Now we're cooking: new successes for shake-and-bake

Steven E Ealick

Shake-and-bake is an automatic procedure for phase determination developed for large molecules. The procedure is based on a minimal function which is optimized through alternate cycles of reciprocal space phase refinement and real-space filtering. The shake-and-bake technique has now been used to determine the structures of several small proteins.

Address: Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA.

Email: see3@cornell.edu

Electronic identifier: 0969-2126-005-00469

Structure 15 April 1997, 5:469–472

Introduction

One of the most remarkable advances in X-ray crystallography was made possible by the work of Hauptman and Karle in the 1950s and 1960s in which they showed that the phases of structure factors could be related to the distribution of normalized structure factor magnitudes [1,2]. The practical consequence of this work was the development of computer algorithms in the 1970s and 1980s to derive phases from experimental X-ray measurements, leading eventually to automatic structure determinations for most structures with less than 100 non-hydrogen atoms. These *ab initio* or direct methods, as they have come to be known, were facilitated by the development of automatic computer programs that incorporated tangent formula refinement, multisolution approaches and powerful figures of merit that can identify correct phase sets [3,4]. Together these methods, along with advances in computer technology, are responsible for an explosion of structural information embodied in the Cambridge Structural Database. In 1985, Hauptman and Karle received the Nobel Prize in chemistry for their fundamental contributions to the development of direct methods of structure determination.

With the phase problem essentially solved for small molecules, investigators began to turn their attention to structures that contained hundreds or even thousands of atoms. Historically, most macromolecular structures have been solved by heavy-atom methods, molecular replacement methods or a combination of the two. More recently, multiple wavelength anomalous diffraction (MAD) techniques have proven to be an effective method for phasing large molecules [5,6]. All of these techniques require the presence of atoms with special scattering properties or prior knowledge of the structure. Although *ab initio* phasing methods have clear advantages over other methods, the phase problem for large molecules is complicated by the nature of the conditional probability distributions of the three phase structure invariants, which serve as the basis for phase determination. For small molecules, these probability distributions are often sharp and reliable. However, as the number of atoms increases the probability distributions becomes broad and unreliable and phase determination by classical direct methods is usually ineffective.

The minimal principle

One particularly promising direct method phasing technique for large molecules, based on optimization of a minimal function, has been developed by Hauptman and coworkers [7–9]. As with traditional direct methods, this technique is based on the use of the joint probability distribution of several diffraction intensities and the corresponding phases. However, it differs from traditional direct methods in that the individual joint probability distributions are cast in terms of a minimal principle which involves all of the phases of interest [7]. Hauptman showed that the resulting minimal function has a minimum value for the correct set of phases [7]. Their technique also differs from traditional direct methods in that it is an iterative process in which phase refinement in reciprocal space is cyclically alternated with real-space filtering so as to impose the phase constraints implicit in real space.

Shake-and-bake and the SnB computer algorithm

Shake-and-bake is an *ab initio*, multiple solution, phase-determining procedure that is global and automatic (Fig. 1). This approach was successfully implemented by Miller and Weeks in the form of a computer algorithm called SnB [10–12] and applied to a number of large, small molecules and proteins [13]. The name of the computer algorithm, SnB, was derived from the two distinct steps: a reciprocal space phase refinement step (shake) and a real-space filtering step (bake). All necessary phases are assigned initial values by first generating a trial structure, consisting of a significant set of randomly positioned atoms, and then computing structure factors. The resultant phases are then refined and Fourier transformed, and a specified number of the largest peaks in the electron-density function (usually equal to the expected number of atoms) are found and used as a new trial structure in the next cycle.

The initial formulation of shake-and-bake used a parameter shift algorithm to reduce the value of the minimal function. The procedure was successfully tested using the experimentally measured, atomic-resolution diffraction data for known structures, ranging in size from n = 25

Figure 1

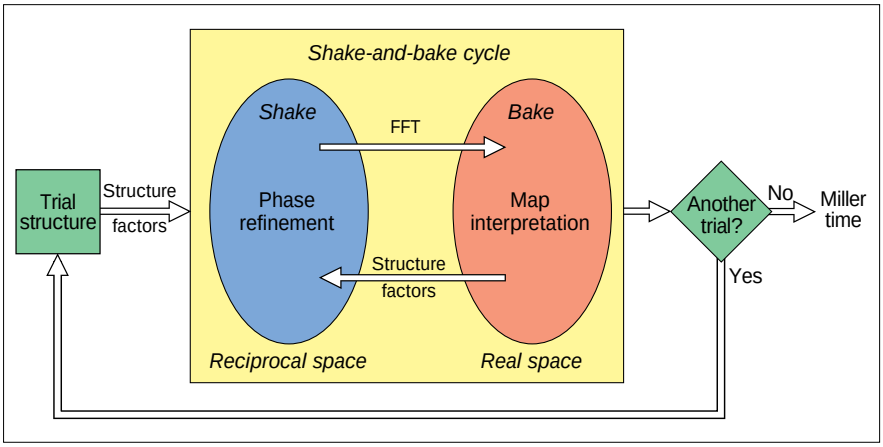


Diagram of the shake-and-bake algorithm. In general, structure factors are generated from a random trial structure. The phases corresponding to the large E values (normalized structure factors) are then processed through cycles of reciprocal space phase improvement and real-space filtering. The SnB computer algorithm uses a fast Fourier transform (FFT) for calculating electron-density maps. The value of the minimum function is used to decide whether success has been achieved or another trial structure is needed.

to n = 500 atoms and crystallized in a variety of space groups (Table 1). The number of trials required for success depends on a number of factors including structure size and composition, data quality and resolution, and the choice of values for several parameters which affect the refinement process. Tests with these structures showed that an atom : phase : triplet ratio of 1:10:100 was a good choice and that a three pass, 90°, 2-step, parameter-shift refinement was optimal.

Shake-and-bake has now been used by many different investigators to solve both known and unknown structures [13–22] (Table 1). Particularly noteworthy are the successful applications to the small proteins gramicidin A [19], Er-1 [20], crambin [21], alpha-1 peptide [23] and rubredoxin [19]. Solutions have been obtained using data with resolution as low as 1.1 Å–1.2 Å.

Structure determination of toxin II

One of the most recent successes of Shake-and-bake was the structure determination of toxin II from the scorpion *Androctonus australis* hector by Smith, Miller and coworkers at the Hauptman-Woodward Institute [22]. Toxin II is a small protein neurotoxin containing 64 amino acids and 4 disulfide bridges. The structure was previously determined using a combination of molecular replacement and heavy-atom methods and refined at 1.8 Å [24]; this was later further refined at 1.3 Å resolution [25]. The asymmetric unit, including ordered water molecules, contains over 600 non-hydrogen atoms. Toxin II crystallizes in the space group $P2_12_12_1$ with a = 45.94 Å, b = 40.68 Å and c = 29.93 Å, and with one molecule in the asymmetric unit. X-ray intensities measured to 0.95 Å resolution were provided by Fontecilla-Camps for the direct methods structure determination.

Initial phases for toxin II were determined by automatic application of the SnB program using default parameters. Based on the assumption of 500 atoms in the asymmetric unit, input parameters were chosen as number of phases = 5000, number of triplets = 50 000 and number of shake-and-bake cycles = 250. With these parameters, the structure determination was distributed over multiple workstations, each performing trials with random starting phases. After a total of approximately 1600 trials, an apparent solution was identified by a sharp decrease in the value of the minimal function residual which is characteristic of a correct set of phases.

The phases for the 5000 reflections from this trial were used to calculate an E map, which revealed clear structural features accounting for most of the toxin II structure. The remainder of the structure was revealed through several cycles of model refinement and Fourier calculations. Figure 2 shows a ribbon diagram of toxin II highlighting the regions that were identified from the initial map interpretation. The structure of toxin II contains 647 non-hydrogen atoms, including 129 ordered water molecules, and is the largest structure ever determined by *ab initio* methods of phase determination. The structure has now been refined at 0.95 Å resolution with an overall R factor of 0.158 for 30 609 reflections [22].

Future directions

The main limitations of shake-and-bake in its current form are the requirement for atomic-resolution data and the large amounts of computer time consumed by the SnB algorithm. Hauptman and coworkers are addressing the first limitation by developing procedures that use less than atomic-resolution data and/or by including anomalous-scattering data [26,27]. Anomalous-scattering data has been successfully used in many crystallographic phasing procedures: Woolfson and coworkers have successfully applied direct methods to single wavelength anomalous-scattering measurements [28,29]. Hauptman is currently developing procedures that combine anomalous-scattering data and

Table 1

Some structures determined by application of the shake-and-bake direct method procedure.

Molecule	Atoms	Space group	Resolution
Ternatin II [13]	105	$P2_12_12_1$	0.82
Ternatin I [13]	110	$P2_12_12_1$	0.94
Alpha-conotoxin PnIA [14]	110	$P2_1$	1.1
Alpha-conotoxin G1 [15]	117	$P2_1$	1.2
Isoleucinomycin analog [13]	127	$P2_12_12_1$	<1.0
Cyclic dodeca peptide [16]	156	P1	1.0
Riboflavin tetrabutyrate*	188	P1	NA
Vancomycin [17,18]†	255	$P4_32_12$	0.9
Gramicidin A [19]	317	$P2_12_12_1$	0.86
Er-1 pheromone [20]	328	C2	1.0
Crambin [21]	400	$P2_1$	0.83
Alpha-1 peptide [23]	471	P1	0.92
Rubredoxin [19]	497	$P2_1$	1.0
Toxin II [22]	647	$P2_12_12_1$	0.96

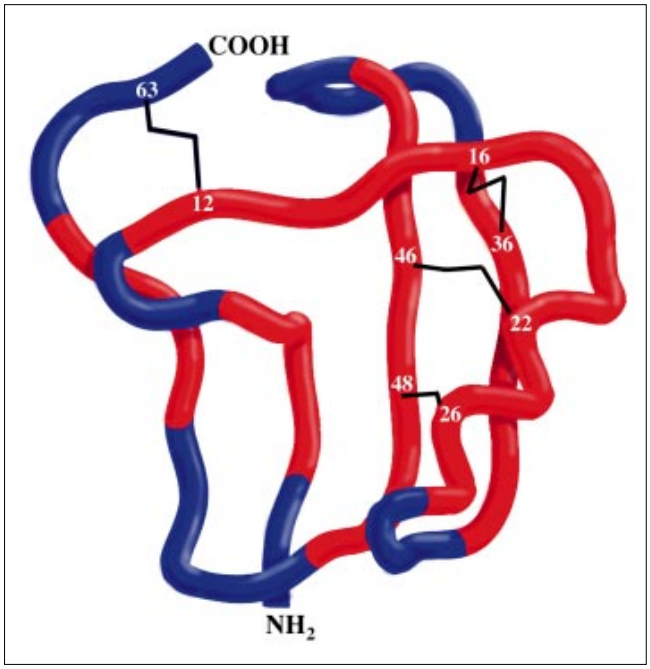
*Beverly Vincent, personal communication. †Solved using a method based on the shake-and-bake algorithm.

the shake-and-bake approach. The procedure has been tested on known structures and shows promise for high- as well as medium-resolution data.

The second limitation, the need for large amounts of computing time, is being addressed both by hardware and software considerations. Computer vendors are designing faster processors and the shake-and-bake procedure is naturally parallel. In addition, Weeks and Miller (personal communication) have now shown that the success rate of SnB can be greatly improved by judicious choice of input parameters. The experience with toxin II showed that correct phases lead to initial structures comprised of only the most ordered atoms. By reducing the expected number of atoms, the success rates have been improved from 1/1600 to approximately 1/100. The newest version of the SnB program also includes improvements, such as the incorporation of fast Fourier transform (FFT) algorithms and density modification modules, that significantly reduce the time per shake-and-bake cycle. These results suggest that the structure of toxin II could now be solved in less than a day using the improved version of SnB and the fastest available commercial workstations.

The SnB program has been described in the Journal of Applied Crystallography [11] and in the User's Manual for Version 1.0.0 [30]. There is also a home page for SnB on the World Wide Web at URL: http://www.hwi.buffalo.edu/SnB. Fundamental information is provided including a brief description, a list of personnel, critical citations, announcements, bug reports/ fixes, the current manual, and general information on how to obtain a copy of SnB. The SnB home page is directly accessible from the American Crystallographic Association home page.

Figure 2



Ribbon diagram of the structure of toxin II. The red ribbon segment corresponds to the parts of the structure that were identified from the initial shake-and-bake phases. The remainder of the structure (shown in blue) was generated from iterative cycles of least-squares refinement and difference Fourier analysis. Disulfide bridges Cys16–Cys36 and Cys22–Cys46 were identified from the shake-and-bake phases while Cys26–Cys48 and Cys12–Cys63 (which was disordered) were located from difference Fourier analysis.

Acknowledgements

The development of the shake-and-bake technique and the SnB computer algorithm is supported by grant GM46733 (H Hauptman, principal investigator) from the National Institute of Health.

References

1. Hauptman, H. & Karle, J. (1953). Solution of the phase problem I. The Centrosymmetric Crystal, ACA Monograph Number 3.
2. Karle, J. & Hauptman, H. (1956). A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, 3P12, 3P22. *Acta Cryst.* 9, 635–651.
3. Germain, G., Main, P. & Woolfson, M.M. (1971). The application of phase relationships to complex structures. III. The optimum use of phase relationships. *Acta Cryst. A* 27, 368–376.
4. Sheldrick, G.M. (1986). SHELX86. Program for the solution of crystal structures. University of Göttingen, Germany.
5. Karle, J. (1980). Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. Quantum Chem.* 7, 357–367.
6. Hendrickson, W.A., Smith, J.L. & Sheriff, S. (1985). Direct phase determination based on anomalous scattering. *Methods Enzymol.* 115, 41–55.
7. Hauptman, H.A. (1991). A minimal principle in the phase problem. In *Crystallographic Computing 5: from Chemistry to Biology, Papers presented at the International School of Crystallographic Computing, Bischenberg, France, July-August (1990)*. (Moras, D., Podnarny, A.D. & Thierry, J.C. eds), pp. 324–332, IUCr Oxford University Press, Oxford, UK.
8. DeTitta, G., Weeks, C., Thuman, P., Miller, R. & Hauptman, H. (1994). Structure solution by minimal function phase refinement and Fourier filtering: I. Theoretical basis. *Acta Cryst. A* 50, 203–210.

9. Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P. & Miller, R. (1994). Structure solution by minimal function phase refinement and Fourier filtering: II. Implementation and applications. *Acta Cryst. A* **50**, 210–220.
10. Weeks, C.M., DeTitta, G.T., Miller, R. & Hauptman, H.A. (1993). Applications of the minimal principle to peptide structures. *Acta Cryst. D* **49**, 179–181.
11. Miller, R., Gallo, S.M., Khalak, H.G. & Weeks, C.M. (1994). SnB: crystal structure determination via shake-and-bake. *J. Appl. Cryst.* **27**, 613–621.
12. Chang, C.-S., DeTitta, G., Hauptman, H., Miller, R., Thuman, P. & Weeks, C. (1993). Using parallel computers to solve the phase problem of X-ray crystallography. *J. Supercomputer Applications* **7**, 25–49.
13. Miller, R., DeTitta, G.T., Jones, R., Langs, D.A., Weeks, C.M. & Hauptman, H.A. (1993). On the application of the minimal principle to solve unknown structures. *Science* **259**, 1430–1433.
14. Hu, S.H., Gehrmann, J., Guddat, L.W., Alewood, P.F., Craik, D.J. & Martin, J.L. (1996). The 1.1 Å crystal structure of the neuronal acetylcholine receptor antagonist, alpha-conotoxin PnIA from *Conus pennaceus*. *Structure* **4**, 417–423.
15. Guddat, L.W., Martin, J.A., Shan, L., Edmundson, A.B. & Gray, W.R. (1996). Three-dimensional structure of the alpha-conotoxin GI at 1.2 Å resolution. *Biochemistry* **35**, 11329–11335.
16. Pletev, V.Z., *et al.*, & Duax, W.L. (1995). Crystal and molecular structure of the centrosymmetric meso-valinomycin analogue - cyclo (D-Val-D-Hyi-D-Val-L-Hyi-L-Val-D-Hyi-L-Val-L-Hyi-L-Val-D-Hyi-D-Val-L-Hyi) (C-60H-102N-6O-18). *Biopolymers* **36**, 615–621.
17. Schaefer, M., Schneider, T.R. & Sheldrick, G.M. (1996). Crystal structure of vancomycin. *Structure* **4**, 1509–1515.
18. Loll, P.J., Bevivino, A.E., Korty, B.D. & Axelsen, P.H. (1997). Crystal structure of the vancomycin : acetate complex. *J. Am. Chem. Soc.* **119**, 1516–1522.
19. Hauptman, H. (1995). Looking ahead. *Acta Cryst. B* **51**, 416–422.
20. Anderson, D.H., Weiss, M.S. & Eisenberg, D.S. (1996). A challenging case for protein crystal structure determination: the mating pheromone Er-1 from *Euplotes raikovi*. *Acta Cryst. D* **52**, 469–480.
21. Weeks, C.M., Hauptman, H.A., Smith, G.D., Blessing, R.H., Teeter, M.M. & Miller, R. (1995). Crambin: a direct solution for a 400 atom structure. *Acta Cryst. D* **51**, 33–38.
22. Smith, G.D., *et al.*, & Miller, R. (1997). The *ab initio* structure determination of a scorpion neurotoxin. *Acta Cryst.*, in press.
23. Prive, G., Ogihara, N., Wesson, L., Cascio, D. & Eisenberg, D. (1995), [abstract M501.03.05]. A designer peptide at high resolution: shake-and-bake solution of a 400 atom structure. American Crystallographic Association Meeting, Montreal, Canada.
24. Fontecilla-Camps, J.C., Habersetzer-Rochat, C. & Rochat, H. (1988). Orthorhombic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* hector. *Proc. Natl. Acad. Sci. USA* **85**, 7443–7447.
25. Housset, D., Habersetzer-Rochat, C., Astier, J.P. & Fontecilla-Camps, J.C. (1994). Crystal structure of toxin II from the scorpion *Androctonus australis* hector refined at 1.3 Å resolution. *J. Mol. Biol.* **238**, 88–103.
26. Hauptman, H.A. (1996). The SAS maximal principle: a new approach to the phase problem. *Acta Cryst. A* **52**, 490–496.
27. Guo, D.Y., Blessing, R.H., Langs, D.A. & Hauptman, H.A. (1997). On integrating the techniques of direct methods with anomalous dispersion. IV. A simplified perturbation treatment for SAS phasing. *Acta Cryst. A* **53**, 74–83.
28. Sha, B.D., *et al.*, & Woolfson, M.M. (1995). Direct phasing of one-wavelength anomalous-scattering data of the protein core streptavidin. *Acta Cryst. D* **51**, 342–346.
29. Xiao-Feng, Z., Fan, H.-F., Hao, Q., Dodd, F.E. & Hasnain, S.S. (1996). Direct-method structure determination of the native azurin II protein using one-wavelength anomalous scattering data. *Acta Cryst. D* **52**, 937–941.
30. Khalak, H.G., Gallo, S.M., Weeks, C.M. & Miller, R. (1994). SnB: a structure determination package. User's manual for version 1.0.0.