Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

*P*1 *Shake-and-Bake*: can success be guaranteed?

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

[a]Hauptman–Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203, USA, and [b]Department of Computer Science, State University of New York, Buffalo, NY 14260, USA

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

Received 19 August 1999
Accepted 19 November 1999

The multi-trial direct-methods procedure known as *Shake-and-Bake* has been applied to three small proteins (alpha-1 peptide, vancomycin and lysozyme) that crystallize in space group *P*1. Phase refinement was accomplished through parameter-shift optimization using both the cosine and exponential forms of the minimal function. By extending error-free data to sufficiently high resolution, 100% convergence of trial structures to solution could be achieved in all three cases by using the exponential minimal function and a shift angle in the range 130–150°. These results suggest optimum parameters for other *P*1 structures and emphasize the importance of collecting data to the highest possible resolution.

1. Introduction

By automatically and repetitively alternating phase refinement in reciprocal space with a peak-picking protocol that imposes constraints in real space, the *Shake-and-Bake* algorithm (Weeks *et al.*, 1994) has increased the size of structures solvable by multi-trial direct methods to more than 1000 unique non-H atoms (Deacon *et al.*, 1998).[1] During the phase-refinement step, a parameter-shift optimization strategy (Bhuiya & Stanley, 1963) is generally used to reduce the value of the cosine minimal function,

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

(Debaerdemaeker & Woolfson, 1983; Hauptman, 1991; DeTitta *et al.*, 1994), written here in a truncated form without negative quartets, since such quartets are expected to contribute little useful phasing information for protein-sized molecules. (1) measures the mean-square difference between (i) the cosine values of the three-phase structure invariants,

$$\varphi_{\mathbf{HK}} = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}, \quad (2)$$

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

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

where the $|E|$s are the normalized structure-factor magnitudes and N is the number of atoms, assumed identical, in the unit cell. More recently, an exponential form of the minimal function,

$$m(\varphi) = \left(\sum_{\mathbf{H},\mathbf{K}} W_{\mathbf{HK}}\right)^{-1} \sum_{\mathbf{H},\mathbf{K}} W_{\mathbf{HK}} \left\{\exp\left[\mu A_{\mathbf{HK}}\cos\eta\cos\left(\varphi_{\mathbf{HK}}+\eta\right)\right] - \frac{I_0(A_{\mathbf{HK}}X)}{I_0(A_{\mathbf{HK}})}\right\}^2, \quad (4)$$

has been proposed, where

$$X = [\mu(\mu+2)\cos^2\eta + 1]^{1/2} \quad (5)$$

$$Y = [4\mu(\mu+1)\cos^2\eta + 1]^{1/2} \quad (6)$$

and

$$W_{HK} = \left[\frac{I_0(A_{\mathbf{HK}}Y)}{I_0(A_{\mathbf{HK}})} - \frac{I_0^2(A_{\mathbf{HK}}X)}{I_0^2(A_{\mathbf{HK}})}\right]^{-1}. \quad (7)$$

This exponential minimal function has been shown to be superior to the cosine form of the minimal function when applied to structures in space group *P*1 (Hauptman *et al.*, 1999). A good value for the parameter η has been determined empirically to be 20° and the parameter μ is dependent on $\max_{\mathbf{H},\mathbf{K}} A_{\mathbf{HK}}$ in a known way.

*P*1 is a unique space group. Since there is no symmetry other than unit-cell translations, the origin can be located anywhere and the response of *P*1 structures to direct-phasing methods is often different from that of structures in other space groups (Chang *et al.*, 1997; Deacon *et al.*, 1998). In particular, *P*1 structures frequently exhibit an unusually high percentage of solutions from *SnB*, especially if a single large shift angle is applied during

[1] Readers unfamiliar with the basic concepts and terminology of direct methods are referred to the recent authoritative book by Giacovazzo (1998).

Table 1
Test data sets and experimental parameters.

Structure	Unique non-H atoms	'Heavy' atoms	Phases	Triplets	Peaks	Cycles
Alpha-1	471	1 Cl	4000	40000	300	500
Vancomycin	547	12 Cl	4000	40000	150	500
Lysozyme	1295	10 S	11100	111000	350	750

parameter-shift optimization. The present study was undertaken to determine the theoretical limits of success for *Shake-and-Bake* applications in *P*1.

2. Materials and methods

The *Shake-and-Bake* procedure, using both the cosine and exponential minimal functions, was applied to the data from *P*1 crystals of three small proteins using the computer program *SnB* version 2.0 (Weeks & Miller, 1999*a*). The overall completeness and maximum resolution of these experimental data sets were as follows: alpha-1 peptide (86% complete, d_{max} = 0.9 Å; Privé *et al.*, 1999), vancomycin (80% complete, d_{max} = 0.97 Å; Loll *et al.*, 1998) and lysozyme (68% complete, d_{max} = 0.85 Å; Deacon *et al.*, 1998). Complete error-free data sets to very high resolutions were also generated for these structures using the known atomic coordinates and the program *EGEN* (R. Blessing, personal communication). Structure sizes and the values used for the major *SnB* parameters (phases, invariants, peaks picked and refinement cycles) are summarized in Table 1. Parameter-shift angles of 10, 20, 30, ..., 180° were applied. Solutions were unequivocally identified on the basis of the cosine-invariant figure of merit (Weeks *et al.*, 1995), which compares the invariant values for trial phase sets to the values for the true phase set. In all cases, solutions identified in this fashion were correlated with solutions identified by constrained minimal-function values. The success rate (SR) was defined as the percentage of trial structures that go to solution; the standard deviation of the success rate was calculated using Bernoulli's distribution, where $\sigma_{SR} = (npq)^{1/2}$, with n being the number of trials, p being the success rate expressed as a fraction and q being the failure rate.

3. Results

The results of the *SnB* applications to the three experimental data sets are summarized in Table 2. When the default double 90° shift conditions for acentric space groups were used, solutions were obtained only for alpha-1 peptide. Fortunately, single shifts of more than 90° were more productive and permitted the solution of all three structures. In fact, by using an optimum shift angle of 150°, the success rate for lysozyme increased to more than 25%. Replacing the experimental data with error-free data (but adding no more reflections) did not lead to any improvement in success rate for alpha-1 or vancomycin, indicating that no significant measurement errors had been made. On the other hand, using *complete* (at the nominal resolution of the experimental data) error-free data resulted in a very significant improvement in all three cases, with lysozyme having an astounding success rate of 99%! It seems likely that the adverse effects of missing phases are magnified by the even larger number of missing invariant relationships.

The foregoing observations served to motivate additional investigation designed to determine whether it was possible to find circumstances where all trials would converge to solution for all three structures. In order to consider the effect that resolution might play, additional error-free high-resolution data were generated for alpha-1 and vancomycin. The success rates for *SnB* applications to error-free data at a variety of resolutions are displayed in Fig. 1 as a function of parameter-shift angle. By optimizing the parameter-shift angle and using sufficiently high-resolution data, it was possible to achieve a 100% success rate for all three structures, provided that a single-shift angle in the neighborhood of 140° was chosen.

4. Discussion

The results of the experiments reported here underscore the importance of measuring data to the highest possible resolution if direct-phasing methods are to be used to solve a structure. These results also suggest that the solution of *P*1 structures can be virtually guaranteed if the requirement for high-resolution data is met and the structure is not too large. An accurate estimate of the maximum structure size accessible by *Shake-and-Bake* cannot be given with certainty, but it has been known for many years that the reliability of the structure invariants, which provide the underpinning for direct methods, decreases as the number of atoms in the unit cell increases (Cochran, 1955). Of the three structures described here, the alpha-1 peptide needs the highest resolution data in order to ensure that conditions can be found where all trial structures converge to solution. This may be related to the fact that the peptide contains only a single chloride ion (information not used in the *SnB* experiments), whereas several independent Cl or S atoms are present in the other two structures. The effects of heavier atoms on the efficiency and proper parameterization of *SnB* have been described elsewhere (Weeks & Miller, 1999*b*). These experiments also suggest that it may be beneficial in some cases to treat structures in other space groups as if the space group were *P*1. However, previous attempts to do this have had mixed results (Chang *et al.*, 1997) in terms of computational efficiency.

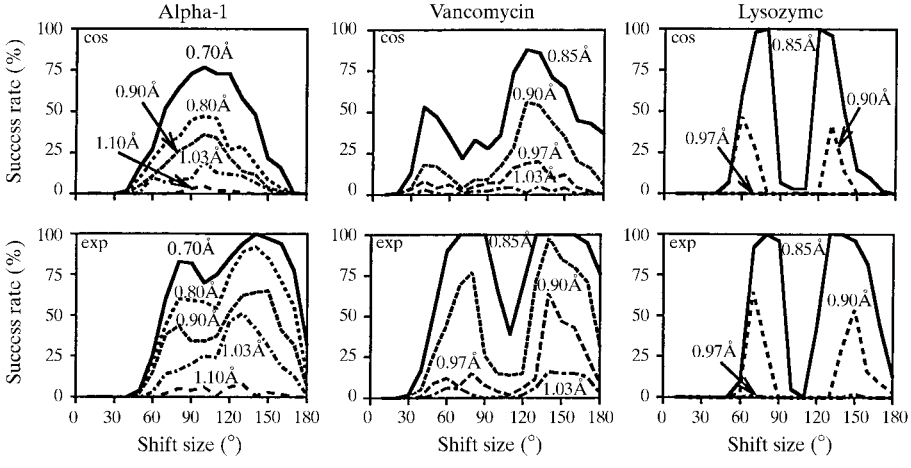

Figure 1
Success rates as a function of single-shift size sampled in increments of 10° for the cosine and exponential minimal functions. Each point represents 100 trial structures. Complete error-free data sets were used.

Table 2
Effects of measurement error and data completeness on success rates (SR) using the cosine minimal function.

In each case, the resolution limit was the same for experimental and error-free data and 500 trial structures were tested. The *SnB* default parameter-shift conditions for acentric non-*P*1 space groups are two 90° shifts. The optimum conditions shown here are single shifts.

Data type	Shift type	Alpha-1		Vancomycin		Lysozyme	
		Angle (°)	SR (%)	Angle (°)	SR (%)	Angle (°)	SR (%)
Experimental	Default	90	13.0 ± 1.5	90	0	90	0
Experimental	Optimum	110	19.8 ± 1.8	120	1.4 ± 0.6	150	25.6 ± 2.0
Error-free	Optimum	110	17.2 ± 1.7	120	0.2 ± 0.2	140	47.2 ± 2.2
Error-free complete	Optimum	110	30.4 ± 2.1	120	19.4 ± 1.7	120	99.0 ± 0.5

We would like to express our appreciation to Melda Tugac for preparing the figures. 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.
Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
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.
Giacovazzo, C. (1998). *Direct Phasing in Crystallography: Fundamentals and Applications*. Chester/Oxford: IUCr/Oxford University Press.
Hauptman, H. A. (1991). *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 324–332. Chester/Oxford: IUCr/Oxford University Press.
Hauptman, H. A., Xu, H., Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* A**55**, 891–900.
Loll, P. J., Miller, R., Weeks, C. M. & Axelsen, P. H. (1998). *Chem. Biol.* **5**, 293–298.
Privé, G. G., Anderson, D. H., Wesson, L., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1400–1409.
Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.
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.