Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 15 July 1999
Accepted 1 November 1999

Ill-conditioned *Shake-and-Bake*: the trap of the false minimum

Hongliang Xu,[a]* Charles M. Weeks,[a] Ashley M. Deacon,[b] Russ Miller[a,c] and Herbert A. Hauptman[a]

[a]Hauptman–Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203, USA, [b]Department of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA, and [c]Department of Computer Science, State University of New York, Buffalo, NY 14260, USA. Correspondence e-mail: xu@hwi.buffalo.edu

The alternation of phase refinement with the imposition of real-space constraints is the essence of the *Shake-and-Bake* procedure. Typically, these constraints prevent trial structures from falling into local minima. Nevertheless, *P*1 structures appear to migrate to false minima with significant frequency. These false minima are characterized by the presence of a large 'uranium' peak on the corresponding Fourier map. Fortunately, they can be recognized and avoided by considering the values of the minimal function both before and after the application of constraints. However, it appears that finding solutions for large *P*1 structures is likely also to require parameter-shift conditions different from those that have been found to work well in other space groups. In fact, these conditions often yield an unusually high percentage of solutions.

1. Introduction

Shake-and-Bake (Weeks *et al.*, 1993) is a multitrial method of crystal-structure determination capable of providing *ab initio* solutions for structures containing as many as 1200 independent non-H atoms (Deacon *et al.*, 1998). Unlike conventional direct methods, *Shake-and-Bake* is a cyclical process that automatically alternates phase refinement in reciprocal space with the imposition, in real space, of physically meaningful constraints through an atomic interpretation of the electron density. As implemented in the computer program *SnB* (Miller *et al.*, 1994; Weeks & Miller, 1999*a*), the *Shake-and-Bake* algorithm first requires the generation of trial structures composed of randomly positioned atoms. The subsequent phase refinement is based either on the tangent formula (Karle & Hauptman, 1956) or on a parameter-shift procedure (Bhuiya & Stanley, 1963) that reduces the value of the minimal function (Debaerdemaeker & Woolfson, 1983; Hauptman, 1991). Peak picking is used to apply the real-space constraints. Cost-effective parameter values and refinement procedures have been described for six small proteins crystallizing in a variety of space groups including $P2_1$, $P2_12_12_1$, $P4_32_12$ and $I4$ (Weeks & Miller, 1999*b*). In the context of the *SnB* program, parameter-shift reduction of the minimal function appears to be the most effective phase-refinement method for these structures.

The most commonly used form of the minimal function,

$$R(\varphi) = \left(\sum_{H,K} A_{HK}\right)^{-1} \sum_{H,K} A_{HK}\left[\cos\varphi_{HK} - \frac{I_1(A_{HK})}{I_0(A_{HK})}\right]^2, \quad (1)$$

measures the mean-square difference between the cosine values of the three-phase structure invariants,

$$\varphi_{HK} = \varphi_H + \varphi_K + \varphi_{-H-K}. \quad (2)$$

computed using a set of trial phases and the expected values of the same invariants based on a ratio of modified Bessel functions (Germain *et al.*, 1970). The associated parameters A_{HK} are defined by

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

where the $|E|$'s are the normalized structure-factor magnitudes and N is the number of atoms, assumed identical, in the unit cell. Although the (unconstrained) global minimum of the minimal function $R(\varphi)$ [equation (1)] was originally thought (incorrectly as it later turned out) to yield the true values of the phases for some choice of origin and enantiomorph (Debaerdemaeker & Woolfson, 1983); it was later recognized (DeTitta *et al.*, 1994) that it is the *constrained* global minimum of $R(\varphi)$ that identifies the solution. The constrained global minimum is equal to the value of the minimal function $R(\varphi)$ when all the phases are equal to their true values for any choice of origin and enantiomorph (the minimal principle). Constraints arise from the existence of identities among the phases, which must of necessity be satisfied, if only approximately. These identities, in turn, are a consequence of the overabundance of available magnitudes $|E|$, usually far more than needed to determine the values of the relatively small number of atomic position vectors r_j. Denoting by $R_{\text{unconstrained}}$ the value of the unconstrained global minimum of $R(\varphi)$ and by $R_{\text{constrained}}$ the value of the constrained global minimum of $R(\varphi)$, it is clear that $R_{\text{unconstrained}}$ is always less than $R_{\text{constrained}}$,

and even unconstrained *relative* minima are often less than $R_{\text{constrained}}$. This observation makes clear the need to avoid the trap of unconstrained local minima, here called false minima, the recognition of which provided the motivation for the initial formulation of *Shake-and-Bake*.

Because of its lack of crystallographic symmetry, $P1$ is a unique space group. Lack of symmetry means that the origin can be located anywhere, and this property may have significant implications with respect to the application of direct phasing methods. Indeed, there have been previous observations that suggest that $P1$ structures require special treatment. In fact, the chances of solving small $P1$ structures by *Shake-and-Bake* are unusually high, especially if the minimal amount of phase refinement is performed in each cycle (Chang *et al.*, 1997). When the *SnB* program was recently applied to data for triclinic hen egg-white lysozyme using default parameters (Deacon *et al.*, 1998), it was observed that virtually all the trial phase sets had values of $R_{\text{unconstrained}}$ that were much less than the corresponding values of $R_{\text{constrained}}$. In addition, Fourier maps computed using these phases exhibited a 'uranium' peak several times larger than any other peak. Once a trial fell into one of these unconstrained local minima, there was no escape no matter how many subsequent *Shake-and-Bake* cycles were performed. In order to avoid such false minima and solve $P1$ lysozyme using *SnB*, it was necessary to change the parameter-shift conditions used to reduce the minimal function. In fact, when a single large parameter-shift angle of about 157° was employed, a surprisingly high percentage (considering that the structure contains more than 1000 unique non-H atoms) of the trial structures converged to solution.

In order to understand more fully the relationship between $P1$ structures and the *Shake-and-Bake* procedure, a study of several such data sets, including three small proteins (alpha-1 peptide, vancomycin and triclinic lysozyme), was undertaken. This study was designed to answer the following questions: Is lysozyme an anomaly or does it behave like a typical $P1$ structure? Can false minima be recognized and routinely avoided? Can computing time be saved by early termination of unproductive trials? What are the best parameter-shift conditions, in general, for $P1$ structures? Finally, the occurrence of false minima in other space groups is discussed.

2. Avoiding the trap

By definition, a false minimum is any value of $R(\varphi)$ that is less than the value of $R(\varphi)$ when all phases are equal to their true values for any choice of origin and enantiomorph. False minima occur frequently in space group $P1$ regardless of which phase-refinement procedure (parameter-shift or tangent-formula) is employed. In view of the minimal principle, a false minimum can occur only when the requisite identities among the phases are not satisfied. Thus, the false minimum cannot serve as a useful figure of merit and it is therefore essential to identify and eliminate those *Shake-and-Bake* trials leading to false minima. Refer to Fig. 1 and recall that each cycle of *Shake-and-Bake* consists of two halves. In the first half (Shake), phases are refined by reducing the values of the minimal function using a parameter-shift technique; in the second half (Bake), the refined phases are used to calculate a Fourier map that is then modified by an appropriate peak-picking protocol, thus leading to a further modification of the phase values. Denote by R_{unc} the value of $R(\varphi)$ after the 'Shake' half cycle and by R_{con} the value of $R(\varphi)$ after the 'Bake' half cycle. Then, R_{unc} and R_{con} depend on the cycle number. After a sufficient number of cycles, the values of R_{unc} and R_{con} approach limiting values, R_{unconstr} and R_{constr}, respectively. Those trials for which $R_{\text{unconstr}} \approx R_{\text{constr}}$ are designated well conditioned trials, and their common value, R_{min}, serves as an effective figure of merit. If solutions are present among a group of well conditioned trials, then those trials with the smallest value of R_{min} are solutions. Those trials for which $R_{\text{unconstr}} \neq R_{\text{constr}}$, the so-called ill-conditioned trials, present the trap of the false minimum because not only is $R_{\text{unconstr}} \ll R_{\text{constr}}$ but R_{unconstr} is also much less than the value of $R(\varphi)$ corresponding to the solution. To summarize, those trials leading to false minima have, in general, the following properties:

(*a*) The value of R_{unconstr} is much less than the value of $R(\varphi)$ when the phases are equal to their true values.

(*b*) The value of R_{constr}, on the other hand, is never significantly less than the value of $R(\varphi)$ when the phases are equal to their true values.

These properties provide clues for avoiding the trap of the false minimum. Define the R-Ratio by means of

$$R\text{-Ratio} = (R_{\text{con}} - R_{\text{unc}})/(R_{\text{con}} + R_{\text{unc}}). \qquad (4)$$

Then the value of the R-Ratio can be used to distinguish false minima from other nonsolution trials as well as the true solutions by the fact that the R-Ratio values for false minima will be much greater than those for either solutions or nonsolutions. Therefore, the R-Ratio can be used as a criterion for early termination of unproductive trials. It is very important to find a default cut-off value for the R-Ratio that will eliminate most false minima without risking rejection of any potential solutions.

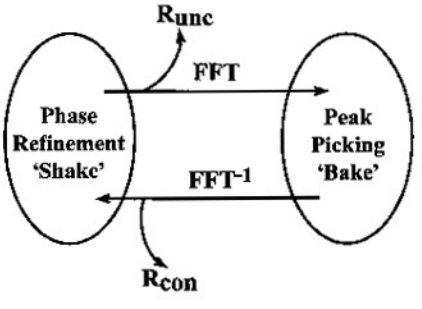

Figure 1
R_{unc} is the value of $R(\varphi)$ after the 'Shake' half cycle, and R_{con} is the value of $R(\varphi)$ after the 'Bake' half cycle.

Table 1
Test data sets used in this investigation.

Structure	Atoms per ASU	'Heavy' atoms	Completeness (%)	d (Å)	Reference
Emerimycin	74	–	98.5	0.91	Marshall *et al.* (1990)
Enkephalin analog	96	–	95.4	0.83	Krstenansky (unpublished)
Alpha-1	471	1Cl	85.6	0.90	Privé *et al.* (1999)
Vancomycin	547	12Cl	80.2	0.97	Loll *et al.* (1998)
Lysozyme	1295	10S	68.3	0.85	Deacon *et al.* (1998)

Table 2
Fixed experimental parameters.

Structure	Phases	Triplets	Peaks	Cycles	Trials
Emerimycin	740	7400	74	50	1000
Enkephalin analog	960	9600	96	100	1000
Alpha-1	4000	40000	300	500	500
Vancomycin	4000	40000	150	500	500
Lysozyme	11100	111000	350	750	500

3. Experiments

The five $P1$ structures listed in Table 1 were tested using *SnB* version 2.0 (Weeks & Miller, 1999*a*). Atomic resolution data sets were available for these structures which range in size from 74 to 1295 non-H atoms in the unit cell. Fixed values of the basic parameters for *SnB* (*i.e.* the numbers of phases, triplet invariants, peaks and refinement cycles) are all dependent on structure size (Weeks & Miller, 1999*b*) and are summarized in Table 2. Error-free data sets were generated using the program *EGEN* (R. Blessing, personal communication) and the known atomic coordinates.

The notation *PS*(90°, 2) denotes the default parameter-shift conditions for optimization of the cosine minimal function (*i.e.* a 90° shift angle applied two times). These conditions were based on a series of previous studies using data sets for known small-molecule structures (Weeks *et al.*, 1994; Chang *et al.*, 1997). A recent application of the *Shake-and-Bake* procedure to triclinic lysozyme (Deacon *et al.*, 1998) demonstrated that a single large shift of 157.5° produced the best results for this structure. Therefore, large single shifts were also tested in this study. The notation COS(S, 1) is used to denote parameter-shift optimization of the cosine minimal function using a single shift of size S. In all cases, each phase was refined only once during a *Shake-and-Bake* cycle.

When performing *post mortem* studies using data for previously known structures, a trial structure subjected to the *Shake-and-Bake* procedure is counted as a solution if there is a close match between the peak positions produced by *Shake-and-Bake* and the true atomic positions for some choice of origin and enantiomorph. A solution in $P1$ is recognized by examining the cosine-invariant figure of merit (Weeks *et al.*, 1995),

$$\mathrm{COSFOM} = \left(\sum_{H,K} A_{HK}\right)^{-1} \sum_{H,K} A_{HK}\left|\cos(\varphi_{HK}) - \cos(\varphi_{HK}^{T})\right|, \quad (5)$$

which is a measure of the average difference between the true cosine invariant values ($\cos\varphi_{HK}^{T}$) and values computed using trial phases. Of course, in actual applications to unknown structures, potential solutions are identified on the basis of minimal function values. In this study, the single-shift procedure (COS) and the double-shift procedure (PS) are compared on the basis of the following two criteria:

(i) *Success rate* (SR) is defined as the percentage of trial structures that go to solution, and the standard deviation of success rate is calculated using Bernoulli's distribution, where $\sigma_{\mathrm{SR}} = (npq)^{1/2}$ with n being the number of trials, p being the success rate expressed as a fraction and q being the failure rate. The measurement of success rates at the end of a fixed number of cycles provides one important indication as to the quality of a particular refinement method. However, this measurement by itself provides an incomplete comparison since it does not take into account the computational effort (running time) needed to produce the solutions.

(ii) The relative efficiency of different procedures can be compared on the basis of the *cost effectiveness* (CE),

$$\mathrm{CE} = 3600S/T\bar{C}t, \quad (6)$$

where T is the number of trial structures, $\bar{C}$ is the average number of cycles per trial structure (with early termination) or the default number of cycles (without early termination), S is the number of solutions produced by T such trials, and t is the running time (in s) for one cycle of one trial. In this communication, CE has units of solutions per hour on a 195 MHz Silicon Graphics R10000 Indigo workstation. All experiments were conducted either on a network of SGI R10000 workstations at the Hauptman–Woodward Medical Research Institute, or on an IBM SP2 at the Center for Computational Science and Technology (CCST) at Argonne National Laboratory.

4. Results

4.1. Recognition and early termination of false minima

Fig. 2 illustrates R_{unc}, R_{con} and R-Ratio as functions of *SnB* cycle for different types of trials for the three small protein structures listed in Table 1. Based on the results presented in this figure, a good default cut-off value for the R-Ratio appears to be 0.2. This value distinguishes false minima from both solutions and conventional nonsolutions for all three structures. The *SnB* program calculates the R-Ratio value at the end of each cycle, and may (optionally) terminate the

current trial as soon as R-Ratio exceeds the specified cut-off value. Such early termination will minimize the computational effort expended on hopeless trials without risking rejection of any potential solutions. In the case of triclinic lysozyme, false-minimum trials could be recognized and terminated, on the average, by cycle 25. Since the default recommendation for the number of cycles is 1000, a substantial saving in CPU time could be realized by utilizing the R-Ratio early termination test.

4.2. Optimal parameter-shift angles

Table 3 summarizes success rates and percentages of false minima obtained with the cosine minimal function for the five $P1$ structures using several different single-shift angles. The following are observed:

(a) A high percentage of false minima occurs for medium or large $P1$ structures.

(b) The optimal shift angle occurs at or near the angle where the percentage of false minima falls to approximately zero.

(c) The optimal shift angle depends on the size of the structure.

(d) For medium and large structures, the success rate for the optimal single-shift angle significantly exceeds that for the default double 90° shift which appears to be best suited for smaller structures.

The results, presented in Table 3, suggest that a fruitful strategy for structures exhibiting a large percentage of false minima (*i.e.* R-Ratio > 0.2) is to run 100 or so trials at each of several shift angles in the range [90°, 180°], find the smallest angle that gives nearly zero false minima, and then use this angle as a single shift for many trials. This assumes, of course, that a solution has not already been found while varying the shift angle. As an application of this strategy, 100 vancomycin trials were tested at each of the shift angles 90°, 100°, ..., 180°. None of these shift angles produced a solution, but 120° was the smallest shift angle that gave almost no false minima. Next, 1000 trials were completed using 120° as the single shift angle, and a 1.1% success rate was obtained. This result was superior to any vancomycin success rate listed in Table 3.

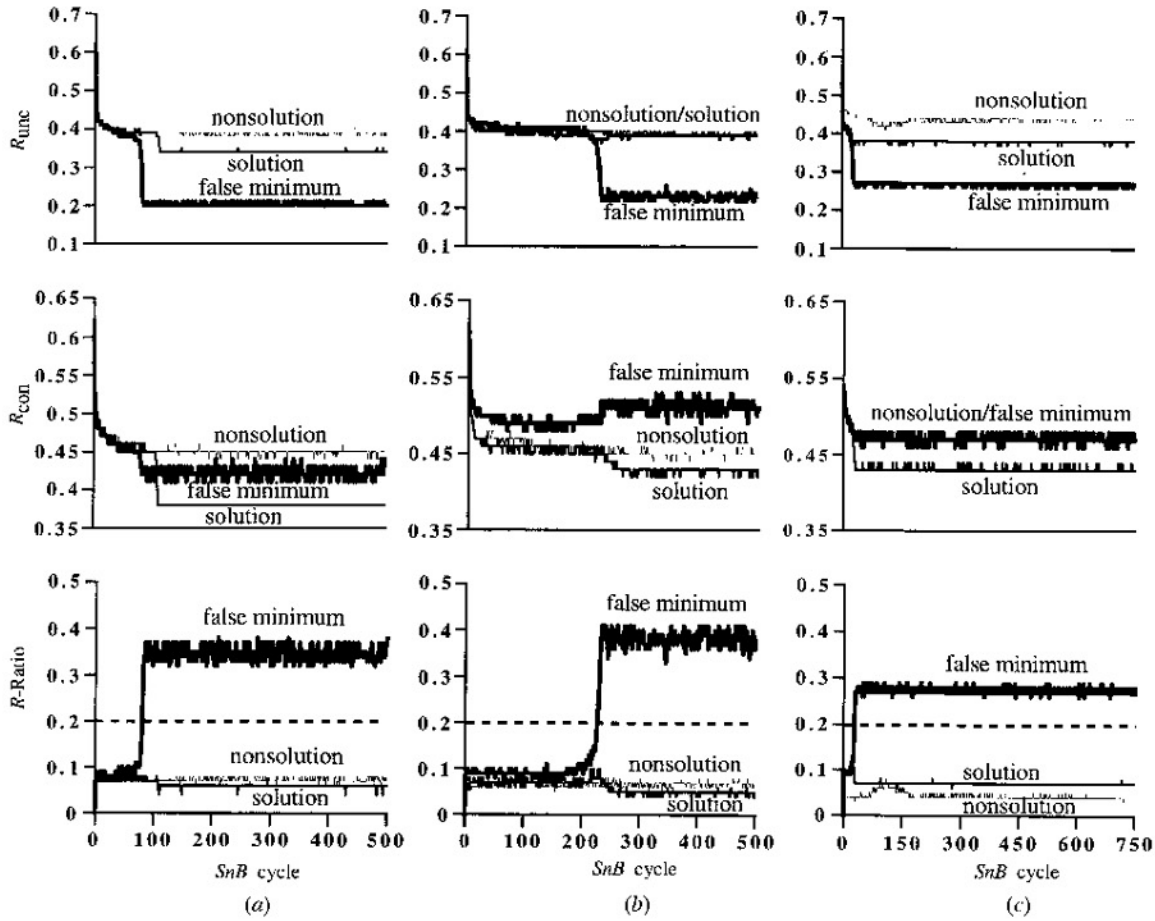

Figure 2
The values of R_{unc}, R_{con} and R-Ratio as a function of *SnB* cycle for solutions, nonsolutions and false minima. The broken line shows how a cut-off value of 0.2 for the R-Ratio can distinguish false minima from other trials. (a) Alpha-1; (b) vancomycin; (c) lysozyme.

Table 3
Correlation of optimum parameter-shift angle with percentage of false minima for five *P*1 structures.

Bold type indicates maximum success rate and corresponding percentage of false minima.

Structure	Default PS(90°, 2)	Cosine single parameter-shift size COS(*S*, 1) 22.5°	45.0°	67.5°	90.0°	112.5°	135.0°	157.5°	180.0°
(*a*) Success rate (%)									
Emerimycin	63.1	2.9	38.1	58.0	**61.6**	42.2	21.7	6.7	2.2
Enkephalin analog	42.7	0.1	14.1	35.1	**35.7**	24.4	8.8	2.7	0.5
Alpha-1	13.7	0.0	0.0	2.8	15.8	**19.8**	8.7	3.3	0.2
Vancomycin	0.05	0.0	0.0	0.10	0.05	**0.25**	0.10	0.0	0.0
Lysozyme	0.0	0.0	0.6	0.2	0.0	0.0	0.8	**13.5**	1.6
(*b*) False minima (%)									
Emerimycin	3.7	0.0	0.0	5.3	**0.0**	0.0	0.0	0.0	0.0
Enkephalin analog	4.6	0.0	0.0	4.2	**1.8**	0.0	0.0	0.0	0.0
Alpha-1	28.7	0.0	2.0	21.0	29.0	**2.0**	0.0	0.0	0.0
Vancomycin	98.8	0.0	28.5	93.0	98.8	**29.6**	0.0	0.0	0.0
Lysozyme	100	0.0	27.0	99.8	100	100	99.0	**0.0**	0.0

Table 4
Comparison of cost effectiveness (solutions h^{-1}) using default double 90° shifts (with or without early termination) and using the optimal single shift.

Bold type indicates optimal cost effectiveness.

	Default *PS*(90°, 2)				Optimum *COS*(*S*°, 1)		
	Do all cycles		Early termination		Do all cycles		
Structure	Cycles	Solutions h^{-1}	(Cycle)	Solutions h^{-1}	*S*°	Cycles	Solutions h^{-1}
Alpha-1	500	0.89	416	1.07	110	500	**1.15**
Vancomycin	500	0.005	116	0.020	120	500	**0.073**
Lysozyme	750	0	25	0	150	750	**0.26**

It is observed that the high percentage of false minima for medium and large *P*1 structures is centered at 90°, which is also the default shift angle for small *P*1 structures and non *P*1 structures. Thus, using the *R*-Ratio early termination test at shift angle 90° will save substantial CPU time and increase cost effectiveness for medium and large *P*1 structures. However, the results presented in Table 4, based on the cost effectiveness (solutions h^{-1}) both for the default double-parameter-shift conditions [PS(90°, 2)] and for optimal single-parameter-shift conditions [COS(*S*, 1)], show that the optimal single-parameter-shift is more efficient than the default double 90° shift coupled with early termination. Although additional computation is needed to find the optimal shift angle, it appears that one would never solve the lysozyme structure under the default double 90° shift conditions.

Using the tangent formula (rather than parameter-shift reduction of the minimal function) as the means for phase refinement did not alleviate the trap of the false minimum for these structures. In all, 500 trials were tested for each of the experimental data sets. In the case of alpha-1, 94.4% of trials went to false minima, and all trials went to false minima for vancomycin and lysozyme. The values of $R_{\rm unc}$ for false minima approached zero and the *R*-Ratio values for false minima were greater than 0.9.

4.3. Effects of measurement errors and data completeness

The relative effects of data accuracy and data completeness on *Shake-and-Bake* success rates and false minima frequency using the cosine minimal function were studied for the three large *P*1 structures. The results of these studies, which are based on at least 100 trials at each of the single-shift angles of 10°, 20°, ..., 180°, are presented in Fig. 3 and show the following:

(*a*) There are strong negative correlations between success rates and the frequency of false minima. Regardless of data accuracy or data completeness, the highest success rate occurs at or near the angle where the percentage of false minima falls to approximately zero.

(*b*) Experimental error appears to be sufficiently small that it has no effect on the success rates for alpha-1 and vancomycin. However, experimental error has a significant impact on the success rate for lysozyme, particularly at small shift angles.

(*c*) Using complete error-free data results in a substantial increase in success rates for all three structures. An astounding success rate of 100% is achieved for lysozyme (~1300 atoms)! In direct-methods applications, the adverse effects of missing data may be magnified because of the number of triplet invariant relationships that become unavailable for use.

Table 5
Optimum shift angle and success rate for various *P*1 data sets.

In each case, 500 trials were tested.

	Alpha-1		Vancomycin		Lysozyme	
No. of atoms	471		547		~1200	
Completeness (%)	85.6		80.2		68.3	
Resolution (Å)	0.90		0.97		0.85	
Data	Angle (°)	SR (%)	Angle (°)	SR (%)	Angle (°)	SR (%)
Experimental	110	19.8 (18)	120	1.4 (6)	150	25.6 (2)
Error-free	110	17.2 (17)	120	0.2 (2)	140	47.2 (22)
Error-free complete	110	30.4 (21)	120	19.4 (17)	120	99.0 (5)

The optimal shift angle for each data set was identified for each structure based on the results presented in Fig. 3. 500 trials for each data set were tested using the optimal shift angle and the results, presented in Table 5, show that data accuracy and completeness had significant impact on the value of the optimal shift angle for lysozyme, but almost no effect on the optimal angle for alpha-1 and vancomycin.

The experimental vancomycin data did not include any data at 10 Å resolution or lower. A total of 4000 reflections from the experimental data were phased in the dual-space loop of *Shake-and-Bake* using the optimal single-shift angle of 120°. Some of these data were then replaced with the largest error-free magnitudes chosen from among the missing reflections at several different resolution limits. The results based on 1000 trials, presented in Table 6, show an eightfold increase in success rate when only 400 of the largest missing magnitudes with a maximum resolution of 1.3 Å were supplied. This simple example illustrates the importance of data completeness for *Shake-and-Bake*.

5. Discussion

The results presented above illustrate that false minima are a significant occurrence during *Shake-and-Bake* applications to large *P*1 data sets. These minima occur regardless of whether the tangent formula or parameter-shift reduction of the minimal-function value is used as the phase-refinement mechanism. Fortunately, however, the frequency of such minima can be reduced by a judicious choice of parameter-shift conditions, namely through use of a single large shift of 120° or more. Thus, solving larger *P*1 structures may require parameter-shift conditions different from the standard

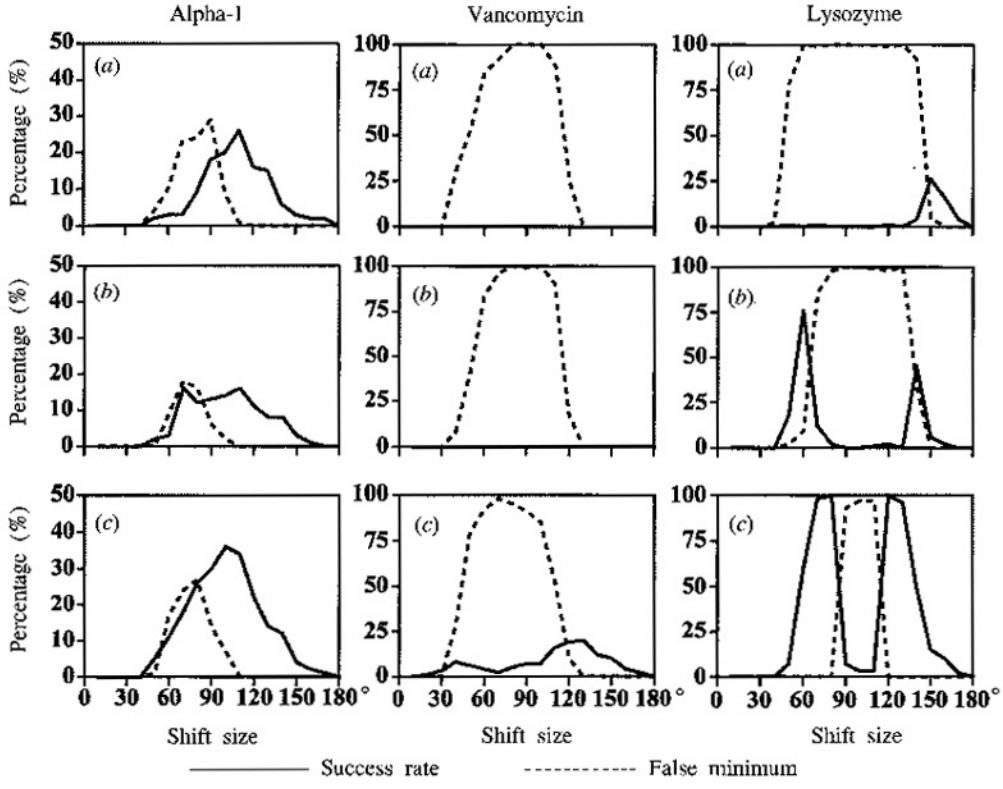

Figure 3
Success rate and frequency of false minima as a function of single-shift size for various data sets. Each point represents 100 trials. (*a*) Experimental; (*b*) error-free; (*c*) error-free complete.

Table 6
Improving success rates by 'completing' the vancomycin data.

In each case, 1000 trials were tested

Error-free reflections added	0	200 (2.8 Å)	300 (2.0 Å)	400 (1.3 Å)	800 (1.1 Å)	4000 (0.97 Å)
Success rate (%)	1.1 (3)	2.1 (5)	2.8 (5)	8.4 (9)	11.3 (10)	18.8 (12)

conditions [PS(90°, 2)] that have been found to work well for structures in other space groups as well as for small structures in space group $P1$. Version 2.0 of the *SnB* program (Weeks & Miller, 1999*a*) automatically chooses an appropriate single-shift angle for large $P1$ structures. In addition, false minima can be recognized on the basis of a function (R-Ratio) of the unconstrained (computed immediately after phase refinement) and constrained (computed after peak picking and structure-factor calculation) minimal-function values. Termination of trial structures failing the R-Ratio test during early *Shake-and-Bake* cycles leads to increased computational efficiency. In most cases, the E maps resulting from false minima are dominated by a single large peak, and the height ratio of the top two peaks (P-Ratio test) can also be used as a criterion for early termination.

False minima have been observed in other space groups such as $P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_1/c$ and $C222$, but the percentage of such occurrences is generally quite low and poses no difficulty for the experimenter provided that R_{con} rather than R_{unc} is used as the figure of merit. *SnB* v2.0 contains a provision for creating an excluded volume about special positions by selectively eliminating all peaks in such regions. This feature provides a powerful real-space constraint for preventing the occurrence of false minima because, if it is not used, there is a tendency for density to build up in such locations leading to a 'uranium' peak. A typical example of this phenomenon was provided by the selenium substructure of homocysteine hydrolase (Turner *et al.*, 1998), which crystallizes in space group $C222$. In a *post mortem* application of *SnB* to the 1,3-bis(heptaphenyl-1-naphthyl)benzene structure of Tong and co-workers (Tong *et al.*, 1997), which crystallizes in $P\bar{1}$, 34 out of 5000 trial structures went to false minima, but all of them were successfully recognized by the R-Ratio test. On the other hand, only 26 of the 34 false minima were identified by the P-Ratio test. Thus, R-Ratio appears to be a more robust indicator for false minima.

An obvious question is: What would be the frequency of false minima if non $P1$ structures were treated as if they were $P1$ structures? To answer this, the default parameter-shift conditions [PS(90°, 2)] were applied to 84-atom isoleucinomycin ($P2_12_12_1$) and 327-atom crambin ($P2_1$) structures. The percentage of false minima increased from 0% (in $P2_12_12_1$) to 24.9% (in $P1$) for isoleucinomycin and from 27.5% (in $P2_1$) to 99.4% (in $P1$) for crambin. All false minima in $P1$ could be recognized by the R-Ratio test in both cases. The results presented above for large $P1$ structures also indicate that the success rate can be anomalously high in this space group. This observation suggests that it might be advantageous to determine all structures with the data expanded to $P1$, and then to locate the symmetry elements afterwards. However, this is more computationally expensive than performing the whole procedure in the true space group, and, in practice, is only competitive in low-symmetry space groups such as $P2_1$, $C2$ or $P\bar{1}$ (Chang *et al.*, 1997; Hauptman *et al.*, 1999). Exceptions to this generalization can occur when expansion to $P1$ also offers opportunities for starting from 'slightly better than random' phases. This possibility was successfully demonstrated by Sheldrick & Gould (1995) who used a rotation search with a small fragment to generate many sets of starting phases. After expansion to $P1$, the translational search usually required for molecular replacement is not needed.

We would like to express our appreciation to Thomas Tang, Andrew Palumbo and Jan Pevzner for assisting in the development of *SnB* v2.0 and to Melda Tugac for designing figures. We also thank the Center for Computational Science and Technology (CCST), Mathematics and Computer Science Division, Argonne National Laboratory, for allowing us to use their High Performance Computing Research Facilities. This research was supported by NIH grant GM-46733 and NSF grant ACI-9721373.

References

Bhuiya, A. K. & Stanley, E. (1963). *Acta Cryst.* **16**, 981–984.
Chang, C.-S., Weeks, C. M., Miller, R. & Hauptman, H. A. (1997). *Acta Cryst.* A**53**, 436–444.
Deacon, A. M., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.
Debaerdemaeker, T. & Woolfson, M. M. (1983). *Acta Cryst.* A**39**, 193–196.
DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.
Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B**26**, 274–285.
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. A., Xu, H., Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* A**55**, 891–900.
Karle, J. & Hauptman, H. A. (1956). *Acta Cryst.* **9**, 635–651.
Loll, P. J., Miller, R., Weeks, C. M. & Axelsen, P. H. (1998). *Chem. Biol.* **5**, 293–298.
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., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Prive, G. G., Anderson, D. H., Wesson, L., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1400–1409.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Tong, L., Ho, D. M., Vogelaar, N. J., Schutt, C. E. & Pascal, R. A. Jr (1997). *J. Am. Chem. Soc.* **119**, 7291–7302.
Turner, M. A., Yuan, C.-S., Borchardt, R. T., Hershfield, M. S., Smith, G. D. & Howell, P. L. (1998). *Nature Struct. Biol.* **5**, 369–375.
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.
Weeks, C. M. & Miller, R. (1999*a*). *J. Appl. Cryst.* **32**, 120–124.
Weeks, C. M. & Miller, R. (1999*b*). *Acta Cryst.* D**55**, 492–500.