Acta Cryst. (1997). A53, 436–444

Incorporating Tangent Refinement in the *Shake-and-Bake* Formalism

CHUN-SHI CHANG,[a,b] CHARLES M. WEEKS,[a]* RUSS MILLER[a,b] AND HERBERT A. HAUPTMAN[a]

[a]*Hauptman-Woodward Medical Research Institute, Inc., 73 High Street, Buffalo, NY 14203-1196, USA, and* [b]*Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, USA. E-mail: weeks@hwi.buffalo.edu*

(Received 2 October 1996; accepted 25 February 1997)

Abstract

Shake-and-Bake is a direct-methods procedure in which phase refinement and Fourier refinement are alternated repetitively, unconditionally and automatically. The traditional *Shake-and-Bake* approach invoked a parameter-shift routine to perform phase refinement in an effort to reduce the value of the minimal function. In this paper, parameter shift is replaced with the tangent formula as a means of phase refinement. This study shows that the tangent formula is more efficient than parameter shift for small structures when the number of refinement cycles and number of applications of the tangent formula per *Shake-and-Bake* cycle are chosen very carefully. For larger structures, including the 400 non-H-atom crambin structure, the two methods generally perform with similar efficiency. However, only parameter shift has successfully produced recognizable solutions for the difficult 317 non-H-atom structure gramicidin A.

1. Introduction

Shake-and-Bake (Weeks, DeTitta, Hauptman, Thuman & Miller, 1994) is a multisolution method of crystal structure determination capable of providing *ab initio* solutions to structures containing as many as 600 independent non-H atoms (Weeks, Hauptman, Smith, Blessing, Teeter & Miller, 1995; Anderson, Weiss & Eisenberg, 1996; Prive, Ogihara, Wesson, Cascio & Eisenberg, 1995; Smith, Blessing, Ealick, Fontecilla-Camps, Hauptman, Housset, Langs & Miller, 1996). Unlike conventional direct methods, *Shake-and-Bake* is a cyclical process that automatically alternates phase refinement in reciprocal space with the imposition of physically meaningful constraints through an atomic interpretation of the electron density in real space. Previously reported applications of *Shake-and-Bake* have also differed from traditional methods, which rely on the tangent formula (Karle & Hauptman, 1956), in that reciprocal-space phase refinement has utilized a parameter-shift procedure (Bhuiya & Stanley, 1963) that reduces the value of the minimal function (Debaerdemaeker & Woolfson, 1983; Hauptman, 1991; DeTitta, Weeks, Thuman, Miller & Hauptman, 1994). Initial constraints are imposed on *Shake-and-Bake* starting phase sets by deriving them from randomly positioned atoms rather than simply assigning random values. *Shake-and-Bake* is contrasted to conventional direct methods in Fig. 1.

Since *Shake-and-Bake* has been used in a routine manner to solve structures that were difficult or impossible by traditional methods operating only in reciprocal space, it is important to develop a more complete understanding of why it has been so successful. Therefore, the present investigation considers the following questions within the framework of *SnB* (Miller, Gallo, Khalak & Weeks, 1994), a computer program that implements *Shake-and-Bake*. What are the relative contributions of (i) alternation between real and reciprocal space and (ii) the particular method of phase refinement used to the success of *Shake-and-Bake*? What is the effect of replacing parameter-shift phase refinement with tangent-formula phase refinement? Can solutions be reliably recognized when this is done? What balance should be sought between the amount of time spent in the two spaces (*i.e.* how many phase-refinement iterations should there be per *Shake-and-Bake* cycle)?

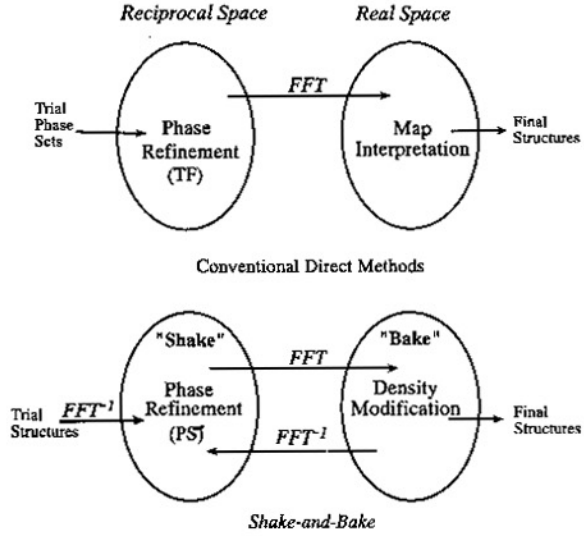

Fig. 1. A comparison of conventional direct methods with *Shake-and-Bake*. TF = tangent formula; PS = parameter shift.

1.1. *Phase refinement*

Direct methods exploit probabilistic relationships among normalized structure-factor magnitudes $|E|$ to derive values for individual phases (Hauptman & Karle, 1953). In practical applications of conventional direct methods, the *tangent formula*,

$$\tan(\phi_{\mathbf{H}}) = \frac{\sum_{\mathbf{K}} |E_{\mathbf{K}}E_{\mathbf{H-K}}| \sin(\phi_{\mathbf{K}} + \phi_{\mathbf{H-K}})}{\sum_{\mathbf{K}} |E_{\mathbf{K}}E_{\mathbf{H-K}}| \cos(\phi_{\mathbf{K}} + \phi_{\mathbf{H-K}})} \quad (1)$$

(Karle & Hauptman, 1956), has played a key role in the phase-determination process. If several pairs of phases, $\phi_{\mathbf{K}}$ and $\phi_{\mathbf{H-K}}$, and their associated $|E_{\mathbf{K}}|$, $|E_{\mathbf{H-K}}|$ are known, (1) can be used to determine the most probable value for $\phi_{\mathbf{H}}$. Phase expansion and/or refinement in reciprocal space is accomplished through successive applications of this relationship. The tangent formula, in either its original or a weighted form (Hull & Irwin, 1978), is the heart of conventional multisolution phasing programs such as *MULTAN* (Germain, Main & Woolfson, 1971; Main, Fiske, Hull, Lessinger, Germain, Declercq & Woolfson, 1980), *RANTAN* (Yao, 1981) or *SHELXS* (Sheldrick, 1985*a*), which refine multiple sets of trial phases by making many iterations or passes through the phase list. Although thousands of small-molecule structures have been solved through the use of these programs, the tangent-refinement process generally fails when the number of independent non-H atoms exceeds 100–150, and successful applications to large structures have been rare. Cases have been reported where the tangent formula had difficulty in refining phases properly (Lessinger, 1976) and, in fact, divergence was observed when the true phases were refined. This problem may arise from the tangent formula's tendency to refine phases to an overly consistent set (*i.e.* the refined cosine-invariant values are significantly greater than the expected values).

The constrained global minimization of an objective function such as the *minimal function*,

$$\begin{aligned} R(\phi) = &\left(\sum_{\mathbf{H,K}} A_{\mathbf{HK}} \{\cos(\phi_{\mathbf{H}} + \phi_{\mathbf{K}} + \phi_{\mathbf{-H-K}}) \right. \\ &- [I_1(A_{\mathbf{HK}})/I_0(A_{\mathbf{HK}})]\}^2 + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \\ &\times \{\cos(\phi_{\mathbf{L}} + \phi_{\mathbf{M}} + \phi_{\mathbf{N}} + \phi_{\mathbf{-L-M-N}}) \\ &\left. - [I_1(B_{\mathbf{LMN}})/I_0(B_{\mathbf{LMN}})]\}^2 \right) \\ &\times \left(\sum_{\mathbf{H,K}} A_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} |B_{\mathbf{LMN}}| \right)^{-1} \end{aligned} \quad (2)$$

(Debaerdemaeker & Woolfson, 1983; Hauptman, 1991; DeTitta, Weeks, Thuman, Miller & Hauptman, 1994), provides an alternative approach to phase refinement. The minimal function expresses a relationship among phases related by triplet and negative quartet invariants that have the associated parameters (or weights)

$$A_{\mathbf{HK}} = (2/N^{1/2})|E_{\mathbf{H}}E_{\mathbf{K}}E_{\mathbf{H+K}}| \quad (3)$$

and

$$\begin{aligned} B_{\mathbf{LMN}} = &(2/N)|E_{\mathbf{L}}E_{\mathbf{M}}E_{\mathbf{N}}E_{\mathbf{L+M+N}}| \\ &\times [(|E_{\mathbf{L+M}}|^2 + |E_{\mathbf{M+N}}|^2 + |E_{\mathbf{N+L}}|^2) - 2], \end{aligned} \quad (4)$$

respectively, where the $|E|$'s are the normalized structure-factor magnitudes and N is the number of atoms, assumed identical, in the unit cell. $R(\phi)$ is a measure of the mean-square difference between the calculated structure invariants and their expected values as given by the ratio of Bessel functions, and it is expected to have a constrained global minimum when the phases are equal to their correct values for some choice of origin and enantiomorph (the minimal principle). Experimentation has thus far confirmed that, when the minimal function is used actively in the phasing process and solutions do indeed exist, the final trial structure corresponding to the smallest value of $R(\phi)$ is a solution.

Parameter shift (Bhuiya & Stanley, 1963) is a seemingly simple search technique that has proven to be quite powerful as an optimization method when used to reduce the value of the minimal function, provided that appropriate choices of parameter values are made. In *SnB*, the phases are considered in decreasing order with respect to the values of the associated $|E|$'s. When considering a given phase ϕ_i, the value of the minimal function [equation (2)] is initially evaluated three times. First with the given set of phase assignments, second with phase ϕ_i modified by the addition of the predetermined phase shift and third with ϕ_i modified by the subtraction of the predetermined phase shift. If the first evaluation yields the minimum of these three values of the minimal function, then consideration of ϕ_i is complete and parameter shift proceeds to ϕ_{i+1}. Otherwise, the direction of search is determined by the modification that yields the minimum value and the phase is updated to reflect that modification. In this case, phase ϕ_i continues to be updated by the predetermined phase shift in the direction just determined so long as the value of the minimal function is reduced, though there is a user-defined predetermined maximum number of times that the shift is attempted. Based on extensive experimentation involving a variety of structures in several space groups, it has been determined that, in terms of running time and percentage of trial structures that produce a solution, a maximum of two 90° phase shifts is optimum except in centrosymmetric space groups where only a single shift of 180° is required for each phase (Weeks, DeTitta, Hauptman, Thuman & Miller,

1994). Refined phase values are used immediately in the subsequent refinement of other phases. It should be noted that the parameter-shift routine is similar to that used in ψ-map refinement (White & Woolfson, 1975) and *XMY* (Debaerdemaeker & Woolfson, 1989).

1.2. *Dual-space phase improvement*

The goal of real-space refinement techniques is to improve the agreement of an electron-density map with a set of physically meaningful constraints. Such techniques are used only in a rudimentary way in conventional small-molecule direct-methods applications. The final phase sets resulting from tangent refinement are ranked according to figures of merit and one, or a few, of the most promising combinations are then transformed to real space. If possible, the corresponding maps are interpreted in terms of atomic structures. The quality of a basically correct model structure may be significantly improved by doing a few cycles of Fourier refinement, a process termed *E-Fourier recycling* (Sheldrick, 1985*b*). Another form of recycling was introduced by Jerome Karle (1968), who recognized that even a relatively small chemically sensible fragment, extracted by manual interpretation of an E map, could be parlayed into a complete solution by transformation back to reciprocal space and then performing additional iterations of tangent-formula refinement.

Historically, real-space phase-improvement methods have played a larger role in macromolecular structure determination where the physical constraints have included atomicity (in high-resolution cases), positivity, solvent flatness, polymer continuity and conformity with known non-crystallographic symmetry [see review by Podjarny, Bhat & Zwick (1987) and references therein]. Macromolecular applications typically involve a single phase set (*e.g.* phases determined by MIR). Density-modification procedures that exploit physical constraints typically consist of the following steps:

(i) compute an electron-density map using the observed magnitudes $|F_{obs}|$ and initial phases ϕ_{init};

(ii) modify the electron density using the known constraint [*e.g.* $\rho_{mod} = \max(O, \rho)$ for negative density truncation, where ρ is the electron-density value at a specific grid point];

(iii) calculate structure factors from ρ_{mod};

(iv) merge the calculated structure factors with the experimental and produce a set of new phases, ϕ_{new}.

The entire process may be repeated as many times as desired beginning with ϕ_{new}. Although the complete cycle consists of two substantive steps, density modification and structure-factor merging, as well as the Fourier transforms, the emphasis in such procedures is on the real-space density modification where the major portion of the refinement is occurring.

The tremendous increases in computer speed in recent years have made it feasible to consider cycling every trial structure, generated with random phases in a multisolution procedure, back and forth between real and reciprocal space many times while performing optimization alternately in each space. This is a computer-intensive task, as it requires the use of two Fourier transforms (forward and inverse) during each cycle. This cyclical process forms the basis of the synergistic *Shake* (phase refinement) and *Bake* (density modification) procedure in which the power of reciprocal-space phase refinement is augmented by filtering to impose the phase constraints implicit in real space. The *Shake-and-Bake* algorithm is diagrammed in Fig. 2 and can be seen to closely resemble the macromolecular density-modification procedure described above. The significant difference is the addition of a refinement process (*e.g.* parameter shift) in reciprocal space. In the generalized procedure, any phase-refinement method can be considered, whether or not the minimal function is used actively or only passively as a figure of merit. The imposition of physical constraints counteracts the tendency of phase refinement to propagate errors or produce overly consistent phase sets.

Automatic real-space electron-density map interpretation consists of selecting an appropriate number of the largest peaks (typically equal to or less than the expected number of atoms) to be used as an updated trial structure without regard to chemical constraints other than a minimum allowed distance between atoms. If markedly unequal atoms are present, appropriate numbers of peaks (atoms) can be weighted by the proper atomic numbers during transformation back to reciprocal space. Thus, *a*

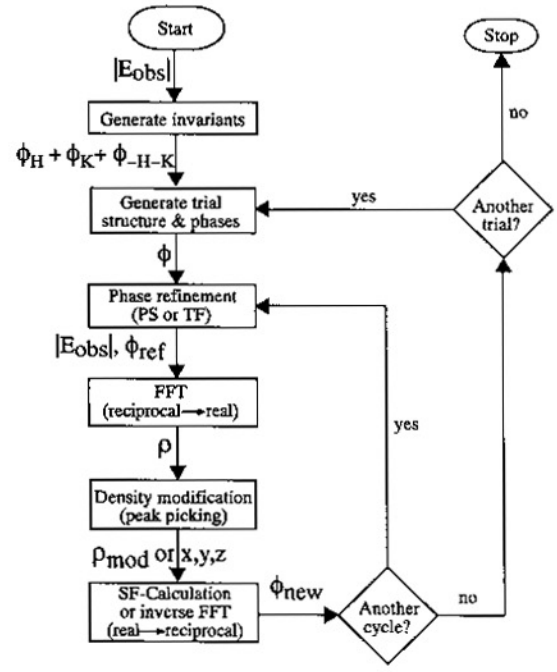

Fig. 2. A flow chart for the *Shake-and-Bake* algorithm.

Table 1. *Test data sets used in this investigation*

Structure	Space group	Atoms/ASU (n)	Chemical formula	Reference
9α-Methoxycortisol	$P2_12_12_1$	28	$C_{22}H_{32}O_6$	Weeks, Duax & Wolff (1976)
Isoleucinomycin	$P2_12_12_1$	84	$C_{60}H_{102}N_6O_{18}$	Pletnev, Galitskii, Smith, Weeks & Duax (1980)
Ternatin	$P2_12_12_1$	104	$C_{74}H_{134}N_{14}O_{16}$	Miller, DeTitta, Jones, Langs, Weeks & Hauptman (1993)
Hexaisoleucinomycin	$P2_12_12_1$	127	$C_{80}H_{136}N_8O_{32}$	Pletnev, Ivanov, Langs, Strong & Duax (1992)
Gramicidin A	$P2_12_12_1$	~300	$C_{228}H_{370}N_{40}O_{49}$	Langs (1988)
Tetrahymanol	$P2_1$	63	$(C_{30}H_{52}O)_2{\cdot}H_2O$	Langs, Duax, Carrell, Berman & Caspi (1977)
Cholesterol butanoate	$P2_1$	132	$C_{124}H_{208}O_8$	Han, Craven & Langs (1994)
Valinomycin dioxane	$P2_1$	176	$(C_{54}H_{90}O_{18}N_6)_2{\cdot}(C_4H_8O_2)_3{\cdot}2H_2O$	Langs, Blessing & Duax (1992)
Crambin	$P2_1$	~400	$C_{202}H_{321}N_{55}O_{64}S_6{\cdot}$ ~$75H_2O$	Hendrickson & Teeter (1981); Teeter, Roe & Heo (1993)
Prostaglandin E_2	$P1$	25	$C_{20}H_{32}O_5$	Edmonds & Duax (1974)
5,16-Pregnadiene	$P1$	48	$C_{44}H_{64}O_4$	Duax, Langs, Strong & Osawa (1979)
Emerimicin-(1-9) benzyl ester	$P1$	74	$C_{51}H_{77}N_9O_{11}{\cdot}3H_2O$	Marshall, Hodgkin, Langs, Smith, Zabrocki & Leplawy (1990)
Enkephalin analog	$P1$	96	$(C_{24}H_{30}N_2O_6)_3$	Krstenansky, Langs & Smith, unpublished

priori knowledge concerning the chemical composition of the crystal is utilized but no knowledge of constitution is required or used during peak selection. It is useful to think of peak picking in this context as simply an extreme form of density modification appropriate when atomic resolution data are available. The entire dual-space refinement procedure is repeated for the desired number of cycles.

2. Methods

Both the parameter-shift and tangent-formula phase-refinement variants of *SnB* were applied to a series of known structures in space groups $P2_12_12_1$, $P2_1$ and $P1$ having atomic resolution data and ranging in size from 25 to ~400 atoms (see Table 1). All except ternatin and crambin were originally solved by traditional direct methods. Ternatin was the first previously unknown structure solved by *Shake-and-Bake* (Miller, DeTitta, Jones, Langs, Weeks & Hauptman, 1993) and previous attempts to solve crambin *ab initio* using pure conventional tangent-based direct methods were unsuccessful (Sheldrick, Dauter, Wilson, Hope & Sieker, 1993).

For each structure, the atom:phase:triplet:negative-quartet ratio of 1:10:100:0 was used, regardless of whether the parameter-shift procedure or the tangent formula was used for phase refinement. Negative quartets were omitted because their inclusion typically resulted in a less-efficient refinement. A sample of 1000 randomly positioned n-atom trial structures (where n is the number of non-H atoms in the asymmetric unit) was generated for each data set. These 1000 trials were refined for n cycles using both phase-refinement methods. When the tangent formula was employed, the minimal-function value for the refined phase set was still computed but used only as a figure of merit. Parameter-shift phase refinement was carried out using 1, 2, 3, 4 and 5 iterations (passes through the phase set) per *Shake-and-Bake* cycle, and a maximum of two 90° shifts per phase was applied in each iteration of refinement. Tangent-formula phase refinement involved 1, 2, 4, 8, 16, 32 and 64 iterations per cycle, although higher numbers of iterations were not tested for some structures when it became apparent that the number of solutions was dropping rapidly with additional iterations.

The tangent formula was implemented in both its original Karle–Hauptman [equation (1)] and weighted forms (Hull & Irwin, 1978). When the new value of a phase ϕ_i is determined by tangent refinement, a decision must be made as to whether or not to update the value of ϕ_i before determining the value of the next phase, ϕ_{i+1}. The term *feedback* is used to refer to the situation where the value of a phase ϕ_i is updated immediately, thus making the new value available for subsequent phase evaluations. The alternative situation is to refine phases based on the current phase set and withhold the new values until all phases are refined. Implementations of the tangent formula both with and without feedback are reported. Based on previously reported experimentation (Weeks, DeTitta, Hauptman, Thuman & Miller, 1994) the parameter-shift routine is always used in a feedback mode.

Solutions are trial structures having a close match between peak positions and the true atomic positions for some choice of origin and enantiomorph, and the success rate is the percentage of trial structures that become solutions over the course of refinement. Solutions typically have mean phase errors of 30° or less. In space groups such as $P2_12_12_1$, where there are

only a few possible discrete origin positions, *Shake-and-Bake* trials for known structures can be rapidly screened for solutions by examining the mean phase error or average absolute value of the deviations of the phases from their known values calculated using final refined coordinates and thermal parameters. In all space groups, similar judgments can be made by examining the cosine-invariant figure of merit,

$$\begin{aligned}\text{COSFOM} &= \left\{ \sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} \left| \cos(\phi_{\mathbf{H}} + \phi_{\mathbf{K}} + \phi_{-\mathbf{H}-\mathbf{K}}) - \cos(\phi^T_{\mathbf{H}} + \phi^T_{\mathbf{K}} + \phi^T_{-\mathbf{H}-\mathbf{K}}) \right| \right. \\ &\quad \left. + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}| \left| \cos(\phi_{\mathbf{L}} + \phi_{\mathbf{M}} + \phi_{\mathbf{N}} + \phi_{-\mathbf{L}-\mathbf{M}-\mathbf{N}}) - \cos(\phi^T_{\mathbf{L}} + \phi^T_{\mathbf{M}} + \phi^T_{\mathbf{N}} + \phi^T_{-\mathbf{L}-\mathbf{M}-\mathbf{N}}) \right| \right\} \\ &\quad \times \left(\sum_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}} + \sum_{\mathbf{L},\mathbf{M},\mathbf{N}} |B_{\mathbf{LMN}}| \right)^{-1}, \end{aligned} \tag{5}$$

which measures the average weighted absolute value of the difference between the values of the invariants computed using the trial (ϕ) and known phases (ϕ^T). Although the values of the individual phases depend on the choice of origin and enantiomorph, the cosine invariants are independent of these choices. Therefore, cosine invariants can be compared without first referring two phase sets to a common origin and enantiomorph. COSFOM values have a bimodal distribution, lying in the range 0.10–0.25 for solutions and being greater than 0.35 for non-solutions. Minimal function values $R(\phi)$ also have a bimodal distribution. Consideration of both COSFOM and $R(\phi)$ permits trials to be categorized as true, false or missed solutions or as non-solutions.

The measurement of success rates at the end of a fixed number of cycles provides an important indication as to the effectiveness of a particular method. However, this measurement by itself provides an incomplete comparison of two refinement protocols because it does not take into account the computational effort (running time) needed to produce the solutions. The relative efficiency of two phase-refinement methods can be compared as a function of cycle and the number of iterations per cycle on the basis of the *cost effectiveness*,

$$\text{CE} = S \times 3600/TCt, \tag{6}$$

where T = the number of trial structures, C = the number of cycles per trial structure, S = the number of solutions produced by T such trials and t = the running time (in seconds) for one cycle of one trial. CE has units of solutions per hour and values reported here were measured on a Silicon Graphics R4000 Indigo workstation.

Table 2. *Maximum tangent-formula cost effectiveness (CE) or solutions/hour*

The number of cycles and iterations/cycle producing peak performance varied.

Structure	Unweighted No feedback	Unweighted Feedback	Weighted Feedback
9α-Methoxycortisol	89.34	101.44	145.00
Isoleucinomycin	4.17	4.10	5.51
Tetrahymanol	2.50	3.01	2.64
Valinomycin dioxane	0.53	0.66	0.56
Crambin	0.01	0.02	0.003
Emerimicin ester	55.93	65.51	61.96
Enkephalin analog	8.17	12.09	11.39

3. Results

The efficiency of the tangent formula under various combinations of weighting and feedback was compared for several data sets and the results are presented in Table 2. In general, the peak performance, as measured by the maximum cost effectiveness as defined in (6), is slightly greater under feedback conditions. Hull–Irwin weights gave better performance for the small $P2_12_12_1$ test structure, but the results were similar or inferior to the unweighted formula for the larger or the $P1$ structures. Consequently, the unweighted feedback conditions were chosen for further study.

Fig. 3 compares tangent-formula and parameter-shift phase refinement by illustrating the cost effectiveness as a function of *Shake-and-Bake* cycle for several of the test structures. The family of curves presented for each structure shows the results for various numbers of phase-refinement iterations per cycle. 9α-Methoxycortisol is a 28-atom steroid that crystallizes in space group $P2_12_12_1$ and is representative of the type of structure easily solved by conventional direct methods. In this case, the tangent formula is seen to be much more cost effective than parameter shift with peak efficiency occurring at 32 phase-refinement iterations after only one cycle. Tangent-formula cost effectiveness is highly dependent on the number of iterations per cycle whereas parameter shift does not exhibit this dependency. The tangent-formula curves for high numbers of iterations per cycle peak quickly and then fall off quite rapidly indicating that the number of cycles must be chosen judiciously if high efficiency is to be achieved. In contrast, the parameter-shift curves rise more slowly to a lower maximum and then decrease very gradually. At one iteration per cycle, the tangent-formula curve resembles the family of parameter-shift curves. Isoleucinomycin is a larger (84-atom) and more difficult $P2_12_12_1$ structure. Although tangent-formula phase refinement is still superior to parameter shift, maximum cost effectiveness occurs with fewer iterations per cycle (*i.e.* 2, 4, 8). This trend continues for the 63-atom $P2_1$ structure tetrahymanol. In this case, however, parameter shift is almost as efficient as the tangent formula. This is also true for the larger (400-atom) $P2_1$ structure crambin,

where the tangent formula is most efficient with only one iteration per cycle.

The results of the tangent-formula and parameter-shift comparison for a variety of structures are summarized in Table 3, including information concerning the cycle (expressed as a function of n) and the number of iterations per cycle at which peak performance occurred. In general, the tangent formula solves small structures more cost effectively but both phase-refinement methods are equally efficient for solving most of the large structures, including crambin. However, only parameter shift has produced recognizable solutions for gramicidin A. Approximately 5000 gramicidin A trial structures have been processed by each method, with parameter shift yielding 12 solutions (success rate of ~0.25%). The tangent formula has, in fact, produced one solution (low COSFOM). However, this solution would not have been found if gramicidin A were an unknown because it had a relatively high value for the minimal function. This suggests that the minimal function is not such a robust figure of merit when it is used only passively to trace the progress of the phasing process.

Structures in space group $P1$ exhibit behavior which, in many respects, differs from that of structures crystallizing in other space groups. As shown by the data for the 74-atom emerimicin ester, the maximum cost effectiveness is anomalously high considering the size of the structure. The tangent formula still does better than parameter shift but only when one iteration per cycle is used. The data for a 96-atom enkephalin analog show the same effects, as do those for two smaller $P1$ structures (see Table 3). It seems clear that, in $P1$, it is always best to do a minimum amount of 'shaking' (phase refinement). Although no rigorous explanation can be given to explain this observation, it can be argued heuristically that, since an infinite number of choices of origin position are available in this space group, it is statistically likely that some subset of the atoms in any trial structure be consistent with some choice of origin, and it is therefore better to allow Fourier refinement to play a larger role. The unexpectedly high success rate observed for $P1$ structures also raises the question of whether or not it would be better to treat all structures as if they were $P1$ structures. Pertinent data are presented

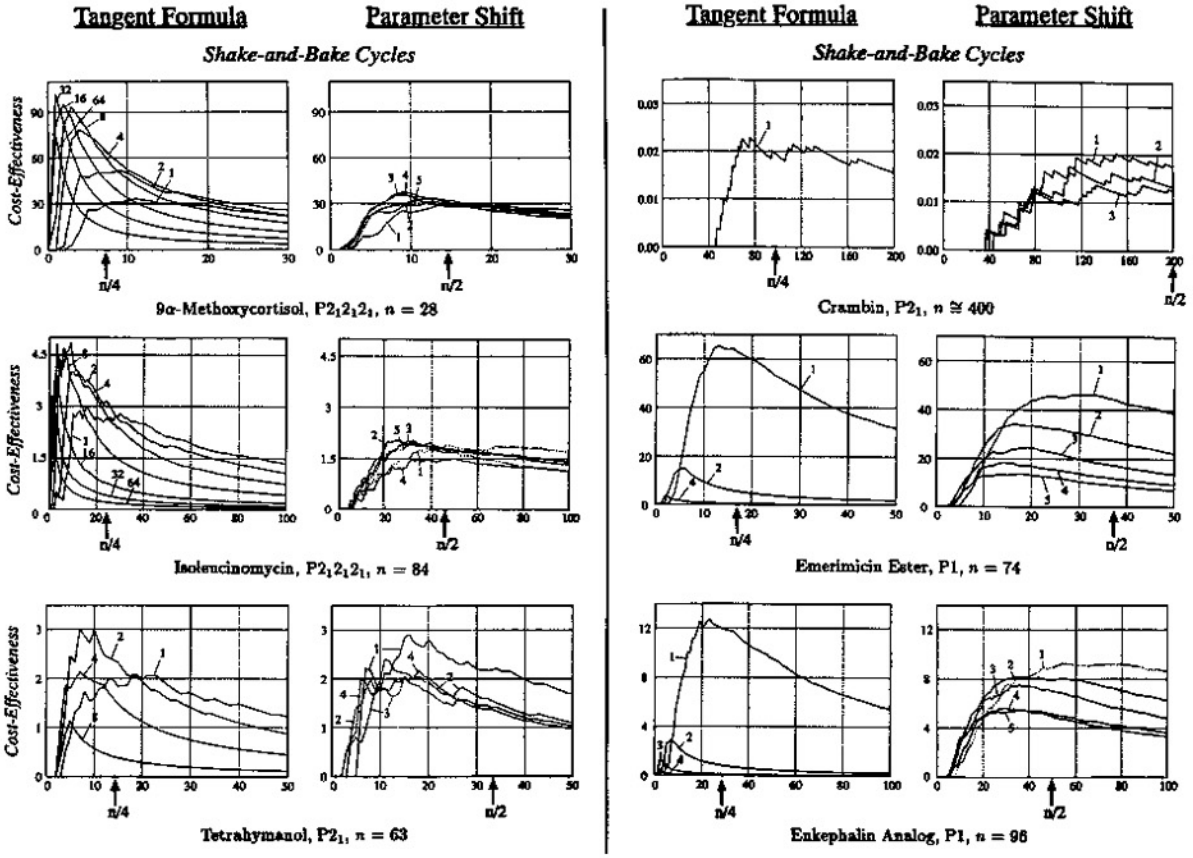

Fig. 3. Cost effectiveness of tangent-formula and parameter-shift phase refinement as a function of *Shake-and-Bake* cycle for several structures. Labels indicate the number of phase-refinement iterations per *Shake-and-Bake* cycle.

Table 3. *Comparison of the maximum tangent-formula (TF) and parameter-shift (PS) cost-effectiveness (CE) values and the numbers of cycles and iterations/cycle producing these values*

The unweighted tangent formula with feedback was used.

		Maximum CE		Cycles		Iterations per cycle	
Structure	n	*TF*	*PS*	*TF*	*PS*	*TF*	*PS*
$P2_12_12_1$							
9α-Methoxycortisol	28	101.4	37.4	$n/28$	$n/3$	32	3
Isoleucinomycin	84	4.1	1.9	$n/9$	$n/2$	4	3
Ternatin	104	0.03	0.06	$n/2$	$n/2$	2	1
Hexaisoleucinomycin	127	0.06	0.05	$n/1.5$	$n/1.3$	1*	3*
Gramicidin A	~300	0.0	0.01	—	$n/3$	—	3*
$P2_1$							
Tetrahymanol	63	3.0	2.9	$n/9$	$n/4$	2	1
Cholesterol butanoate	132	0.4	0.3	$n/5$	$n/3$	1*	3*
Valinomycin dioxane	176	0.7	0.5	$n/5$	$n/3$	1	2
Crambin	~400	0.02	0.02	$n/4$	$n/3$	1	2
$P1$							
Prostaglandin E_2	25	194.6	80.7	$n/4$	$n/2$	1	2
5,16-Pregnadiene	48	405.3	209.3	$n/10$	$n/5$	1	1
Emerimicin ester	74	65.5	46.4	$n/6$	$n/2$	1	1
Enkephalin analog	96	12.1	9.3	$n/4$	$n/2$	1	1

* Only values tested.

Table 4. *Comparison of success rate and maximum cost effectiveness (CE) in P1 and the actual space group*

(*a*) 9α-Methoxycortisol

	Space group $P1$ ($n = 112$) Success rates			Actual space group $P2_12_12_1$ ($n = 28$) Success rates		
Optimization method	115 cycles	500 cycles	Max CE (solutions/h)	30 cycles	500 cycles	Max CE (solutions/h)
Parameter shift	76.8%	92.0%	13.4	19.4%	27.0%	37.4
Tangent formula	51.2	52.7	21.2	16.5	17.6	78.5

(*b*) Tetrahymanol

	Space group $P1$ ($n = 126$) Success rates			Actual space group $P2_1$ ($n = 63$) Success rates		
Optimization method	130 cycles	500 cycles	Max CE (solutions/h)	50 cycles	500 cycles	Max CE (solutions/h)
Parameter shift	47.0%	66.4%	7.5	4.4%	6.8%	2.9
Tangent formula	24.6	26.6	10.0	3.1	4.2	2.1

in Table 4 for 28-atom 9α-methoxycortisol ($P2_12_12_1$) and 63-atom tetrahymanol ($P2_1$). When refinement is performed in $P1$, success rates after $\sim n$ cycles increase dramatically for both structures using either phasing procedure, and the maximum cost effectiveness is also increased for the $P2_1$ structure which requires only a twofold increase in computational effort. With continued refinement, only parameter shift produces a significant number of additional solutions. In their implementation of an algorithm closely related to *Shake-and-Bake*, but employing only tangent-formula phase refinement, Sheldrick & Gould (1995) have chosen to treat all structures in space group $P1$. Since these authors use a rotation search to provide starting coordinates when the structure contains a relatively rigid fragment, working in $P1$ is also advantageous since no translation is required.

4. Conclusions

As a consequence of the experiments described above, it is possible to make the recommendations presented in Table 5 for optimum use of the *Shake-and-Bake* procedure as implemented in *SnB* version 1.5. It is clear that, regardless of the phase-refinement method used, alternate refinement in reciprocal and real space makes an important contribution to the successful application of direct methods to structures larger than those routinely solved by such methods in the past. Tangent-formula phase refinement is ideally suited to provide quick answers to smaller structure problems. The immediate feedback of tangent-refined phases appears to give the best results in the *Shake-and-Bake* context and Hull–Irwin weights do not improve performance. On

Table 5. *Phase-refinement recommendations*

Recommendation	Method	Cycles	Shake/Bake (iterations/cycle)
$n < 100$ atoms	TF	$n/4$	4 or 1 ($P1$)
$n > 100$ atoms	PS	$n/2$	1
Always safe	PS	n	1

the other hand, parameter shift appears to be more robust – sometimes rather slow, but dependable and capable of producing recognizable results in difficult circumstances. Since efficiency often decreases rapidly with increasing numbers of *Shake-and-Bake* cycles, the numbers of cycles and phase-refinement iterations per cycle must be selected carefully if full advantage is to be taken of the tangent formula's potential speed. It is interesting to note that, if optimum parameter choices are made, the expected time to solution using the present *SnB* program (version 1.5) running on an R4000 Indigo workstation is on the order of a weekend for crambin and a week for gramicidin A.

It is important to remember that $P1$ structures are special and respond best to a minimum amount of phase refinement. In particular, $P1$ structures should never be subjected to more than one iteration of phase refinement per *Shake-and-Bake* cycle. If the data for structures in other space groups are treated in $P1$, the success rate continues to rise if more than n parameter-shift (but not tangent-formula) cycles are performed. Finally, it should be noted that all conclusions regarding relative efficiency or cost effectiveness are dependent on the particular computer program (*SnB* version 1.5) used. Parameter-shift running times can be made competitive with the tangent formula by coding in a way that takes greater advantage of trigonometric relationships but only permits shifts in some multiple of 90°.

The authors wish to thank Drs George T. DeTitta and David A. Langs for many helpful discussions, Dr Robert H. Blessing for assistance in preparing some of the data sets, Steven Gallo and Hanif Khalak for their help in developing the computer programs, and Drs Martha Teeter and Hakon Hope for use of their crambin data. This research was supported in part by NSF grant IRI-9412415 and NIH grant GM-46733.

References

Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* D**52**, 469–480.

Bhuiya, A. K. & Stanley, E. (1963). *Acta Cryst.* **16**, 981–984.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *Acta Cryst.* A**39**, 193–196.

Debaerdemaeker, T. & Woolfson, M. M. (1989). *Acta Cryst.* A**45**, 349–353.

DeTitta, G. T., Weeks, C. W., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.

Duax, W. L., Langs, D. A., Strong, P. & Osawa, Y. (1979). *Cryst. Struct. Commun.* **8**, 565–568.

Edmonds, J. W. & Duax, W. L. (1974). *Prostaglandins*, **5**, 275–281.

Germain, G., Main, P. & Woolfson, M. M. (1971). *Acta Cryst.* A**27**, 368–376.

Han, G. W., Craven, B. M. & Langs, D. A. (1994). *J. Lipid Res.* **35**, 2069–2082.

Hauptman, H. A. (1991). *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 324–332. IUCr/Oxford University Press.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem I. The Centrosymmetric Crystal. Am. Crystallogr. Assoc. Monogr.* No. 3. Ann Arbor, MI: Edwards Brothers.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Hull, S. E. & Irwin, M. J. (1978). *Acta Cryst.* A**34**, 863–870.

Karle, J. (1968). *Acta Cryst.* B**24**, 182–186.

Karle, J. & Hauptman, H. A. (1956). *Acta Cryst.* **9**, 635–651.

Langs, D. A. (1988). *Science*, **241**, 188–191.

Langs, D. A., Blessing, R. H. & Duax, W. L. (1992). *Int. J. Pept. Protein Res.* **39**, 291–299.

Langs, D. A., Duax, W. L., Carrell, H. L., Berman, H. & Caspi, E. (1977). *J. Org. Chem.* **42**, 2134–2137.

Lessinger, L. (1976). *Acta Cryst.* A**32**, 538–550.

Main, P., Fiske, S. J., Hull, S. E., Lessinger, L., Germain, G., Declercq, J.-P. & Woolfson, M. M. (1980). *MULTAN*80. *A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*. Universities of York, England, and Louvain, Belgium.

Marshall, G. R., Hodgkin, E. E., Langs, D. A., Smith, G. D., Zabrocki, J. & Leplawy, M. T. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 487–491.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.

Miller, R., Gallo, A. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.

Pletnev, V. Z., Galitskii, N. M., Smith, G. D., Weeks, C. M. & Duax, W. L. (1980). *Biopolymers*, **19**, 1517–1534.

Pletnev, V. Z., Ivanov, V. T., Langs, D. A., Strong, P. & Duax, W. L. (1992). *Biopolymers*, **32**, 819–827.

Podjarny, A. D., Bhat, T. N. & Zwick, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351–373.

Prive, G., Ogihara, N., Wesson, L., Cascio, D. & Eisenberg, D. (1995). Proceedings of ACA Meeting, Montreal, Canada. Abstract W008.

Sheldrick, G. M. (1985*a*). *SHELXS*86. *Program for the Solution of Crystal Structures*. University of Göttingen, Germany.

Sheldrick, G. M. (1985*b*). *Crystallographic Computing 3: Data Collection, Structure Determination, Proteins, and Databases*, edited by G. M. Sheldrick, C. Kruger & R. Goddard, pp. 184–189. Oxford: Clarendon Press.

Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). *Acta Cryst.* D**49**, 18–23.

Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.

Smith, G. D., Blessing, R. H., Ealick, S. E., Fontecilla-Camps, J. C., Hauptman, H. A., Housset, D., Langs, D. A. & Miller, R. (1996). *Acta Cryst.* A**52**, C64.

Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

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. & Wolff, M. E. (1976). *Acta Cryst.* B**32**, 261–263.
Weeks, C. M., Hauptman, H. A., Chang, C.-S. & Miller, R. (1997). ACA Transactions Symposium, Vol. 30. In the press.
Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* **D51**, 33–38.
White, P. S. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 53–56.
Yao, J.-X. (1981). *Acta Cryst.* A**37**, 642–644.