Acta Cryst. (1995). A51, 81–87

Use of the Minimal Function for Partial Structure Development in Direct Methods

BY DAVID A. LANGS, RUSS MILLER* AND HERBERT A. HAUPTMAN

Medical Foundation of Buffalo, 73 High Street, Buffalo, NY 14203, USA

AND GYE WON HAN

Department of Crystallography, University of Pittsburgh, Pittsburgh, PA 15260, USA

(*Received* 3 *February* 1994; *accepted* 15 *June* 1994)

Abstract

The shake-and-bake procedure, which is based on the minimal function, has been tested and shown to be extremely effective in molecular-fragment recycling applications. Correctly positioned fragments as small as 5% of the scattering power of the structure typically have a 50% chance of producing a solution in a single recycling trial. While starting models for tangent-formula recycling methods normally require an average r.m.s. displacement error of less than ~0.25 Å from the refined structure to ensure an adequate chance of success, the shake-and-bake method often tolerates r.m.s. model errors well in excess of 0.5 Å. Tests indicate that the new method can outperform traditional tangent-formula procedures in difficult structural applications involving multiple copies of pseudosymmetrically related molecules or low-resolution data.

Introduction

More than a decade has passed since it was reported that a multisolution random-phasing algorithm [*RANTAN*, see Yao (1981, 1983)] based on the tangent formula could substantially extend the range and sensitivity of partial-structure recycling procedures (Karle, 1968) to less than 10% of the scattering power of the unit cell. The original recycling method often required two or more cycles to fully develop the entire structure from a fragment and success depended on skill in recognizing and correctly building upon potentially good fragments in marginally phased E maps produced along the way. The *RANTAN* procedure, however, once given the initial trial fragment, could normally identify a successfully converged phase set in a small number of multisolution trials without human intervention. This paper reports on an alternative fragment-recycling strategy that is used in *SnB* (Miller, Gallo, Khalak & Weeks, 1994), a computer program based on the 'shake-and-bake' method (Weeks, DeTitta, Hauptman, Thuman & Miller, 1994), which alternates phase refinement in reciprocal space with density modification in real space in an effort to minimize a function of the structure invariants (Hauptman, 1991). This study indicates that shake-and-bake methods can produce successful results from smaller starting models having larger r.m.s. displacement errors than can normally be tolerated by the earlier tangent-based techniques.

Fragment recycling

Partial-structure tangent-formula recycling methods are based on the assumption that, although the phases computed from a partial structure may not be sufficiently accurate to complete the structure by standard Fourier techniques, a certain subset of the phases – those with the largest relative computed structure amplitudes – may have phase values sufficiently good to initiate phase extension by the tangent formula to determine the remaining phase values (Karle, 1968). Sim (1960), for example, showed that the phase-error distribution of a structure-factor amplitude computed from some fraction of the total structure could be expressed as

$$P(\varphi_{\mathbf{h}}) = \exp(X\cos\varphi)/2\pi I_0(X), \qquad (1)$$

where $X = 2|F_{\mathbf{h}}(\text{obs.})F_{\mathbf{h}}(\text{calc.})|/\Sigma_L$ and Σ_L was defined as the sum of the squares of the atomic scattering factors of the remainder of the structure not included in the calculation of $F_{\mathbf{h}}$(calc.). To be more precise, Σ_L should be multiplied by $\exp[-2B(\sin\theta/\lambda)^2]$ to normalize X with respect to the thermal fall-off of the data. Weights proportional to the average expected magnitude of the phase error of the partial structure may be obtained from the average expected cosine of $\Delta\varphi = \varphi_{\mathbf{h}}(\text{true}) - \varphi_{\mathbf{h}}(\text{calc.})$:

$$W_{\mathbf{h}} = \varepsilon[\cos(\Delta\varphi)] = I_1(X)/I_0(X), \qquad (2)$$

where I_0 and I_1 are the zeroth- and first-order modified Bessel functions of the first kind. It should be noted that $W_{\mathbf{h}}$ approaches 1 as the product $F_{\mathbf{h}}(\text{obs.})F_{\mathbf{h}}(\text{calc.})$ becomes larger and the residual scattering power Σ_L becomes smaller so that the

* Permanent address: Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, USA.

value of X increases. In practice, one rarely computes (2) to select the most reliable starting-set phases; it is usually sufficient to choose phases for which the ratio $p = F_{\mathbf{h}}(\text{calc.})/F_{\mathbf{h}}(\text{obs.})$ exceeds some threshold value, *e.g.* $p \simeq 0.25$, which will be sufficient to select about 20% of the total number of E's used in the tangent-formula refinement for the starting-basis phase set. Slightly better phase selection can be achieved if p is multiplied by $|E_{\mathbf{h}}(\text{obs.})|^2$. Generally, if more than 50% of the phases are included in the basis set, the refinement will return the initial trial fragment and not much more. If too small a percentage of phases is selected, for example less than 5%, one may not have a sufficient number of triple contributors to ensure that errors do not propagate from the basis set during tangent-formula phase extension. *RANTAN* recycling differs from the traditional tangent-formula method in that after the basis-set phases are selected and given some uniform high refinement weight, *e.g.* $W_{\mathbf{h}} \geq 0.75$, multiple refinement trials are generated by assigning to the remaining phases different sets of random values and a low uniform refinement weight, usually in the vicinity of 0.25.

The minimal function

A new phasing formula based on a minimum-variance structure-invarient residual was first reported by Hauptman in 1988.

$$R(\varphi) = \sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{hk}}[\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{\mathbf{k}-\mathbf{h}}) - I_1(A_{\mathbf{hk}})/I_0(A_{\mathbf{hk}})]^2 \Big/ \sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{hk}}, \quad (3)$$

where $A_{\mathbf{hk}} = 2|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|/N^{1/2}$ and N is the number of equivalent non-H atoms in the primitive reduced cell. Subsequent work has established the shake-and-bake method, in the form of the *SnB* program, for systematically sampling the polydimensional phase space of $R(\varphi)$ and finding solutions. A key ingredient of these studies has been that it is beneficial to initiate these random-phase multisolution methods not with random phase values but rather with phases computed from randomly chosen atomic structural models. A typical iterative cycle of *SnB* for an N-atom structure involves (i) assigning an initial set of phases corresponding to a set of randomly chosen atoms, (ii) performing one or two passes of phase optimization to reduce the value of (3) and (iii) synthesizing a Fourier map from which the positions of the N largest peaks are used to compute phases for the next cycle. The Fourier syntheses in step (iii) typically account for 80 to 90% of the total *SnB* computation time. Moreover, the interpolation of discrete atom peaks will usually fail at less than 1.2 Å resolution such that the structure-factor calculation must be replaced by a second Fourier to back transform the modified electron density representing the molecular envelope.

Practical applications have demonstrated that structures that have resisted solution by tangent-formula direct methods and other sophisticated methodologies can be determined *ab initio* by *SnB* in a fairly routine manner (Miller, DeTitta, Jones, Langs, Weeks & Hauptman, 1993), often with phases that have been initially computed from a single randomly chosen atom. In fact, the small macromolecular structures of crambin (Weeks *et al.*, 1995) and rubredoxin (Miller, 1993), each containing more than 400 non-H atoms, were routinely determined in this manner when high-resolution data sets were available.

The original impetus behind this work was to investigate whether phases associated with random atoms, which had proven to be very effective on trial sets for *SnB*, might also show a similar advantage if employed in a random phasing mode in tangent-formula calculations. We used a local version of the *RANTAN* algorithm, written in Fortran (Langs, 1985), for this purpose. Subsequently, we investigated the performance of *SnB* for obtaining phase solutions from partial structural models of various sizes and conformational fidelity.

Trial computations

Four test structures were selected upon which to perform trial calculations to compare the random-start tangent-formula and minimal-function fragment recycling techniques. These included the ionophoric antibiotic isoleucinomycin (ILED), $C_{60}H_{102}N_6O_{18}$, $P2_12_12_1$, $N = 84$ (Pletnev, Galitskii, Ivanov, Smith, Weeks & Duax, 1980), cholesteryl butanoate (CBT), $C_{31}H_{52}O_2$, $P2_1$, $N = 132$ (Han, Craven & Langs, 1994), and the orthorhombic – $P2_12_12_1$, $N = 317$ (Langs, 1988) – and monoclinic – $P2_1$, $N = 314$ (Langs, Smith, Courseille, Précigoux & Hospital, 1991) – forms of gramicidin A, $C_{198}H_{280}N_{40}O_{34}$, which crystallize with 15 and 21 molecules of ethanol and methanol, respectively. The room-temperature data of monoclinic gramicidin (MONOGRAM) diffracted to only 1.5 Å resolution, the other three data sets all diffracted to better than the 0.86 Å resolution limit of the orthorhombic gramicidin (GRAMA) structure.

In our initial investigation, a series of 8192 *RANTAN* phasing trials for the ILED structure (500 phases, 5000 triples, random phase weight = 0.25) were refined to convergence and resulted in 20 solutions (Nqest $\simeq -0.12$, Resid $\simeq 0.33$, Cosav $\simeq 0.53$). This success rate of about 0.25% was about the same whether *MULTAN* weights or Hull–Irwin weights (Hull & Irwin, 1978) were used. Changing the

random phase weight from 0.25 to either 0.20 or 0.30 reduced the number of solutions by nearly 50%, so the value of this weight appears to be optimal. Higher success rates could be achieved with various selected origin and Σ_1 phases included as a known starting set. By way of comparison, *SnB* calculations based on phases from a single randomly generated atom were previously shown to have an average solution rate of 6.4% after 50 cycles for the ILED structure, or about 25 times greater than the *ab initio* *RANTAN* trials. *RANTAN*, however, was five to ten times more cost effective than *SnB* in the average time it took to produce a solution. The r.m.s. phase errors of the initial starting sets of the 20 *RANTAN* solutions ranged from 101.1 to 107.1°, with an average value of 105.1°, where the best fit among the 16 possible origin and enantiomorph choices for the $P2_12_12_1$ cell is reported. It is interesting to note that these r.m.s. phase errors are very close to that expected for randomly selected unrestricted phase values, *i.e.* $\pi/3^{1/2}$ rad or 103.92°.

Although *RANTAN* recycling is not normally expected to succeed for fragments as small as 1/84 of the ILED structure, it was important to determine whether on occasion it might, and whether, as was demonstrated for the SnB trials, the random atom need not correspond to an actual position of an atom in the real structure. This point may not have been fairly tested in earlier investigations of the *RANTAN* procedure. Therefore, a *RANTAN* calculation of 8192 trials based on phases generated from a single randomly selected atom were performed for the ILED structure with the expectation that the success rate might conceivably be improved by a minor change in the phasing protocol. This calculation, however, produced no solutions. Neither did a second calculation involving another 8192 different sets of similarly generated phases.

Calculations were next performed for starting sets of phases that were computed from the coordinates of each of the 84 non-H atoms of the structure. Although these 84 phase sets had an initial average r.m.s. phase error appreciably lower (99.6° *versus* 105.1°) than the 20 solutions that had converged from randomly generated phases, none of these 84 trials converged to produce a structural solution. It is now more understandable why none of the previous 16000 sets produced a solution.

By way of comparison, *SnB* produced 26 solutions from the aforementioned set of 84 trials, *i.e.* a 31% success rate for a fragment size of 1.2%. The modified *RANTAN* procedure was subsequently found to require fragment sizes of the order of three correctly positioned atoms in order to achieve a similar success rate. Although this represents only 3.6% of the molecular structure, and convincingly demonstrates the power of the *RANTAN* recycling method, the probability of randomly generating such a three-atom fragment in a $P2_12_12_1$ cell to within a r.m.s. precision of ± 0.2 Å is conservatively estimated to be one in 13×10^6! The probability of correctly positioning one atom to within ± 0.2 Å by random selection is quite reasonable, however, and is estimated to be slightly less than one chance in 40, or about 2.5%.

In the second CBT example, we now consider the effectiveness of *SnB* for fragment recycling in an application involving input from a molecular-replacement search. The structure has four independent pseudosymmetrically related copies of the molecule in the asymmetric unit and was difficult to solve using both molecular replacement and various multisolution direct-methods techniques (Han, 1993). The structure was eventually solved in a traditional *RANTAN* run that included a basis set of 36 zonal restricted phases defined by seven symbols (0 or π) that were sampled by 16 permuted phase combinations using a scheme outlined by Woolfson (1954), referred to as substantialization. Each of these 16 phase sets was subjected to ten random phasing trials affecting the remaining phases before the first solution was indicated. An additional 308 random phase trials were required on each of the 16 sets before the second solution occurred using this phasing strategy.

Previous attempts to solve the CBT structure included more than 16000 *ab initio* random phasing trials, plus many unsuccessful attempts to fragment recycle various tentative molecular replacement positions provided by the semirigid 20-atom fused-ring nucleus of the cholesterol molecule. The basis set for the fragment-recycling trials consisted of ~140 fragment-defined phases; the remainder of the 720 phases used were assigned different random values for each of 50 separate trials; each trial persisted for 50 cycles of refinement before it was terminated.

Since none of the dozens of potential molecular-replacement positions tested by the random tangent-formula recycling procedure produced a solution, we reasoned that either none of these positions was sufficiently close to the true structure to effect a convergence or the fragment was too small to compensate for the unknown deleterious effects of pseudosymmetry. After the structure had been solved and refined, it was later discovered that a number of the earlier molecular-replacement trials were indeed close to a correct position in the cell. In fact, one of the replacement models that was tested had a r.m.s. fit of 0.094 Å to the refined structure and failed to produce a solution after a run of 50 separate random-phasing trials (Han, 1993). These results are quite atypical for *RANTAN* recycling. In an earlier structural application, a model that had a r.m.s. fit of 0.55 Å with the refined structure successfully converged to a solution (Langs, Blessing & Duax, 1991).

Table 1. *Comparison of tangent-formula and minimal-function success rates for CBT based on a partial structure of* 20/132 *atoms*

500 or more separate *RANTAN* recycling trials were performed using error-free coordinates (r.m.s. = 0 Å) and three additional models with random r.m.s. errors of 0.1, 0.15 and 0.2 Å. The *SnB* calculations used similarly generated r.m.s. deformed models (column 3) and, in addition, rigid molecular-replacement models having varying degrees of angular (column 5) and positional (column 7) misfit with the known structure. 100 random trials were run for each of ten ranges of r.m.s. errors between 0.2 and 1.4 Å. The average numbers of cycles in which *SnB* trials produced solutions are given in parentheses in columns 4, 6 and 8.

	Tangent formula	Minimal function		
R.m.s. error (Å)	Deformed r.m.s. model solutions (%)	Deformed r.m.s. model solutions (%) ⟨cycles⟩	Angular r.m.s. model solutions (%) ⟨cycles⟩	Positional r.m.s. model solutions (%) ⟨cycles⟩
0.0	19.8	†	†	†
0.1	7.2	†	†	†
0.15	2.2	†	†	†
0.2	0.15*	†	†	100 (3.6)
0.3		100 (4.7)	100 (4.3)	96 (7.2)
0.4		86 (11)	96 (11)	72 (12)
0.5		50 (17)	72 (16)	22 (33)
0.6		12 (13)	66 (20)	0 (100)‡
0.7		2 (52)	57 (21)	
0.8		0 (100)‡	40 (26)	
1.0			32 (26)	
1.2			6 (37)	
1.4			0 (100)‡	

* Three successful convergences in 2000 random trials were noted.

† Trials not performed – solution rate assumed to be ~100% in view of rows 4 and 5.

‡ 25 solutions were obtained in 1000 trials using random 20-atom fragments, which establishes a base level of success of 2.5%. The surprising fact that no solutions were found for these three trailing entries indicates that, for sufficiently large perturbations, the perturbed search model is actually worse than a randomly chosen model.

In most instances, r.m.s. positional errors not exceeding 0.25 to 0.35 Å are required.

Normal *RANTAN* recycling runs were re-performed on the 20-atom CBT fragment described above. First, the error-free coordinates of the least-squares refined structure were tested; then, three other starting models, to which Gaussian r.m.s. errors of 0.1, 0.15 and 0.2 Å had been added, were tested. Sets of 500 random trials were run for the 0.0, 0.1 and 0.15 Å r.m.s. fragments and 2000 trials were run for the 0.2 Å r.m.s. fragment. The percentages of successful convergent solutions are recorded in column 2 of Table 1.

SnB was then applied to similarly Gaussian-deformed fragments (column 3) and then to the undeformed 0.093 Å r.m.s. molecular-replacement search fragment, which was either orientationally (column 5) or positionally (column 7) mis-set to create starting models with overall r.m.s. displacement errors ranging from 0.2 to 1.4 Å. A set of 100 random trials was generated for each of these three kinds of models over ten r.m.s. error ranges between 0.2 and 1.4 Å. The results of these calculations are summarized in columns 4, 6 and 8 of Table 1 for comparison.

The *RANTAN* and *SnB* phasing protocols are *not* equivalent. The phases in the *SnB* trials were not partitioned into fragment-defined and randomly assigned phases. All phases were assigned by the fragment and multiple solution trials were generated by deforming the fragment in a different manner or perturbing its position and orientation to achieve the desired misfit of the model. Additional *SnB* trials could have been generated using the *RANTAN* protocol but this was not systematically explored.

The third example, GRAMA, represents a structure determination in which an idealized helical fragment of 60 atoms was positioned to span three small peptide fragments that in aggregate totalled 19 atoms. These fragments were found in an E map produced in a multisolution direct-methods run. The 19 atoms were insufficient to produce a solution by recycling methods but the fitted 60-atom fragment produced two *RANTAN* solutions in only ten trials. Here, we investigate the average minimum size for a single *contiguous* fragment by which this 317-atom structure could have been determined.

For this analysis, the solvent molecules and flexible amino acid side chains were excluded from the GRAMA trial fragments that were selected from contiguous sections of the two helical peptide backbones that totalled 80 atoms each. The largest 1500 E values were used to generate 17500 triples. Roughly 300 of the 1500 basis-set phases were assigned by the fragment using an F(calc.)/F(obs.) criterion. Initial test calculations determined that fragment sizes of between 10 and 16 atoms roughly bracket the range marking *RANTAN* success or failure. Each *RANTAN* fragment calculation consisted of 100 trials of 20 refinement cycles in which the

Table 2. *Fragment-recycling results obtained from the 0.86 Å low-temperature GRAMA data*

100 tangent-formula recycling trials were performed for each fragment; only one *SnB* run was calculated in comparison. The two peptide chains are labeled *A* and *B*. The percentage of *RANTAN* successes is noted, as is the cycle number for which *SnB* successfully converged or failed (*F*).

	RANTAN (%)		*SnB* cycle no.	
Atoms	*A*	*B*	*A*	*B*
(*a*) Ten 16-atom fragments				
1–16	0	0	10	7
17–32	0	87	4	3
33–48	100	96	4	3
49–64	0	12	5	5
65–80	82	91	4	7
(*b*) 16 ten-atom fragments				
1–10	0	0	*F*	6
11–20	0	0	5	*F*
21–30	0	11	*F*	6
31–40	12	5	6	4
41–50	0	0	4	8
51–60	0	0	11	6
61–70	0	0	5	*F*
71–80	0	0	8	*F*

fragment assigned phases were held fixed for the first five cycles. Table 2(*a*) lists results obtained using ten separate 16-atom fragments that were obtained by cutting the two helical chains into five equal segments. The columns in Table 2 denote the sequence numbers of the atoms in the *A* or *B* chains and the number of successful convergences in each set of 100 random trials. A single *SnB* calculation of ten cycles was then performed for each of the ten fragments; no fraction of the phases was held fixed at any time. A second series of calculations employing 16 smaller ten-atom fragments was performed and the results are recorded in Table 2(*b*). Successful *SnB* trials are noted by the cycle number of convergence; failures are identified by the letter *F*.

The fourth example, the 1.5 Å room-temperature MONOGRAM structure, helps illustrate some difficulties that may arise when working with low-resolution data. This structure was originally solved by molecular replacement using a 193-atom helical backbone fragment taken from the GRAMA structure. *RANTAN* refinement failed to reveal the side-chain atoms even though the molecular fragment represented more than 60% of the total structure. Moreover, many of the details of the original fragment were lost in subsequent *E* maps as tangent-formula refinement diverged from the solution. Constrained least-squares refinement of the original fragment, however, allowed the phases to improve to the point at which the side-chain positions and solvent molecules could be determined. The 193-atom fragment was later shown to have a r.m.s. fit of 0.52 Å with the refined structure.

17000 triples were generated from the largest 750 *E* values. *RANTAN* trials were computed for each of two series of calculations, the first using *MULTAN* α weights, the second using Hull–Irwin weights. Only the results from the best trial of the 100 is noted in Table 3. The optimal tangent-formula results were obtained when between 50 and 65% of the basis phases were defined by the fragment and held fixed for the first five refinement cycles. The refinements rapidly deteriorated with a worse r.m.s. fit to the final structure-refined phases if the fractions of fragment-selected phases were much less than 50%. In the *SnB* calculations, several sophisticated density-modification schemes (Hoppe & Gassmann, 1968) were tested prior to map inversion to re-estimate the phases but simple low-density elimination appeared to be the most effective (Shiono & Woolfson, 1992). Slightly better results were noted when the basis-set fraction of phases was held fixed for the first *SnB* refinement cycle. *RANTAN* results using 1500 *E* values and 28000 triples were considerably worse than the 750 phase calculations but the *SnB* results were slightly better. Table 3 compares the cycle-by-cycle r.m.s. phase errors of *RANTAN* and *SnB* for the 750 phase refinements for the MONOGRAM structure.

Table 3. *Recycling results obtained from the 193-atom GRAMA replacement model using the 1.5 Å room-temperature MONOGRAM data*

100 tangent formula recycling trials were computed for each of two series of calculations, the first using *MULTAN* α weights, the second using Hull–Irwin (HI) weights. The fragment-defined phases were held fixed (*) for five cycles; only the results from the best trial among the 100 are noted. Only two *SnB* trials were performed; slightly better results were noted when the basis-set fraction was held fixed for a single cycle (compare last two columns). The r.m.s. phase errors are reported in degrees.

No. of tangent cycles	α weights r.m.s. (°)	HI weights r.m.s. (°)	No. of *SnB* cycles	Trial 1 r.m.s. (°)	Trial 2 r.m.s. (°)
0	62.9	62.9	0	62.9	62.9
1–5*	58.0–57.0	58.0–58.6	1	59.9	58.9*
6	68.3	64.3	2	55.6	53.5
10	62.6	74.6	3	58.5	57.2
15	67.0	91.8	4	60.9	60.2

Discussion of results

Despite the wealth of experience that large numbers of laboratories have accumulated with regard to tangent-formula recycling applications over the past 25 years, some of the results we obtained were quite unexpected.

Our initial ILED tests have shown that the advantages offered to *SnB* by random-atom phasing cannot be successfully transferred to tangent-formula recycling methods. Replacing randomly generated phase sets with random atom phases does not improve *RANTAN* success rates, rather it worsens

them, even though the average r.m.s. phase error in the starting sets is appreciably lower (99.6° as compared to 105.1°) for the 20 sets of randomly generated phase values that ultimately produce structural solutions.

In the case of the 132-atom CBT structure, pseudosymmetry may have been a contributing factor to why tangent-formula procedures were unable to recycle a 20-atom fragment with a r.m.s. fit of 0.093 Å with the refined structure. Table 1 shows that *RANTAN* has a success rate of 7.2% for similarly deformed fragments having a r.m.s. error of 0.1 Å but the rate of success rapidly falls to 2.2 and 0.15% for r.m.s. error models of 0.15 and 0.2 Å, respectively. No solutions were obtained for 20-atom fragments having random Gaussian displacement errors in excess of 0.20 Å.

In contrast, the subsequent *SnB* results for the CBT structure are quite remarkable. Although we know of no analysis that has suggested conditions under which the sensitivity of fragment-recycling methods toward larger displacement errors in the model may be tolerated, from Table 1 it may be noted that the degrees of positional, orientation and deformational r.m.s. misfit of the 20-atom CBT fragment are not equivalent. Moreover, these data indicate that it is better to have a model that is poorly oriented but correctly positioned than a model that is better oriented but more miscentered with regard to its true location in the cell. *This observation strongly suggests that one should sample additional points in the vicinity of each molecular replacement position, say within* ±0.5 Å, *if initial recycling attempts do not succeed.* We caution that tangent-formula-based methods may not share the same degree of success as the *SnB* method with regard to retrieving solutions using this latter strategy.

RANTAN was seen to produce solutions for 60% (6/10) of the 16-atom GRAMA fragments as compared to 100% for *SnB* as shown in Table 2(*a*). Five of the six *RANTAN* fragments that produced solutions exhibit abnormally high solution rates in that 82% or more of the 100 random trials succeeded. This strongly indicates that smaller segments of these five fragments will also produce solutions. For the 16 smaller ten-atom GRAMA fragments, the *RANTAN* success rate falls to 19% (3/16) as compared to 69% (11/16) for *SnB* in Table 2(*b*). No *RANTAN* solutions are found upon further dividing the two chains into 32 five-atom fragments; however, *SnB* still retains a success rate of 19% (6/32). It took an average of 15 *SnB* refinement cycles for these six solutions to converge. The smallest subfragment in this later group to produce an *SnB* solution is comprised of atoms 25–27 on the *B* chain and gives convergence in 19 refinement cycles. *This three-atom fragment represents a mere* 1% *of the* 317 *non*-H *atoms to be found in this structure.*

The best results thus far to solve GRAMA *ab initio* by using single randomly chosen atoms have located only three solutions among 240396 potential sets that were investigated (Weeks, DeTitta, Miller & Hauptman, 1993). The solution rate may be significantly higher than the 0.0012% reported as it was not feasible to rigorously test all these sets since an average of 450 *SnB* refinement cycles were required for convergence [subsequently, with random 300-atom starting models, *SnB* produced 35 *ab initio* solutions in 12000 trials for a success rate of about 0.3% (Hauptman, Miller & Weeks, 1994)]. Given that correctly chosen five-atom GRAMA fragments require an average of 15 cycles to reach a solution 19% of the time, one could survey 30 potential five-atom fragments from a previous *E* map in the same time it would take to test a single randomly generated *N*-atom structure for 450 cycles. This strategy could conceivably outperform the random-atom approach by many orders of magnitude with regard to finding phase solutions for larger structures.

The MONOGRAM structure, by comparison, is a severe test for traditional small-molecule direct methods. Firstly, the 1.5 Å resolution restricts us to fewer *E* magnitudes from which to generate triples. Secondly, the point-atom approximation upon which direct methods are based tends to break down when the atoms of the structure have a wide distribution of thermal motion as is the case when many solvent molecules and amino-acid side chains are thermally disordered (Langs, 1993). The result is that the average values of the phase-invariant triples are less than their expected values and the tangent formula tends to drive the refinement to higher than average phase-invariant values since the ratio of triples to phases is much lower than for higher-resolution data.

SnB refinement proves to be much more stable than the tangent formula since it forces the average phase invariant value toward its I_1/I_0 estimate in (3) and thus avoids overly consistent refinements that diverge from the expected phase-invariant values. The results in Table 3 clearly show the instability of the tangent formula. The r.m.s. phase error for the best set among the 100 *RANTAN* trials based on the 193-atom fragment is seen to jump nearly 10° after the basis-set fraction is allowed to refine (compare rows 2 and 3 of columns 2 and 3). The r.m.s. phase errors for the two *SnB* calculations, however, are seen to improve by nearly 5°, in marked contrast to the *RANTAN* results (compare rows 2 and 3 of columns 5 and 6). The reduction of the *SnB* r.m.s. phase error to ~55° as compared to ~65° (row 3) for the best of the 100 *RANTAN* trials has another advantage. The traditional figures of merit by which

direct-methods solutions are indicated often fail at low resolution (Woolfson & Yao, 1990; Gilmore, Henderson & Bricogne, 1991), making it difficult to identify the more likely solution sets among the 100 that were computed. *SnB* recycling has been formulated to produce a single density map with the likelihood that it is a marked improvement over the best of those computed by multisolution tangent-formula recycling methods.

Concluding remarks

The minimal function has been shown to be a powerful new tool in fragment-recycling applications involving small fragments, large unknown r.m.s. displacement errors in the model and low-resolution data. Although *SnB* fragment-recycling calculations are five to ten times more computer costly than tangent-formula recycling methods, they may provide a solution when the more economic tangent-based solution methods fail.

This research has been supported in part by NIH grant GM-46733 and NSF grant IRI-9108288. The authors thank Professor Bryan M. Craven of the University of Pittsburgh for providing the cholesteryl butanoate data and for additional research support to GWH under NIH grant HL-20350. Help provided by G. David Smith in reconstructing the MONOGRAM molecular-replacement model and assisting with computer-graphics evaluations is appreciated.

References

GILMORE, C., HENDERSON, A. N. & BRICOGNE, G. (1991). *Acta Cryst.* A**47**, 842–846.
HAN, G. W. (1993). PhD thesis, Department of Crystallography, Univ. of Pittsburgh, PA,USA.
HAN, G. W., CRAVEN, B. M. & LANGS, D. A. (1994). *J. Lipid Res.* Submitted.
HAUPTMAN, H. A. (1988). Proc. Am. Crystallogr. Assoc. Meet., Philadelphia, PA, USA. Abstract R4.
HAUPTMAN, H. A. (1991). *Crystallographic Computing* 5: *from Chemistry to Biology*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 324–332. New York: Oxford Univ. Press.
HAUPTMAN, H. A., MILLER, R. & WEEKS, C. M. (1994). Unpublished work.
HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97–107.
HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* A**34**, 863–870.
KARLE, J. (1968). *Acta Cryst.* B**24**, 182–186.
LANGS, D. A. (1985). *RANTAN: Fortran Program for Implementing the Random-Start Tangent-Formula Phase Refinement and Fragment-Recycling Schemes.* Medical Foundation of Buffalo, 73 High Street, Buffalo, NY 14203, USA.
LANGS, D. A. (1988). *Science*, **241**, 188–191.
LANGS, D. A. (1993). *Acta Cryst.* D**49**, 158–167.
LANGS, D. A., BLESSING, R. H. & DUAX, W. L. (1991). *Int. J. Pept. Protein Res.* **39**, 291–299.
LANGS, D. A., SMITH, G. D., COURSEILLE, C., PRÉCIGOUX, G. & HOSPITAL, M. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 5345–5349.
MILLER, R. (1993). Unpublished work.
MILLER, R., DETITTA, G. T., JONES, R., LANGS, D. A., WEEKS, C. M. & HAUPTMAN, H. A. (1993). *Science*, **259**, 1430–1433.
MILLER, R., GALLO, S. M., KHALAK, H. G. & WEEKS, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
PLETNEV, V. Z., GALITSKII, N. M., IVANOV, V. T., SMITH, G. D., WEEKS, C. M. & DUAX, W. L. (1980). *Biopolymers*, **19**, 1517–1534.
SHIONO, M. & WOOLFSON, M. M. (1992). *Acta Cryst.* A**48**, 451–456.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
WEEKS, C. M., DETITTA, G. T., HAUPTMAN, H. A., THUMAN, P. & MILLER, R. (1994). *Acta Cryst.* A**50**, 210–220.
WEEKS, C. M., DETITTA, G. T., MILLER, R. & HAUPTMAN, H. A. (1993). *Acta Cryst.* D**49**, 179–181.
WEEKS, C. M., HAUPTMAN, H. A., SMITH, G. D., BLESSING, R. H., TEETER, M. M. & MILLER, R. (1995). *Acta Cryst.* D**51**, 33–38.
WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 65–67.
WOOLFSON, M. M. & YAO, J.-X. (1990). *Acta Cryst.* A**46**, 409–413.
YAO, J.-X. (1981). *Acta Cryst.* A**37**, 642–644.
YAO, J.-X. (1983). *Acta Cryst.* A**39**, 35–37.

Acta Cryst. (1995). A**51**, 87–91

Determination of Laue Class from Diffraction Data of Polycrystalline Materials

BY MASAAKI OHMASA

Department of Life Science, Himeji Institute of Technology, Kanaji 1479-1, *Kamigori, Akogun, Hyogo* 678-12, *Japan*

AND KAZUMASA OHSUMI

Photon Factory, National Laboratory for High Energy Physics, Oho 1, *Tsukuba, Ibaraki* 305, *Japan*

(*Received* 7 *February* 1994; *accepted* 4 *July* 1994)

Abstract

It has long been believed that superposition of Debye–Scherrer lines caused by symmetry of the reciprocal lattice should suppress true information on Laue symmetry of crystal systems with high symmetry: trigonal, tetragonal, hexagonal and cubic. An interpretation of intensities at superposed reciprocal-lattic points reveals that Laue classes of polycrystalline materials can be identified from con-