SnB: A Direct Methods Program

Russ Miller

Hauptman-Woodward Med. Res. Inst.

Principal Contributors:

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

Partial funding from NIH and NSF.

Outline of Talk

SnB

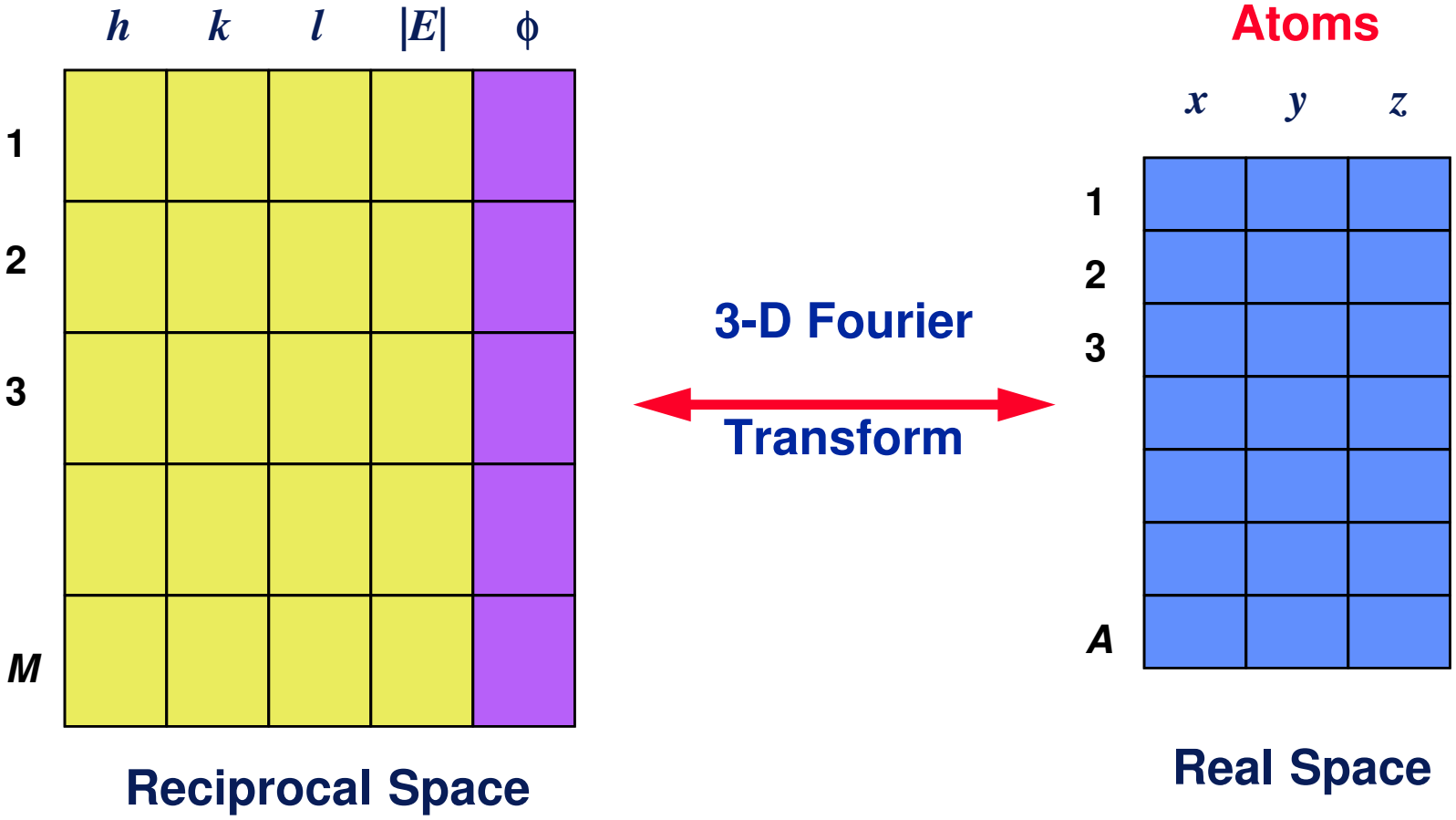
- Direct Methods
- *Shake-and-Bake:* The Method
  - Cyclical Dual Space Procedure
  - The Minimal Function
- *SnB:* The Program
- Applications of *SnB*
- *SnB* version 2.0

The Phase Problem

SnB

- **Experiment yields:**
  - **reflections**
  - **associated intensities**
- ***Phase angles are lost in experiment.***
- **Underlying atomic arrangement is related to the reflections by a 3-D Fourier transform.**
- ***Phase Problem*: determine the set of phases corresponding to the reflections.**

Data Structures



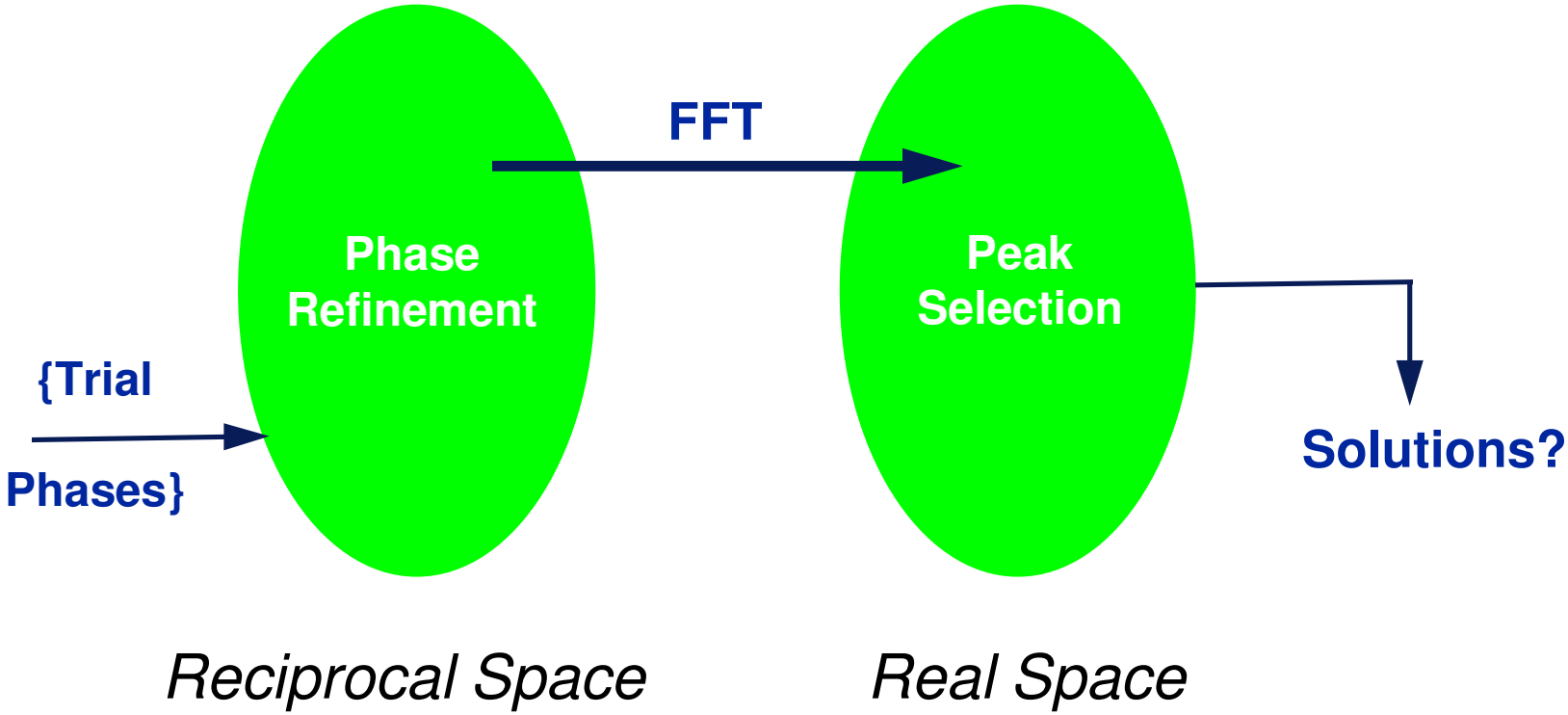
Direct Methods

SnB

- *Direct Methods* use probabilistic theories to exploit linear relationships among phases.
- Resolution of 1.2Å or better.
- Routinely applied to structures with 150 or fewer atoms.
- Standard packages:
  - SHELX
  - teXsan
  - SIR92/96

Conventional Direct Methods

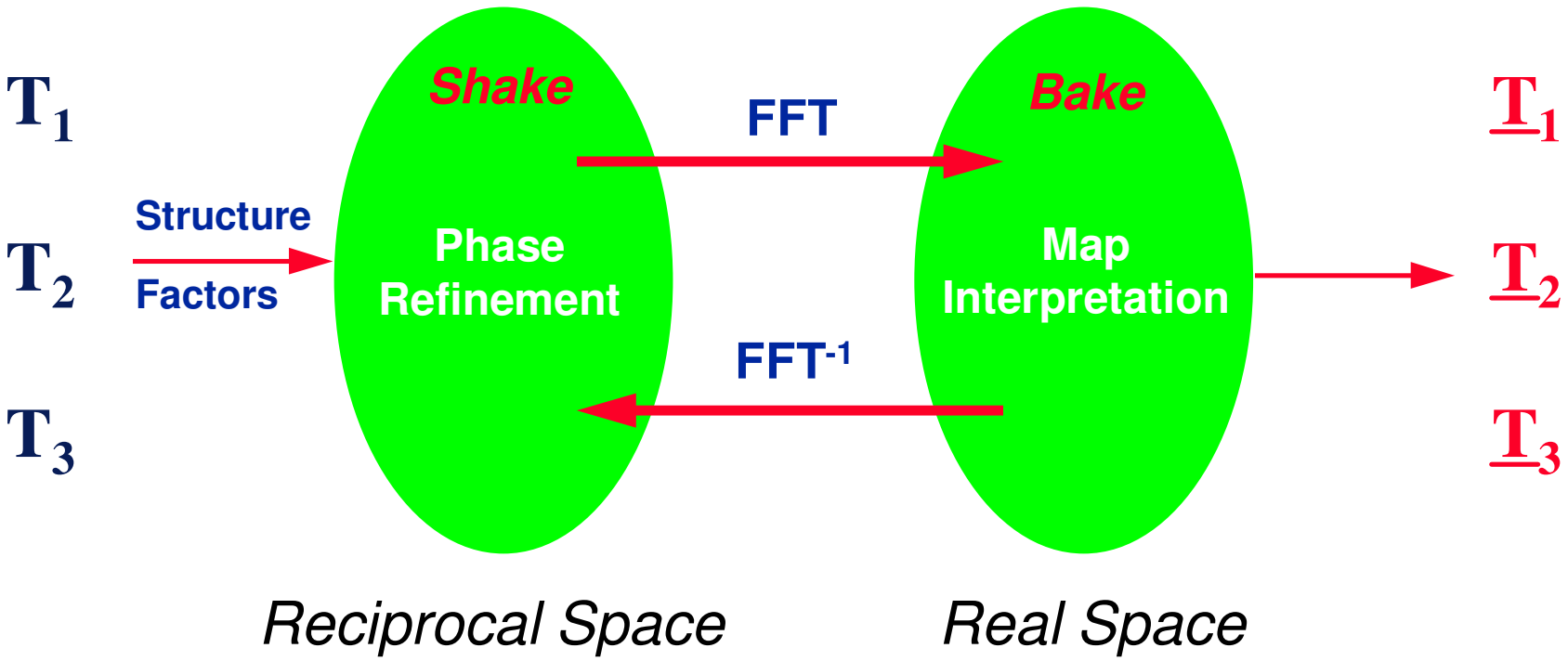
SnB



Shake-and-Bake

Shake-and-Bake

SnB



Shake-and-Bake

SnB

- Direct Methods Optimization Technique
- Multiple Trial Structures
- Real Space ⇔ Reciprocal Space
- Phase Refinement Techniques
  - Parameter Shift (Minimal Function)
  - Tangent Formula
- Minimal Function as FOM

The Minimal Function

SnB

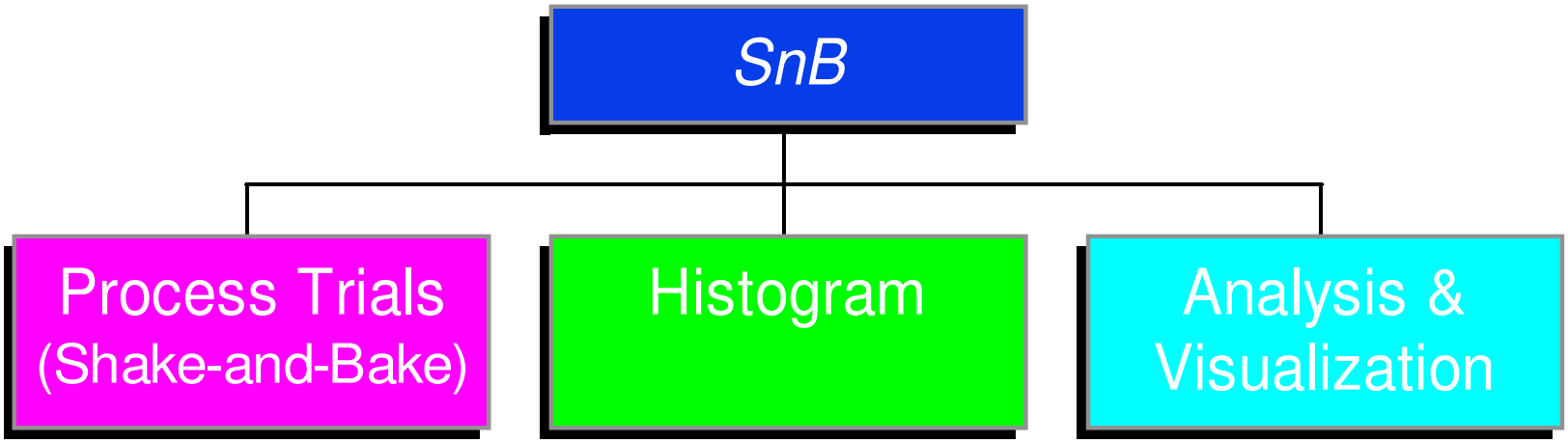
$$R = \frac{\sum_T W_T \left(\cos\phi_T - est_T\right)^2}{\sum_T W_T}$$

Triple: $\phi_T = \phi_h + \phi_k + \phi_{-h-k}$

$$W_T = \left(\frac{2}{N^{1/2}}\right) \left| E_h E_k E_{-h-k} \right|$$

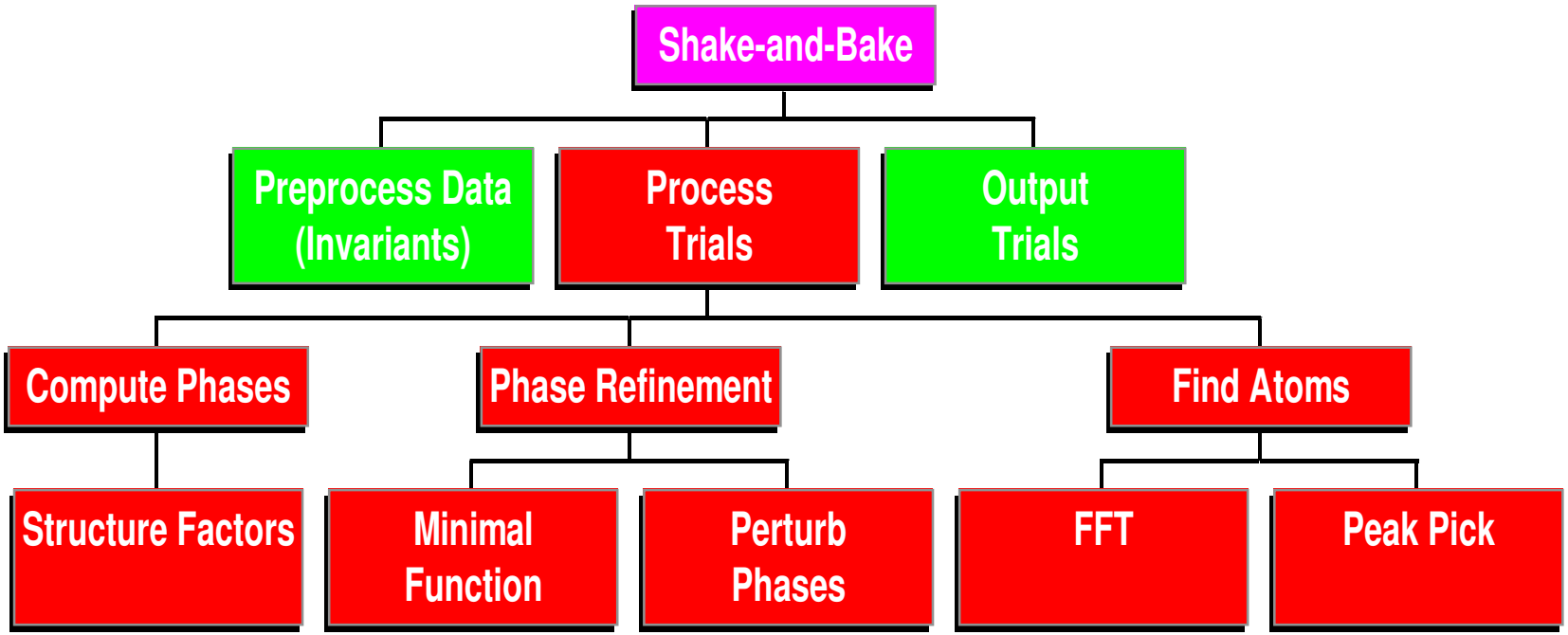
est_T is the known expected value of $\cos\phi_T$

Structure of *SnB*



Shake-and-Bake

SnB



SnB Parameters

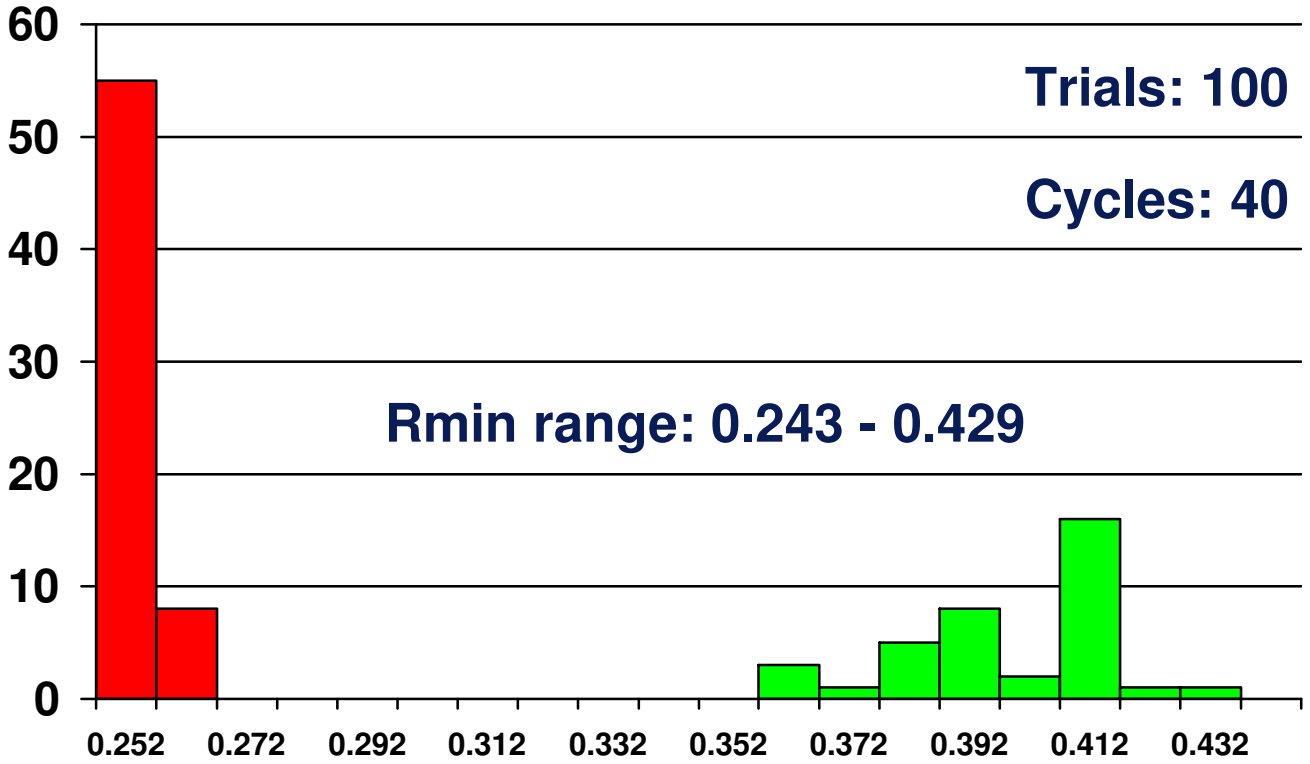
	Default	Ph8755	ToxII
Atoms (a.u.)	n	74	508
Phases	8n - 10n	740	5,000
Triples	70n - 100n	7,400	50,000
Cycles (PS)	n/2	40	255
Peaks recycled	0.8n - n	74	400
E-Fourier Steps	2	2	5

Sample Histogram (Ph8755)

SnB

Atoms: 74
Space Group: P1

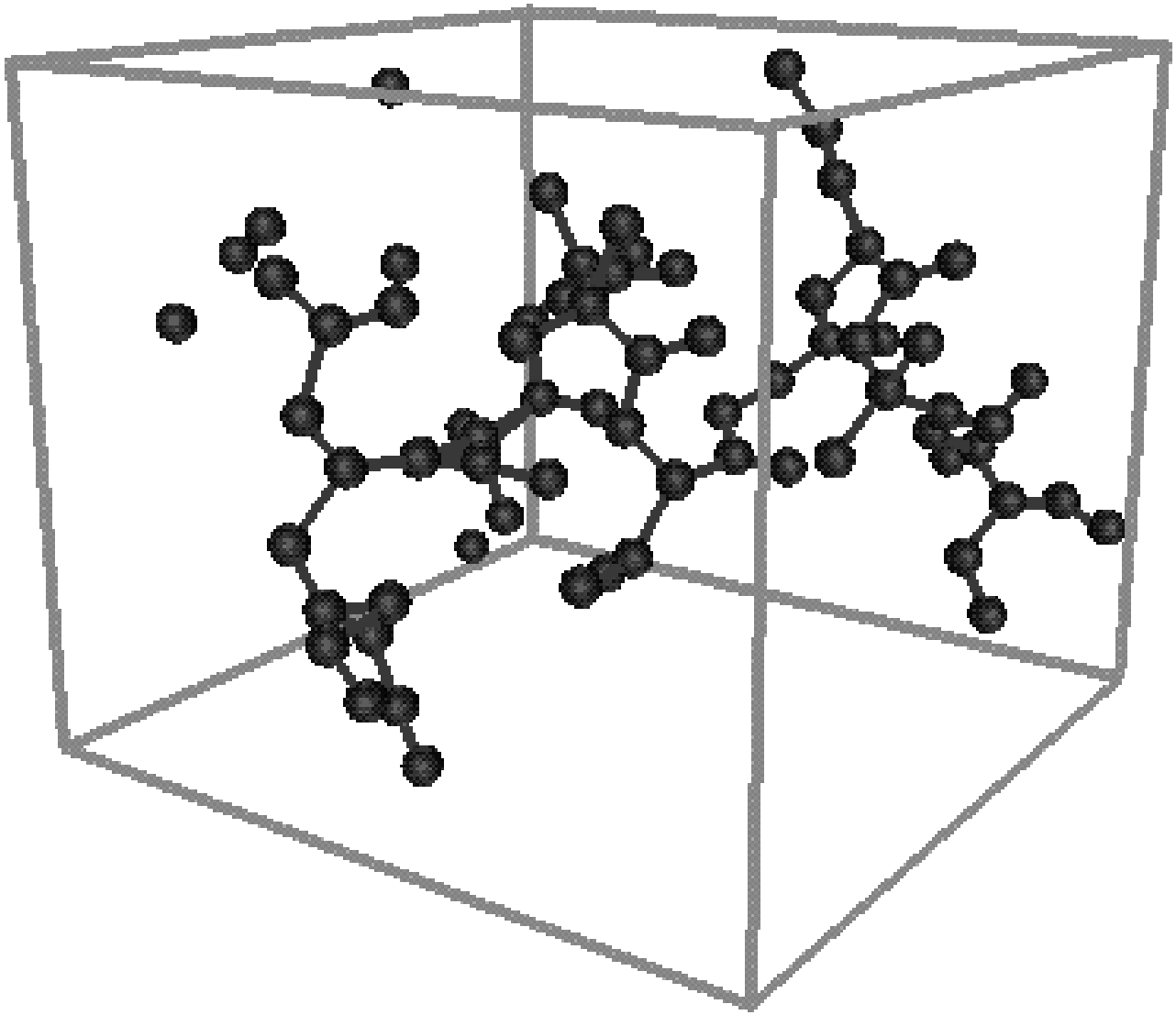
Phases: 740
Triples: 7,400



Visualization in *SnB* (Ph8755)

Geomview: Geometry Center, U. Minn.

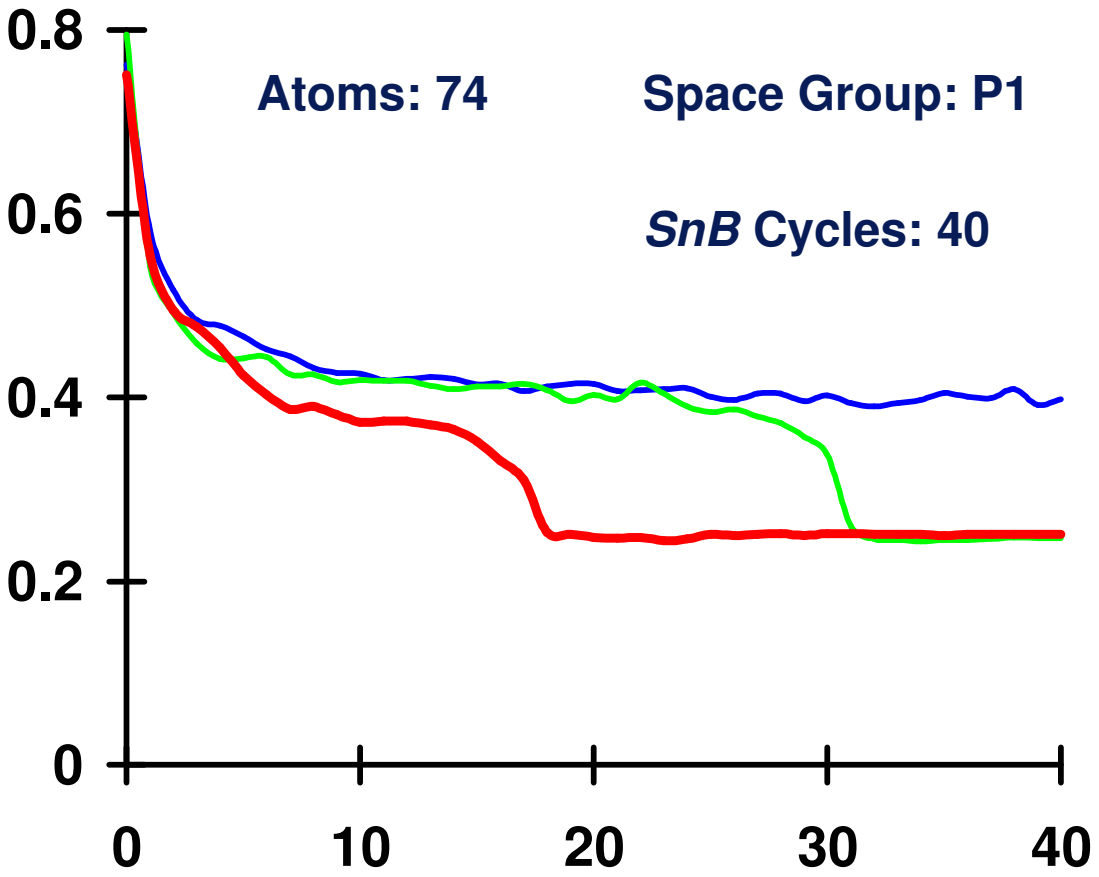
SnB



The Minimal Function is Diagnostic

Structure: Ph8755

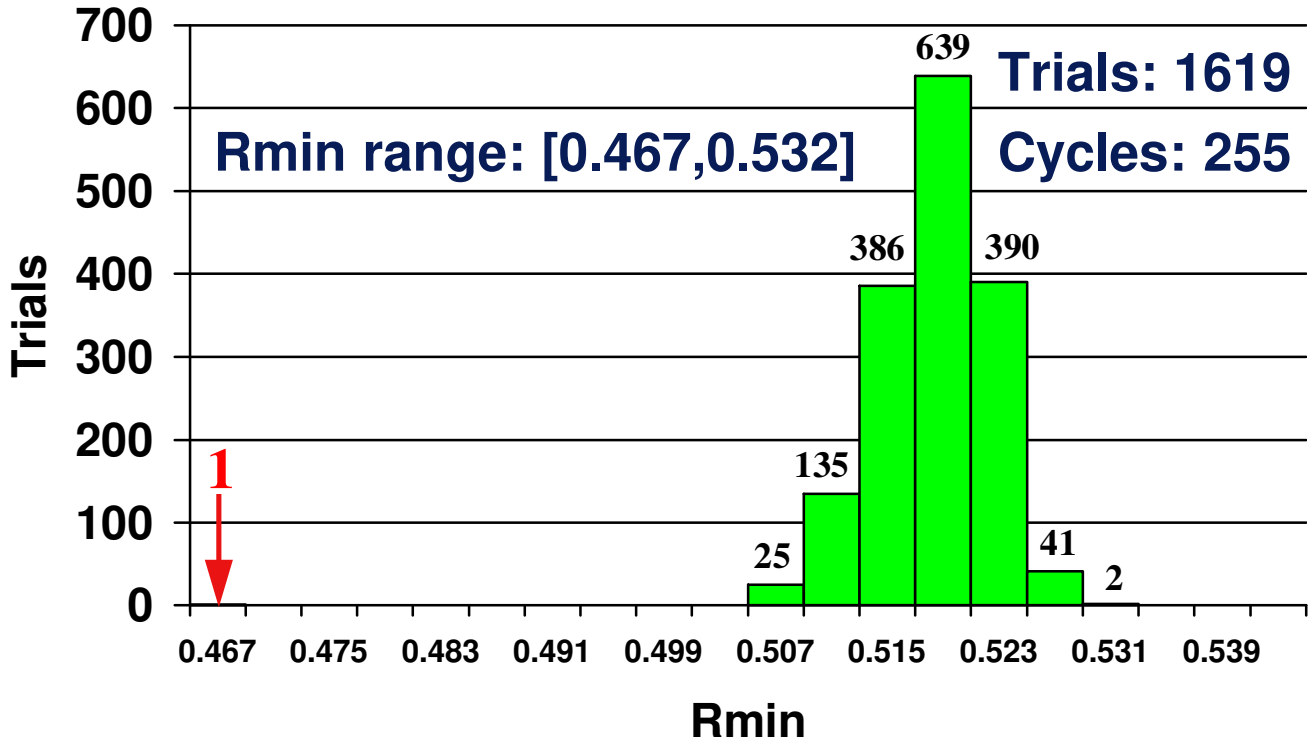
SnB



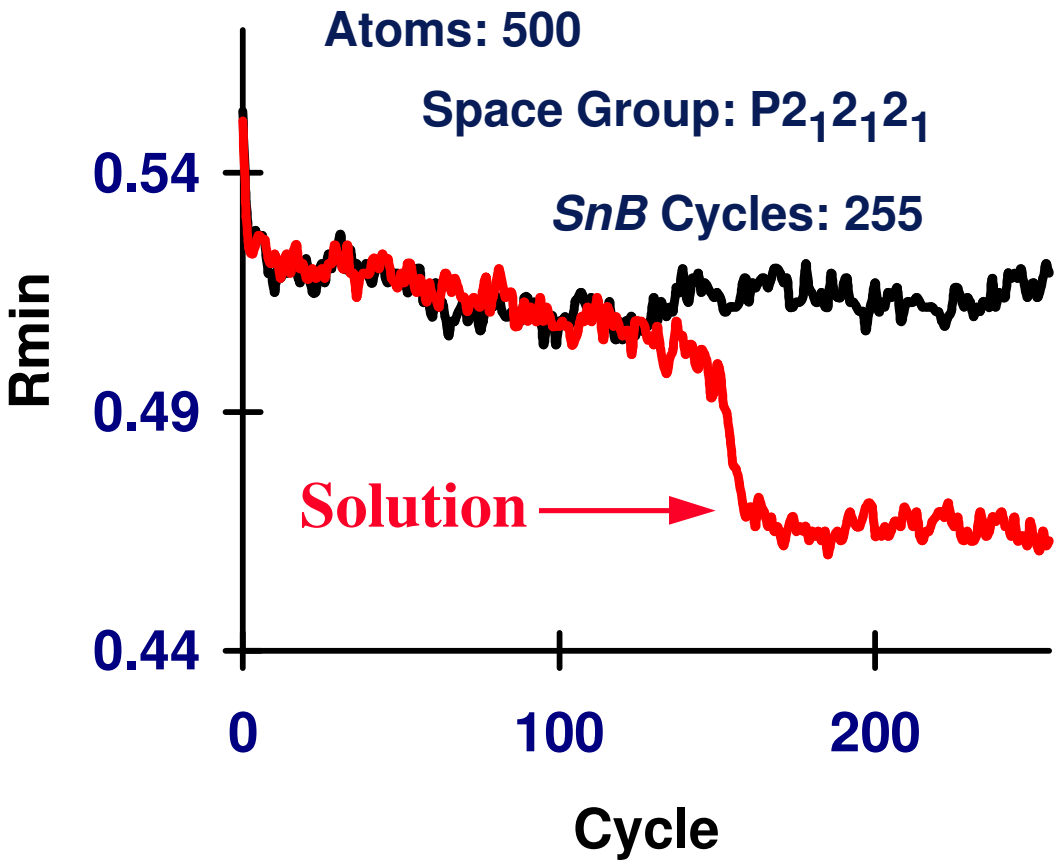
ToxII: *SnB* Histogram

Atoms: 500
Space Group: $P2_12_12_1$

Phases: 5,000
Triples: 50,000



Tox II: Trace of *SnB* Solution



Some *SnB* Solutions

Structure	Location	Atoms	Space Group
Scripps	Scripps	144	P1
Riboflavin tetrabutyrate	MSC	180	P1
Vancomycin	Penn	255	$P4_32_12$
Er-1 pheromone	UCLA	302	C2
Gramicidin A	HWI	317	$P2_12_12_1$
Crambin	HWI	400	$P2_1$
Alpha-1 peptide	OCI/U. of T.	450	P1
Rubredoxin	HWI	500	$P2_1$
ToxII	HWI	630	$P2_12_12_1$

Additional *SnB* Solutions

STRUCTURE	LOCATION	ATOMS	RES	SPACE GROUP
α -conotoxin PnIA	Queensland	110	1.1Å	$P2_1$
α -conotoxin PnIB	Queensland	111	1.1Å	C2
α -conotoxin G1	Queensland...	117	1.1Å	

Notes on Structures

SnB

STRUCTURE	ATOMS	SPACE GROUP	SUCCESS RATE
Ternatin	110	$P2_12_12_1$	2%
Scripps	144	P1	
Riboflavin tetrabutyrate (MSC)	180	P1	(8 years with traditional methods)
Er-1 pheromone	302	C2	
Gramicidin A	317	$P2_12_12_1$	0.3%
Crambin	400	$P2_1$	4%
Alpha-1 peptide	408	P1	5%
Rubredoxin	500	$P2_1$	2.7%
ToxII	630	$P2_12_12_1$	1/1619

Computing Platforms

SnB

- Unix Workstations
  - SGI, Sun, DEC/Alpha
  - IBM RS6000
  - Wintel/Linux
- Parallel Computers
  - Thinking Machines Corp. CM-5 (NCSA)
  - Cray T3D (PSC)
  - IBM SP2 (CTC)
- Cray C90 (PSC)

SnB Sites

Abbott Labs	Clemson	IVIC (Venezuela)	Pittsburgh	Windsor
Adelaide	CUHK (Hong Kong)	Keene State	Roche	Yale
Ames	Dalhousie	Kentucky, Louisville	Scripps	York
Ariad	Delaware	Loughborough	Shell	Zurich
ANU	Dow	McMaster	Southampton	Kodak
Berkeley	Dupong	Michigan State	South Carolina	LLNL
Birmingham	Duquesne	Minnesota	St. Andrews	SmithKline Beecham
Boehringer Ingelheim	East Anglia	MIT	Stanford	NIH
Boston College	Genetics Institute	NIU - Illinois	Sydney	Cornell
Brandeis	Griffith	Northwestern	SYSNU (Taiwan)	UCSF
BYUH	Guelph	Oak Ridge (ORNL)	UBC	Imerial College (London)
Calgary	Helsinki	Pavia	UCLA	Caltech
Cambridge	Hoffman Laroche	Univ. Penn.	UGA	Madras
Chinoin (Hungary)	Hong Kong University	Pfizer	USC	Univ. Queensland (Australia)

Looking Ahead: *SnB* vers. 2.0

- Speed: 5 - 10X
- Includes Inverse FFT
- Additional Density Modification Options
- Improved Fourier Recycling: “Twice Baking”
- Input Improvements: |E| Calculation
- Output Improvements: Compatibility
- (SIR/SAS/MAD Invariants with est. values)

Summary

SnB

- *Shake-and-Bake:* Enhanced Direct Methods
- Targeted at 100-800 atom structures
- Publicly available
- LEVY / EVAL (Bob Blessing) recommended for calculating |E|'s in *SnB* vers. 1.5
- *SnB* version 2.0
- Addresses:
  - http://www.hwi.buffalo.edu/SnB/
  - snb-requests@hwi.buffalo.edu
  - snb-comments@hwi.buffalo.edu

SnB

Running Times

(Average Seconds/Cycle on Iled)

SnB

Machine	Total	StrFac	Refine	FFT	Pk Pk
DEC/Alcor 266	0.672	0.160	0.248	0.196	0.029
SGI Pwr Ind 2	1.247	0.210	0.426	0.485	0.068
DEC/Alpha 150	1.822	0.437	0.682	0.520	0.075
SGI Indigo	2.975	0.881	0.760	1.037	0.163
Sparc 10	3.140	0.821	0.919	1.111	0.162
Pentium 75	3.535	1.181	0.995	1.205	0.154
Sparc 5	3.759	0.765	1.087	1.661	0.210
i486dx2-66	11.396	2.175	2.423	5.467	0.711

ToxII: Running Times

SnB

- **SGI Indigo (R4000):**
  - **50 seconds per cycle**
  - **200 minutes per trial**
  - ***7 CPU months for 1600 trials***
- **IBM SP2 Projection:**
  - ***4 hours for 1600 trials***