	1.00	0.32–0.36	0.36 0.54
Aldosterone	0.18	0.98	0.22 0.25	0.30 0.53	0.33	0.78	0.27 0.29	0.31–0.51
9α-Methoxycortisol	0.24	0.88	0.27 0.29	0.40–0.60	0.36	0.71	0.32–0.34	0.41 0.56
AZET (equal atoms)	0.31	0.77	0.29–0.34	0.38 0.50	0.40	0.64	0.39 0.41	0.44–0.52
AZET (as Cl_2C_{46})			0.26–0.27	0.32–0.48			0.37–0.38	0.40 0.51
Tetrahymanol	0.19	0.97	0.21	0.24 0.43	0.37	0.70	0.35	0.40–0.49
APAPA (equal atoms)	0.39	0.65	0.39	0.44–0.56	0.44	0.58		0.48 0.54
APAPA (as P_2C_{67})			0.37	0.43 0.55				0.48–0.53
Antibiotic A204A	0.30	0.79	0.28	0.30 0.47	0.40	0.64	0.38	0.42 0.50
Isoleucinomycin	0.36	0.70	0.33 0.35	0.45–0.51	0.42	0.62	0.42–0.43	0.47–0.54
meso-Valinomycin	0.61	1.39	0.53	0.60 0.94	0.79	1.21	0.75	0.85 1.00
Non-peptidic enkephalin analog	0.27	0.82	0.25 0.26	0.30–0.42	0.40	0.65	0.36–0.37	0.42–0.48
Hexaisoleucinomycin	0.36	0.70	0.43–0.45	0.49–0.56	0.40	0.64	0.45	0.47–0.54

valinomycin were calculated using the formulas appropriate for the space group $P\bar{1}$ [equations (10)–(12)]. $R(\varphi)$ values for all other structures were calculated using the $P1$ formulas [equations (7)–(9)] regardless of the identity of the actual non-centrosymmetric space group.

Table 10. *Effects of invariant set size and composition on success rate after* 100 *cycles for* 100 *isoleucinomycin trials having an initial mean phase error in the range* 65–75° *for* 600 *phases (approximately* 7 *phases per atom)*

Triplets per atom	Quartets per atom	Total number of invariants	Solutions (%)
107	714	69000	40
71	536	51000	50
57	393	37800	39
36	250	24000	32
14	107	10200	27
71	0	6000	43

The data in Table 8 show that the success rate is 10–15% for structures in the 25–50 atom range and falls to 1–5% for structures in the 60–100 atom range. In many cases, better success rates might be obtained by using different parameter-shift conditions (*e.g.* two steps of 90° each). The success rates for the *P*1 structures (prostaglandin E_2 and enkephalin analog) are unusually high. Although further experimentation is clearly needed, the explanation for this may be related to the fact that space group *P*1 has great flexibility with respect to the location of the origin. This may mean that the atoms in more trials start out sufficiently close to consistent atomic positions, which, based on the results in Table 4, might contribute to the high success rate.

When moderately heavy atoms such as chlorine are present, as they are in the AZET structure, the success rate is improved if appropriate numbers of the largest peaks on the maps are treated as heavier atoms in the structure-factor calculations. When this is done, the $R(\varphi)$ values also improve. The discrepancy between the equal- and unequal-atom treatments of the AZET data is most striking when the results for relatively few cycles are compared. It seems that the presence of heavier atoms allows several trials to converge to solution more rapidly than would be expected. In fact, the structure appears to behave as if the effective number of atoms is less.

Generally, but not always, inclusion of negative quartets improves the success rate; however, this improvement is often not cost effective in terms of computer time. In most cases, triplets and quartets were combined in such a way that $\sum A \simeq \sum |B|$, the footnote to Table 8 indicating when this was not the case. Although a phase-invariant ratio of 10:20:100 is clearly adequate for 9α-methoxycortisol, it is possible that more solutions may be obtained for larger structures by utilizing more invariants as indicated by the data for isoleucinomycin trials with low initial mean phase errors as presented in Table 10. For this reason, ratios such as 10:100:500 were used for the larger structures. In any event, the use of negative quartets is certainly not obligatory for $P2_12_12_1$ and $P2_1$ structures.

It is well known that, when conventional direct methods are used, the inclusion of negative quartets often alleviates the problem of an over-consistent solution, with loss of enantiomorph resolution or a single dominant peak, which occurs in symmorphic and polar space groups (Schenk, 1972; Sheldrick, 1990). This type of situation did not prove to be a major concern even when the shake-and-bake method was used without negative quartets. It did occur, however, in the case of prostaglandin $F_1\beta$, a 25 atom structure that crystallizes in space group *C*2 and has unusually high thermal motion for a structure of its size. As shown in Table 9, the $R(\varphi)$ values for some non-solutions (30 in all) are lower than the values for any of the 10 solutions when only triplet invariants are used. The problem is not resolved when a small number of quartets (phase-invariant ratio 10:100:20) is used because seven non-solutions still have $R(\varphi)$ values less than the values for any of the seven solutions. When equal numbers of triplets and quartets are used (ratio 10:60:60), one of the 14 solutions does have the lowest value of $R(\varphi)$, but the next four lowest values are for non-solutions. However, when more negative quartets are included to give a ratio of 10:20:100 and $\sum A \simeq \sum |B|$, the lowest $R(\varphi)$ values obtained are for 11 of the 16 solutions.

A few non-solution trials for the larger (71 atom) *C*2 test structure, antibiotic A204A, also have $R(\varphi)$ values close to the values for solutions when only triplets are used. Although the $R(\varphi)$ values are always diagnostic for the enkephalin analog solutions, there is a very significant increase in success rate when negative quartets are used for this 96 atom *P*1 structure. This increase, though dramatic, is not cost effective when the phase-invariant ratio is 10:100:350. However, the less dramatic increase obtained with a ratio of 10:100:20 is computationally efficient. These observations, as well as the findings for prostaglandin $F_1\beta$, indicate that it is probably wise to include some negative quartets when making applications in space groups lacking a screw axis.

Another problem that frequently arises with tangent-formula-based direct methods is the occurrence of translated molecules that may have few, if any, atoms missing but that will not respond to least-squares refinement. Tetrahymanol, a steroid with much internal symmetry and two molecules in the asymmetric unit, was a classic example of this problem. Of the 128 trials refined using triplets alone for 100 cycles, five were solutions with $R(\varphi)$ in the range 0.208–0.210, five were non-solution translated molecules with $R(\varphi)$ in the range 0.244–0.270 and the remaining 118 non-solutions had $R(\varphi)$ in the range 0.274–0.427. The actual solutions could be distinguished with no difficulty.