Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Sine-enhanced *Shake-and-Bake*: the theoretical basis and applications to Se-atom substructures

Hongliang Xu,* Herbert A. Hauptman and Charles M. Weeks

Hauptman–Woodward Medical Research Institute and Department of Structural Biology, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, 73 High Street, Buffalo, NY 14203 USA

Correspondence e-mail: xu@hwi.buffalo.edu

Received 12 July 2001
Accepted 17 October 2001

Shake-and-Bake is a dual-space direct-methods procedure for crystal structure determination capable of providing *ab initio* solutions for structures containing as many as 1200 independent non-H atoms, as well as for heavy-atom substructures containing as many as 160 Se atoms in the asymmetric unit. In traditional *Shake-and-Bake*, phase refinement in reciprocal space utilizes the technique of parameter shift to reduce the value of a minimal function that considers only the mean-square differences between the current values of the cosine structure invariants and their expected values. A new type of minimal function, termed the sine-enhanced minimal function, considers both cosine and sine values of the structure invariants. Exhaustive tests on six Se-atom substructures, ranging in size from 12 to 160 Se atoms in the asymmetric unit, have shown that a two- to eightfold increase in the percentage of trials that converge to solution is attainable with the technique of sine-enhanced parameter shift. The corresponding sine-enhanced *Shake-and-Bake*, with suitable default parameter values, is being incorporated into a new distributed version of the *SnB* computer program.

1. Introduction

Shake-and-Bake (Weeks *et al.*, 1994) is a multisolution or multitrial direct-methods procedure that automatically and repetitively alternates phase refinement in reciprocal space with peak picking in real space to impose constraints through a physically meaningful interpretation of the electron density. Typically, the phase-refinement portion of the *Shake-and-Bake* cycle utilizes the technique of parameter shift (Bhuiya & Stanley, 1963; Chang *et al.*, 1997) to reduce the value of the minimal function (Debaerdemaeker & Woolfson, 1983; Hauptman, 1991; DeTitta *et al.*, 1994). The *Shake-and-Bake* method, as implemented in the computer program *SnB* (Weeks & Miller, 1999*a*), has successfully provided *ab initio* solutions for structures containing as many as 1200 independent non-H atoms (Deacon *et al.*, 1998) as well as for large substructures such as the 160-site selenomethionine derivative of ketopantoate hydroxymethyltransferase from *E. coli* (F. von Delft, personal communication).

If **H** is an arbitrary reciprocal-lattice vector, then the normalized structure factor E_H is defined by

$$E_H = |E_H| \exp(i\varphi_H) = N^{-1/2} \sum_{j=1}^{N} \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j), \qquad (1)$$

where N is the number of atoms, here assumed for simplicity to be identical, in the unit cell and $\mathbf{r}_j$ is the position vector of

the atom labeled *j*. For every pair of reciprocal-lattice vectors (**H**, **K**), the structure invariant (triplet) φ_{HK} and its associated parameter A_{HK} are defined by means of

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

The conditional probability distribution $P(\varphi|A_{HK})$ of the triplet φ_{HK} given the A_{HK} is known to be

$$P(\varphi|A_{HK}) = [2\pi I_0(A_{HK})]^{-1} \exp(A_{HK}\cos\varphi) \quad (3)$$

and the expected value of $\cos\varphi$ is

$$\langle\cos\varphi\rangle = I_1(A_{HK})/I_0(A_{HK}), \quad (4)$$

where I_0 and I_1 are modified Bessel functions (Cochran, 1955).

The phase problem can be formulated as a problem in constrained global minimization. The commonly used cosine minimal function (DeTitta *et al.*, 1994; Weeks *et al.*, 1994)

$$R(\varphi) = \left(\sum_{H,K} A_{HK}\right)^{-1} \sum_{H,K} A_{HK}\left[\cos(\varphi_{HK}) - \frac{I_1(A_{HK})}{I_0(A_{HK})}\right]^2 \quad (5)$$

measures the mean-square difference between the current values of the cosine structure invariants, $\cos(\varphi_{HK})$, and their expected values. The minimal function reaches its constrained global minimum when the phases are equal to their true values no matter what the choice of origin or enantiomorph (the traditional minimal principle). This principle is the theoretical basis of traditional *Shake-and-Bake*.

The cosine minimal function uses only information about the cosine structure invariants. However, owing to the relationship

$$\cos^2\varphi + \sin^2\varphi = 1,$$

it follows that

$$\sin\varphi = \pm(1 - \cos^2\varphi)^{1/2}$$

and the expected value of $\sin\varphi$ is approximated by

$$\langle\sin\varphi\rangle = \pm(1 - \langle\cos\varphi\rangle^2)^{1/2} = \pm[1 - I_1^2(A_{HK})/I_0^2(A_{HK})]^{1/2}. \quad (6)$$

If the correct signs could be selected, the enantiomorph of the targeted structure would be fixed. The values of the three-phase structure invariants (triplets) would then be uniquely determined and the values of the individual phases could be easily estimated. In this paper, we propose to incorporate information about the sine invariants, $\sin(\varphi_{HK})$, into the minimal function. To do so, we first define $\varepsilon_{HK} = \pm 1$ accordingly as $\sin\varphi_{HK} \gtrless 0$, respectively, and then define a new minimal function, termed the sine-enhanced minimal function, by means of

$$m(\varphi;\varepsilon) = \left(2\sum_{H,K} A_{HK}\right)^{-1} \sum_{H,K} A_{HK}\left(\left[\cos\varphi_{HK} - \frac{I_1(A_{HK})}{I_0(A_{HK})}\right]^2 + \left\{\sin\varphi_{HK} - \varepsilon_{HK}\left[1 - \frac{I_1^2(A_{HK})}{I_0^2(A_{HK})}\right]^{1/2}\right\}^2\right) \quad (7)$$

(where the factor 2 in the divisor is to ensure that the value of the minimal function always lies between 0 and 1). We note that reversing the sign of ε is equivalent to switching enantiomorphs. The sine-enhanced minimal function measures the sum of the mean-square differences of the cosine and the sine invariants from their expected values, respectively. It is expected to have a constrained global minimum when the phases and the εs are equal to their true values no matter what the choice of origin or enantiomorph (sine-enhanced minimal principle). This minimal principle provides the theoretical basis for a sine-enhanced *Shake-and-Bake* procedure. Unlike the cosine minimal function, the sine-enhanced minimal function treats both phases φ and ε as unknown variables. It is anticipated that with proper modifications to the parameter-shift procedure, the refinement of the εs in (7) would lead to the resolution of the twofold phase ambiguity, thus strengthening the ability to determine phases *ab initio*. This expectation is, in fact, realised here. Specifically, a sine-enhanced minimal function and its minimal principle are formulated and the corresponding sine-enhanced *Shake-and-Bake* and its associated parameter-shift procedure, here termed sine-enhanced parameter shift, are proposed. The initial sine-enhanced *Shake-and-Bake* applications to anomalous difference data for six Se-atom substructures, ranging in size from 12 to 160 Se atoms in the asymmetric unit, demonstrate that a two- to eightfold increase in the percentage of trials that converge to solution is attainable with proper choice of parameters such as the size of the shift angle.

1.1. Parameter shift (PS)

In traditional *Shake-and-Bake*, the phases are the only unknown variables. Phase refinement in reciprocal space utilizes the technique of parameter shift to reduce the value of the cosine minimal function (5). The phases are sorted in decreasing order with respect to the values of the associated $|E|$s and initial values of phases are calculated based on trial structures with randomly positioned atoms. When considering a given phase φ_H, the value of the cosine minimal function (5) is initially evaluated three times: firstly with the current values of the phases, secondly with phase φ_H modified by the addition of the predetermined phase shift (shift size) and thirdly with phase φ_H modified by the subtraction of the predetermined phase shift. If the first evaluation yields the minimum of these three values of the cosine minimal function, then consideration of φ_H is complete and parameter shift proceeds to the next phase. Otherwise, the direction of search is determined by the modification that yields the minimal value and the phase is updated to reflect that modification. In this case, phase φ_H continues to be updated by the predetermined phase shift in the direction just determined as long as the value of the minimal function continues to be reduced, although there is a user-defined predetermined maximum number of times (steps) that the shift is attempted. Refined phase values are used immediately in the subsequent refinement of other phases. The notation PS(S, m) is used to denote the parameter-shift optimization of the cosine minimal function using

shift size S and a maximum of m steps. Based on extensive experimentation involving a variety of structures in several space groups, it has been confirmed, in terms of running time and the percentage of trial structures that produce a solution, that (i) PS(180°, 1) is optimum for centrosymmetric space groups (Weeks *et al.*, 1994), (ii) PS(S°, 1) is optimum for space group $P1$ where S° is related to the size of the structure (Hauptman *et al.*, 1999) and (iii) PS(90°, 2) is optimum for all other space groups (Weeks & Miller, 1999*b*).

1.2. Sine-enhanced parameter shift (SEPS)

The parameter-shift procedure needs to be modified to accommodate the fact that both phases φ and εs are unknown variables in the sine-enhanced minimal function (7). For each fixed reciprocal-lattice vector **H**, we define two partial minimal functions

$$m_1(\varphi_H) = \sum_K A_{HK}\left(\left[\cos\varphi_{HK} - \frac{I_1(A_{HK})}{I_0(A_{HK})}\right]^2 + \left\{\sin\varphi_{HK} - \varepsilon_{HK}\left[1 - \frac{I_1^2(A_{HK})}{I_0^2(A_{HK})}\right]^{1/2}\right\}^2\right) \quad (8)$$

and

$$m_2(\varphi_H) = \sum_K A_{HK}\left(\left[\cos\varphi_{HK} - \frac{I_1(A_{HK})}{I_0(A_{HK})}\right]^2 + \left\{\sin\varphi_{HK} + \varepsilon_{HK}\left[1 - \frac{I_1^2(A_{HK})}{I_0^2(A_{HK})}\right]^{1/2}\right\}^2\right), \quad (9)$$

where starting values of ε_{HK} are assigned as described in §2. The sine-enhanced parameter-shift optimization of the sine-enhanced minimal function involves three steps. (i) Firstly, parameter-shift optimization PS(S, m) is performed on the partial minimal function $m_1(\varphi_H)$ and the corresponding optimal phase $\varphi_H^\star$ is obtained. (ii) Next, the values of $m_1(\varphi_H^\star)$ and $m_2(\varphi_H^\star)$ are calculated and $m^\star = \min[m_1(\varphi_H^\star), m_2(\varphi_H^\star)]$. (iii) Finally, another parameter-shift optimization PS(S, m) is performed on the partial minimal function $m_2(\varphi_H)$ using $\varphi_H^\star$ and $m^\star$ as initial phase and initial minimum, respectively, and the optimal phase $\varphi_H^{\star\star}$ is obtained. The sine-enhanced parameter-shift for φ_H is then complete and the optimization procedure proceeds to the next phase. Again, refined phase values are used immediately in the subsequent refinement of other phases. The notation SEPS(S, m) is used to denote the sine-enhanced parameter-shift optimization of the sine-enhanced minimal function using shift size S and a maximum of m steps. Fig. 1 illustrates optimization of φ_H using SEPS(90°, 2) as an example.

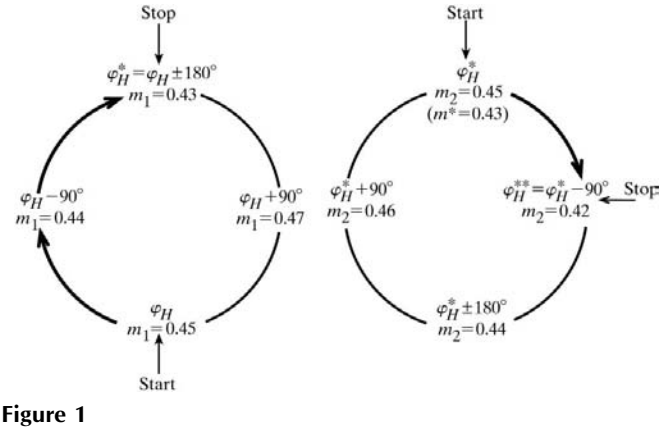

Figure 1
Example of sine-enhanced parameter-shift optimization, SEPS(90°, 2). First, PS(90°, 2) is performed on the left circle, resulting in an optimal phase $\varphi_H^\star = \varphi_H \pm 180°$. Since $m_1(\varphi_H^\star) = 0.43$ and $m_2(\varphi_H^\star) = 0.45$, $m^\star = 0.43$. Another PS(90°, 2) is then performed on the right circle using $\varphi_H^\star$ and $m^\star$ as initial phase and minimum, respectively, and the optimal phase $\varphi_H^{\star\star} = \varphi_H^\star - 90° = \varphi_H + 90°$ is obtained.

Table 1
Se-atom anomalous difference substructure data sets used in this investigation.

Asterisks indicate the number of sites sought in the original determination.

Substructure	Se atoms	Space group	d (Å)	Reference
CTLA4	12	$P4_32_12$	2.97	Ostrov *et al.* (2000)
AdoHcy	30	$C222$	2.80	Turner *et al.* (1998)
PDHC E1	40 (42*)	$P2_1$	3.50	Arjunan *et al.* (2000)
AEPT	66	$P2_1$	2.55	Chen *et al.* (2000)
Epimerase	70	$P2_1$	2.91	Deacon *et al.* (2000)
KPHMT	160 (180*)	$P2_1$	3.50	von Delft (personal communication)

2. Materials and methods

In this study, alternative computational procedures are compared on the basis of two criteria. The success rate (SR) is defined as the percentage of trial structures that go to solution and cost effectiveness (CE) is defined by

$$CE = \frac{3600B}{TCt}, \quad (10)$$

where T is the number of trial structures, C is the number of cycles per trial structure, B 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 Silicon Graphics R10000 Indigo workstation. All experiments were conducted either (i) on a network of SGI R10000 workstations at the Hauptman–Woodward Medical Research Institute or (ii) on an SGI Origin 2000 or an IA32 Linux Cluster at the Center for Computational Research (CCR), State University of New York at Buffalo. 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. Of course, in actual applications to unknown structures, potential solutions are identified on the basis of minimal function values.

Both the cosine minimal function and the sine-enhanced minimal function were applied to six known Se-atom substructures using a modification to version 2.1 of the computer program *SnB* (Weeks & Miller, 1999*a*). Basic information concerning the anomalous difference data sets is listed in Table 1. A sample of 1000 (for the first five

Table 2
Values of experimental parameters.

Substructure	Phases	Triplets	Peaks	Cycles
CTLA4	360	3600	12	24
AdoHcy	600	6000	30	60
PDHC E1	1260	12600	42	84
AEPT	1980	19800	66	132
Epimerase	1400	14000	70	140
KPHMT	3600	36000	144	360

Table 3
Number of solutions per 1000 *SnB* trial structures for protocols P1–P5 applying C&S(S, 2, 2) to the Se-atom substructure PDHC E1.

S (°)	P1	P2	P3	P4	P5
10	0	0	0	1	0
20	0	0	0	3	10
30	0	0	0	14	38
40	0	0	0	19	97
50	0	0	0	35	139
60	0	0	0	51	167
70	0	0	0	88	173
80	0	0	0	127	182
90	0	0	110	134	166
100	0	0	102	83	109
110	0	0	51	46	55
120	0	0	29	28	32
130	0	0	21	28	23
140	0	0	18	8	8
150	0	0	6	9	8
160	0	0	4	2	3
170	0	0	1	0	1

substructures) or 10 000 (for KPHMT) randomly positioned N_μ-atom trial structures (where N_μ is the number of independent Se atoms in the asymmetric unit) was generated for each data set. For each substructure, the default values of the important size-dependent *SnB* parameters are summarized in Table 2. Unless specified otherwise, these values were used in the experiments reported here.

The notations COS(S, m, k) and C&S(S, m, k) are used to denote parameter-shift (PS) optimization of the cosine minimal function or sine-enhanced parameter-shift (SEPS) optimization of the sine-enhanced minimal function, respectively, using shift size S, a maximum of m shifts or steps and k iterations (passes through the phase set) of phase refinement per *Shake-and-Bake* cycle. Based on previous studies of small proteins (Weeks & Miller, 1999*b*) and small Se-atom substructures (Howell *et al.*, 2000), COS(90°, 2, 3) was chosen as the default condition for non-$P1$ (sub)structures. However, in this study, COS(S, 2, 3) with S = 10, 20, ..., 160, 170° was applied to provide a basis for comparison with the results of the sine-enhanced minimal function. The latter was applied in the form C&S(S, 2, k) with S = 10, 20, ..., 160, 170° and k = 1, 2, 3, 4 to investigate the effects of varying the shift angle and the number of iterations.

In the case of the sine-enhanced minimal function, it is necessary to address the issue of assigning the initial values of ε_{HK} for every pair of lattice vectors (**H**, **K**). The following series of experiments document the actual steps that lead us to a successful outcome: an improved *Shake-and-Bake* procedure utilizing the sine-enhanced method. These experiments involve five different protocols (P1–P5) that vary the dependence of ε_{HK} on φ_{HK}.

P1: for each *SnB* cycle, initially assign $\varepsilon_{HK} = \pm 1$ according as $\sin(\varphi_{HK}) \gtrless 0$ and update ε_{HK} during each *SnB* iteration if the optimal phase is obtained on the right circle (Fig. 1).

P2: for each *SnB* cycle, assign $\varepsilon_{HK} = \pm 1$ according as $\sin(\varphi_{HK}) \gtrless 0$ and hold them fixed for all *SnB* iterations.

P3: for each *SnB* trial structure, assign $\varepsilon_{HK} = \pm 1$ according as $\sin(\varphi_{HK}) \gtrless 0$ and hold them fixed for all *SnB* cycles.

P4: for each *SnB* trial structure, assign $\varepsilon_{HK} = \pm 1$ randomly and hold them fixed for all *SnB* cycles.

P5: for each *SnB* cycle, assign $\varepsilon_{HK} = \pm 1$ randomly and hold them fixed for all *SnB* iterations.

Table 4
Number of solutions per 1000 *SnB* trial structures using protocol P3 with C&S(S, 2, 2) for Se-atom substructure PDHC E1.

Identified by	82°	83°	84°	85°	86°	87°	88°	89°	90°
Post mortem	0	4	4	25	42	91	104	100	110
Minimal function	0	0	0	0	0	0	0	100	110

3. Results

The initial sine-enhanced parameter-shift experiment was undertaken to compare the efficacy of the five protocols using medium-sized PDHC E1 (40 Se sites in the asymmetric unit) as the test structure. Table 3 summarizes the number of solutions per 1000 *SnB* trial structures for each protocol employing C&S(S, 2, 2) with various shift angles. These results can be summarized as follows.

(i) Protocols P1 and P2 produced no solutions.

(ii) Protocol P3 produced solutions only for large shift angles.

(iii) Protocol P5 outperformed the other four protocols; it had a wide range of shift angles that yielded large numbers of solutions.

Furthermore, protocol P3 shows a sudden increase in the number of solutions, going from no solution at 80° to 110 solutions at 90°. Results for this protocol are shown in greater detail in Table 4. It is observed that although the number of solutions identified by *post mortem* analysis gradually increases as shift angle increases, solutions can be identified by the values of the minimal function only when $S \geq 89°$. The evidence presented in Tables 3 and 4 emphasizes the need to treat ε and φ as independent variables during sine-enhanced parameter shift. The less the dependence between ε_{HK} and φ_{HK}, the higher the number of solutions. Therefore, P5 was chosen as the default protocol for assigning the initial values of the εs and it was used throughout the remainder of these experiments. The sine-enhanced parameter-shift procedure alternately refines φ and ε in such a way that the refinement of

φ strengthens the ability to select the correct value of ε. This, in turn, enhances the enantiomorph separation and further improves the phase refinement.

Table 5 lists the number of solutions per 1000 *SnB* trial structures (10 000 for KPHMT) produced by COS(S, 2, 3) for various shift angles using the default *SnB* parameters given in Table 2. It provides a basis for comparing the results of the traditional minimal function with those of the sine-enhanced minimal function and the data shown here reconfirm that $S = 90^\circ$ is a good choice of default shift angle for the cosine minimal function. Fig. 2 illustrates the number of solutions per 1000 *SnB* trial structures (10 000 for KPHMT) produced by C&S(S, 2, k) as a function of shift angle S and number of iterations k. The family of four curves shows the results for various numbers of phase-refinement iterations (k = 1, 2, 3, 4). The following can be observed from Fig. 2.

(i) The optimal number of iterations, in terms of success rate and computing time, is $k = 3$ for small substructures (CTLA4, AdoHcy), $k = 2$ for medium substructures (PDHC E1, AEPT) and $k = 1$ for large substructures (epimerase, KPHMT).

(ii) With the optimal number of iterations, the highest success rate of each substructure occurs when the shift angle is between 50° and 100°.

(iii) Appropriate choice of shift angle S and number of iterations k is critical for achieving an optimal success rate.

(iv) Replacing the traditional minimal function with the sine-enhanced minimal function leads, for Se-atom substructures, to a two- to eightfold increase in the number of solutions, provided that appropriate choices have been made for parameters S and k.

Based on this information, C&S(80°, 2, k) with

$$k = \begin{cases} 3 & \text{if } N_\mu < 35 \\ 2 & \text{if } 35 \leq N_\mu < 70 \\ 1 & \text{if } 70 \leq N_\mu < 200 \end{cases} \qquad (11)$$

is recommended as the default condition for sine-enhanced *Shake-and-Bake* for Se-atom substructure determination.

Table 6 summarizes and compares success rates and cost effectiveness using default conditions for traditional and sine-enhanced *Shake-and-Bake*. A sample of 10 000 randomly positioned N_μ-atom trial structures was generated for each of the six test data sets and the success rate is reported in the form of SR $\pm$ σ_{SR}, where σ_{SR} is the standard deviation calculated by Bernoulli's distribution, $\sigma_{SR} = (npq)^{1/2}$, with n being the number of trials, p being the success rate expressed as a fraction and q being the failure rate. It is clear from Table 6 that sine-enhanced *Shake-and-Bake* outperformed the traditional *Shake-and-Bake* in terms of success rate and cost-effectiveness. Specifically, the sine-enhanced *Shake-and-Bake* has a twofold increase in success rate for substructures CTLA4, AdoHcy and AEPT, a threefold increase for substructure epimerase, and an eightfold increase for substructures PDHC E1 and KPHMT.

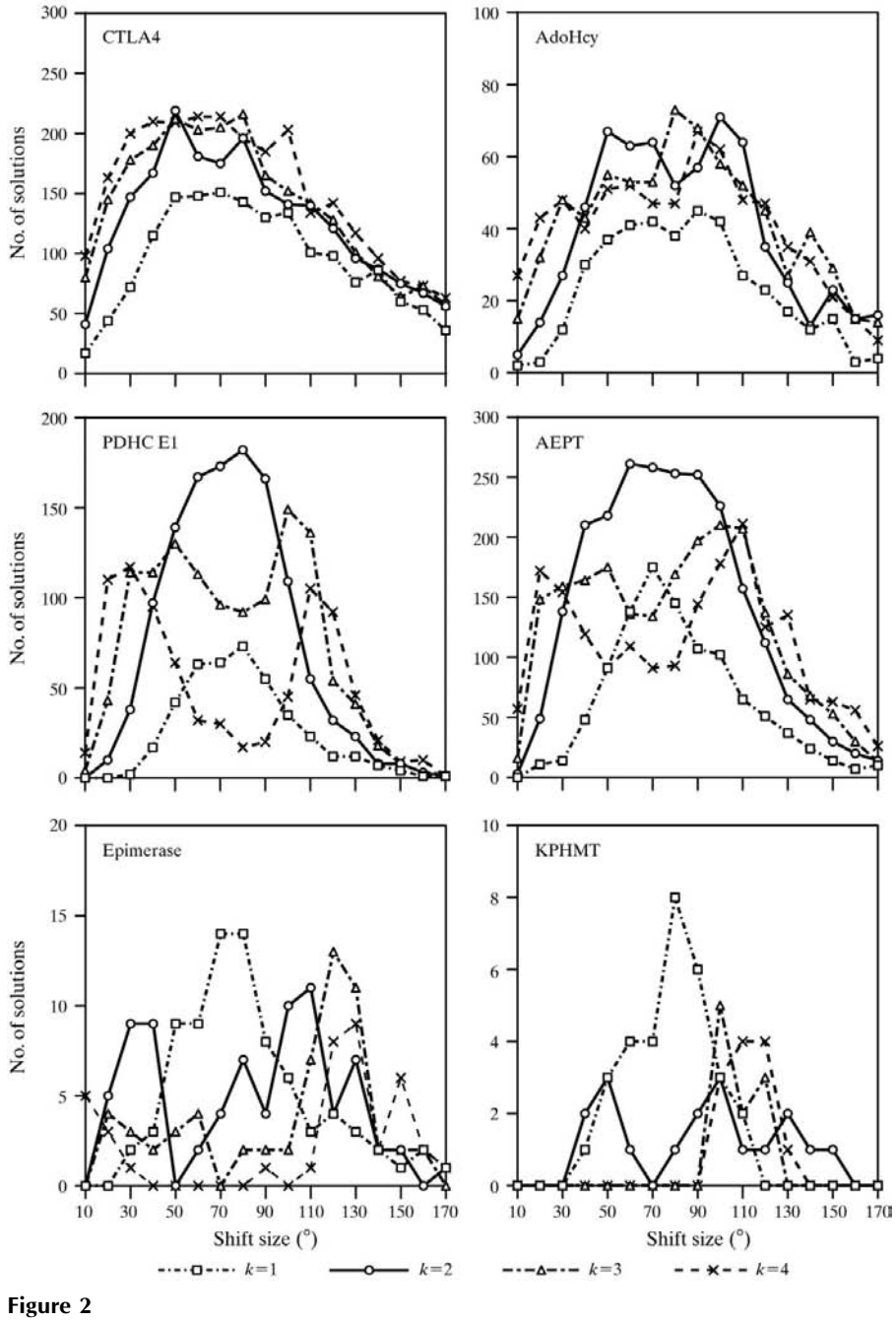

Figure 2
Number of solutions per 1000 *SnB* trial structures (10 000 for KPHMT) produced by sine-enhanced *Shake-and-Bake* as function of shift angles (S) and number of iterations (k).

Table 5
Number of solutions per 1000 *SnB* trial structures (10 000 for KPHMT) produced by COS(*S*, 2, 3) for various shift angles.

The bold numbers in each column indicate that these numbers are within 20% of the largest value.

S (°)	CTLA4	AdoHcy	PDHC E1	AEPT	Epimerase	KPHMT
10	58	3	1	4	0	0
20	67	14	4	18	2	0
30	100	19	4	37	3	0
40	**107**	21	6	62	3	0
50	**109**	25	9	77	4	0
60	**109**	25	15	**89**	**5**	0
70	**132**	25	**17**	**105**	**6**	**1**
80	**124**	**35**	**19**	**108**	**5**	**1**
90	**124**	**36**	**20**	**105**	**5**	**1**
100	**119**	**35**	**21**	**101**	**5**	0
110	**103**	**32**	**20**	79	4	0
120	92	**30**	**20**	76	4	0
130	95	28	10	50	4	0
140	76	28	3	40	1	0
150	44	17	1	28	1	0
160	57	8	2	13	0	0
170	47	8	2	12	0	0

Table 6
Comparison of SR (success rate) and CE (cost-effectiveness) for six Se-atom substructures using either traditional or sine-enhanced *Shake-and-Bake* with default conditions.

The comparison clearly shows the superior performance of sine-enhanced *Shake-and-Bake*.

	Traditional *Shake-and-Bake*		Sine-enhanced *Shake-and-Bake*	
Substructure	Success rate (%) ± s.d.	Solutions per hour	Success rate (%) ± s.d	Solutions per hour
CTLA4	11.48 ± 0.32	13.05	19.09 ± 0.39	21.37
AdoHcy	3.52 ± 0.18	0.54	5.66 ± 0.23	0.86
PDHC E1	2.03 ± 0.14	0.92	17.59 ± 0.38	7.21
AEPT	10.49 ± 0.31	0.68	21.96 ± 0.41	1.40
Epimerase	0.47 ± 0.07	0.04	1.42 ± 0.12	0.11
KPHMT	0.01 ± 0.01	0.00072	0.08 ± 0.03	0.0078

4. Discussion and conclusions

The sine-enhanced minimal function and its minimal principle have been formulated and the corresponding sine-enhanced *Shake-and-Bake* and its associated sine-enhanced parameter-shift procedure have been proposed and tested. As a consequence of the experiments described above, it is possible to recommend that COS(90°, 2, 3) (*i.e.* a maximum of two 90° phase shifts together with three passes through the phase list) be used in traditional *Shake-and-Bake* and C&S(80°, 2, *k*) [*i.e.* a maximum of two 80° phase shifts together with *k* passes through the phase list, where *k* is defined by (11)] be used in the sine-enhanced *Shake-and-Bake* for Se-atom substructure determination. Experiments on six substructures ranging in size from 12 to 160 Se atoms in the asymmetric unit showed that sine-enhanced *Shake-and-Bake* outperformed traditional *Shake-and-Bake* by a two- to eightfold increase in success rate. It is worth pointing out that, in the case of 160-site KPHMT, a solution can be obtained in 1250 trials (on average) using sine-enhanced *Shake-and-Bake*, whereas a solution may be found in 10 000 trials (on average) using traditional *Shake-and-Bake*. A user who has limited computing power might routinely solve similar substructures using sine-enhanced *Shake-and-Bake*, but be more likely to give up if traditional *Shake-and-Bake* were used.

Table 7
Number of solutions per 1000 *SnB* trial structures employing minimal function (12) with various *P* values for Se-atom substructure PDHC E1.

C&S(*S*, 2, 2) with various shift angles was used.

S(°)	*P* = 1.0	*P* = 0.8	*P* = 0.6	*P* = 0.5	*P* = 0.4	*P* = 0.3
30	2	7	33	38	34	15
40	4	11	58	97	62	2
50	7	15	93	139	42	0
60	13	31	117	167	14	0
70	12	32	128	173	2	0
80	18	25	118	182	1	0
90	20	25	93	166	5	0
100	14	21	46	109	48	0
110	13	23	33	55	154	0
120	12	13	15	32	131	0
130	7	10	7	23	77	0
140	2	3	7	8	50	1
150	3	0	0	8	20	87

The exact reason why sine-enhanced *Shake-and-Bake* outperformed traditional *Shake-and-Bake* is still unknown. It seems clear that the use of supplementary information concerning sine structure invariants is helpful in some way. The following question is then suggested: what is the optimal percentage of sine information to be used in the sine-enhanced minimal function? To answer this question, we define a new minimal function

$$M(\varphi;\varepsilon,P)=\left(\sum_{H,K}A_{HK}\right)^{-1}\sum_{H,K}A_{HK}\left(P\left[\cos\varphi_{HK}-\frac{I_1(A_{HK})}{I_0(A_{HK})}\right]^2+(1-P)\left\{\sin\varphi_{HK}-\varepsilon_{HK}\left[1-\frac{I_1^2(A_{HK})}{I_0^2(A_{HK})}\right]^{1/2}\right\}^2\right) \quad (12)$$

where P is a parameter, $P\in[0,1]$. The minimal function (12) coincides with the cosine minimal function when $P = 1$ [*i.e.* $M(\varphi;\varepsilon,1.0) = R(\varphi)$] and with the sine-enhanced minimal function when $P = 0.5$ [*i.e.* $M(\varphi;\varepsilon,0.5) = m(\varphi;\varepsilon)$]. An experiment was performed in which the sine-enhanced minimal function $m(\varphi;\varepsilon)$ (7) was replaced by the new minimal function $M(\varphi;\varepsilon,P)$ defined by (12) and C&S(*S*, 2, 2) with various shift angles was performed on the substructure PDHC E1. The results are summarized in Table 7. It is clear from Table 7 that (i) for every shift angle tested, the number of solutions increases when *P* decreases from 1.0 to 0.5, and (ii) $P = 0.5$ is the optimum. The precise reason for higher success rates at $P = 0.5$ remains unknown.

Both traditional and sine-enhanced *Shake-and-Bake* were also applied to seven known full structures ranging in size from 28 to 327 atoms in the asymmetric unit. Basic information regarding these data sets is listed in Table 8. A sample of 5000 randomly positioned trial structures was generated for

Table 8
Test data sets of full structures used in this investigation.

Structure	Atoms	Space group	d (Å)	Reference
9α-Methoxycortisol	28	$P2_12_12_1$	0.80	Weeks *et al.* (1976)
Isoleucinomycin	84	$P2_12_12_1$	0.94	Pletnev *et al.* (1980)
Pseudo-peptide	96	$P1$	0.83	Langs (personal communication)
Ternatin	104	$P2_12_12_1$	0.94	Miller *et al.* (1993)
Hexaisoleucinomycin	113	$P2_12_12_1$	1.00	Pletnev *et al.* (1992)
Gramicidin A	272	$P2_12_12_1$	0.86	Langs (1988)
Crambin	327	$P2_1$	0.83	Hendrickson & Teeter (1981)

Table 9
Comparison of success rates of traditional and sine-enhanced *Shake-and-Bake*.

Structure	Traditional *Shake-and-Bake* (%)	Sine-enhanced *Shake-and-Bake* (%)
9α-Methoxycortisol	11.68	18.34
Isoleucinomycin	4.72	9.60
Pseudo-peptide	19.30	59.60
Ternatin	0.76	1.42
Hexaisoleucinomycin	1.06	4.16
Gramicidin A	0.72	0.52
Crambin	3.54	4.00

each data set. For each structure, the default *SnB* parameters (reflections, invariants, peaks, cycles *etc.*) were used (Weeks & Miller, 1999*b*). Phase-refinement method COS(90°, 2, k), with $k = 1$ for $P1$ and $k = 3$ for non-$P1$ structures, was used in the traditional *Shake-and-Bake* procedure and C&S(80°, 2, k), with k defined by (11), was used in the sine-enhanced *Shake-and-Bake* procedure. The comparison of success rates listed in Table 9 shows that the sine-enhanced *Shake-and-Bake* procedure improves performance for small structures (less than 200 atoms in the asymmetric unit). However, for the two larger test structures, the sine-enhanced procedure either did not result in a significant improvement in success rate (crambin) or the success rate actually decreased (gramicidin A). Consequently, the sine-enhanced procedure is recommended for all structures (full structures as well as substructures) containing fewer than 200 unique atoms, but the traditional procedure is still recommended for larger structures.

We wish to express our appreciation to Jason Rappleye and Andrew Palumbo for assisting in the development of *SnB* v2.1. We also thank the Center for Computational Research (CCR), State University of New York at Buffalo for allowing us to use its High Performance Computing Research Facilities. This research was supported by NIH grant GM-46733 and NSF grant ACI-9721373.

References

Arjunan, P., Brunskill, A., Chandrasekhar, K., Sax, M. & Furey, W. (2000). Abstr. Am. Annu. Meet., Abstract P006.
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.
Chen, C. H., Kim, A., Zhong, H., Howard, A. J., & Sheldrick, G. M. (2000). Abstr. Am. Annu. Meet., Abstract 02.06.03.
Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
Deacon, A. M., Ni, Y. S., Coleman, W. G. Jr & Ealick, S. E. (2000). *Structure*, **8**, 453–462.
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.
Hauptman, H. A. (1991). *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A. D. Podnarny & 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.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Howell, P. L., Blessing R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
Langs, D. A. (1988). *Science*, **241**, 188–191.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Ostrov, D. A., Shi, W., Schwartz, J. C., Alomo, S. C. & Nathenson, S. G. (2000). *Science*, **290.** 816–819.
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.
Turner, M. A., Yuan, C. S., Borchardt, R. T., Hershfield, M. S., Smith, G. D. & Howell, P. L. (1998). *Nature Struct. Biol.* **5**, 369–376.
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. & Miller, R. (1999*a*). *J. Appl. Cryst.* **32**, 120–124.
Weeks, C. M. & Miller, R. (1999*b*). *Acta Cryst.* D**55**, 492–500.