Shake-and-Bake:
Applications and Advances

Russ Miller & Charles M. Weeks

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

- *Shake-and-Bake*
  - The Minimal Function
- *SnB*
  - Results
- *SnB* v2.0
  - Rationale
  - Results
- Summary

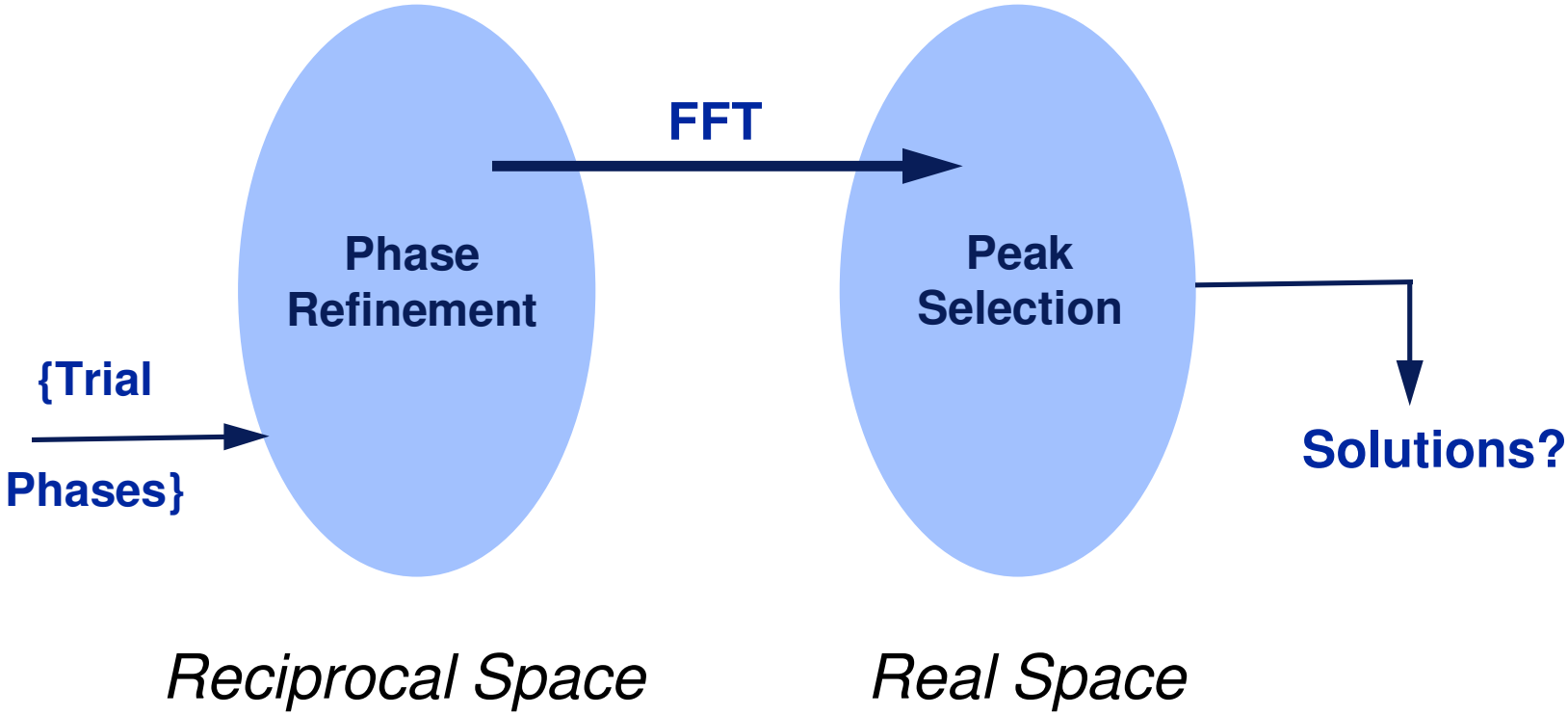
Direct Methods

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

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 $\cos\phi_T$

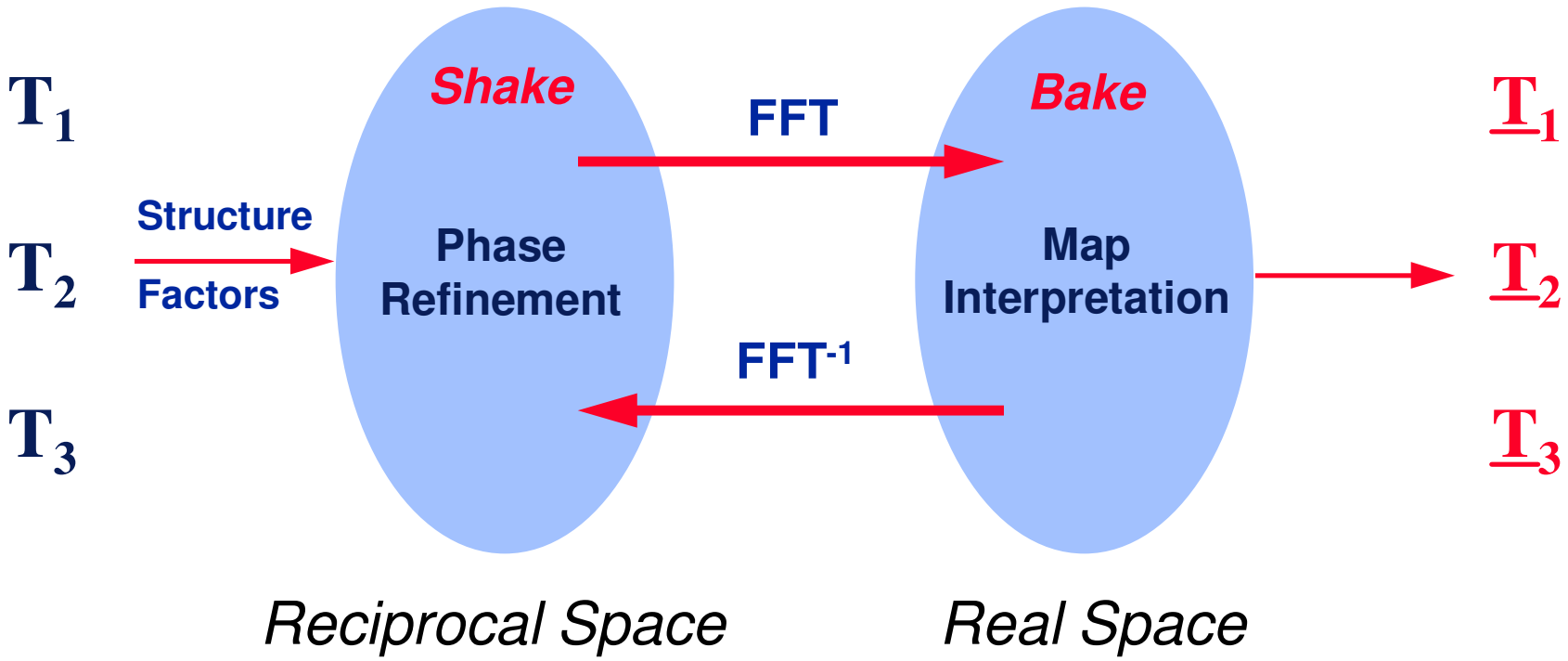
Shake-and-Bake

SnB

- **Direct Methods Optimization Technique**
- **Multiple Random-Atom Trial Structures**
- **Real/Reciprocal Space Cycling**
- **Phase Refinement Techniques:**
  - **Parameter Shift**
  - **Tangent Formula**
- **Minimal Function as FOM**

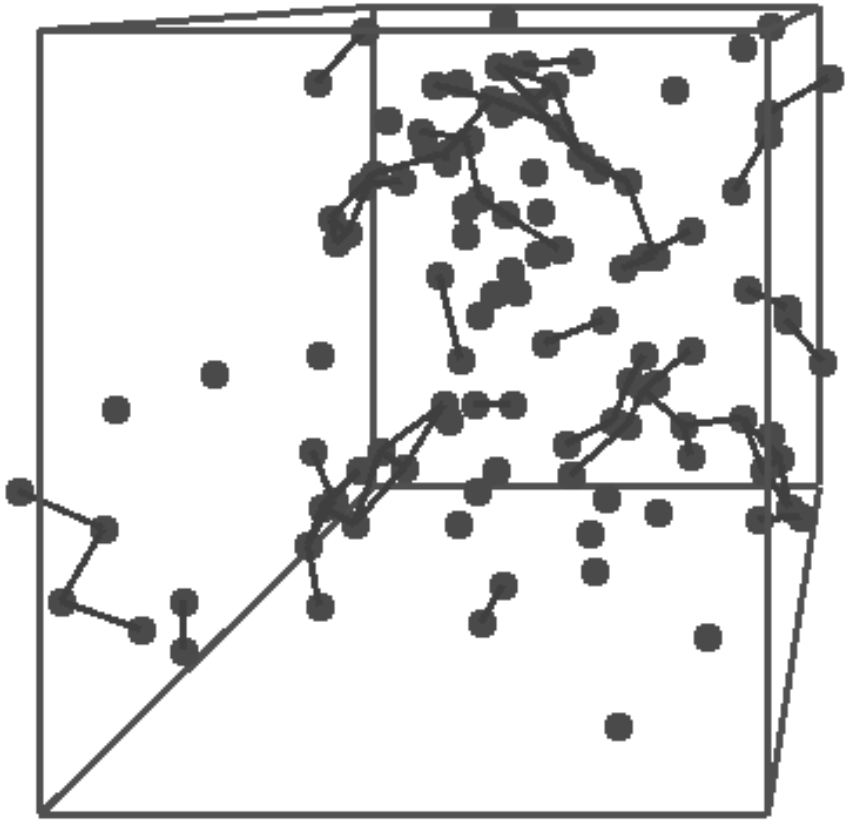
Shake-and-Bake

SnB



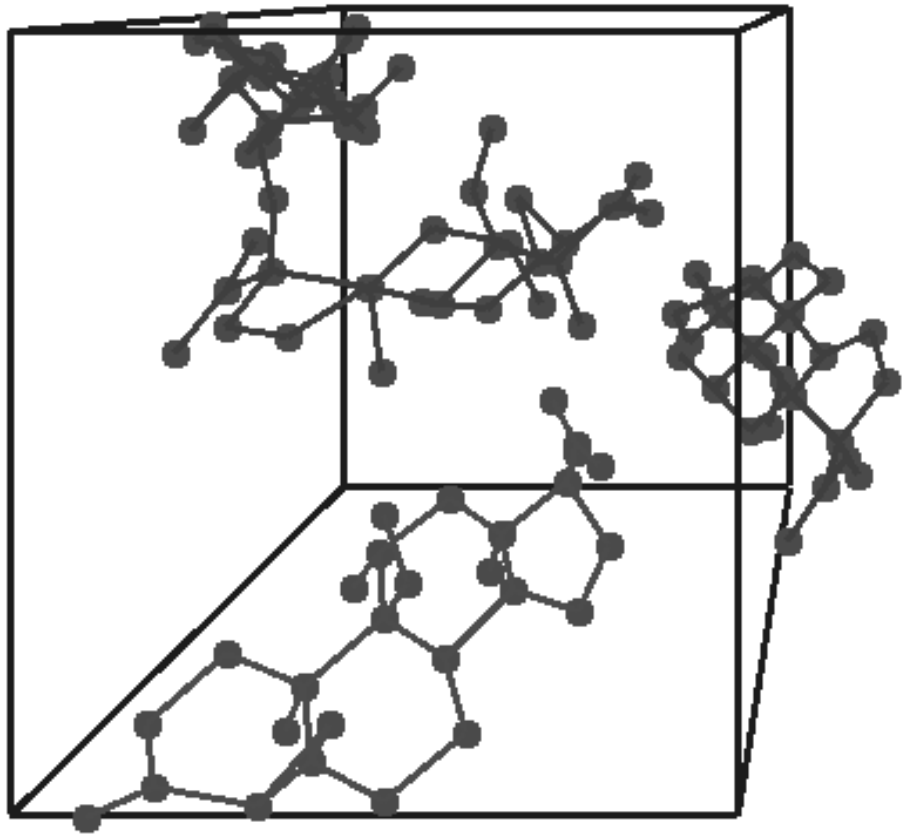
SnB: Random Start

SnB

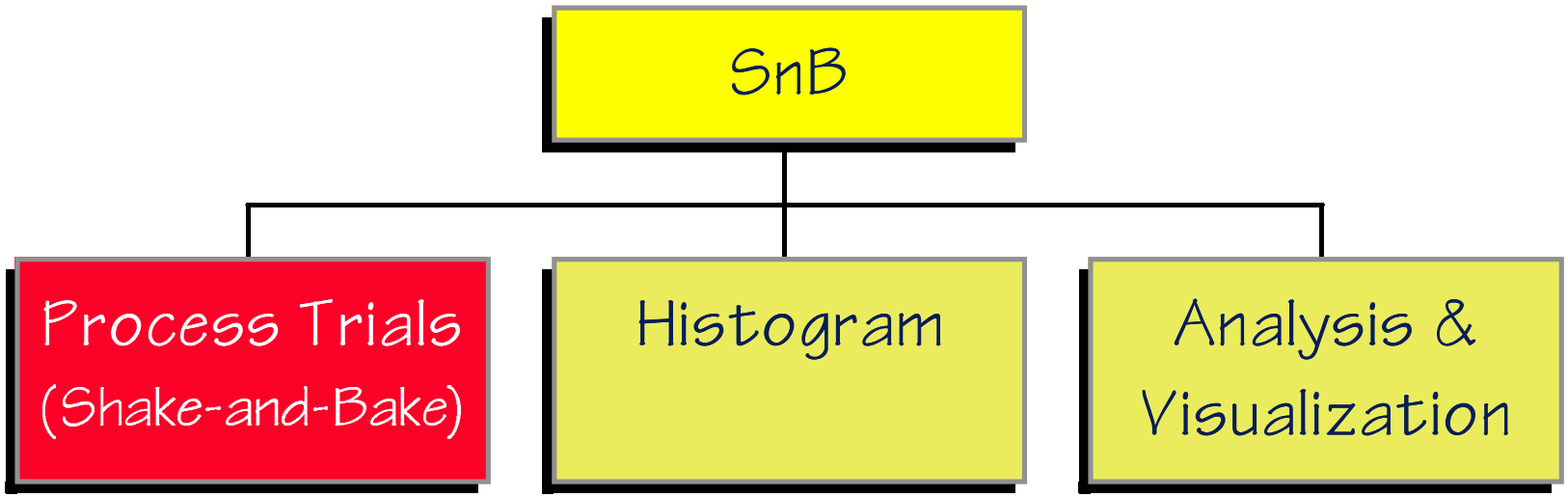


SnB: Final Structure

SnB

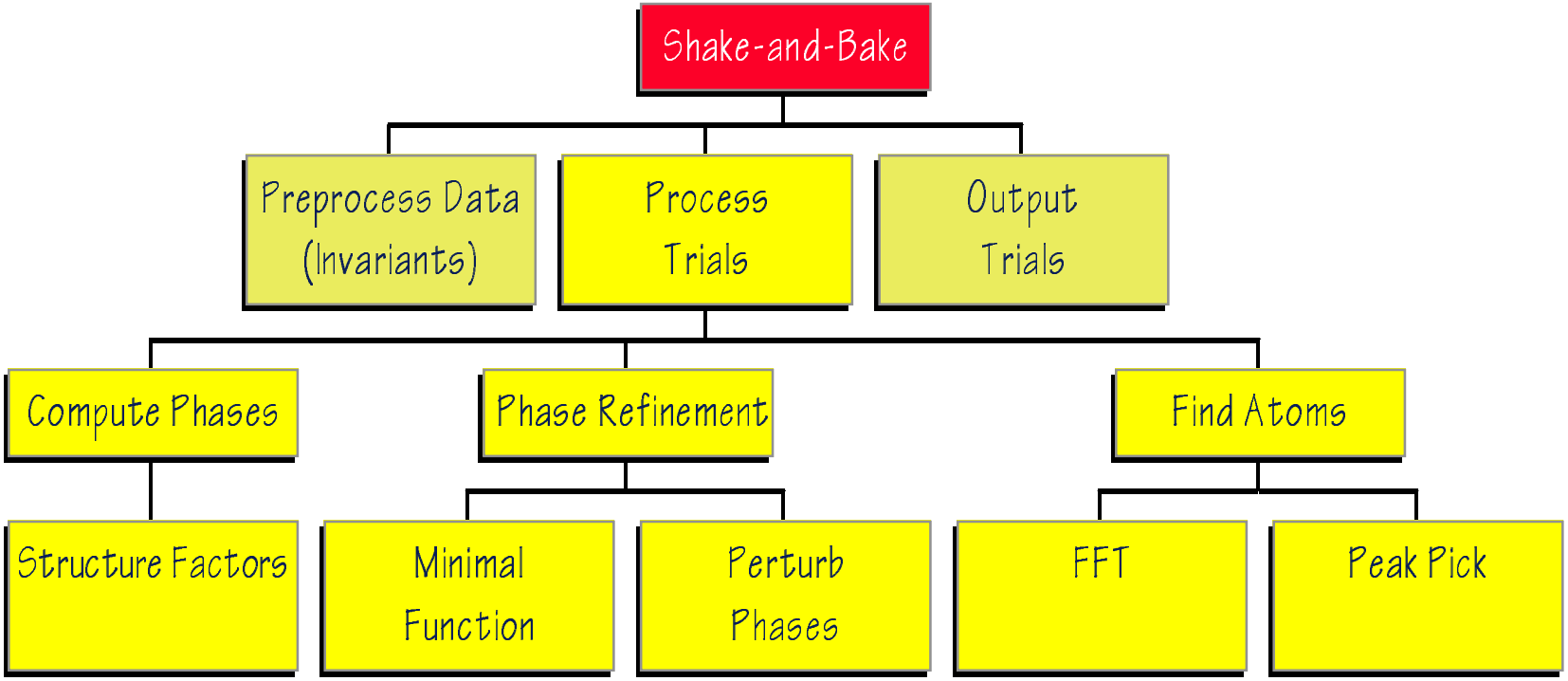


Structure of *SnB*



Shake-and-Bake

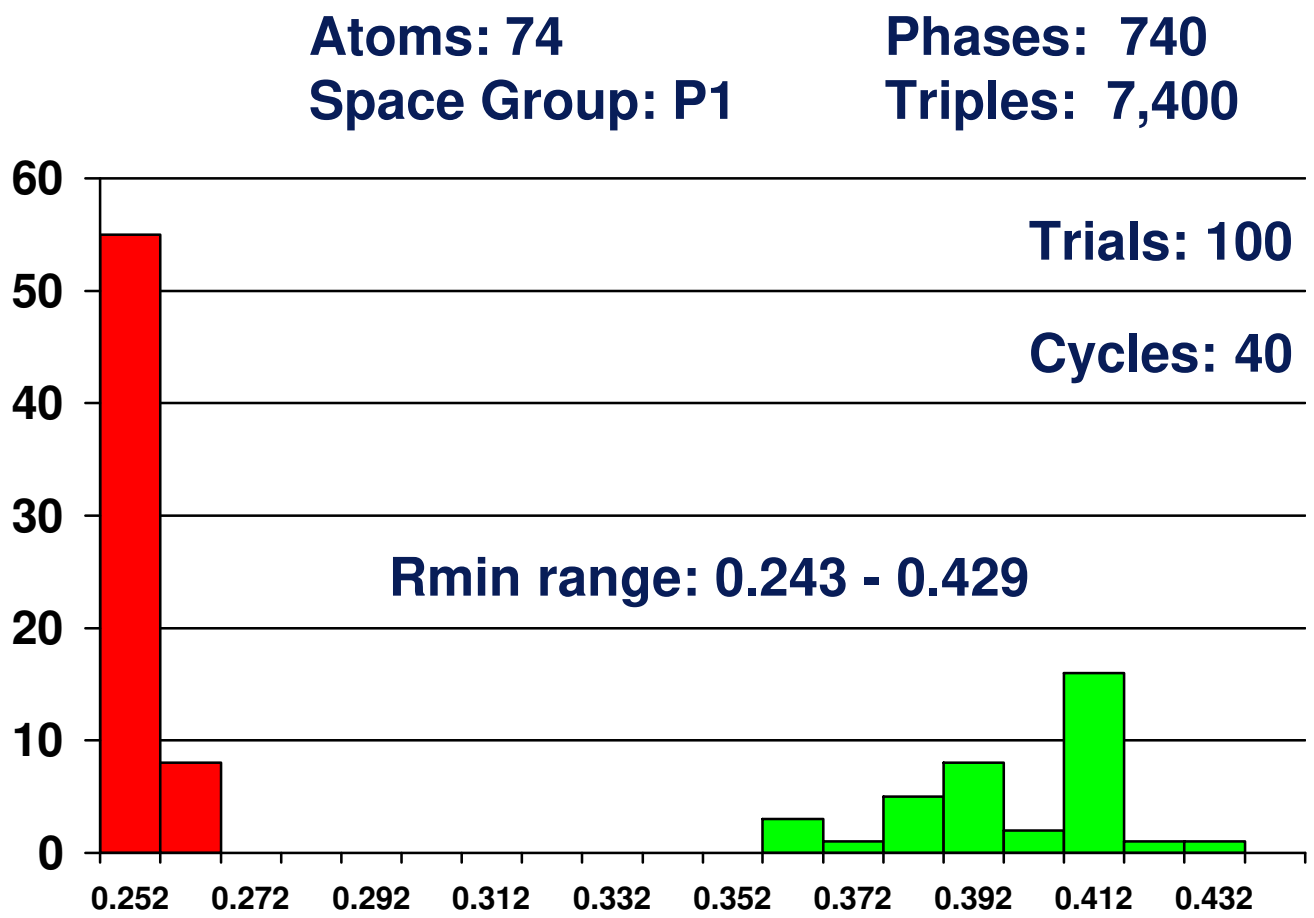
SnB



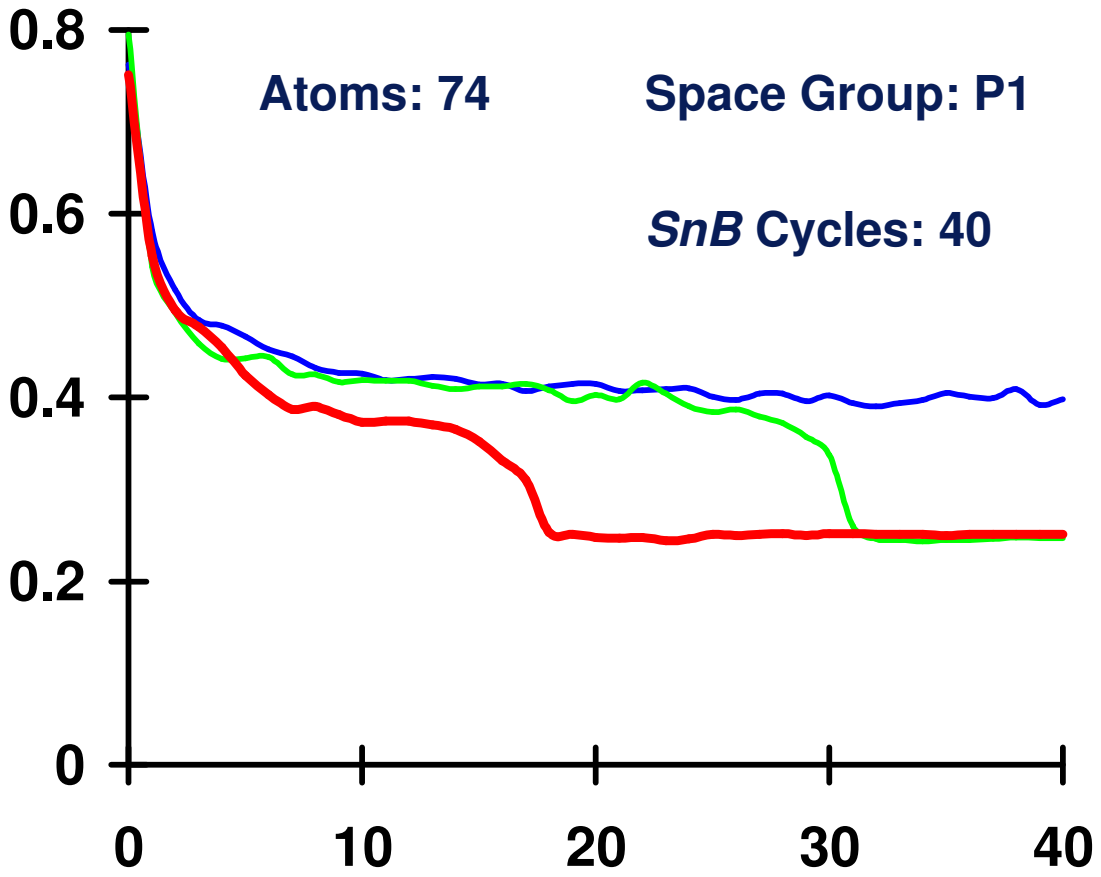
SnB Parameters

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

Ph8755: *SnB* Histogram



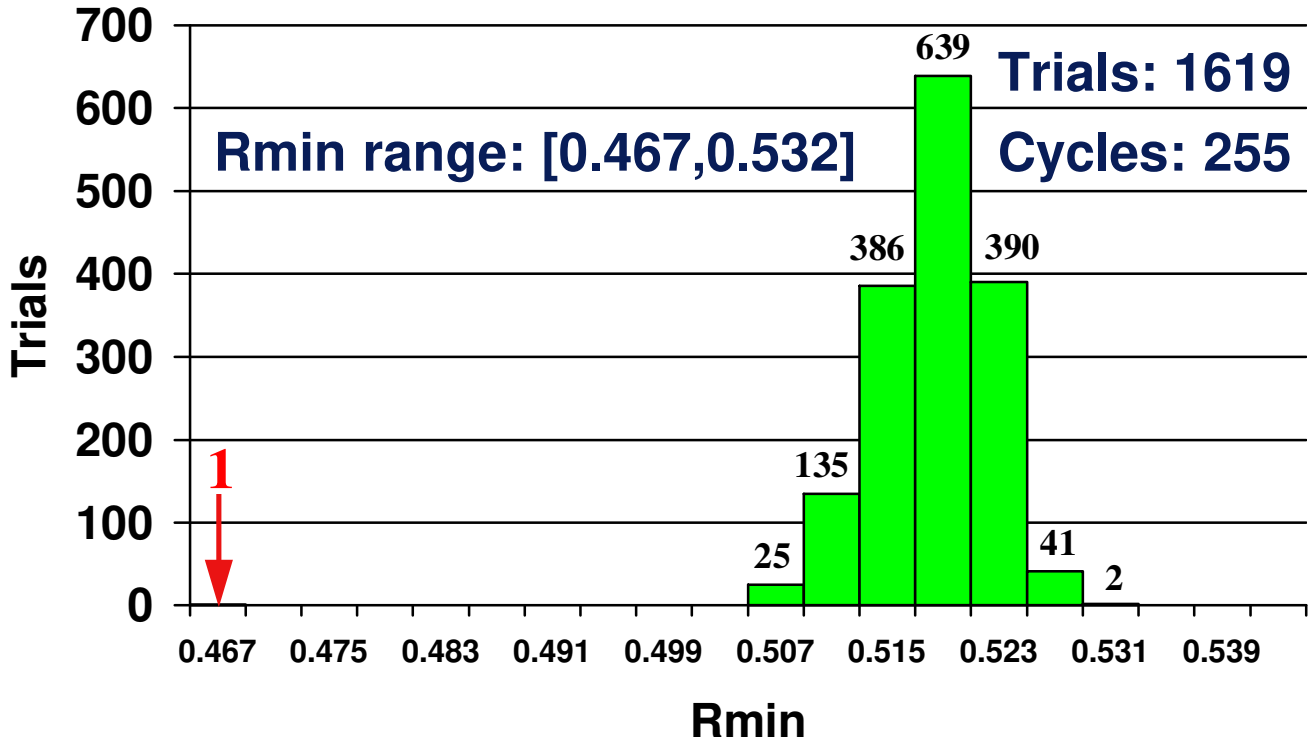
Ph8755: Trace of *SnB Solution*



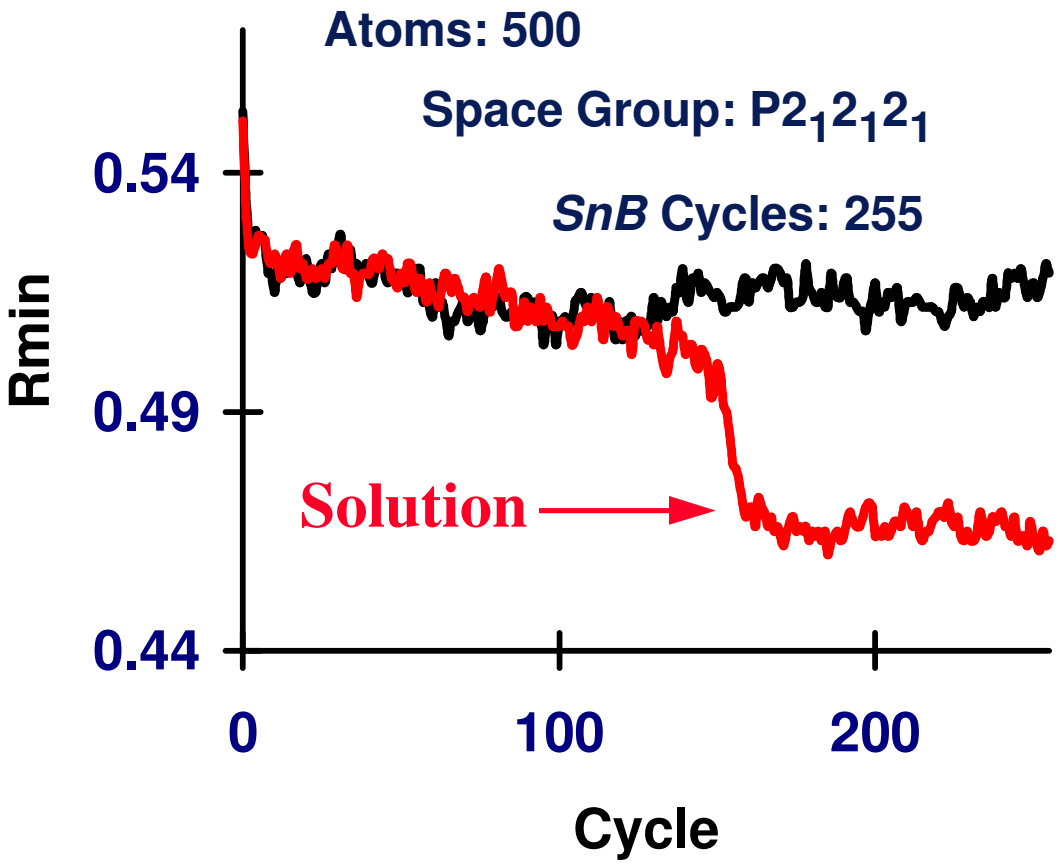
ToxII: *SnB* Histogram

Atoms: 500
Space Group: $P2_12_12_1$

Phases: 5,000
Triples: 50,000



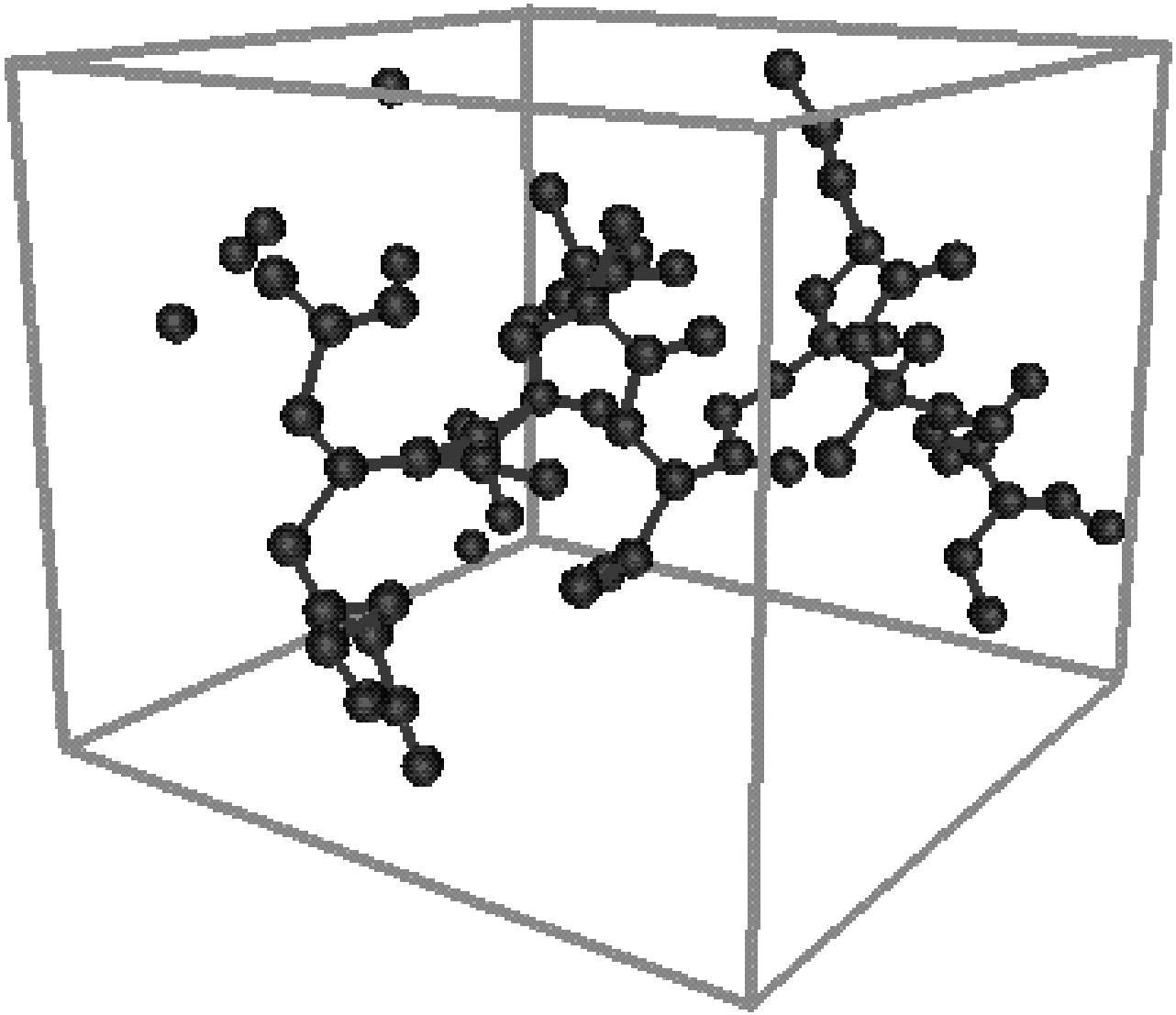
Tox II: Trace of *SnB* Solution



Visualization in *SnB* (Ph8755)

Geomview: Geometry Center, U. Minn.

SnB



Some *SnB* Applications

Structure	Location	Atoms	Sp Grp	Res
Vancomycin	Penn	258	$P4_32_12$	0.9Å
I4 Peptide	HWI	289	I4	1.1
Microlide	France	296	$P2_1$	1.1
Gramicidin A	HWI	317	$P2_12_12_1$	0.86
Er-1 pheromone	UCLA	328	C2	1.0
Crambin	HWI	~400	$P2_1$	0.83
Alpha-1 peptide	OCI/U. of T.	471	P1	0.92
Rubredoxin	HWI	497	$P2_1$	1.0
Scorpion Toxin II	HWI	624	$P2_12_12_1$	0.96

Factors Determining Success Rate

SnB

- **Data quality**
- **Resolution**
- **Complexity and connectivity of structure**
- **Space group**
- ***Presence of heavy atoms***

An Interesting I4 Structure

SnB

- **Structure:**
  - Peptide with 10 Sulfurs
  - 289 nonH atoms total
  - 1.1Å resolution data
- **Bugs: Special Positions & Refinement**
- **Results (*SnB* 2.0)**
  - PS: 53% success rate
  - PS/Rest: 44% success rate
  - Tan: 25% success rate

Extending Resolution: the I4 Structure

SnB

- **Truncate to 1.2Å - 1.5Å**
- **Solutions at all resolutions**
  - **1.2Å:** Standard bimodal distribution
  - **1.3Å:** Standard bimodal distribution
  - **1.4Å:** Good, but some mixing of solutions and nonsolutions
  - **1.5Å:** Solutions (~50 deg. phase error) but *not* recognizable by FOM
- **Recognize low-resolution solutions??**

SnB 2.0: Rationale

- **Improve running time**
  - Build from ground up
- **Provide additional features**
  - Inverse Fourier
  - Density modification
  - Grid size
  - "Twice Baking"
  - Peaks at special positions

SnB v2.0 Parameters/Proteins

Structure	*Atoms (n)*	*Heavy Atoms*	*Phases*	*Cycles*	*Max Succ Rate*
Vancomycin	202	*Cl* 8	2000	200	0.8%
I4 Peptide	248	S10	1900	250	53.0
Gramicidin A	272	-	3000	275	1.1
Crambin	327	S6	3000	300	4.8
Rubredoxin	395	FeS6	4000	400	6.2
Scorpion Toxin II	508	S8	5000	500	1.4

Note: *n* = independent protein atoms

SnB 2.0: Varying Peaks

Structure	50	100	200	300	400
Vancomycin	0.4%	0.6%	0.2%	---	---
I4 Peptide	53.0	52.0	45.0	---	---
Gramicidin A	0.0	0.3	1.1	0.7%	---
Crambin	4.3	4.8	3.3	3.4	---
Rubredoxin	5.7	6.2	5.4	3.9	3.4%
Scorpion Toxin II	---	1.0	1.4	0.4	0.1

SnB 2.0: Varying Cycles

Success Rates while varying *SnB* Phase Refinement Cycles

Structure	0.25*n*	0.5*n*	0.75*n*	*n*	1.25*n*	1.5*n*
Vancomycin	0.1%	0.4%	0.4%	0.6%	0.7%	-
I4 Peptide	27.0	40.0	48.0	53.0	-	-
Gramicidin A	0.0	0.4	0.6	0.9	1.2	2.0%
Crambin	3.1	4.1	4.6	4.8	-	-
Rubredoxin	4.6	5.5	5.9	6.0	-	-
Scorpion Toxin II	0.05	0.5	1.0	1.4	-	-

SnB 2.0: Phase Refinement

Structure	*Peaks*	*PS Standard*	*PS Restricted*	*Tangent*
Vancomycin	100	0.6%	0.4%	0.3%
I4 Peptide	50	53.0	44.0	25.0
Gramicidin A	200	1.1	0.5	0.0
Crambin	100	4.8	3.7	2.2
Rubredoxin	150	6.0	5.2	4.0
Scorpion Toxin II	200	1.4	1.0	0.7

SnB 2.0 Parameters (>1.1Å)

- **Peaks**
  - **0.4*n* if "heavy" atoms present**
  - **0.8*n* if all *C,N,O***
- **Phase Refinement**
  - ***Unrestricted* Parameter Shift**
- **Cycles**
  - ***n*/2 if *n*<400 and "heavy" atoms present**
  - ***n* otherwise**

SnB 2.0: Timings
(SGI R10000 Workstation)

Structure	*non-H Atoms*	*Space Group*	*n/2 Cycles*	*Trials/ Day*	*Solns/ Day*
Vancomycin	258	$P4_32_12$	100	391	1.5
I4 Peptide	289	I4	125	274	110
Gramicidin A	317	$P2_12_12_1$	135	572	2
Crambin	~400	$P2_1$	150	1029	42
Rubredoxin	497	$P2_1$	200	294	16
Scorpion Toxin II	624	$P2_12_12_1$	250	109	0.5

Note: For each structure, optimum no. of peaks used.

Computing Platforms

SnB

- Unix Workstations
  - SGI, Sun, DEC/Alpha
  - Wintel/Linux
- Parallel Computers
  - Cray T3D/E, TMC CM-5, IBM SP2
  - SGI Origin 2000
  - HP-Convex Exemplar
- Cray C90

Summary

SnB

- ***Shake-and-Bake:* Dual-Space Direct Methods**
- **Targeted at 100-800 atom structures**
- ***SnB* version 2.0**
  - Optimized code with Inverse FFT
  - Additional Density Modification Options
  - Improved Fourier Recycling: “Twice Baking”
  - I/O: |E| calculation and visualization interface
  - (SIR/SAS/MAD Invariants with estimated values)
- **http://www.hwi.buffalo.edu/SnB/**

SnB