An Introduction to *SnB*

(Version 0.9)

Russ Miller

SUNY-Buffalo

Medical Foundation of Buffalo

Partial funding from NIH and NSF.
Computing from TMC, PSC, Intel, and NIH.

Principal Contributors:

C.-S. Chang	G.T. DeTitta	S.M. Gallo
H.A. Hauptman	R. Jones	H. Khalak
D.A. Langs	R. Miller	S. Potter
P. Thuman	C.M. Weeks	

Selected Publications

1. R. Miller, S.M. Gallo, H.G. Khalak, C.M. Weeks, SnB: Crystal structure determination via Shake-and-Bake, *Journal of Applied Crystallography*, in press.

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

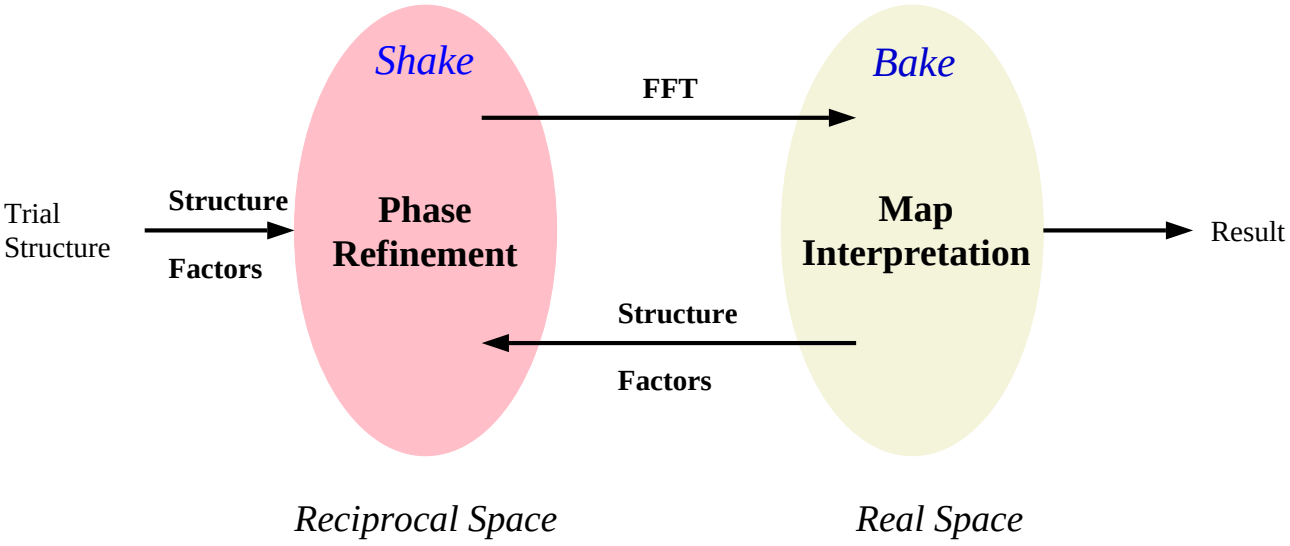
3. C.M. Weeks, G.T. DeTitta, H.A. Hauptman, P. Thuman, and R. Miller, Structure solution by minimal function phase refinement and Fourier filtering II: implementation and applications, *Acta Crystallographica* **A50**, 1994, pp. 210-220.

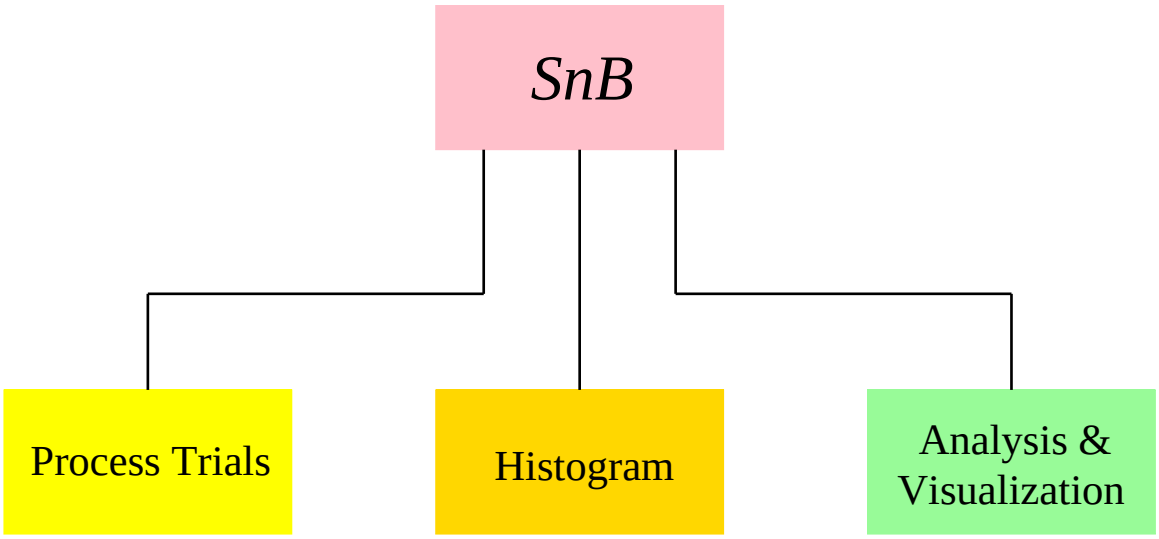
4. R. Miller, G.T. DeTitta, R. Jones, D.A. Langs, C.M. Weeks, and H.A. Hauptman, On the application of the minimal principle to solve unknown structures, *Science*, vol. 259, March, 1993, pp. 1430-1433.

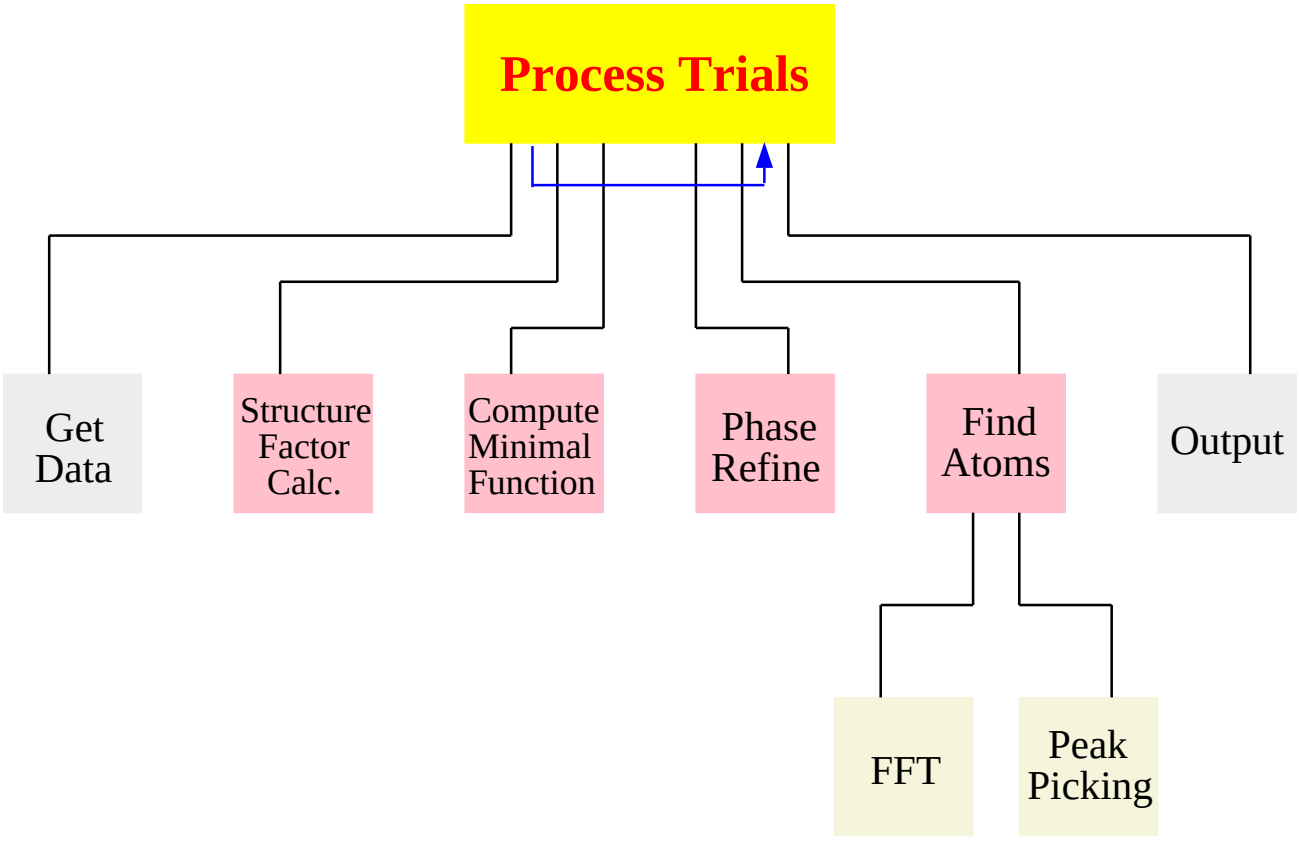
5. C.M. Weeks, G.T. DeTitta, R. Miller, and H.A. Hauptman, Applications of the minimal principle to peptide structures, *Acta Crystallographica* **D49**, 1993, pp. 179-181.

6. C.-S. Chang, G. DeTitta, H. Hauptman, R. Miller, P. Thuman, and C. Weeks, Using parallel computers to solve the phase problem of x-ray crystallography, *The International Journal of Supercomputer Applications*, vol. 7, no. 1, Spring 1993, pp. 25-49.

Shake-and-Bake




SnB
Process Trials
Histogram
Analysis &
Visualization


Process Trials
Get Data
Structure Factor Calc.
Compute Minimal Function
Phase Refine
Find Atoms
Output
FFT
Peak Picking

Shake-and-Bake

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, $R(\phi)$

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 $R(\phi)$
end Do

Store final $R(\phi)$ for histogram purposes

If $R(\phi) < R_{best}$
$R_{best} = R(\phi)$
Store atoms
end If

end For-Do

Critical Input Files

structure.ref

This is the only file required to be generated by the user. This file contains the reflections for the structure under consideration. For each reflection, the user must supply E, F, and σ(F).

structure.inv

This file contains the invariants generated from a subset of the reflections defined in the reflection file. It consists of triplets and quartets. Each invariant is given in canonical form and consists of a set of indices into the reflection list, the associated |A| or |B| value, the expected cosine of the invariant, and the phase shift.

structure.trials

This file contains a set of input trial structures. All trials in a given file consist of the same number of atoms. Every atom is represented by a set of fractional coordinates.

structure.ref

H	K	L	E	F	Sig(F)
1	-3	12	3.084	30.64	0.65
4	2	13	3.061	19.09	0.40
3	0	14	3.015	19.94	0.42
3	-1	8	2.974	44.06	0.91
3	1	14	2.873	18.47	0.39
5	-4	12	2.826	17.42	0.37
0	3	-13	2.813	25.63	0.54
4	-7	8	2.798	22.06	0.46
3	-5	10	2.764	25.28	0.53
5	4	5	2.749	29.74	0.63
6	2	8	2.736	21.96	0.47
6	-2	10	2.681	18.70	0.40
1	3	-6	2.675	66.79	1.35
6	-6	3	2.643	23.23	0.49
1	-8	-2	2.637	29.42	0.62
0	0	13	2.630	24.69	0.52
4	0	-2	2.612	69.33	1.47

Format: 3I5, F7.3, F9.2, F8.2

structure.ref

(updated by *SnB*)

```
1 0
    1   -3   12 0.4269  3.084    30.64    0.65 -1.318 0
    4    2   13 0.5415  3.061    19.09    0.40  2.281 0
    3    0   14 0.5267  3.015    19.94    0.42  1.206 0
    3   -1    8 0.3387  2.974    44.06    0.91  0.492 0
    3    1   14 0.5339  2.873    18.47    0.39 -0.647 0
    5   -4   12 0.5444  2.826    17.42    0.37  0.627 0
    0    3  -13 0.4466  2.813    25.63    0.54 -0.681 0
    4   -7    8 0.4825  2.798    22.06    0.46  3.544 0
    3   -5   10 0.4456  2.764    25.28    0.53  2.802 0
    5    4    5 0.4065  2.749    29.74    0.63  1.484 0
    6    2    8 0.4780  2.736    21.96    0.47  3.782 0
    6   -2   10 0.5128  2.681    18.70    0.40  4.154 0
    1    3   -6 0.2337  2.675    66.79    1.35 -1.003 0
    6   -6    3 0.4553  2.643    23.23    0.49  3.062 0
    1   -8   -2 0.3991  2.637    29.42    0.62 -0.477 0
    0    0   13 0.4398  2.630    24.69    0.52 -0.770 0
    4    0   -2 0.2214  2.612    69.33    1.47  0.763 0
```

structure.inv

ϕ	T	Q	Tot
740	7400	0	7400

R1	R2	R3	R4	PS	A/B	Exp_Cos
3	-9	-25	0	0	465	0.885229
7	10	-21	0	0	449	0.880821
2	-20	-25	0	0	432	0.875823
2	-17	-43	0	0	413	0.869579
6	13	-36	0	0	404	0.866603
1	17	-62	0	0	393	0.863894
4	-13	-57	0	0	393	0.863772
3	-5	124	0	0	382	0.857835
8	25	-27	0	0	376	0.854839
4	-5	127	0	0	376	0.854451
3	-8	123	0	0	373	0.852863
4	17	-69	0	0	372	0.852444
1	-4	235	0	0	363	0.847894
1	-12	134	0	0	360	0.846173
7	16	-71	0	0	353	0.842378
1	7	-207	0	0	352	0.841730
4	-14	114	0	0	351	0.841368
4	37	-51	0	0	350	0.840607
8	-9	-116	0	0	344	0.837680
4	-9	-172	0	0	342	0.836757
2	-23	-124	0	0	341	0.836067
12	-24	-52	0	0	340	0.835238

Format: 6I5, F9.6

structure.trials

```
Trls Atms
1000  74
    X         Y         Z
0.859624 0.805765 0.970482
0.023051 0.133481 0.850767
0.182767 0.410052 0.020193
0.400339 0.760908 0.766624
0.208911 0.325480 0.059339
0.122625 0.011623 0.798625
0.448195 0.018768 0.649339
0.929195 0.804623 0.000051
0.455337 0.716909 0.646768
0.996054 0.008767 0.007195
0.661051 0.528338 0.781909
0.724054 0.767053 0.126338
0.577051 0.168479 0.379909
0.045051 0.919765 0.005907
0.597196 0.003196 0.731196
0.115480 0.514053 0.606053
0.395767 0.653908 0.014481
0.758052 0.462481 0.732051
0.376196 0.201481 0.752339
0.812910 0.604337 0.407480
0.774768 0.642195 0.194622
0.521194 0.912624 0.447910
```

Critical Output Files

structure.SnB_review

SnB stores the parameter settings in this file for review by the user.

structure.SnB_Rmins

SnB stores the final minimal function value of every trial structure in this file. This file is updated upon the completion of every trial structure.

structure.SnB_best

This file contains the initial trial structure and final set of coordinates corresponding to the best trial structure (i.e., smallest final minimal function value) processed thus far.

structure.SnB_trace

This file gives the value of the minimal function at the end of every cycle of *Shake-and-Bake* for every trial structure.

structure.SnB_times

Running times of the trials are written to this file only after **all** trials are complete.

structure.SnB_review

```
Shake-and-Bake runtime parameter file

Prefix for files containing results: demo
Reflection file prefix: ph8755
Space group: P1
Invariant file used: ./ph8755.inv
Number of reflections used to generate file: 740
Number of triples used: 7400
Number of negative quartets used: 0
Number of atoms in structure: 74
Number of starting atoms in trial structures: 74
Number of trials to process: 100
Cycles of Shake-and-Bake to run: 40
Number of initial trial structure: 1
Random number seed for generating trials: 13579
Atoms saved to file: ./ph8755.trials
Refinement method: Parameter Shift
  Number of complete passes through phase set: 1
  Number of attempted phase shifts per pass: 2
    Phase shift used in pass #1: 90
Exploit restricted phases: No
Exploit heavy atoms: No
Number of peaks to use: 74
Contents of the asymmetric unit: O14,N9,C51,H83
Continuation of previous run: No
```

structure.SnB_Rmins

0.248248
0.249226
0.248746
0.400157
0.409556
0.249998
0.412340
0.404515
0.247349
0.403407
0.249182
0.262525
0.251926

structure.SnB_best

```
0.247349

0.996482 0.951480 0.025051
0.809336 0.509481 0.764482
0.519481 0.506624 0.876337
0.735624 0.066337 0.198196
0.937624 0.115622 0.944622
0.869195 0.763054 0.347194
0.755910 0.334194 0.399337
...
0.370765 0.724624 0.183622
0.479481 0.788766 0.446339
0.368196 0.840625 0.731622
0.740196 0.534337 0.176480
0.967908 0.090482 0.897053

0.185223 0.841127 0.288835
0.801047 0.877601 0.024816
0.724279 0.905735 0.958776
0.251702 0.830124 0.363324
0.847681 0.822354 0.287030
0.812136 0.555463 0.140590
...
0.072492 0.441646 0.074690
0.335448 0.801100 0.312223
0.468750 0.031250 0.359375
0.804165 0.192537 0.729534
0.375000 0.968750 0.812500
0.781250 0.750000 0.031250
```

structure.SnB_trace

```
# ./demo.SnB_trace

# Structure    1
# initial atom #1: 0.859624 0.805765 0.970482
# initial atom #2: 0.023051 0.133481 0.850767
# initial atom #3: 0.182767 0.410052 0.020193
...
# initial atom #72: 0.843624 0.823910 0.733767
# initial atom #73: 0.831054 0.974624 0.702625
# initial atom #74: 0.730908 0.577480 0.893482
#
# i       Rmin
  0:     0.761
  1:     0.543
  2:     0.478
...
 11:     0.380
 12:     0.372
 13:     0.367
 14:     0.316
 15:     0.247
 16:     0.246
 ...
 36:     0.248
 37:     0.248
 38:     0.250
 39:     0.248
 40:     0.248
```

structure.SnB_times

Cumulative Times for 40 cycles

Trial#	TOTAL	STRFAC	RCALC	SIMPLX	EDMAP
1	38.570	7.360	0.880	8.490	21.830
2	38.580	7.320	0.910	8.440	21.910
3	38.620	7.310	0.870	8.550	21.890
4	38.650	7.320	0.870	8.490	21.950
5	38.730	7.350	0.840	8.490	22.030
6	38.610	7.350	0.850	8.480	21.920
7	38.690	7.300	0.870	8.540	21.980
8	38.730	7.340	0.860	8.450	22.080
9	38.530	7.330	0.880	8.460	21.860
10	38.640	7.340	0.860	8.550	21.890
...					
95	38.670	7.350	0.880	8.490	21.940
96	38.610	7.310	0.900	8.470	21.920
97	38.690	7.340	0.860	8.490	22.000
98	38.800	7.350	0.910	8.490	22.040
99	38.610	7.360	0.860	8.500	21.890
100	38.750	7.410	0.890	8.510	21.940

Average times per cycle:

0.965	0.183	0.022	0.212	0.548

The average time per trial is 38 seconds. The total time to complete 100 trials using 40 cycles of Shake-and-Bake per trial is: 64.33 minutes.

SnB Default Parameters

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
Number of SnB cycles for parameter shift, binary search, and centrosymmetric search	$\lceil n/2 \rceil$
Number of SnB cycles for tangent refinement	$\lceil n/4 \rceil$

SnB

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

Version 0.9

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

MFB is 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.

SnB

Crystal Structure Determination by Shake-and-Bake

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 :

SnB

Crystal Structure Determination by Shake-and-Bake

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 : help

Main menu selection

Enter:

(1) To run the Shake-and-Bake structure determination procedure; This choice requires that a reflection file be available in the appropriate format. The user must be prepared to enter cell constants, contents of the asymmetric unit and make, choices (or use defaults) for some basic SnB parameters. In addition, this option allows the user to create a set of invariants (triples and/or quartets) and initial trial structures.

(2) To visualize a histogram of final minimal function (Rmin) values corresponding to a currently running or completed Shake-and-Bake process. The user may generate several histograms before returning to the main menu.

(3) To generate bond distances and angles corresponding to the best best structure available from a currently running or completed Shake-and-Bake process. A text plot of this final structure is created and saved to a file.

(4) To display any currently running Shake-and-Bake processes in the selected directory.

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

Press RETURN to resume operation.

SnB

Crystal Structure Determination by Shake-and-Bake

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 : 1

SnB Structure Determination Procedure

Please enter your name []: Miller

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

Do you want to use info. from a previously stored structure [no]:

The current directory is "/usr2/home/miller/Datasets/PH8755.demo"

  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:

ph8755

Please choose a reflection file or enter "quit" to return
to the main menu [ph8755]:

Confirm that this is what you want [yes]:

Enter the space group [P212121]: P1

Enter the cell constants as decimal numbers. Give angles in degrees.
A   : 9.13
B   : 10.508
C   : 15.077
ALPHA: 96.73
BETA : 99.29
GAMMA: 89.09

Contents of the asymmetric unit (eg., H2,0): O14,N9,C51,H83

Generate new invariant set [yes]:

Save invariant set to a file [yes]: y
Enter the file prefix to use when saving invars. [ph8755]:

Enter the number of phases to be used [740]:

Enter the number of triples to be generated [7400]:

Enter the number of negative quartets to be generated [0]:

```
Generate initial trial structures [yes]:

Number of trials to generate [1000]:

Enter an integer seed [11909]: 13579

Atoms per initial trial structure [74]:

Save initial trial structures to file [yes]:
Enter the file prefix for these trials [ph8755]:

Start processing at trial structure [1]:

Number of trial structures to process [100]:

Number of Shake-and-Bake cycles to be performed [40]:
```

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

Enter refinement method [1]: 1

Exploit restricted phases [no]:

Enter the number of complete passes through the phase set [1]:

Enter the maximum number of attempted shifts per phase [2]:

Enter the phase shift to be used during the single pass:
   Enter, in degrees, the phase shift for pass #1,  [90]:

Enter the number of peaks to use [74]:

Keep trace file containing Minimal Function values [yes]:

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

SnB Structure Determination Procedure

1. Search path: ./
2. Reflection file prefix: ph8755
3. Space group: P1
4. Cell constants:

   A: 9.1300 ALPHA: 96.7300
   B: 10.5080 BETA : 99.2900
   C: 15.0770 GAMMA: 89.0900
5. Contents of the asymmetric unit: O14,N9,C51,H83
6. Generate new invariant set: Yes
   Number of phases to use: 740
   Number of triples to use: 7400
   Number of negative quartets to use: 0
   Save invariants to file: ./ph8755.inv
7. Generate random trial structures: Yes
   Number of trials to generate: 1000
   Random number seed: 13579
   Starting atoms per trial: 74
   Save random trials to file: ./ph8755.trials
8. Trial processing information
   Number of trials to process: 100
   Beginning at trial number: 1
   Number of Shake-and-Bake cycles: 40
9. Exploit knowledge of heavy atoms: No
10. Refinement method: Parameter Shift
    Exploit knowledge of restricted phases: No
    Number of complete passes through phase set: 1
    Number of attempted phase shifts per pass: 2
        Phase shift used in pass #1: 90
11. Number of peaks to select: 74
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]:

Would you like to save this information to a file [y]:
Enter the file prefix for this information [ph8755]:

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

SnB Structure Determination Procedure

Do you wish to continue a previous run [no]:

Please enter a prefix name that will be used to identify the files associated with this run. Enter "quit" to return to the main menu [Jul04.2132]: demo

Running SnB with file prefix "demo"...

Please wait:
Processing reflections . . . done.
Generating invariants . . . . done.

Hit RETURN to go back to the main menu.

SnB

Crystal Structure Determination by Shake-and-Bake

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 : 2

Histogram Procedure

```
The current directory is "/usr2/home/miller/Datasets/PH8755.demo"

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

Path [./]:

Files containing Minimal Function (Rmin) values:

demo

Please choose a file to display or enter "quit" to return
to the main menu [demo]:

Enter the number of histogram buckets [20]:

Would you like to see the histogram on the screen [y]:
```

Reflection file name: ph8755.ref
Prefix Filename: demo

Number of atoms in structure: 74
Number of trials processed: 100
Lowest Rmin value: 0.243
Highest Rmin value: 0.429

Number of Shake-and-Bake cycles: 40
Number of phases used: 740
Number of triples used: 7400
Number of quartets used: 0

Minimal Function Range			Trials in range	
0.243	to	0.252	55	**********
0.253	to	0.262	8	****
0.263	to	0.272	0	
0.273	to	0.282	0	
0.283	to	0.292	0	
0.293	to	0.302	0	
0.303	to	0.312	0	
0.313	to	0.322	0	
0.323	to	0.332	0	
0.333	to	0.342	0	
0.343	to	0.352	0	
0.353	to	0.362	3	**
0.363	to	0.372	1	*
0.373	to	0.382	5	***
0.383	to	0.392	8	****
0.393	to	0.402	2	*
0.403	to	0.412	16	********
0.413	to	0.422	1	*
0.423	to	0.432	1	*
0.433	to	0.442	0	

Hit RETURN go continue.

SnB

Crystal Structure Determination by Shake-and-Bake

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 : 3

Display Bond Distances and Angles

You are currently in directory "/usr2/home/miller/Datasets/PH8755.demo"

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

Path [./]:

Files available for display:

demo

Please choose a file to display or enter "quit" to
return to the main menu [demo]:

Existing information files:

ph8755

Please choose an information file related to your display file or enter "quit" to return to the main menu [ph8755]:

Please enter a file prefix used to represent the output of this procedure [demo]:

Enter a description of the output [ph8755 plot]:
Number of connected blocks of atoms [4]:
Minimum number of atoms defining a connected block [4]:
Maximum distance for connected atoms in angstroms [1.75]:

* Output has been placed in a file called ’./demo.SnB_display’
* In order to view this file, exit the main menu and type
* "more ./demo.SnB_display" at the operating system prompt.

Hit RETURN to go back to the main menu.

ph8755 plot

```
 9.13000  10.50800  15.07700  96.73000  99.29000  89.09000
0.00000   1.   0.   0.   0.00000   0.   1.   0.   0.00000   0.   0.   1.
```

FROM THE INPUT LIST ATOMS REJECTED/ACCEPTED 0 74

NO./NAME		X	Y	Z	NO./NAME		X	Y	Z	NO./NAME		X	Y	Z
1	1	0.6831	0.5214	0.0137	2	2	0.7517	0.5129	0.0879	3	3	0.2258	0.5888	0.6820
4	4	0.3123	0.2374	0.8646	5	5	0.1041	0.4803	0.3483	6	6	0.4710	0.3072	0.7026
7	7	0.2265	0.5049	0.8180	8	8	0.5374	0.4662	0.8203	9	9	0.3496	0.5074	0.0106

CONNECTED PEAKS IN PASS 1

ITERATION 1. NUMBER PICKED 70

ITERATION 2. NUMBER PICKED 1

NO./NAME		X	Y	Z	NO./NAME		X	Y	Z	NO./NAME		X	Y	Z
1	1	0.6831	0.5214	0.0137	2	2	0.7517	0.5129	0.0879	3	3	0.2258	0.5888	-0.3180
4	4	0.3123	0.2374	-0.1354	5	5	0.1041	0.4803	-0.6517	6	6	0.4710	0.3072	-0.2974
7	7	0.2265	0.5049	-0.1820	8	8	0.5374	0.4662	-0.1797	9	9	0.3496	0.5074	0.0106
10	10	0.7966	0.2442	0.0831	11	11	0.2988	0.5592	-0.2523	12	12	0.0498	0.5324	-0.5116

BOND DISTANCES AND ANGLES

```
      1      2 1.20         1     57 1.45
      2     57   53


      2      1 1.20         2     16 1.45         2     46 1.41         2     57 1.22
      1     16  123         1     46  113         1     57   73        16     46  121        16     57   51
     46     57  172
```

<Parameter Shift vs. Tangent Formula slide>

<Cost-Effectiveness>

<Electron Diffraction Data (Known)>

<Fragment Recycling>