J. Appl. Cryst. (1999). **32**, 120–124

The design and implementation of *SnB* version 2.0

CHARLES M. WEEKS[a]* AND RUSS MILLER[a,b] at [a]*The Hauptman-Woodward Medical Research Institute, 73 High Street, Buffalo, New York 14203, USA, and* [b]*Department of Computer Science and Engineering, State University of New York, Buffalo, New York 14260, USA. E-mail: weeks@hwi.buffalo.edu*

(*Received 24 April 1998; accepted 3 August 1998*)

Abstract

SnB is a direct-methods program based on the *Shake-and-Bake* methodology. It has been used to solve difficult or large structures that could not be solved by traditional reciprocal-space routines based on the tangent formula. Recently, it has also been used to determine the Se sites in large selenomethionyl-substituted proteins. *SnB* version 1.5 has been available for several years and is being used regularly in many laboratories. In this paper, we introduce *SnB* version 2.0, which incorporates a graphical user interface written in Java, a dynamic histogram display, and an interactive Java/VRML-based visualization facility. In addition, it provides the user with several utility routines and a variety of new algorithmic options.

1. Introduction

SnB (Miller *et al.*, 1994) is a direct-methods package based on the *Shake-and-Bake* method of structure determination (Miller *et al.*, 1993; DeTitta *et al.*, 1994; Weeks *et al.*, 1994). The *Shake-and-Bake* algorithm has been the subject of several recent review articles (Weeks & Miller, 1996, 1997; Miller & Weeks, 1998). *SnB* has been publicly available since 1994, and has been available from the World Wide Web (WWW) *via* http://www.hwi.buffalo.edu/SnB/ since 1995. At the time of its introduction, tangent-based programs such as *RANTAN* (Yao, 1981) and *MULTAN* (Germain *et al.*, 1971) were capable of routinely solving structures containing less than 100 non-H atoms and of occasionally providing solutions for problems in the 100–200 atom range. Consequently, *SnB* represented a significant advance in *ab initio* direct-methods phasing because, as illustrated in Table 1, it can routinely solve structures containing several hundred non-H atoms. These structures include 200- and 400-atom variants of vancomycin (Loll *et al.*, 1997, 1998), a 450-atom designer peptide (Prive *et al.*, 1995), and the 1001-atom triclinic structure of hen egg-white lysozyme (Deacon *et al.*, 1998). In fact, because of the success of *SnB*, Sheldrick & Gould (1995) have recently exploited the *Shake-and-Bake* philosophy in a related algorithm which employs peak-list optimization and have announced the direct-methods solution of several structures ranging in size from 300 to 2500 atoms (Schäfer *et al.*, 1996; Sheldrick, 1997*a*). In addition to solving more complex structures than had previously been possible, *SnB* has been used to increase the number of Se sites that can be located for selenomethionyl-substituted proteins. For example, *SnB* was used to initiate the structure determination process for 190 kDa human placental *S*-adenosylhomocysteine (AdoHcy) hydrolase by finding the 30 Se atoms using peak anomalous difference data (Turner *et al.*, 1998).

SnB version 1.5 (v1.5) is the latest public release of the program, which can be obtained *via* the aforementioned WWW site. In this paper, we describe a major new version of the program, *SnB* v2.0, which has been constructed in an effort to (i) improve the overall performance over *SnB* v1.5, which is important if one is to be able to solve routinely structures with several hundred atoms within a reasonable timeframe, (ii) provide a modern graphical user interface (GUI), as opposed to the menu-driven ASCII interface provided with *SnB* v1.5, (iii) provide a dynamic histogram facility, which is used to guide the user in determining when a potential solution has been obtained, (iv) provide an easy means for porting the program to a variety of platforms, including UNIX workstations, PCs, multiprocessor machines, and networks of workstations (NOWs), (v) correct deficiencies recently detected in *SnB* v1.5 with regard to its handling of atoms in special positions, (vi) provide the user with the ability to generate the $|E|$s that are required to initiate direct phasing, and (vii) provide the user with a graphical visualization tool that can be used to examine potential solutions, as well as to modify graphically the on-screen solution and produce a higher-quality structure to be used in subsequent refinement procedures.

2. *SnB* v2.0

SnB consists of two major pieces of code, namely, the front-end interface and the back-end crystallographic package. The menu-driven ASCII-based front-end of *SnB* v1.5 was written in C, while its back-end was written in Fortran (Gallo *et al.*, 1996). *SnB* v2.0 includes a GUI front-end written in Java, and a significantly improved back-end, again written in Fortran. The file formats have been changed to facilitate interfacing with standard crystallographic programs, such as *SHELX* (Sheldrick, 1997*b*).

The core crystallographic routines were re-implemented 'from the ground up', which permitted a complete and thorough rethinking of the data structures in an effort to maximize efficiency. It should be noted that, when standard parameter settings are used for large structures, the new version of the program is significantly faster. *SnB* v1.5 provides only a structure-factor calculation for transforming from real to reciprocal space, whereas *SnB* v2.0 also includes an inverse fast Fourier transform (FFT). In addition to peak picking, which was the only density modification scheme provided with *SnB* v1.5, low-density elimination is now provided as an optional density modification scheme.

A major deficiency of *SnB* v1.5 was that it did not include a routine to generate $|E|$s. This deficiency has been alleviated in *SnB* v2.0 by incorporating the *LEVY/EVAL* suite of data-processing routines (Blessing *et al.*, 1996). This provides the

Table 1. *A selection of structures solved by SnB*

The empirical formula represents non-H atoms in the asymmetric unit cell (ASU). Success rates are typically based on using the recommended parameter values. Note that such parameter values are designed to minimize the time to solution, not to maximize the success rate. For a more complete list of structures solved by *SnB*, please visit the aforementioned WWW site (see §1), which also contains citations.

Structure	Atoms	ASU (protein)	Space group	Resolution (Å)	Success rate (%)
Vancomycin (Tetr)	258	$C_{132}Cl_4N_{18}O_{48}$	$P4_32_12$	0.9	0.6
Conotoxin EpI	289	$C_{138}N_{50}O_{50}S_{10}$	$I4$	1.1	53.0
Gramicidin A	317	$C_{198}N_{40}O_{34}$	$P2_12_12_1$	0.86	1.1
Er-1 pheromone	328	$C_{183}N_{46}O_{67}S_7$	$C2$	1.0	0.25
Crambin	~400	$C_{202}N_{55}O_{64}S_6$	$P2_1$	0.83	4.8
Alpha-1 peptide	471	$C_{290}ClN_{62}O_{118}$	$P1$	0.92	5.0
Rubredoxin	497	$C_{245}FeN_{58}O_{181}S_5$	$P2_1$	1.0	6.0
Vancomycin (Tric)	547	$C_{264}Cl_8N_{36}O_{96}$	$P1$	1.0	N/A
Scorpion Toxin II	624	$C_{313}N_{88}O_{96}S_8$	$P2_12_12_1$	0.96	1.4
Lysozyme	~1100	$C_{613}N_{193}O_{185}S_{10}$	$P1$	0.85	27.5*

† This success rate is the result of a non-standard parameter-shift condition.

user with the capability of automatically generating the $|E|$s that are required before invoking the *Shake-and-Bake* procedure.

A second significant deficiency of *SnB* v1.5 was discovered during an investigation of the conotoxin EpI peptide (Hu *et al.*, 1998). The structure of this peptide, which crystallizes in space group $I4$, could not be solved until a patch was put into *SnB* v1.5 that eliminated all peaks within 1.5 Å of any rotation axis. The Se substructure of AdoHcy hydrolase (space group $C222$) was similarly unsolvable until peaks near special positions were eliminated. In addition, once the appropriate patch was in place, the success rate (percentage of trial structures going to solution) for tetragonal vancomycin increased dramatically. It is interesting to note that none of these structures actually has a protein atom located near a special position. The effect of including incorrect peaks at special positions in *SnB* v1.5 is magnified by the fact that there is no provision for assigning proper weights based on multiplicity during the structure-factor calculations. These problems are addressed in *SnB* v2.0 in a manner valid for all space groups by the addition of two new parameters. These parameters are (i) a minimum distance between symmetry-related peaks such that peaks violating this restriction are eliminated, and (ii) a maximum number of the highest peaks permitted as exceptions to (i). The first parameter has a default value of 3.0 Å, and no exceptions are permitted unless some atoms are expected to be in special positions. In situations where such atoms are permitted, they are weighted properly.

The new *SnB* v2.0 interface is shown in Fig. 1. The Java language was chosen for this interface due to its extreme portability and ease of management. Once the basic information is typed into the appropriate boxes on the 'General

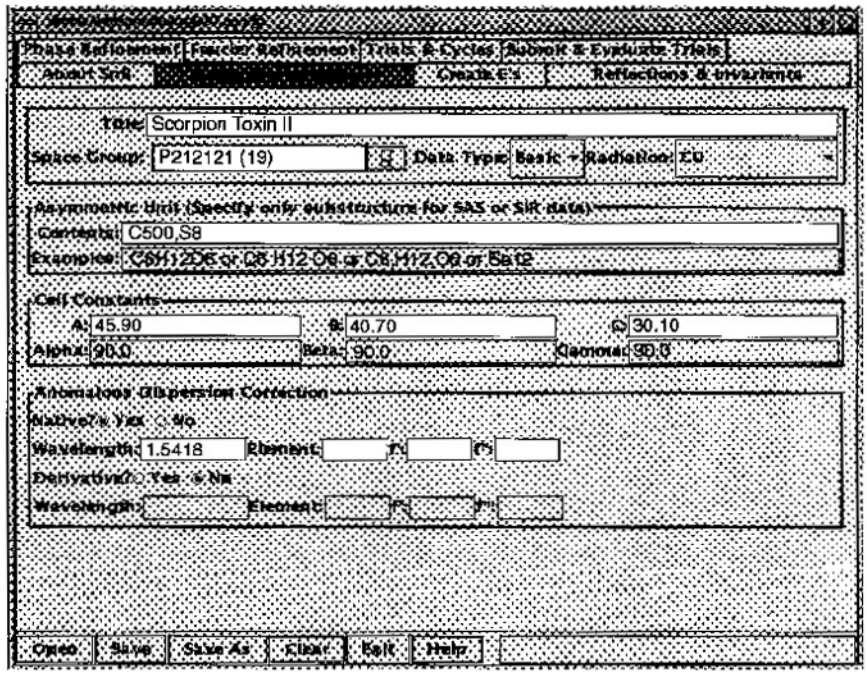

Fig. 1. A snapshot of the main ('General Information') page from the new GUI interface for *SnB* v2.0. This page contains all of the necessary structure-specific information that the user must enter in order to run the *Shake-and-Bake* algorithm.

Information' screen, the user is provided with default values for the other necessary parameters. The information given in Fig. 2 was generated by the system based on extensive experimentation carried out by the *SnB* research team to determine appropriate values. Of course, the user has the freedom to change any of the default values provided.

In *SnB* v1.5, an ASCII-based histogram was provided that proved extremely useful in determining whether or not a potential solution existed. A modern GUI-based histogram is provided with *SnB* v2.0, as shown in Fig. 3. In addition to being graphical, this histogram is dynamic in that it is updated in real time as additional trial structures are processed. The output of

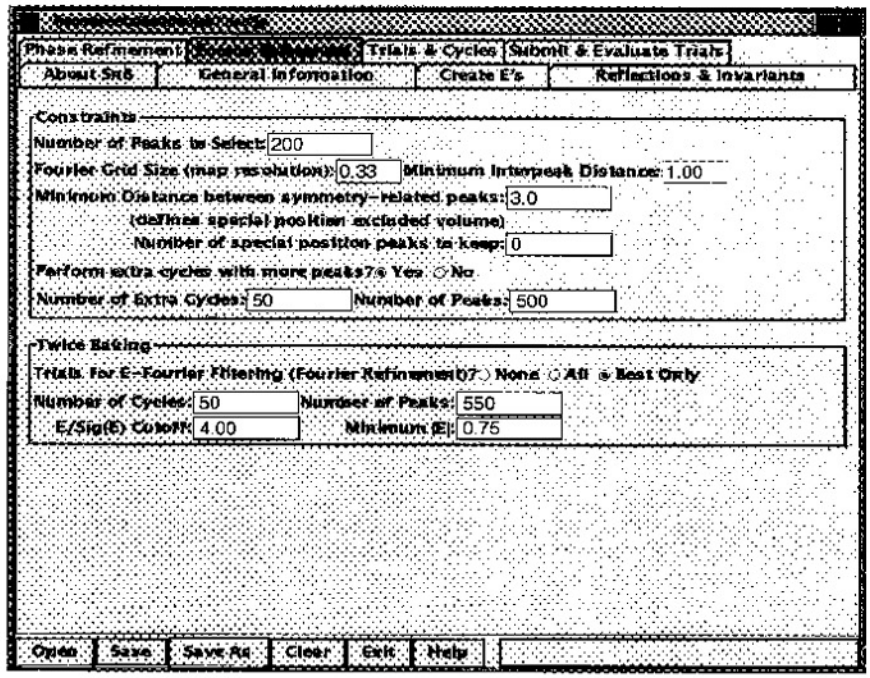

Fig. 2. The 'Fourier Refinement' screen. Note that all of the information on this screen is generated based on the user input to the 'General Information' screen shown in Fig. 1.

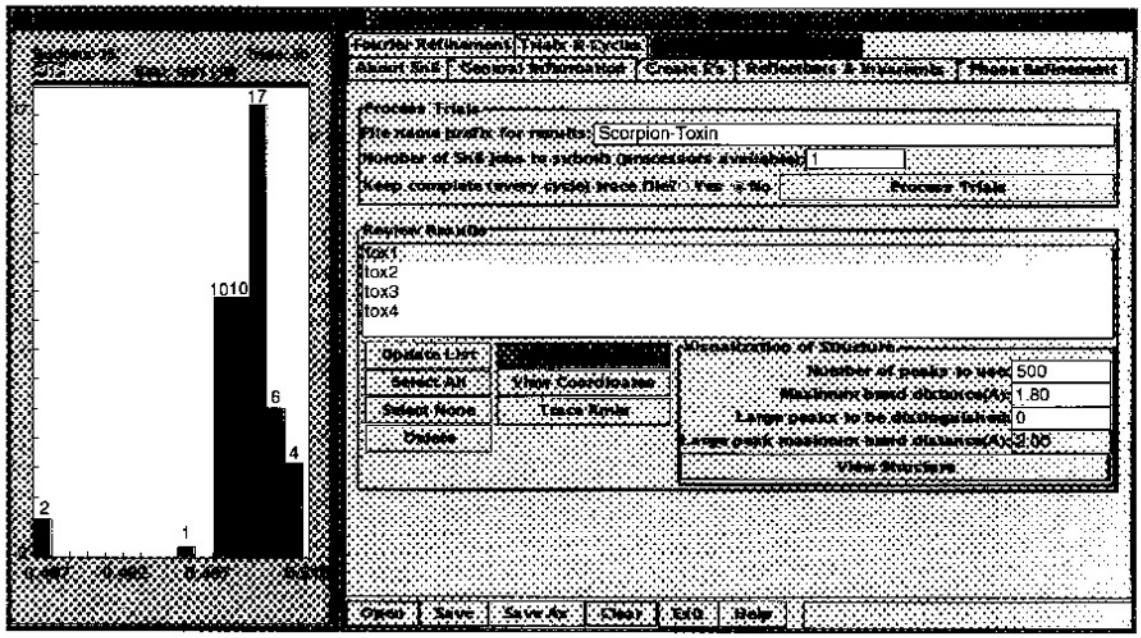

Fig. 3. A dynamic histogram generated by the user with the new *SnB* v2.0 interface. Note that the histogram is updated in real time as new trials are processed.

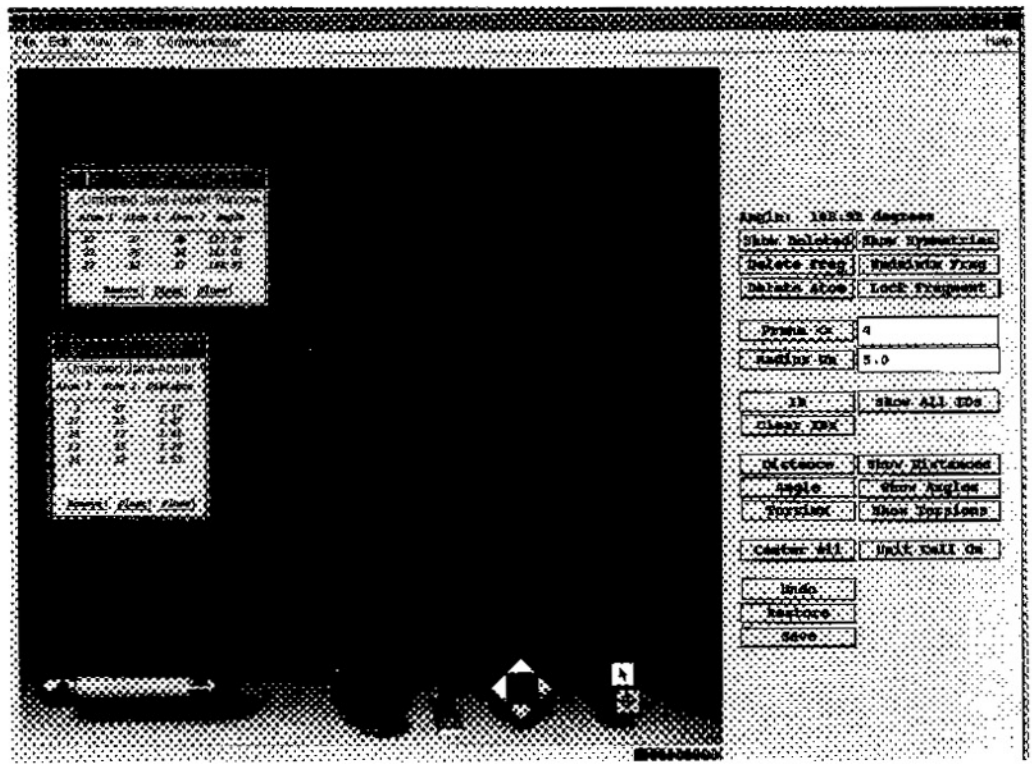

Fig. 4. A snapshot of the Java/VRML visualization tool provided with *SnB* v2.0 as applied to an 84-atom structure.

the *SnB* v2.0 program has also been made more useful and convenient by the provision of a Java/VRML visualization package, as illustrated in Fig. 4. This routine has the benefit of not only allowing the user to view the potential solution as it comes out of *SnB*, but also allowing on-screen editing of the peak/atom file. The revised file can be saved and used as input to either *SnB* or another program for further structure refinement.

3. Discussion

SnB v2.0 is currently under Beta-test, both inside and outside of the *SnB* laboratory at the Hauptman-Woodward Institute. Stripped-down versions of it have already been used to solve some of the structures previously mentioned, including the triclinic forms of vancomycin and lysozyme, the conotoxin EpI peptide, and the Se substructure of AdoHcy hydrolase. The *SnB* research team continues to tune parameters. As improved parameter settings are determined, they are posted to the WWW and are also incorporated into subsequent versions of the program. In addition, new postprocessing routines, targeted at cleaning up the initial map, are under development. Those that show promise will be included in subsequent releases of the program and will supplement the *E*-Fourier filtering routine that was present in *SnB* v1.5.

At present, an enhanced WWW site is under construction. This site has a dedicated server and will provide users with access to *SnB* v2.0. Most importantly, users will be allowed to run *SnB* v2.0 directly on this WWW server.

We would like to thank Andrew Palumbo, Thomas Tang and Jan Pevzner for their contributions to the program, and Herb Hauptman for his continued inspiration and support. We would also like to thank Jenny Martin and Pat Loll for allowing us to use their data to identify, and eventually alleviate, problems with atoms at special positions. This research was supported by grants GM-46733 (NIH) and IRI-9412415 (NSF).

References

Blessing, R. H., Guo, D. Y. & Langs, D. A. (1996). *Acta Cryst.* D**52**, 257–266.

Deacon, A., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.

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

Gallo, S. M., Miller, R. & Weeks, C. M. (1996). *Supercomputing '96 Conference Proceedings*, see http://www.supercomp.org/sc96/proceedings/.

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

Hu, S.-H., Loughman, M., Miller, R., Weeks, C. M., Blessing, R. H., Alewood, P. F., Lewis, R. J. & Martin, J. L. (1998). *Biochemistry*, **37**, 11425–11433.

Loll, P. J., Bevivino, A. E., Korty, B. D. & Axelsen, P. H. (1997). *J. Am. Chem. Soc.* **119**, 1516–1522.

Loll, P. J., Miller, R., Weeks, C. M. & Axelsen, P. H. (1998). *Chem. Biol.*, **5**, 293–298.

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

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

Miller, R. & Weeks, C. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 387–398. Dordrecht: Kluwer Academic Publishers.

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

Schäfer, M., Schneider, T. R. & Sheldrick, G. M. (1996). *Structure*, **4**, 1509–1515.

Sheldrick, G. M. (1997*a*). *Proceedings of the ACA Meeting*, St Louis, MO, USA. Abstract TuA03.

Sheldrick, G. M. (1997*b*). *The SHELX97 Manual*. University of Göttingen, Germany.

Sheldrick, G. M. & Gould, R. O. (1995). *Acta Crys.* **B51**, 423–431.

Turner, M. A., Yuan, C.-S., Borchardt, R. T., Hershfield, M., 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.* **A50**, 210–220.

Weeks, C. M. & Miller, R. (1996). *Crystallographic Computing 7: Proceedings of the Macromolecular Crystallographic Computing School*, edited by P. Bourne & K. Watenpaugh. Oxford University Press. Also available *via* http://www.sdsc.edu/Xtal/IUCr/CC/School96/.

Weeks, C. M. & Miller, R. (1997). *Recent Advances in Phasing: Proceedings of the CCP4 Study Weekend*, DL-CONF-97-001, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 139–146. Warrington: Daresbury Laboratory. Also available *via* http://www.dl.ac.uk.CCP/CCP4/main.html.

Yao, J.-X. (1981). *Acta Cryst.* **A37**, 642–644.