SnB [1]: A Structure Determination Package
User's Manual for Version 1.5.0 [2]

Russ Miller
Steven M. Gallo
Hanif G. Khalak
Charles M. Weeks

June 18, 1996

[2] Partial support for this this project was provided by the National Institutes of Health under grant GM-46733 and the National Science Foundation under grant IRI-9108288

Contents

Just the Facts: *SnB*

SnB has been successfully tested on several dozen structures, ranging from approximately 20-600 atoms in the asymmetric unit cell, covering the following space groups:

- $P1$, $P\bar{1}$
- $P2_1$, $C2$, $P2/c$
- $P2_12_12_1$, $I222$, $Pna2_1$, $Fdd2$
- $P4_1$, $P4_12_12$
- $P3_1$, $R3$, $P3_212$, $P3_221$, $R32$
- $P6_5$, $P6_522$, $P2_13$, $P4_132$

Note: If there are concerns regarding the application of *SnB* to a structure not in one of these space groups, please contact the authors to verify that the symmetry is being handled properly.

Citations to *SnB* should include the following:

- R. Miller, S.M. Gallo, H.G. Khalak, C.M. Weeks, *SnB*: Crystal structure determination via Shake-and-Bake, *Journal of Applied Crystallography*, **27**, 1994, pp. 613-621.

SnB is based on the *Shake-and-Bake* method of structure determination, which was motivated by the *Minimal Principal*. The foundation of this work can be found in the following:

- G.T. DeTitta, C.M. Weeks, P. Thuman, R. Miller, and H.A. Hauptman, Structure solution by minimal function phase refinement and Fourier filtering: theoretical basis, *Acta Crystallographica*, **A50**, 1994, pp. 203-210.
- C.M. Weeks, G.T. DeTitta, H.A. Hauptman, P. Thuman, and R. Miller, Structure solution by minimal function phase refinement and Fourier filtering: implementation and applications, *Acta Crystallographica*, **A50**, 1994, pp. 210-220.

The first application of *SnB* to a difficult, previously unknown, structure can be found in the following:

- R. Miller, G.T. DeTitta, R. Jones, D.A. Langs, C.M. Weeks, and H.A. Hauptman, On the application of the minimal principle to solve unknown structures, *Science*, **259**, March 1993, pp. 1430-1433.

Electronic Communication

We are interested in comments, criticisms, and suggestions with respect to *SnB* and the documentation. All such correspondence may be sent electronically to:

snb-comments@hwi.buffalo.edu

SnB maintains a home page on the World Wide Web at the URL address

http://www.hwi.buffalo.edu/SnB/

This home page contains information pertaining to *SnB*, including

- a brief introduction to the *Shake-and-Bake* philosophy,
- a list of the principal contributors to the *SnB* project,
- a list of major citations for *SnB* and *Shake-and-Bake*,
- a current version of this manual,
- a list of frequently asked questions (FAQs) with responses,
- information on obtaining a copy of *SnB*,
- announcements of updates or changes to *SnB*, and
- a list of reported bugs with their associated fixes.

The *SnB* home page may also be reached from the home page of the Hauptman-Woodward Medical Research Institute (formerly the Medical Foundation of Buffalo), which has a URL address of http://www.hwi.buffalo.edu/ or from the home page of the American Crystallographic Association (ACA), which has a URL address of http://www.hwi.buffalo.edu/ACA/.

1 Introduction

The *Shake-and-Bake* method of crystal structure determination is a direct methods phasing algorithm[1] based on the *minimal principle*[2]. As illustrated in Figure 1, the *Shake-and-Bake* method alternates phase refinement in reciprocal space with density modification in real space to reduce the value of the *minimal function*. As with traditional direct methods techniques, numerous trial structures or sets of phases are evaluated. Unlike traditional direct methods, however, the reciprocal space refinement typically involves a *parameter shift algorithm*, and there is an automated, repeated transformation between reciprocal and real space refinement.

Experimentation has thus far confirmed that:

1. the *minimal function* is diagnostic in that a histogram of final values corresponding to the resultant structures can be used with high confidence for the purpose of determining whether or not a solution exists, and

2. for a set of trial structures that yields a solution, the final structure corresponding to the smallest value of the *minimal function* is typically a solution.

This manual describes the computer program, *SnB* [3], which implements the *Shake-and-Bake* method. Section 2 gives an overview of the structure of the program, including the coordination of the critical routines. Section 3 discusses hardware, language and operating system requirements. Section 4 describes the program installation procedure. Section 5 focuses on the files that are manipulated during execution of the program. Section 6 discusses the help facility available in *SnB*. Section 7 lists the default set of parameters. Finally, Section 8 discusses the user interface, including the crystallographic input required, default values, and options for evaluating potential solutions.

---

[1] DeTitta, G.T., Weeks, C.M., Thuman, P., Miller, R., & Hauptman, H.A., *Acta Cryst.* **A50**, 1993, 203-210; Miller, R., DeTitta, G.T., Jones, R., Langs, D.A., Weeks, C.M. & Hauptman, H.A., *Science*, **259**, 1993, 1430-1433; Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P. & Miller, R., *Acta Cryst.* **A50**, 1994, 210-220.

[2] Hauptman, H.A., Am. Crystallogr. Assoc. Meet., 1988, Philadelphia, USA, Abstract R4; Hauptman, H.A., In *Crystallographic Computing 5: From Chemistry to Biology*, ed. by D. Moras, A.D. Podjarny, and J.C. Thierry, pp. 324-332. Oxford: IUCr Oxford Univ. Press; Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P. & Miller, R., *Acta Cryst.*, **A50**, 1994, 210-220.

[3] Miller, R., Gallo, S.M., Khalak, H.G., Weeks, C.M.,*Journal of Applied Crystallography*, **27**, 1994, pp. 613-621.

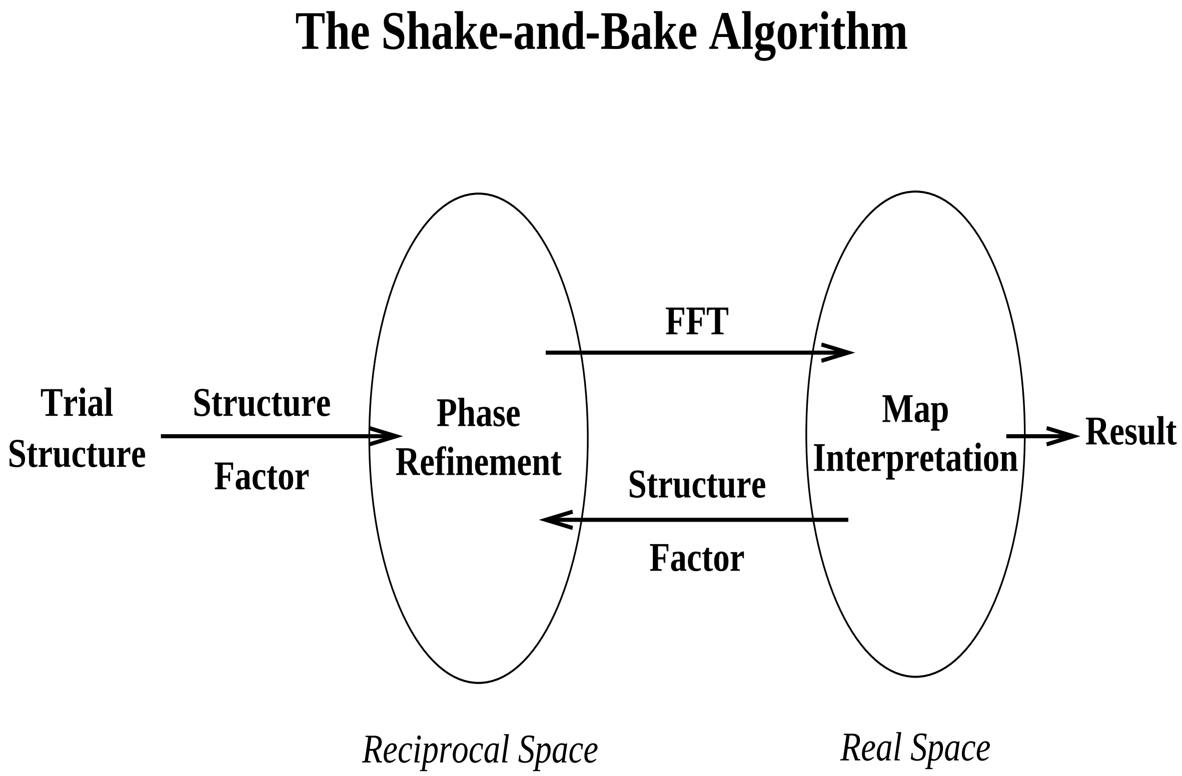

Figure 1: The *Shake-and-Bake* method of structure determination alternates between real and reciprocal space. In the current version of *SnB*, phase refinement is typically performed by parameter shift, and density modification consists of peak picking.

2 Structure of the *SnB* Program

The structure diagram presented in Figure 2 illustrates the basic organization of the *SnB* package, which consists of three major components. The first component performs the actual Shake-and-Bake structure determination procedure by generating and processing trial structures. The second component provides the user with a facility for interactively examining the progress of a structure determination run. This allows the user to produce a histogram of the final *minimal function* values for the trial structures that have been processed. This histogram can be used to decide whether or not a probable solution has been obtained. The third component allows the user to examine the geometry of the current best trial structure for any given run.

Pseudo-code for the *Shake-and-Bake* component of the program is presented in Figure 3. Note that before the actual *Shake-and-Bake* process can begin, triplet and (optionally) negative quartet structure invariants, as well as the coordinates for the trial structures, are generated. For a given trial structure, the first operation performed is a structure factor calculation which yields the corresponding set of phases. These phases are used to determine the corresponding value of the *minimal function.* The cyclical *Shake-and-Bake* phasing procedure then begins. The phases are adjusted so as to reduce the value of the *minimal function.* These phases are then passed to a fast Fourier routine which produces an electron density map. This map is examined by a peak picking routine which identifies the m largest peaks (typically, for an n-atom structure, $m \leq n$) subject to the constraint that no two peaks are closer than a specified distance. These peaks are then treated as atoms, and the process of structure factor calculation, phase refinement, and density modification via peak selection is repeated for a predetermined number of *Shake-and-Bake* cycles.

For every completed trial structure, the final value of the *minimal function* is stored in a file for histograming purposes. Further, a separate file is used to maintain the final minimal function value, as well as the initial and final peak or atom coordinates, associated with the best trial structure processed so far.

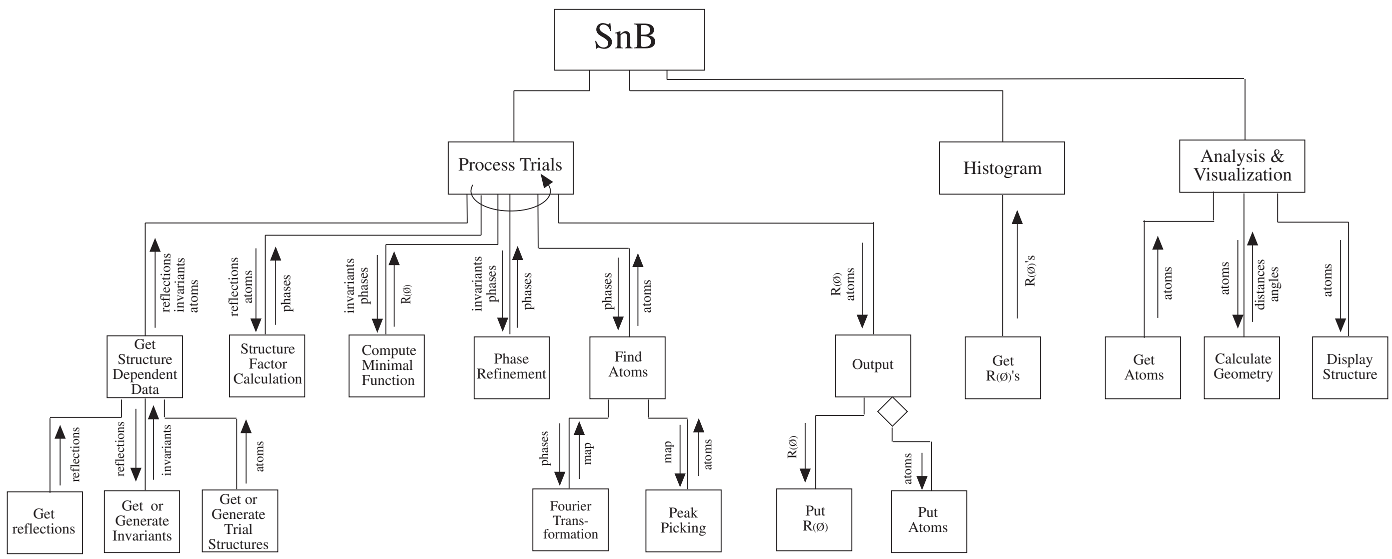

Figure 2: A structure diagram for the *SnB* program.

```
Generate triplet and (optionally) negative quartet invariants
Generate trial structures with randomly positioned atoms
For every trial structure do
    Structure factor calculation {produce phase set φ}
    Compute the initial value of the minimal function, R(φ)
    Do for a specified number of Shake-and-Bake cycles
                Phase refinement {modify phase set φ}
                Fourier transformation {produce electron density map}
                Peak selection {select largest peaks as new set of atoms}
                Structure factor calculation {produce revised phase set φ}
                Compute new R(φ)
    end Do
    Store final R(φ) for histogram purposes
    If R(φ) < R_best
        R_best = R(φ)
        Store atoms
    end If
end For
```

Figure 3: Pseudo-code for the *Shake-and-Bake* structure determination routine in *SnB*.

3 System Requirements

SnB is written in a combination of C and Fortran. The fundamental crystallographic routines, including routines for performing Fourier transformations, peak picking, and structure factor calculations, are written in Fortran. The interactive routines are written in C. Detailed system requirements are given in Figure 4.

Hardware:	Silicon Graphics Incorporated (SGI) 4D series Sun Micro-systems SparcStation Cray Research C90 and T3D Digital Equipment Corporation Alpha Thinking Machines Corporation CM-5 Intel 80x86 International Business Machines (IBM) RS/6000 International Business Machines (IBM) SP2
Memory:	16MB RAM physical, plus 16MB RAM virtual
Storage:	source files = 0.9 MB executables = 1.0 MB input data files = 0.3 MB - 4.0 MB per structure output data files = 0.1 MB - 2.0 MB per structure
Compilers:	C compiler 'cc' or GNU C compiler 'gcc' FORTRAN compiler 'f77' ('cf77' on Cray machines) libraries: m, c, malloc (C) F77, I77, isam (Fortran)
OS (UNIX):	AIX 3.0 IRIX 4.0 + OSF/1 3.0 + SunOS 4.1 + Solaris 2.3 + UNICOS 7.C.3 ULTRIX 4.3A

Figure 4: System requirements for *SnB*

4 Program Installation

The following procedure should be followed in order to install *SnB* Version 1.5.0.

1. In your '.cshrc' file (UNIX system file), define the environment variable SNB_HOME to be the directory where the *SnB* executables will be stored. That is, include the line "setenv SNB_HOME <directory_name>" where <directory_name> is the path to the directory where the executables will be stored. If you are using sh, include the line "export SNB_HOME=<directory_name>".

2. Enter the SnB_v1.5.0/ directory and type "configure". This will generate a makefile for your operating system and hardware configuration. In the event that a particular aspect of your system cannot be detected automatically, type "configure −−help" for a list of options so that the proper system settings can be entered manually. Examine the makefile to make sure that the options are compatible with your particular setup, namely the lines starting with XCFLAGS, XFFLAGS, SNB_LIBS, and MENU_LIBS. See the updated README.SnB file in that directory for more information.

3. Type 'make install'. This will compile both the Menu and Shake-and-Bake procedures. The make routine will install the executables in the SnB_v1.5.0/bin/ directory. The Menu executable is called 'SnB' and the Shake-and-Bake executable, which is invoked by SnB, is called 'SnB_main'. (SnB_v1.5.0/bin also contains the file space.dat.)

4. Place a copy of SnB, SnB_main, and space.dat in the directory defined in item 1.

5. If you are installing the Geomview package as a suppliment to *SnB*, untar the distribution and follow the installation instructions. Geomview is not required to operate *SnB*. However, it provides a much improved visualization routine for evaluating possible solutions. See section 8.4 for more information about Geomview and the visualization options available in *SnB*.

6. Log off and then back on, so as to allow the system to initialize the environment variable defined in item 1. If you are using csh or tcsh you can simply type "source .cshrc", and then "rehash".

7. The user can now run the program by typing the command 'SnB' within the directory containing the reflection file of a structure to be considered. We recommend maintaining a separate directory for each such structure.

5 Structure of Data Files

For each molecular structure of interest, we recommend creating a separate directory containing the reflection file, '*structure*.ref', where *structure* is a user defined prefix that identifies the molecular structure. *SnB* should be run from within such a directory.

The user should verify that they have permission to access the executables 'SnB' and 'SnB_main', as well as 'space.dat', from the directory defined in Section 4. In addition, the user should verify that they have permission to access the '*structure*.ref' file of interest in the current directory.

For each structure, there are several associated *input* data files. For each run initiated, there are several *output* files which contain the results of the computation.

5.1 Input Files

There are five (5) input files which are used by *SnB* ('*structure*' is a user defined ID):

space.dat	*structure*.info
structure.ref	*structure*.trials
structure.inv	

space.dat

This file contains equivalent position and asymmetric unit limit information for *all* space-groups. A single copy of this file (see Section 4, Program Installation) should exist in the directory containing the executables 'SnB' and 'SnB_main'.

*structure***.ref**

This is a **user supplied** file containing the reflections of the structure under consideration. In fact, this is the only file that the user *must* supply. It is in text format, containing one line for each available reflection. Each line contains six (6) values, in format 3I5, 3F9.4, as follows:

```
   H    K    L        E     Fobs      SIG
──────────────────────────────────────────── file begins below
   5    0    9   3.8360  44.4000   1.4300
   0    5    8   3.6140  51.3000   1.6600
  11    9    7   3.2490  17.6000   0.5800
   2    6   12   3.2040  18.7000   0.6200
  15    0    5   3.0930  17.0000   0.5600
   ⋮
────────────────────────────────────────────
```

where

H	= 1st Miller index (along 'a' cell-axis)
K	= 2nd Miller index (along 'b' cell-axis)
L	= 3rd Miller index (along 'c' cell-axis)
E	= measured E-intensity value for each reflection
Fobs	= observed F-intensity value for each reflection
SIG	= standard deviation of samples for each reflection's measurement, $\sigma(F_{obs})$

Note that when *SnB* is first run with such a reflection file, it will prepend a header line containing control information and augment every line in the file corresponding to a reflection with the value of $sin(\theta)/\lambda$ and a reflection code indicator. *SnB* will also sort this reflection file into non-increasing order on $|E|$-values. Note that the original user supplied reflection file will be copied by *SnB* to *structure*.ref.orig prior to these computations being performed. Finally, in subsequent runs of a structure, the augmented reflection file is automatically used.

Recommendation: We recommend that Bob Blessing's data processing routines (LEVY / EVAL) be used to generate E's. Blessing may be reached by e-mail at `blessing@hwi.buffalo.edu`, by phone at (716) 856-9600, or by surface mail at Hauptman-Woodward Medical Research Institute, 73 High Street, Buffalo, New York, 14203.

structure.**inv**

This file contains invariants generated from the reflection data. It is best to use *SnB* to generate invariants, but the user does have the option of supplying their own set of invariants. The format of the invariant file follows.

```
 REF1  REF2  REF3   REF4   PS    NW     EC
——————————————————————————————————————————————file begins below
280 2800 0 2800
    1    -2    13     0     0   695 0.924977
    1     3     7     0     6   671 0.922147
    2     7     8     0     6   552 0.904269
    1    -6    28     0     6   510 0.895768
    3    -8   -13     0     0   462 0.883881

   ⋮
——————————————————————————————————————————————
```

The first line of the file contains, in order, the number of phases used to generate the invariants, the number of triples in the file, the number of negative quartets in the file, and the total number of invariants. These four integers are given in free format.

The subsequent lines, in format 6I5, F9.6, contain the set of triples followed by the set of quartets, where the following information is provided for each invariant:

REF1 = serial number of first reflection
REF2 = serial number of second reflection
REF3 = serial number of third reflection
REF4 = serial number of fourth reflection (negative quartets only)
PS $= \frac{\text{invariant phase shift}}{30}$ (in degrees)
NW = invariant weighting factor (A-value for triples; B-value for negative quartets), as an integer, stored as $100 * \text{value}$.
EC = expected invariant cosine value

Serial number i corresponds to the reflection with the i^{th} largest $|E|$-value. A 'triple' is a linear combination of phases, $\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}}$, with $\underline{h} + \underline{k} + \underline{l} = \underline{0}$. Converting all reflections to canonical form (thereby eliminating redundant phases) requires the total phase-shift, PS, where $\phi^c_{\underline{h}} + \phi^c_{\underline{k}} + \phi^c_{\underline{l}} + PS$ is the new form of the triple. A similar construction occurs for a negative quartet, but with four reflection index vectors $(\underline{h}, \underline{k}, \underline{l}, \underline{m})$, instead of three.

structure.**trials**

This file contains the coordinates of the trial structures. We recommend using *SnB* to generate initial trial structures, although the user does have the option of supplying their own initial structures, which may be useful in terms of fragment recycling. The format of the file follows.

```
     X        Y        Z
——————————————————————————————file begins below
100 28
0.529053 0.723768 0.807909
0.210766 0.965908 0.606480
0.928481 0.964337 0.884053
0.718052 0.977337 0.964482
0.484910 0.326196 0.009337

   ⋮
——————————————————————————————
```

The first line of the file, containing two integers in free format, contains the number of trial structures contained in the file and the number of atoms per structure, respectively. Each atom, with format 3F8.6, is given as a three-tuple (X,Y,Z), where each coordinate is represented as the fractional distance of the unit cell-length in that axis direction.

structure.**info**

This file is created by *SnB* after the user initially enters information about a particular structure, including cell constants, space group, parameters for operation, and so forth. This file **should not be modified** by the user. Once the information is stored in this file, innumerable runs can be initiated with it by appropriately revising information through the queries in option 1 (the structure determination component) of the program.

5.2 Output Files

SnB may produce eight (8) files ('*structure*' is a user-defined ID):

structure.SnB_review	*structure*.SnB_times
structure.SnB_output	*structure*.SnB_trace [available in expert mode]
structure.SnB_Rmins	*structure*.SnB_struc [available in expert mode]
structure.SnB_best	*structure*.SnB_list [used by Geomview]

structure.**SnB_review**

When a run is initiated, certain parameters that are useful in identifying specific characteristics of that particular run are written to this file in a user friendly format.

structure.**SnB_output**

This file contains information pertaining to the progress of an *SnB* run and is periodically updated by the program.

structure.**SnB_Rmins**

This file contains the final *minimal function* value for each completed trial structure. *SnB* updates this file at the completion of every trial. This file is used by *SnB* to provide a histogram of results at any stage of processing. (See Option 2 in Section 8.)

structure.**SnB_best**

This file contains the current lowest *minimal function* value over all trials processed so far for the given run, along with the corresponding initial and final structures,respectively. *SnB* updates this file at the completion of every trial. This information is used when examining the geometry of a best final structure. (See Option 3 in Section 8.)

structure.**SnB_times**

This file contains timing statistics covering a variety of computational modules within the *Shake-and-Bake* procedure of the *SnB* code. After all trials have been processed and the run is complete, the amount of CPU time required to process each trial is given with respect to the following:

Struc Factor	=	structure factor calculation routine
Minimal Funct	=	*minimal function* evaluation function
Phase Refine	=	phase refinement routine
FFT	=	electron density map FFT
Peak Pick	=	peak picking routine

Further, the total cpu time for each trial is given, as well as the average times per cycle for each trial, and the total cpu elapsed running time.

structure.**SnB_trace [available in expert mode]**

This file contains a record of the *minimal function* values for every trial structure at each cycle of the *Shake-and-Bake* process. Note that *SnB* updates this file at the completion of every trial.

structure.**SnB_struc [available in expert mode]**

This file contains the final *minimal function* value, as well as the initial and final structures, for every trial processed.

structure.**SnB_list [used by Geomview]**

This file is created when the user opts to use Geomview to view the best structure for a given run and will be overwritten each time that a structure is viewed. It contains a list of directives to Geomview and should not be edited or changed by the user.

5.3 Sample Files

In the subdirectory 'SnB_v1.5.0/Examples/pr435', there exists a sample *.ref* and *.info* file for pr435, a 28 atom structure in $P2_12_12_1$. Please place a copy of pr435.ref and pr435.info in a clean directory for experimentation. The pr435.info file can be used to run the program as follows:

```
               SnB Structure Determination Procedure

Please enter your name []: steve

Would you like to work in expert or novice mode [novice]: expert

Do you want to use information from a previously stored structure [no]: y

Existing information files:

pr435

Please enter the name of the information file that you wish to use or
enter "quit" to go back to the main menu [pr435]: <RETURN>
```

The program automatically searches the current directory for files with a '.info' extension and lists the possible choices, giving the most recently modified file as the default. The user may then choose one of these files or return to the main menu if a mistake was made.

The subdirectory SnB_v1.5.0/Examples/pr435 also contains a file with a sample histogram for this structure, indicating the range of *minimal function* values and the general characteristics one would expect of a typical run. Note: you may not reproduce the results exactly due to differences in compilers, CPUs, and random number seeds.

6 *SnB* Help Facility

The *SnB* package provides an extensive on-line help facility. At any prompt, enter either '?' or 'help' to elicit a help screen addressing the topic pertinent to the input required at the prompt. After viewing the help message, the user will be returned to the prompt at which help was invoked.

7 *SnB* Default Parameters

The following table serves as a guideline of default values for the basic parameters used in the *Shake-and-Bake* structure determination proce dure. These values, which have been experimentally determined, are targeted at minimizing the “time-to-solution” for an unknown structure.

Parameter	**Default**
Atoms in asymmetric unit	n
Number of phases	$10n$
Number of triples	$100n$
Number of negative quartets	0
Number of trials to generate	1000
Number of trials to process	100
Phase Refinement Routine	Parameter Shift
Number of *SnB* cycles for parameter shift, binary search, and centrosymmetric search	$\lceil n/2 \rceil$
Number of *SnB* cycles for tangent refinement	$\lceil n/5 \rceil$
Number of E-Fourier filtering cycles	2
Number of peaks to select from the electron density map if the number of atoms, n, is greater than 300	$0.8n$

8 Running *SnB*

After the installation has been performed, the *SnB* program may be run within a directory containing a reflection file by typing 'SnB' at the UNIX prompt. This produces the following title screen:

```
                              SnB

      A Program for Crystal Structure Determination by Shake-and-Bake

           COPYRIGHT 1993-1996 by Russ Miller and Charles M. Weeks

                           Version 1.5.0


Principal Contributors: C.-S. Chang (SUNYAB), G.T. DeTitta (HWI),
  S.M. Gallo (SUNYAB), H.A. Hauptman (HWI), R. Jones (TMC), A. Khalak (HWI),
  H. Khalak (HWI), D.A. Langs (HWI), R. Miller (SUNYAB), S. Potter (HWI),
  P. Thuman (BCM), C.M. Weeks (HWI)


HWI is the Hauptman-Woodward Medical Research Institute
    (formerly the Medical Foundation of Buffalo)
SUNYAB is State University of New York at Buffalo
TMC is Thinking Machines Corporation
BCM is Baylor College of Medicine

Please hit RETURN to continue.
```

The user should press the 'Return' (or 'Enter') key in order to proceed to the MAIN MENU.

8.1 Main menu

```
                              SnB

        Crystal Structure Determination by Shake-and-Bake

      COPYRIGHT 1993-1996 by Russ Miller and Charles M. Weeks


 MAIN MENU

   1.  Initiate Shake-and-Bake on trial structures.
   2.  Produce a histogram of completed trial structures.
   3.  Display the current best trial structure.
   4.  List active Shake-and-Bake jobs.
   5.  Exit.

   Please enter your selection :
```

Enter the following:

1 to run the *Shake-and-Bake* structure determination procedure; This choice requires that a reflection file, as defined in Section 5.1, is available in the current directory for the structure under consideration. Also, the user must be prepared to enter the cell constants, contents of the asymmetric unit, and make choices (or use defaults) for some basic parameters of the *Shake-and-Bake* procedure. This option allows the user to create or use an existing set of invariants and/or initial trial structures.

2 to view a histogram of completed trial structures; Final *minimal function* values from a selected run are used to construct a histogram for the purpose of identifying structures that represent potential solutions.

3 to create a file which displays the geometry of the current best structure for a run. The user may select from; *i*) a text plot, which represents the final atoms corresponding to the best structure of a particular run, including bond distances and angles. This plot allows the user to manually "connect the dots." and is sent to a file so that it may be printed; *ii*) a graphical representation of the final atoms using Geomview. This method allows the user to view a 3-dimensional representation of the final atoms corresponding to the best structure which may be rotated and scaled in real time.

4 to display all currently running processes; In a selected directory, all of the currently executing *SnB* runs are listed.

5 to exit from the SnB program and return to the operating system prompt.

8.2 Option 1: Shake-and-Bake Structure Determination

```
                 SnB Structure Determination Procedure

 Please enter your name []: steve
```

Type in your user identification.

```
 Would you like to work in expert or novice mode [novice]:
```

This choice of modes allows for two levels of operation. The novice mode allows for a straightforward, simple and often quicker use. Expert mode involves more selections, with correspondingly greater control of the program for executing the *Shake-and-Bake* routine.

Example:

```
  Would you like to work in expert or novice mode [novice]: ex
```

By entering a string beginning with 'e' or 'E', the expert mode will be invoked. The novice mode will be invoked if the response begins with 'n' or 'N' or if the user simply hits the RETURN key to accept the default.

```
                 SnB Structure Determination Procedure

  Do you want to use information from a previously stored structure [no]:
```

If information about a structure has already been entered and saved, and the user wants to use this information, the user should type 'y' or 'Y'. The system will then display all available information files in the current directory and prompt the user to select one. Typing either 'n' or 'N' will initiate a series of queries for generating the cell/structure information from scratch, as shown in Section 8.2.1. Note that the default is 'no'.

Example:

```
Do you want to use information from a previously stored structure [no]: yes

Existing information files:

pr435

Please enter the name of the information file that you wish to use or
enter "quit" to go back to the main menu [pr435]:
```

The information file chosen, 'pr435', is located in the current directory. The contents of this file will be read, and the information displayed for the user to review.

8.2.1 New Structure

```
Do you want to use information from a previously stored structure [no]: no
```

In this section, sample dialogue is given and defined for inputting the cell constants and basic structure information for the user operating in the expert mode. The user operating in the novice mode will see a proper subset of these queries. Be aware that the user is not queried for some parameters depending on the values of previous parameters. For instance, the user operating in expert mode is not queried as to whether or not to exploit heavy atoms if there are no such atoms (i.e., atoms with atomic weight greater than 10).

8.2.1.1 Identify Reflection File

```
                 SnB Structure Determination Procedure

The current directory is "/usr2/home/sgallo/Datasets/pr435"

  Enter the Unix path from this directory to the directory containing
  the reflection file for the structure you wish to process.

Enter "quit" to return to the main menu.
Path [./]:
```

Enter the path (relative to the current directory) in which the necessary data files (reflection, invariant, trial) can be found. Note that the default is the current directory (the one from which this *SnB* program was executed). Typing 'quit' will exit from this option and return the user to the main menu.

```
Existing reflection files:

pr435

Please enter the name of the reflection file that you wish to use or
enter "quit" to go back to the main menu [pr435]: <RETURN>
```

The program will search the specified directory for all reflection files and display them in alphabetical order, giving the most recently modified file as the default. The user may then enter the name of the file corresponding to the structure they wish to process. This name is used to identify files containing data for the specific structure under investigation. In the above case, the files for structure 'pr435' will be used. Typing 'quit' would have returned the user to the main menu.

```
Confirm that this is what you want [yes]: <RETURN>
```

Confirm your selection of the file prefix. If you entered the information correctly, just press RETURN (default 'yes'), otherwise, type 'no', and you will able to choose another structure name (i.e., file prefix).

8.2.1.2 Basic Crystallographic Input

```
Enter space group [P212121]: <RETURN>
```

Type the name of the space group for the structure under consideration as it appears in the *International Tables for X-Ray Crystallographic*, Volume A, with the exception that superscripts and subscripts appear as normal text (e.g., $P2_12_12_1$ is entered as P212121) and $P\bar{1}$ is entered as P-1. Please make sure that upper/lower-case letters are accurate. Only the 230 standard space groups are allowed (e.g., do not use $P21/a$ or $P21/n$).

NOTE: The b-axis must be the unique axis for monoclinic structures and rhombohedral structures should be referred to hexagonal axes.

Examples:

```
Enter space group [P212121]: p212121

  ** Invalid space group, please re-enter.

Enter space group [P212121]: <RETURN>   [correct]

Enter space group [P212121]: P-1        [correct for 'P 1 bar']
```

```
Enter the cell constants as decimal numbers.  Give angles in degrees.
A     : 13.1
B     : 14.075
C     : 10.658
ALPHA : 90.
BETA  : 90.
GAMMA : 90.
```

Enter lengths (A,B,C) of and angles (alpha, beta, gamma) between the unit cell axes. Be sure to enter a REAL number (i.e., one with a decimal point), and not an integer, for each of the six values.

Example:

```
Contents of the asymmetric unit (eg., O18,N6,C60,H102) : O6,C22,H32
```

Type in the 'empirical formula' of the asymmetric unit, using the element symbols (upper-case only) found in the periodic table, along with the number of individual atoms, separated by commas. In order to handle the situation where there are a fractional number of atoms in the asymmetric unit (e.g., atoms on boundaries, etc.), the number of atoms may be entered in either integer or decimal format.

Example:

```
Contents of the asymmetric unit (eg., O18,N6,C60,H102) : o18,H66,C5

  ** Please re-enter the cell contents (upper-case only)
     Numbers of atoms may be in integer or fractional format
     (i.e.  O18,N6.5,C40,H102.0)

Contents of the asymmetric unit (eg., O18,N6,C60,H102) : O18,H66,C5
```

Example:

```
Contents of the asymmetric unit (eg., O18,N6,C60,H102) : O18.5,H66,C5.2
```

8.2.1.3 Invariant File and Associated Parameters

```
Generate new invariant set [yes]: <RETURN>
```

Type 'y' or 'Y' to generate an invariant set at the outset of execution. (Note: the invariant set will not actually be generated until all of the parameters are set up, and execution begins.) If the user is entering information from scratch, the default is to generate a new invariant file.

```
Save invariant set to a file [yes]: <RETURN>
Enter the file prefix to use when saving invariants [pr435]: <RETURN>
```

After the program generates the invariants, it is often useful to save the invariant file for future use. The naming convention used throughout SnB employs the '.inv' suffix for invariants. Also, by default, the file is located in the current directory with the other input data files.

```
Enter the number of phases to be used [280]: <RETURN>
```

Enter the number of reflections to be used in generating the invariant file. In addition, this is the number of phases that will be used during the structure factor calculation, *Minimal Function* calculation, and Fourier transform. Make sure that the reflection data file ('.ref') contains enough reflections to accommodate your choice. An error message will be displayed if there are not enough phases in the reflection file, and the user will be prompted for the value again. Note that the experimentally determined default value for the number of phases is 10 times the number of non-hydrogen atoms in the asymmetric unit. E.g., for pr435, a 28-atom structure, one might select 280 phases.

```
Enter the number of triples to be generated [2800]: <RETURN>
```

Enter the number of triple invariants (triples) to be generated and stored in the invariant file. The initial run will use this number of triples. During subsequent runs, however, the user has the option of using a subset of the triples in the invariant file. The experimentally determined default value is 10 times the number of reflections. E.g., for pr435, a 28-atom structure the default combination is 280 phases and 2800 triples.

```
Enter the number of negative quartets to be generated [0]: <RETURN>
```

Enter the number of negative quartet invariants to be generated and stored in the invariant file. The initial run will use this number of negative quartets. During subsequent runs, however, the user has the option of using a subset of the negative quartets in the invariant file. Based upon experimentation, it has been determined that *the use of quartets is not cost-effective*. However, if one wishes to use quartets, we suggest using a number of quartets equal to 100 times the number of reflections. E.g., for pr435, a 28-atom structure, if one wishes to use quartets, we suggest that a combination of 280 phases, 2800 triples, and 28000 negative quartets be used.

If the user chooses to use an existing invariant file, they will be presented with a list of such files and asked to choose one. Once an invariant file has been chosen, the number of phases used to generate the file will be displayed along with the number of triples and negative quartets that it contains. The user will then be prompted for the number of triples and quartets to use in the computation. If the number of triples or the number of quartets in the file is zero, the user will not be prompted for that value. An error message will be displayed and the user will be prompted again if they request to use more invariants than are contained in the file.

Example (expert mode):

```
Generate new invariant set [yes]: no

Existing invariant files:

pr435                              pr435.398
pr435.338                          pr435.910

Please choose an invariant file or enter "quit" to return
to the parameter modification menu [pr435]:  <RETURN>

This invariant file was generated using 280 reflections.
There are 2800 triples and 0 quartets in this file.

Enter the number of triples to use [2800]: <RETURN>
```

8.2.1.4 Density Modification Parameter

```
Enter the number of peaks to select [28]: <RETURN>
```

Enter the number of peaks to be selected from the electron density map during each cycle of the Shake-and-Bake program. In general, the default (i.e., the number of non-hydrogen atoms in the asymmetric unit) is sufficient, although for larger structures it is often advisable to select 80% of the default. For example, for a 300 atom structure, it may be advisable to select only 240 atoms during the *Shake-and-Bake* process and attempt to pick up an additional 60 or so atoms during the E-Fourier recycling stage, as described in Section 8.2.1.8

8.2.1.5 Generate Initial Trial Structures

```
Generate initial trial structures [yes]: <RETURN>
```

Type 'y' or 'Y' to generate trial structures or type 'n' or 'N' to use an existing trial structure file. If an existing file is to be used, the user will be presented with a list of valid files and asked to choose one.

Example:

```
Existing initial trial structure files:

pr435

Please choose a trial file to use or enter "quit" to return to the
parameter modification menu [pr435]: <RETURN>
```

In order to generate random trial structures, several parameters are required, including the total number of trials to generate and the name of the file to store the trials in. In addition, we suggest using the following default values (as provided) for the associated trial structure parameters.

Number of trials to generate: 1000
Random number seed: 11909
Starting atoms per trial: n (for n-atom starts)
Save random trials to file: <*structure*>

```
Number of trials to generate [1000]: <RETURN>
```

Enter the desired number of random trial structures to generate. Since the trials generated can be saved for later use, we recommend initially generating a large number of trials (e.g., 1000). It should be noted that the trials will not be generated until all parameters are entered and execution begins.

```
Enter an integer seed [11909]: <RETURN>
```

Enter an integer seed to the random number generator that is used to produce trial structures. This number should be less than 2^{32}. As with all random number generators, it is best to use a large prime number as a seed.

```
Atoms per initial trial structure [28]: <RETURN>
```

Enter the number of atoms to be generated for each of the initial trial structures. This choice typically approximates the number of atoms in the asymmetric unit cell. In any event, experimentation has shown that this number should be at least the minimum number of atoms necessary to fix the origin and enantiomorph in the space group under consideration.

```
Save initial trial structures to file [yes]: <RETURN>
Enter the file prefix for these trials [pr435]: <RETURN>
```

Type 'y' or 'Y' to save the trial structures to a data file. Unless only a small number of trials are being processed, we recommend saving the trial structures for subsequent processing. The name of the file can be used to identify different runs on the same structure. For example, a user generated trial file containing a possible molecule fragment might be called 'pr435.frag.trials'.

8.2.1.6 Trial Processing Information

```
Start processing at trial structure [1]: <RETURN>
```

Enter the number of the first trial structure to be processed. This is useful for splitting a large job over several machines where each machine will process a specific subset of the trials.

```
Number of trials to process [100]: 500
```

Enter the number of trial structures you wish to process. This number is estimated from the expected success rate, which is in turn a function of the size of the structure, the space group, and quality of the diffraction data. If a user selects 0 trials to process, then only the invariant and trial structure generation, as well as the reflection file processing, if necessary, will be performed.

```
Enter the number of Shake-and-Bake cycles to be performed [15]: <RETURN>
```

Enter the number of cycles (structure factor calculation, phase refinement, FFT/peak picking) that *each* trial will proceed through. The experimentally determined default value is equal to approximately one-half the number of atoms in the asymmetric unit when parameter shift, binary search, or centrosymmetric search is used and approximately one-fifth the number of atoms when using tangent refinement.

```
Exploit heavy atoms [yes]: <RETURN>
```

This option is concerned with the weight assigned to the atoms during the structure factor calculation. The user may choose to assign uniform weights to all atoms by answering negatively ('n' or 'N' will suffice). By answering in the affirmative ('y' or 'Y' suffices), all atoms with atomic weights greater than 10 are assigned their proper weights, while all atoms of weight 10 or less are assigned a weight of 6. Experimentation has shown that for structures containing atoms with atomic weights greater than 10, it is most effective to exploit this information in the structure-factor calculation. The user will not be queried for this parameter if there are no atoms with an atomic weight greater than 10.

8.2.1.7 Phase Refinement

```
Number of phases to actively refine [280]: <RETURN>
```

Enter the number of phases to refine during the phase refinement stage. The default is the number used to generate the invariant file, but the user may wish to refine only a subset of these phases.

```
    1 - Parameter Shift
    2 - Tangent Refinement
    3 - Binary Search
    4 - Centrosymmetric Search

Enter refinement method [1]: <RETURN>
```

For each of the phase refinement techniques, the phases are considered in decreasing order according to the values of the associated $|E|$'s.

1. **Parameter shift** is a deceptively simple search technique which has proven to be quite powerful when the appropriate choices of parameter values are made. This routine can be briefly described as follows. When considering a given phase θ_i, the value of the *minimal function* is evaluated three times. First, with the given set of phase assignments, second, with phase θ_i modified by the addition of the predetermined phase shift, and third, with θ_i modified by the subtraction of the predetermined phase shift. If the first evaluation yields the minimum of these three values of the *minimal function*, then consideration of θ_i is complete, and the algorithm proceeds to θ_{i+1}. Otherwise, the direction of search is determined by the modification that yields the minimum value, and the phase is updated to reflect that modification. In this case, phase θ_i continues to be updated by the predetermined phase shift in the direction just determined, so long as the value of the *minimal function* is reduced. Note, however, that there is a predetermined maximum number of times that such a shift is attempted. Default values are presented based on extensive experimentation in terms of structure size and space group. See Figure 5.

   Parameters:
   - number of complete passes through the phase set
   - maximum number of attempted phase shifts per phase
   - phase shift for **each** pass

2. **Tangent refinement** applies the traditional tangent formula as a means of performing phase refinement. For each phase which is involved in a number of triplet relationships, its new value is determined by accumulating the summation of the remaining two phases in each triplet and applying the tangent formula. Once the new value of a phase is determined, either the phase set is updated immediately before determining the value of the next phase (termed *feedback*), or the value of the phase is stored for updating after all phases have been processed. The decision as to whether or not to use feedback is made by the user. Notice that if feedback is chosen, the new value of a phase is immediately available for subsequent phase evaluations.

   Parameters:
   - number of complete passes through the phase set
   - whether or not to use feedback

3. **Binary search** is a multiple iteration, single shift variant of parameter shift, in which the shift size is halved in each subsequent iteration. Experimentation has shown that the cost-effective choice of parameters is to make five (5) complete passes through the phase set, with an initial phase shift of 120°. However, experimentation has also shown that parameter shift is generally a more cost-effective approach.

   Parameters:
   - number of complete passes through the phase set
   - initial phase shift (used in the first pass)

4. **Centrosymmetric Search** is applied in centrosymmetric space groups, where only a single shift of 180° is attempted for each phase.

```
Exploit restricted phases [no]: <RETURN>
```

This option is concerned with whether or not all phases should be treated as general during the phase refinement process. Theoretically, it would seem as if restricted phases in acentric space groups should be handled in a fashion similar to that of the centrosymmetric search. In practice, however, this is not necessarily true. Higher success rates have been obtained for $P2_12_12_1$ structures when all phases are treated as general phases. The default value 'n' is given based on repeated experiments which show that if restricted phases are only assigned their permissible values, then the percentage of trials that go to solution is reduced. This query is not presented when using centrosymmetric search or tangent refinement.

```
Enter number of complete passes through phase set [3]: <RETURN>
```

Enter a value corresponding to the number of complete passes made through the phase set during each invocation of a phase refinement procedure. Experimentation has shown that for Parameter Shift this value is typically three, for Tangent Formula this value is typically one, two or four, depending on the size of the structure and the space group, for Centrosymmetric Search this value is typically one, and for the global Binary Search procedure this value is typically at least three.

```
Enter maximum number of attempted shifts per phase [2]: <RETURN>

Enter the phase shift to be used during the single pass:
   Enter, in degrees, the phase shift for pass #1 [90]: <RETURN>
   Enter, in degrees, the phase shift for pass #2 [90]: <RETURN>
   Enter, in degrees, the phase shift for pass #3 [90]: <RETURN>
```

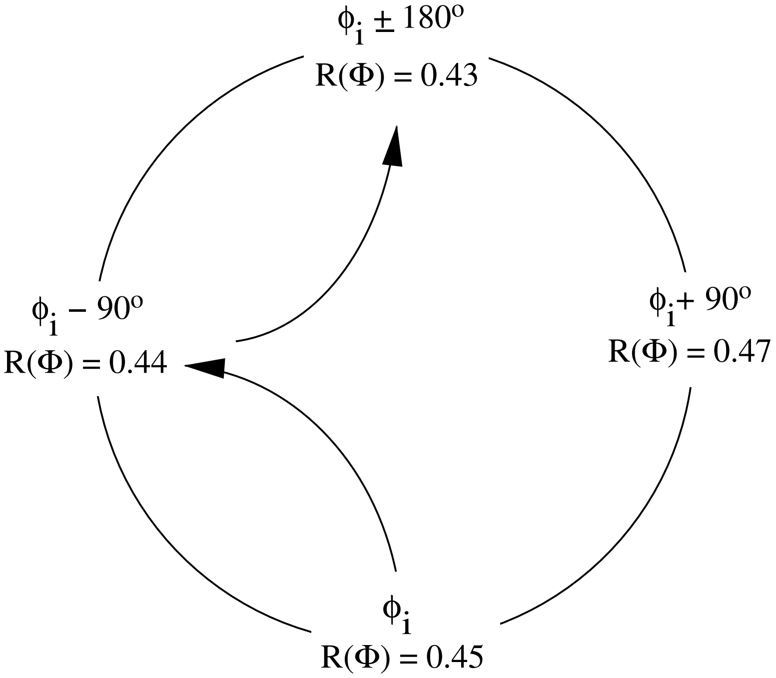

Figure 5: An example of Parameter Shift with a maximum of two 90° phase-shifts. Notice that initially, the *minimal function* is calculated with the current set of phases, yielding a value of 0.45. The *minimal function* is then re-evaluated at $\phi_i + 90°$ and $\phi_i - 90°$, yielding values of 0.47 and 0.44, respectively. Therefore, ϕ_i is updated to $\phi_i - 90°$. Finally, the *minimal function* is evaluated once more by subtracting another 90° from ϕ_i, which yields 0.43 and causes ϕ_i to be updated once again.

The parameter shift routine requires two unique parameters, namely, the maximum number of phase shifts per phase per pass, and the size of the phase shift used during each pass. For example,

```
Number of complete passes through phase set: 3
Number of attempted phase shifts per pass: 2
  Phase shift used in Pass #1:  90
  Phase shift used in Pass #2:  90
  Phase shift used in Pass #3:  90
```

corresponds to making three complete pass through the phase set, where each phase can be adjusted by either $-90°$, $90°$, or $180°$, during each pass.

8.2.1.8 E-Fourier Filtering

```
Number of E-Fourier filtering cycles [2]: <RETURN>

  Enter the number of peaks to select during E-Fourier filtering:
     Enter number of peaks to pick in cycle #1  [28]: <RETURN>
     Enter number of peaks to pick in cycle #2  [28]: <RETURN>
```

This option allows each structure that has resulted from a trial structure to be filtered through the selected number of cycles of structure factor, Fourier transform, and peak picking without the imposition of phase refinement. This procedure serves to clean up the initial map for the user and is useful in identifying solutions, via histograming, in difficult cases. The default is to perform 2 cycles of E-Fourier filtering. (The user has the option of selecting zero cycles of E-Fourier filtering, which effectively disables this option).

8.2.1.9 Output File Options

```
Keep trace file containing Minimal Function values [yes]: <RETURN>

Store final structures for all trial structures in a file [no]: y
```

These expert mode queries allow the user to decide which optional files should be maintained.

8.2.1.10 Remarks

After the queries have been answered, the user is presented with a summary screen of user accessible settings. At this point, the user has the ability to modify any of the choices before initiating a run. The following section describes the process of modifying old structure/run information, as well as saving it to a specified file.

8.2.2 Revisions and Execution

After either (*i*) reading in the information from the specified information file, or (*ii*) answering the queries presented in the previous section for a new structure, the user is provided with a synopsis of the choices made. An example of this screen is given in Figure 6.

If there are any modifications or revisions to be made, type 'y' or 'Y'. The user will be prompted for the item to revise. After all revisions are complete, the user should type either 'n' or 'N' in order to initiate the *Shake-and-Bake* procedure with the current information.

Select the number of the item to revise from the list. In some cases, several queries will be presented. The following example demonstrates changing a cell constant.

```
Enter the number of the item to change (0 to change nothing) [0]: 4

Enter the cell constants as decimal numbers.  Give angles in degrees.
A     [13.200000]: 13.1
B     [14.075000]: <RETURN>
C     [10.658000]: <RETURN>
ALPHA [90.000000]: <RETURN>
BETA  [90.000000]: <RETURN>
GAMMA [90.000000]: <RETURN>
```

Note that by typing <RETURN>, the existing information, which is provided as the default, is retained. Once complete, the revised list is displayed and the user is prompted for further modifications.

After all changes have been made, the new/revised information can be saved.

```
Would you like to save this information to a file? [y]: <RETURN>
```

After generating data or making revisions, it is often a good idea to save the information for future runs. Note that it is necessary to save this information if the user intends to use the structure visualization routine (presented in Section 8.4).

```
Enter the file prefix for this information [pr435]: <RETURN>
```

The *SnB* parameters will be saved to this file so they can be used again at a later time. The default file prefix is the same as that used for the reflection file.

```
             SnB Structure Determination Procedure

 1.  Reflection file: ./pr435.ref
 2.  Space group: P212121
 3.  Cell constants:
       A:   13.1000     ALPHA:   90.0000
       B:   14.0750     BETA :   90.0000
       C:   10.6580     GAMMA:   90.0000
 4.  Contents of the asymmetric unit: O6,C22,H32
 5.  Generate new invariant set: Yes
       Number of phases to use: 280
       Number of triples to use: 2800
       Number of negative quartets to use: 0
       Save invariants to file: ./pr435.inv
 6.  Generate random trial structures: Yes
       Number of trials to generate: 1000
       Random number seed: 11909
       Starting atoms per trial: 28
       Save random trials to file: ./pr435.trials
 7.  Trial processing information
       Number of trials to process: 100
       Beginning at trial number: 1
       Number of Shake-and-Bake cycles: 25
 8.  Exploit knowledge of heavy atoms: No
 9.  Refinement method: Parameter Shift
       Number of phases to actively refine: 280
       Exploit knowledge of restricted phases: No
       Number of complete passes through phase set: 3
       Number of attempted phase shifts per pass: 2
         Phase shift used in pass #1:  90
         Phase shift used in pass #2:  90
         Phase shift used in pass #3:  90
10.  Density Modification
       Number of peaks to select: 28
11.  E-Fourier filtering
       Number of E-Fourier filtering cycles: 2
         Number of peaks picked in cycle #1: 28
         Number of peaks picked in cycle #2: 28
12.  Optional information storage
       Keep trace file containing Minimal Function values: Yes
       Store all final structures in file: No

Would you like to make any changes [n]:
```

Figure 6: A sample of the synopsis provided in expert mode.

Once the structure/run information has been established, the user can proceed to initiate the execution of a run based on the specified parameters.

```
Would you like to run the program **now** with the parameters defined
as above? [y]: <RETURN>
```

If the user chooses not to initiate a run at this time, the following query is displayed.

```
Would you like to return to the main menu? [n]: <RETURN>
```

The user can now return to the main menu by answering in the affirmative, or to the revision list for additional modifications by answering in the negative.

A run can be processed in two ways: either continuing a previous run (appending results to existing files) or initiating a new one.

```
                 SnB Structure Determination Procedure

Do you wish to continue a previous run [no]: <RETURN>
```

The user may occasionally choose to process more trials representing a particular set of parameters, in which case the user should respond in the affirmative and provide the name of the file to which the program should append results to. For example, if this query is answered in the affirmative, the following will be displayed.

```
                 SnB Structure Determination Procedure

Existing SnB runs:

Jul09.1316                           run2
Jul10.1456                           test1

Please choose a file to which the results of this continuation will be
appended.  Enter "quit" to go back [Jul09.1316]: <RETURN>


Continuing previous run with file prefix 'Jul09.1316'...

Hit RETURN to go back to the main menu.
```

However, if a run is being initiated with a new set of parameters, only the path and prefix are required.

```
Please enter a prefix name that will be used to identify the
files associated with this run.  Enter "quit" to go back to
the main menu [Aug11.1240]: results_prefix
```

Enter the prefix name (including path relative to the current directory) that will be used to identify the files connected with this run. As a default file prefix, the user is provided with the current date and time. From the example above, the resulting output files will be generated:

	results_prefix.SnB_review	- summary of parameters
	results_prefix.SnB_output	- run information and error messages
	results_prefix.SnB_Rmins	- final *minimal function* values
	results_prefix.SnB_best	- best structure to date
	results_prefix.SnB_times	- running times per module
(expert option)	*results_prefix*.SnB_trace	- cycle/trial history
(expert option)	*results_prefix*.SnB_struc	- all final structures

8.3 Option 2: Producing a Histogram of *Minimal Function* Values

In order to produce a histogram of *minimal function* (R_{min}) values for a given run, a '.SnB_Rmins' file is required. The '.SnB_review' file corresponding to the '.SnB_Rmins' file is optional and if it is present some important information that it contains will be displayed in the histogram header. The script of a sample session and explanation follows.

```
The current directory is "/usr2/home/sgallo/Datasets/pr435"

  Enter the Unix path from this directory to the directory containing
  the data from which a histogram will be produced.

Path [./]: <RETURN>
```

Enter the path (relative to current directory) where results of SnB calculations can be found, specifically the '.SnB_Rmins' file of interest.

```
Files containing Minimal Function (Rmin) values:

Jul09.1316                         run2
Jul10.1456                         test1
results_prefix

Please choose a file to display or enter "quit" to return
to the main menu [Jul09.1316]: <RETURN>
```

Enter the prefix name of the file in the directory list to process. Note that a '.SnB_Rmins' file and a '.SnB_review' file are needed to produce a histogram.

Pressing RETURN exits this option and returns the user to the main menu.

```
Enter the number of histogram buckets [20]: <RETURN>
```

Enter an integer number of buckets in which to accumulate the *minimal function* data for the selected run. The first bucket will begin with the lowest *minimal function* value, and the last (n^{th}) bucket will end with the largest *minimal function* value. The output can be displayed on the screen or written to a file. The file prefix defaults to the selected run.

```
Would you like to see the histogram on screen [y]: n
Enter a name for the output file [results_prefix]: <RETURN>
```

Type 'y' or 'Y' to display a histogram of *minimal function* values on screen. Type 'n' or 'N' to send the histogram to a text file (name solicited).

A sample histogram is given in Figure 7.

Notice that a potential solution is typically recognized by a multimodal distribution consisting of a "bell-shaped curve" of high *minimal function* values, representing non-solutions, and one or more distinct groups of lower values, where the group with the lowest range of *minimal function* values may be regarded as potential solutions, and typically warrant investigation. (In the case of the histogram in Figure 7, the 69 or so trials in the first few buckets represent potential solutions.)

```
                              Structure Name: pr435
                              Prefix Filename: pr435


Number of Atoms: 28                        Number of Shake-and-Bake cycles: 30
Number of trials processed: 500            Number of Phases: 280
Lowest Rmin value:   0.272                 Number of Triples: 2800
Highest Rmin value:  0.582                 Number of Quartets: 0

                              Trials
Minimal Function Range       in range

   0.272  to   0.287        64       **********************
   0.288  to   0.303         3       *
   0.304  to   0.319         1       *
   0.320  to   0.335         1       *
   0.336  to   0.351         0
   0.352  to   0.367         0
   0.368  to   0.383         0
   0.384  to   0.399         5       **
   0.400  to   0.415        20       *******
   0.416  to   0.431        20       *******
   0.432  to   0.447        41       **************
   0.448  to   0.463        73       *************************
   0.464  to   0.479       103       ***********************************
   0.480  to   0.495        82       ****************************
   0.496  to   0.511        54       ******************
   0.512  to   0.527        22       ********
   0.528  to   0.543         5       **
   0.544  to   0.559         4       **
   0.560  to   0.575         0
   0.576  to   0.591         2       *
```

Figure 7: Sample histogram of *minimal function* results from a pr435 run.

8.4 Option 3: Displaying the Best Structure

```
                    SnB: Visualization

  1.  Generate a text plot suitable for printing on a line printer.
  2.  Display the best structure using GeomView (Graphical).
  3.  Return to the main menu.

Please enter your selection :
```

The user is provided with two options for viewing the current best structure: *i*) a text plot suitable for printing on a line printer or *ii*) a graphical method using GeomView, [4], suitable for an X-Windows environment. In order for the user to view the current best structure for a given run, a ".SnB_best" file is required. A script of a sample session for each display option and an explanation follows.

8.4.1 Display Best Structure Using a Text Plot

```
                    Display Bond Distances and Angles

You are currently in directory "/usr2/home/sgallo/Datasets/pr435"

   Enter the Unix path from this directory to the directory containing
   the data from which a plot will be produced.

Path [./]:
```

Enter the directory to search for the information needed to produce the plot. The default is the current directory.

```
Files available for display:

Jul09.1316                          run2
Jul10.1456                          test1
results_prefix

Please choose a file to display or enter "quit" to
return to the main menu [Jul09.1316]: <RETURN>
```

[4]Geomview was developed by the Geometry Center and a binary version is distributed with *SnB*. Geomview can be obtained by anonymous ftp at ftp.geom.umn.edu or on the World Wide Web at http://www.geom.umn.edu/

Choose the file prefix corresponding to the run to be examined.

```
Existing information files:

pr435

Please choose an information file related to your display file or enter
"quit" to return to the main menu [pr435]: <RETURN>
```

Enter the name of the previously stored file containing structure and parameter information. This file contains data which is necessary for computing the geometry of the structure under consideration.

```
Please enter a file prefix used to represent the output of this
procedure [results_prefix]: <RETURN>
```

Enter the prefix of the file to which a text plot of the best structure of the selected run will be sent. The default name is the name of the run with '.SnB_display' appended to it, which will be written in the current directory.

```
Enter a description of the output [pr435 plot]: <RETURN>
Number of connected blocks of atoms [4]: <RETURN>
Minimum number of atoms defining a connected block [4]: <RETURN>
Maximum distance for connected atoms in angstroms [1.750000]: <RETURN>
```

Structure plot parameters:

Title of plot:	header for output file
Number of connected blocks of atoms:	maximum number of allowable segments
Min. # of atoms/block:	minimum segment size (atoms)
Max. distance within blocks:	maximum bond distance permitted between pairs of atoms

```
 * The output has been placed in a file called './results_prefix.SnB_display'
 * In order to view this file, exit the main menu and type
 * "more ./results_prefix.SnB_display" at the operating system prompt.

Hit RETURN to go back to the main menu.
```

The output file will contain a textual representation of the best structure of this run, including the atoms and molecular segments, but excluding all hydrogen atoms.

8.4.2 Display Best Structure Using Geomview

```
                    Display Structure using GeomView

You are currently in directory "/usr2/home/sgallo/Datasets/pr435"

  Enter the unix path from this directory to the directory containing
  the data from which a plot will be produced.

Path [./]:
```

Enter the directory to search for the information needed to produce the plot. The default is the current directory.

```
Files available for display:

Jul09.1316                    run2
May06.1241

Please choose a file to display or enter "quit" to
return to the main menu [May06.1241]:
```

Choose the file prefix corresponding to the run to be examined.

```
Existing information files:

pr435

Please choose an information file related to your display file or enter
"quit" to return to the main menu [pr435]:
```

Enter the name of the previously stored file containing structure and parameter information. This file contains data which is necessary for computing the geometry of the structure under consideration.

```
Maximum distance for connected atoms in angstroms [1.750000]:
```

Maximum bond distance permitted between pairs of atoms.

```
Display atoms in addition to bonds [y]:
```

The default is to display the atoms in addition to the bonds for structures with less then 75 atoms. Atoms are not displayed for larger structures to facilitate faster rendering.

```
Forking geomview...

Geomview is starting.

Please hit RETURN to continue
```

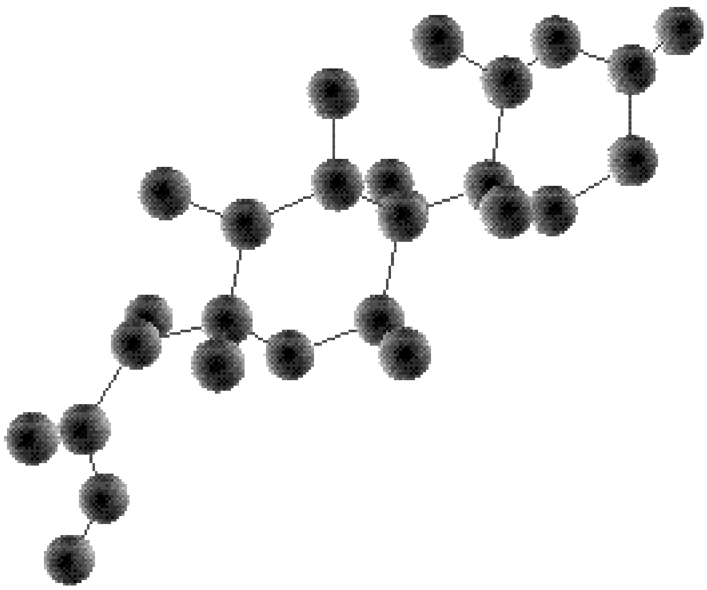

Figure 8: Sample Geomview output for pr435, a 28 atom structure in $P2_12_12_1$. Atoms are displayed as fuzzy circles.

8.5 Option 4: Displaying Active Processes

This option allows the user to determine whether or not any *SnB* runs are currently executing in a given directory. If such a process is found, the prefix is listed as one of the running processes.

```
                 Currently Running SnB Programs

 The current directory is "/usr2/home/sgallo/Datasets/pr435"

   Enter the Unix path from this directory to the directory containing
   the data files of interest.  Only those processes corresponding to data
   files in the selected directory will be displayed.

 Path [./]: <RETURN>
```

Enter the path (relative to current directory) to use when searching for files representing executing *SnB* processes. Note that the default is the current directory. Typing 'quit' will exit from this option back to the main menu.

For example, if *SnB* is processing files in the current directory, the following may be produced:

```
 There are SnB programs running using the following file prefixes:

         File prefix: 'Aug11.1515'
         File prefix: 'current1'
```

Alternatively, if there are no *SnB* processes executing, the program will display the following:

```
 ** There are no SnB programs currently running.
```

8.6 Option 5: Exit SnB

When the user has completed the necessary processing from within the *SnB* package, the session may be terminated by selecting option 5 from the main menu. This option requires confirmation in case of a typographical error.

```
Please enter your selection: 5


Do you really want to quit? (y/n) [n]: y
```

Note that pressing <RETURN> places the user back at the main menu of SnB while responding to the query in the affirmative will place the user back at the UNIX prompt in the directory from which SnB was executed.

```
                SnB Structure Determination Procedure

The current directory is "/usr2/home/sgallo/Datasets/pr435"

  Enter the Unix path from this directory to the directory containing
  the reflection file for the structure you wish to process.

Enter "quit" to return to the main menu.
Path [./]:

Existing reflection files:

pr435

Please choose a reflection file or enter "quit" to return
to the main menu [pr435]: <RETURN>

Confirm that this is what you want [yes]: <RETURN>

Enter the space group [P212121]: <RETURN>

Enter the cell constants as decimal numbers.  Give angles in degrees.
A    : 13.1
B    : 14.075
C    : 10.658
ALPHA: 90.
BETA : 90.
GAMMA: 90.

Contents of the asymmetric unit (eg., O18,N6,C60,H102): O6,C22,H32

Generate new invariant set [yes]: n

Existing invariant files:

pr435                         pr435.398ref

Please choose an invariant file or enter "quit" to return
to the parameter modification menu [pr435]:

Generate initial trial structures [yes]: n

Existing initial trial structure files:

pr435                         pr435.test

Please choose a trial file to use or enter "quit" to return to the
parameter modification menu [pr435]: pr435.test

This file contains 1000 initial trial structures,
each of which consists of 28 atoms.

Start processing at trial structure [1]: <RETURN>
Number of trial structures to process [100]: <RETURN>
Number of Shake-and-Bake cycles to be performed [15]: 25
```

Figure 9: A sample query session in novice mode.