Molecular Structure Determination via *SnB*

Russ Miller

SUNY-Buffalo
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.

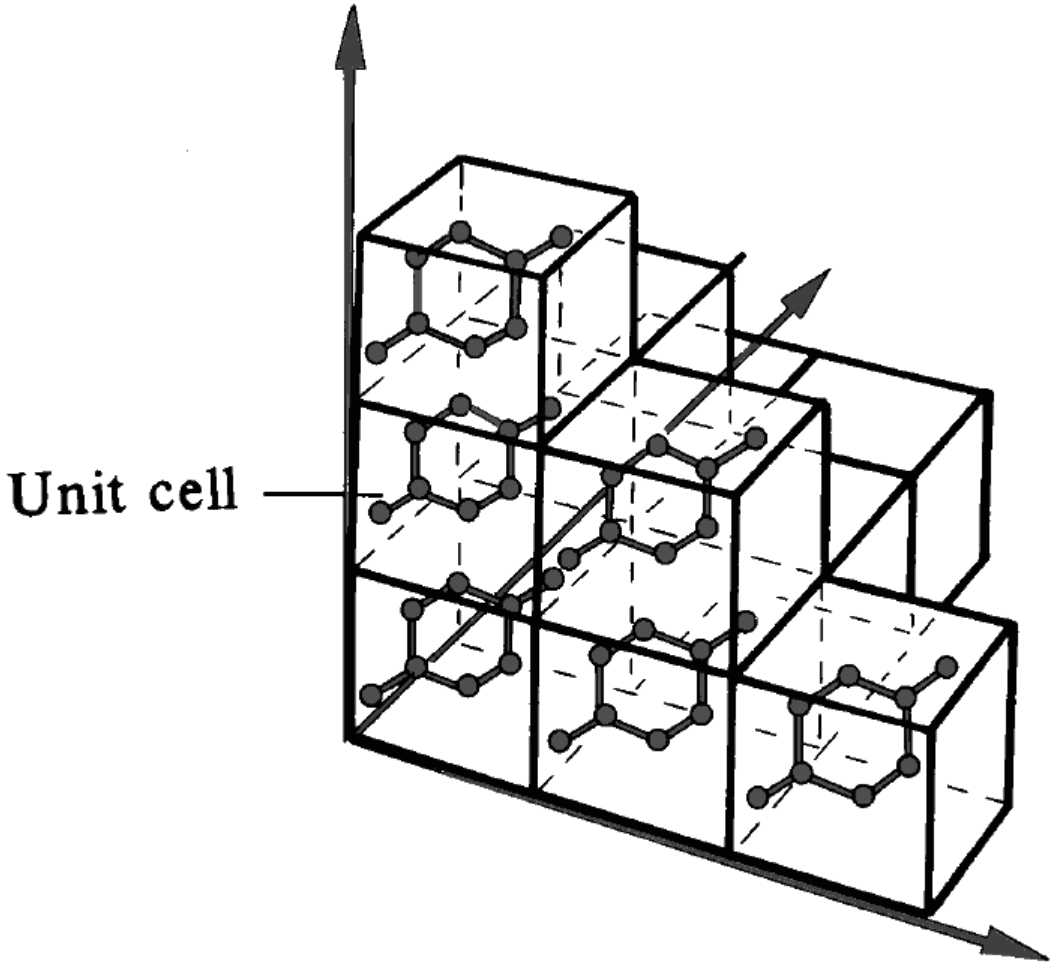
Outline of Talk

SnB

- X-Ray Crystallography
  - The Phase Problem
- *Shake-and-Bake*
  - The Minimal Function
- *SnB*
  - Results
- Summary
- *Demonstration*

A Crystal Structure

SnB



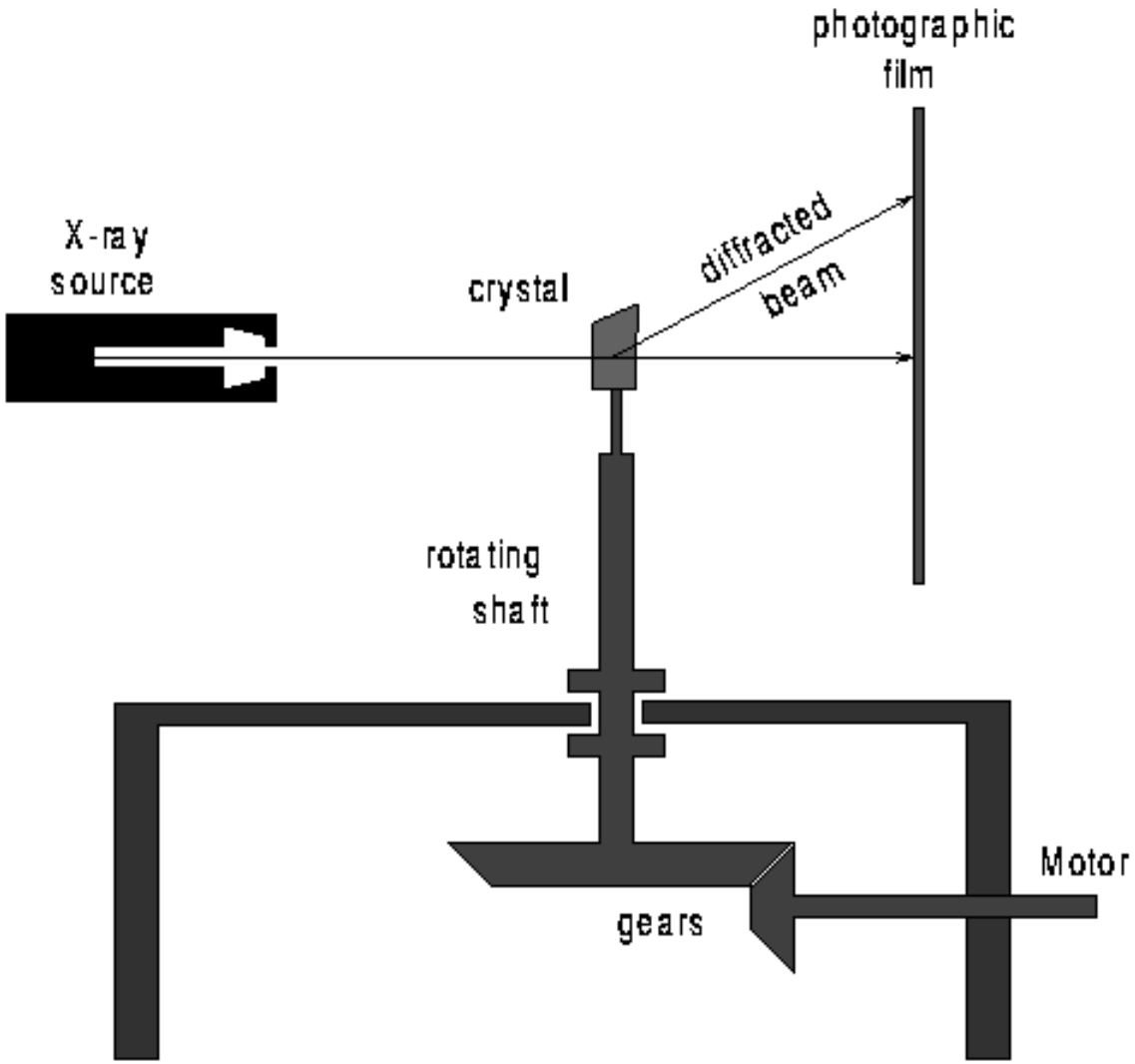
X-Ray Crystallography

SnB

- **Objective: Provide a 3-D mapping of the atoms in a crystal.**
- **Procedure:**
  - **Isolate a single crystal.**
  - **Perform the X-Ray diffraction experiment.**
  - **Determine molecular structure that agrees with diffration data.**

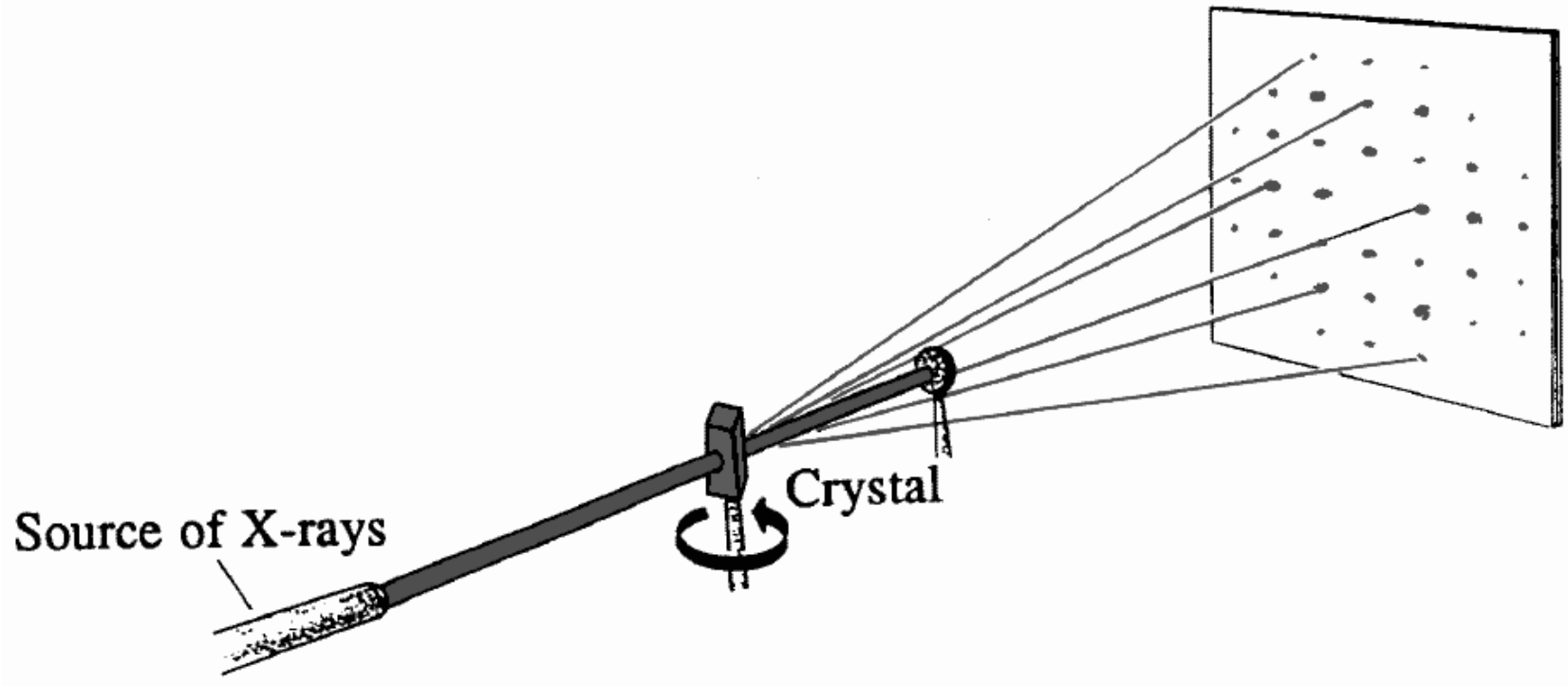
Diffractometer

SnB



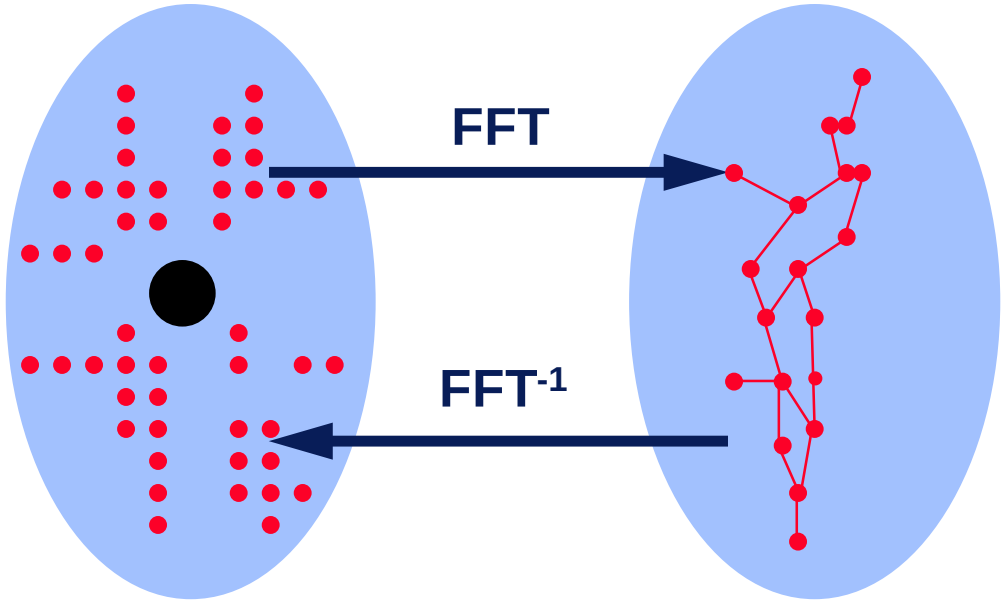
The Diffraction Pattern

SnB



X-Ray Data & Molecular Structure

SnB

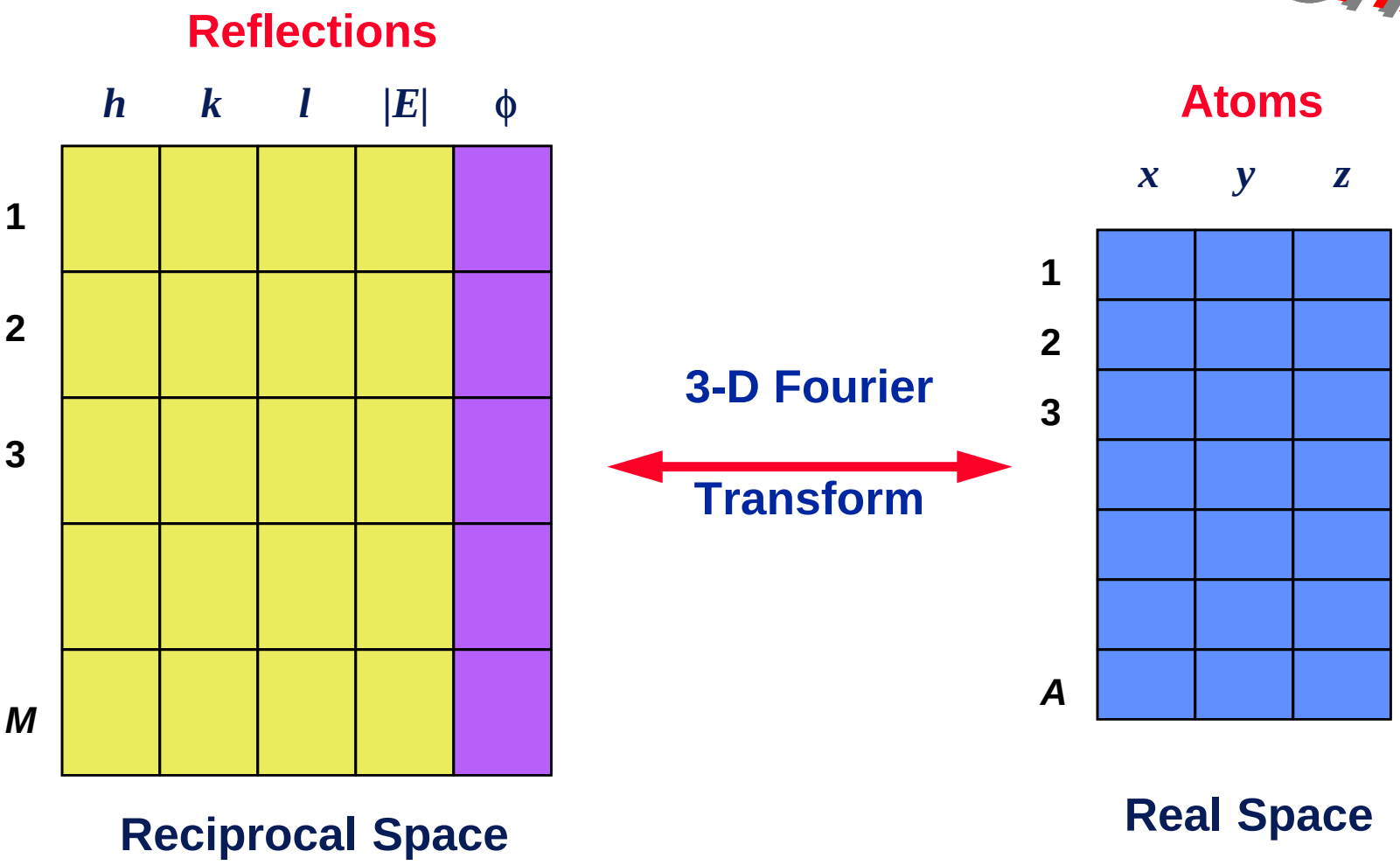


The Phase Problem

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**

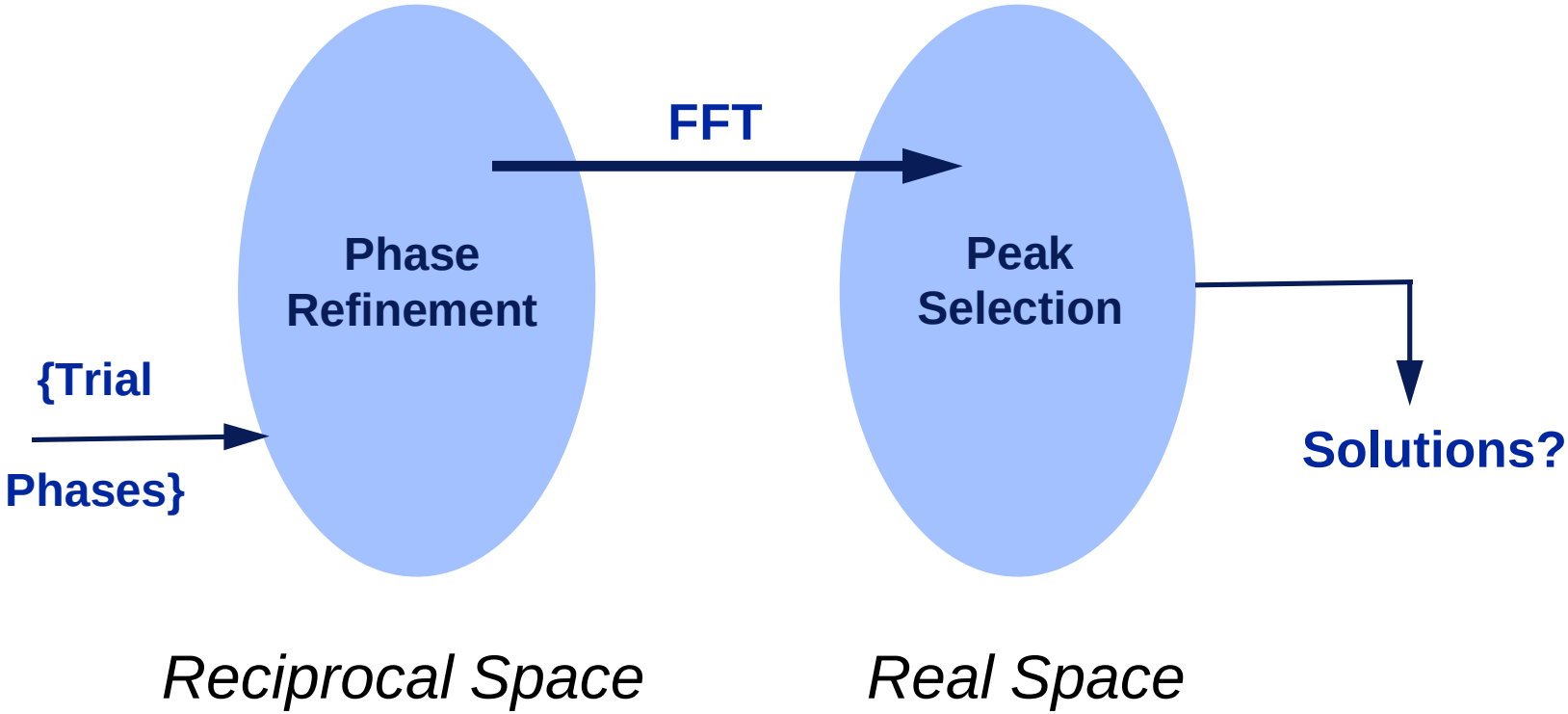
Structure Invariants

SnB

- ***Direct Methods*** **exploit probabilistic theories to exploit linear relationships among phases.**
- **A *triplet*** $(\phi_h + \phi_k + \phi_{-h-k})$ **has a most probable value of** $0 \mod 2\pi$**, given that *h* and *k* are distinct reciprocal vectors.**

Conventional Direct Methods

SnB



Shake-and-Bake

The Minimal Function

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

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

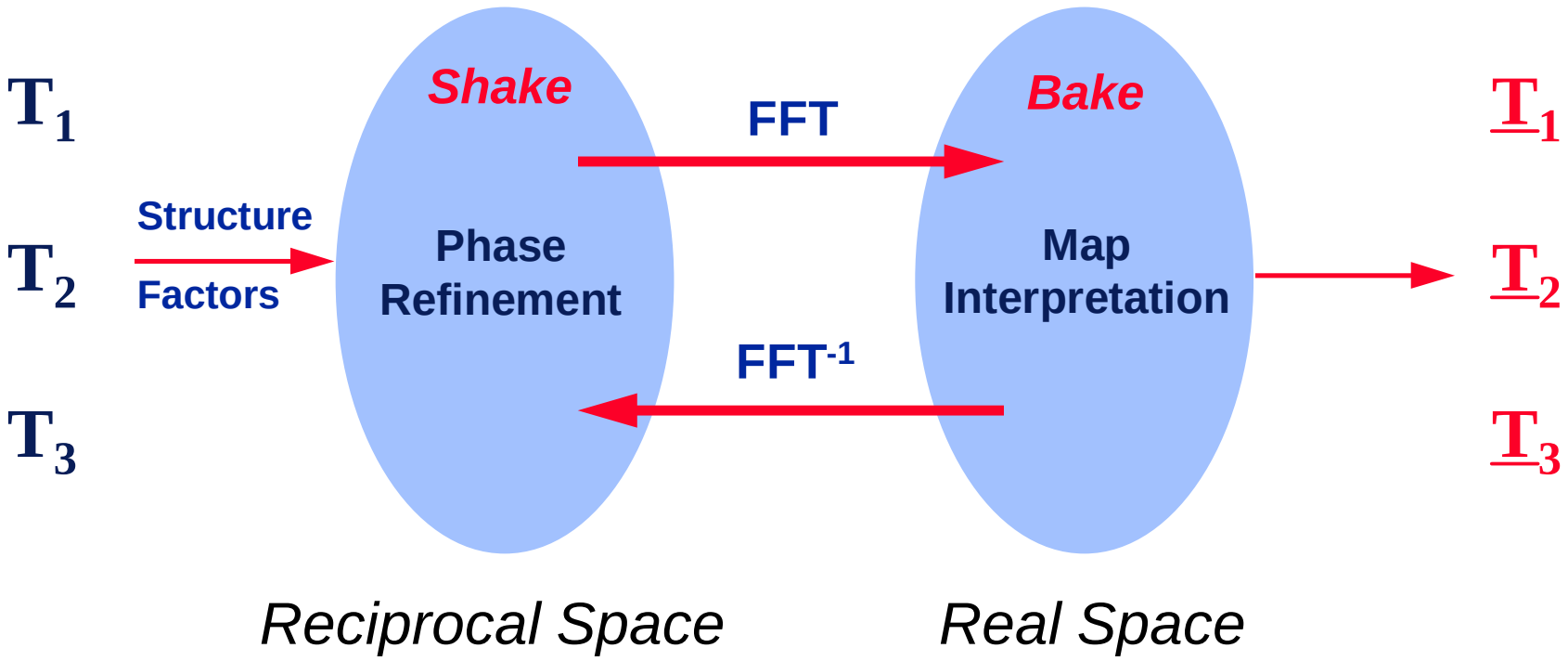
Shake-and-Bake

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**

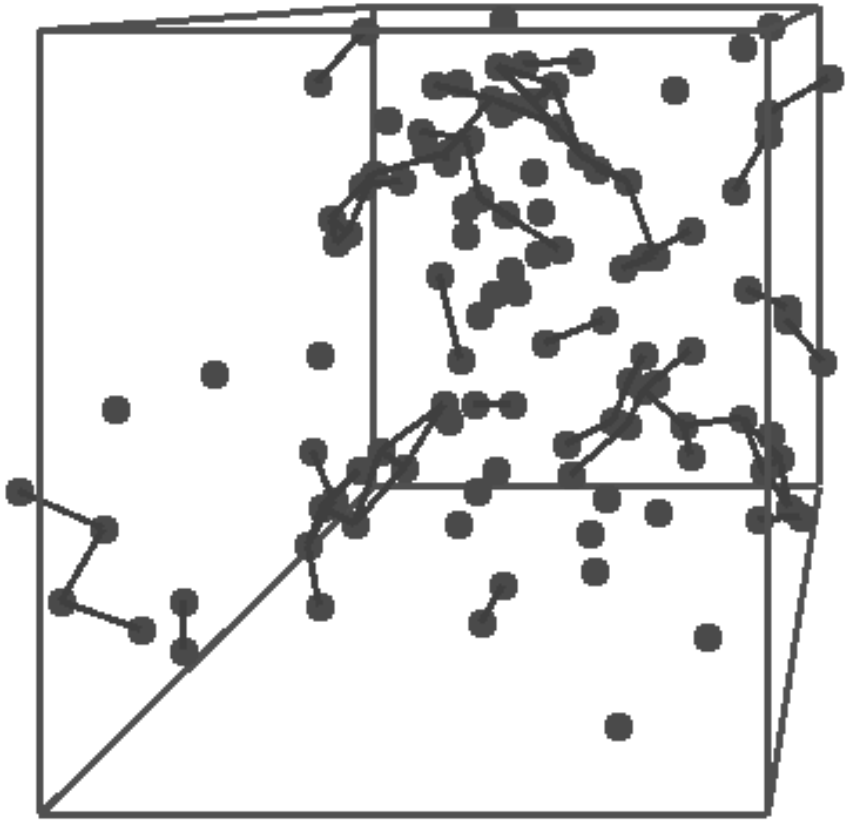
Shake-and-Bake

SnB



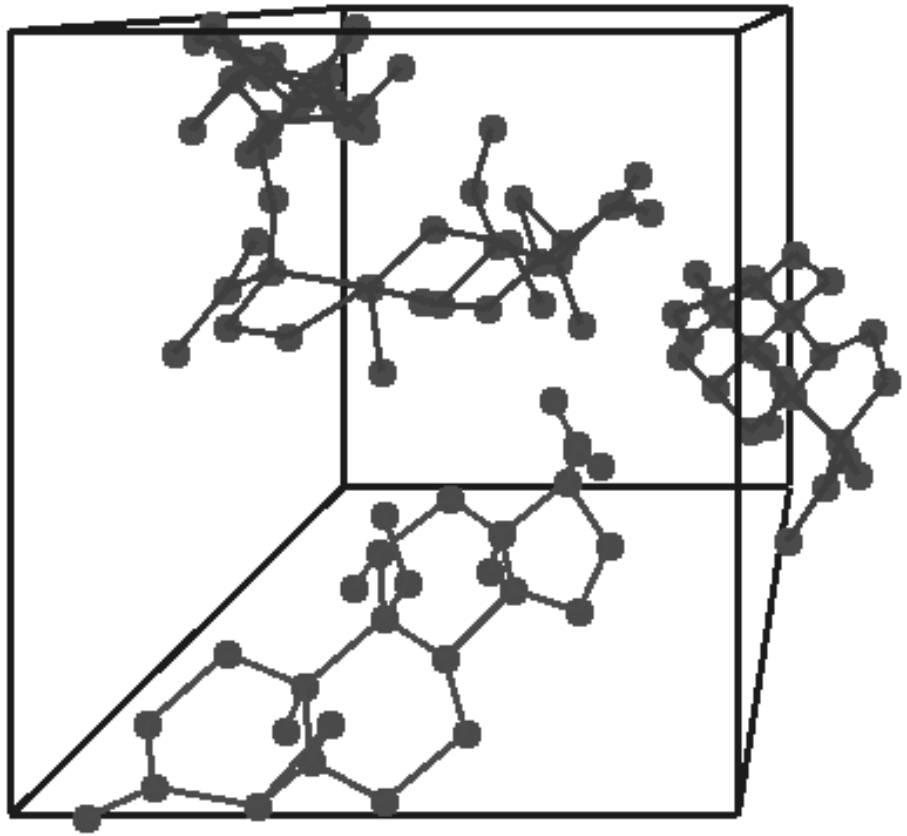
SnB: Random Start

SnB

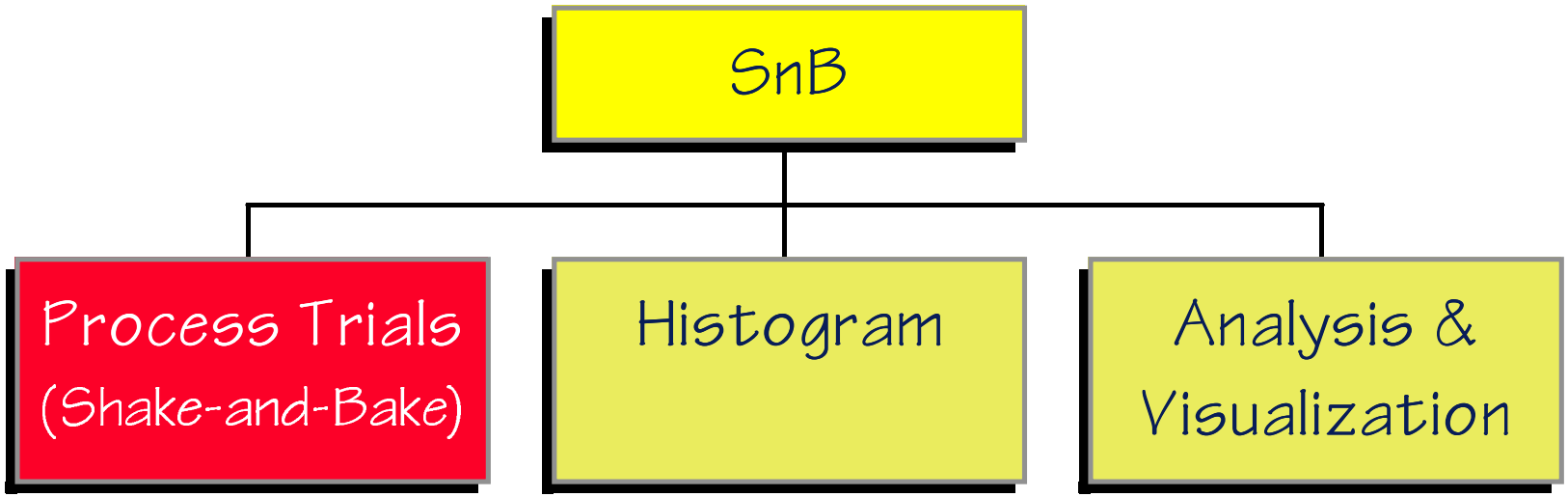


SnB: Final Structure

SnB

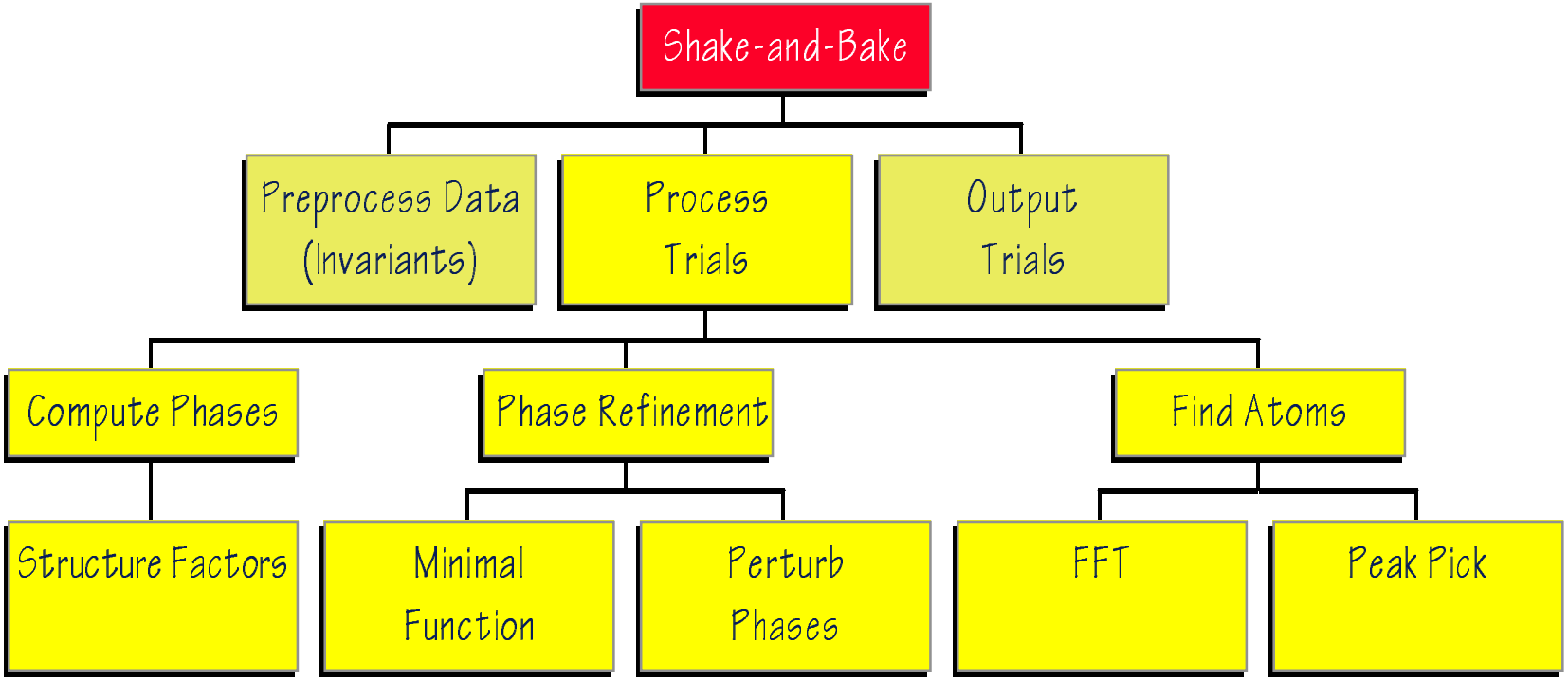


Structure of *SnB*



Shake-and-Bake

SnB



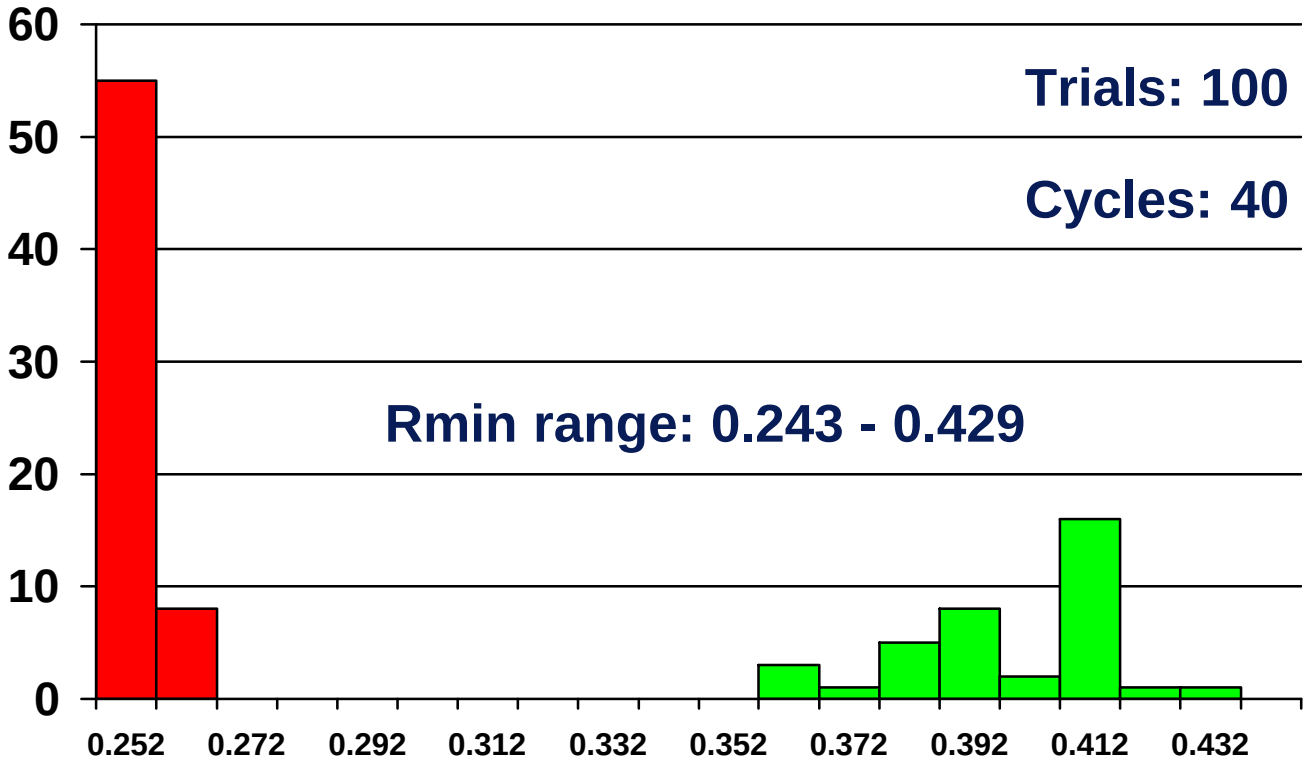
SnB Parameters

	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

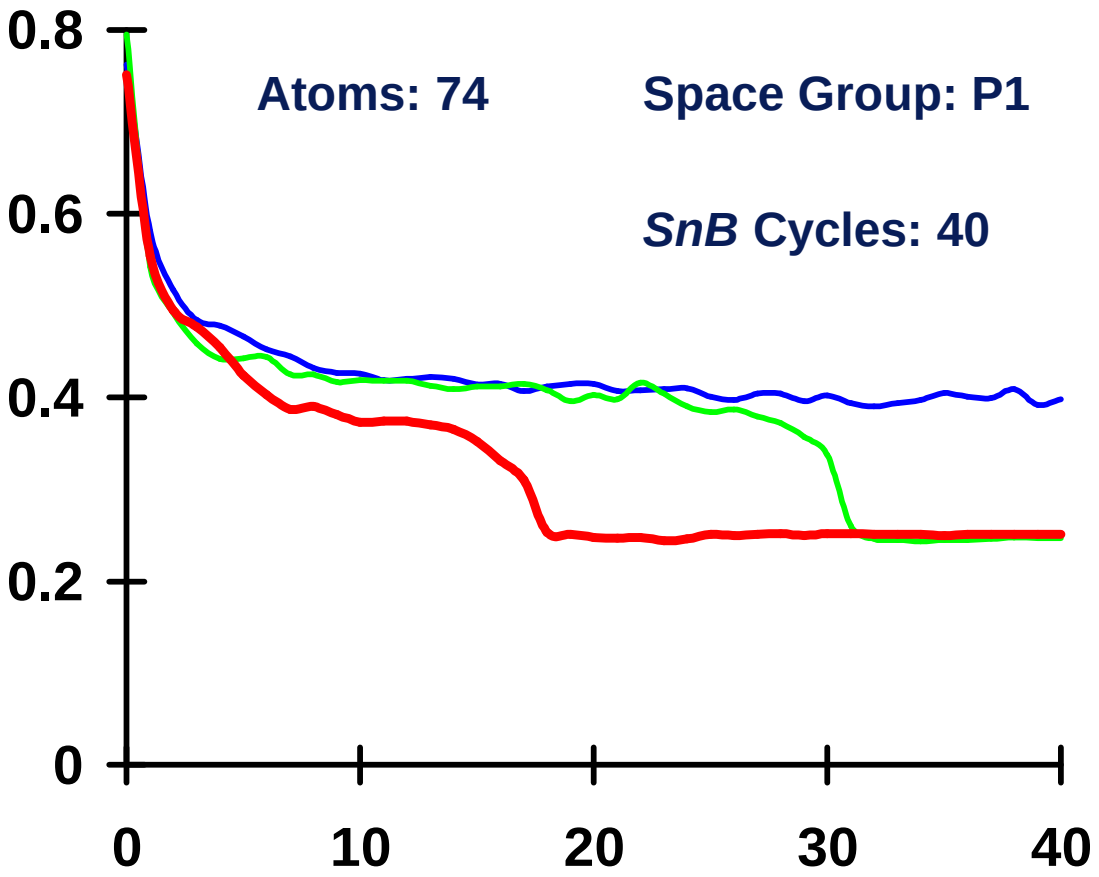
Ph8755: *SnB* Histogram

Atoms: 74
Space Group: P1

Phases: 740
Triples: 7,400



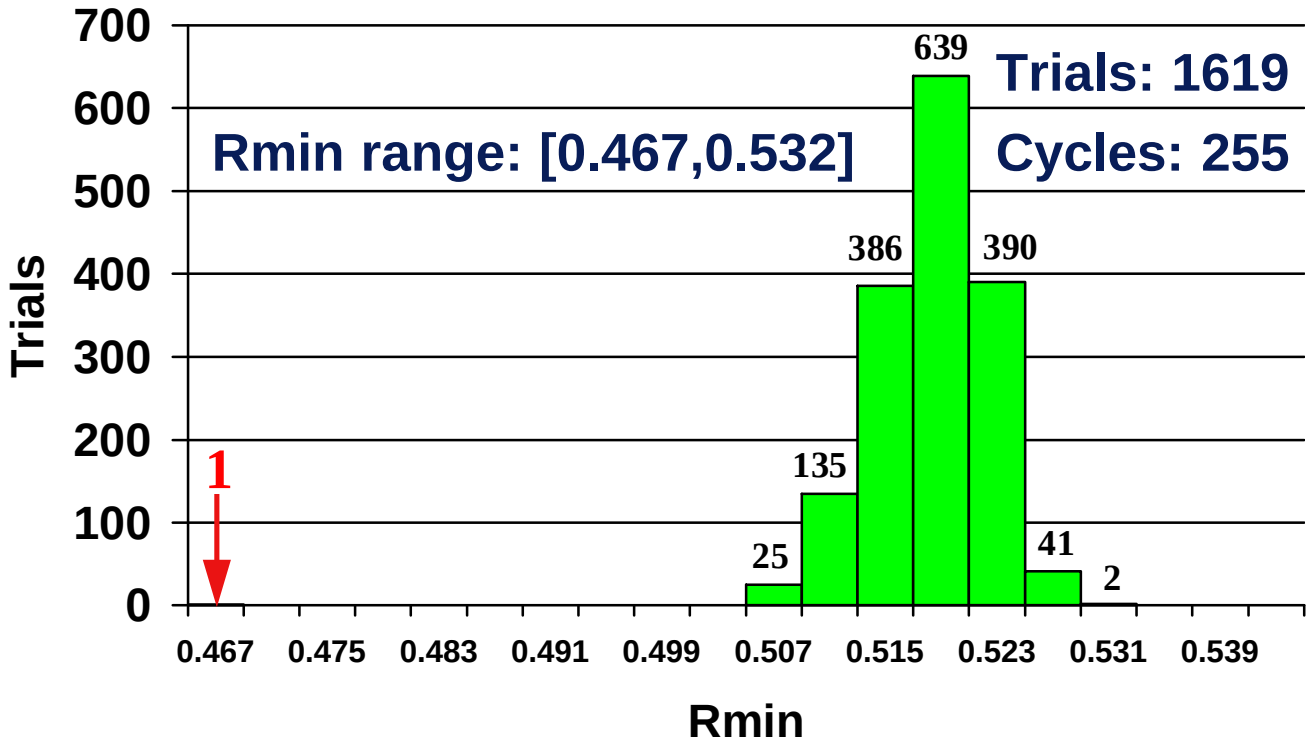
Ph8755: Trace of *SnB* Solution



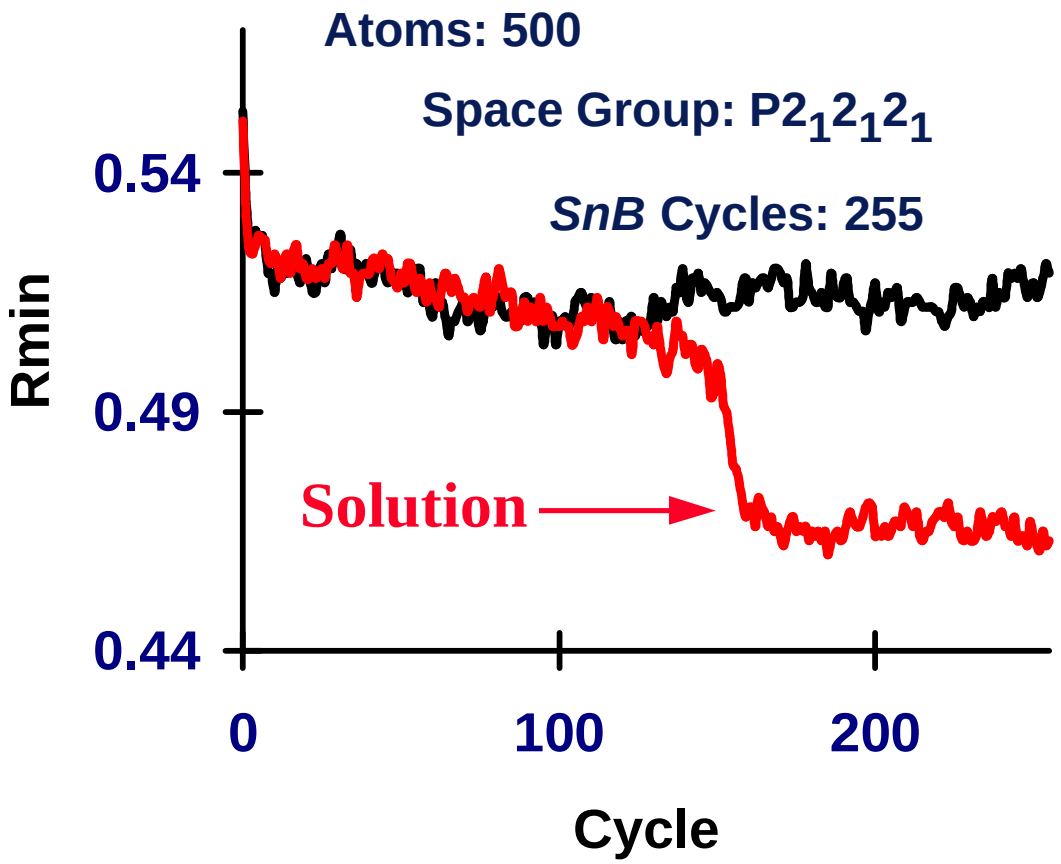
ToxII: *SnB* Histogram

Atoms: 500
Space Group: $P2_12_12_1$

Phases: 5,000
Triples: 50,000



Tox II: Trace of *SnB* Solution

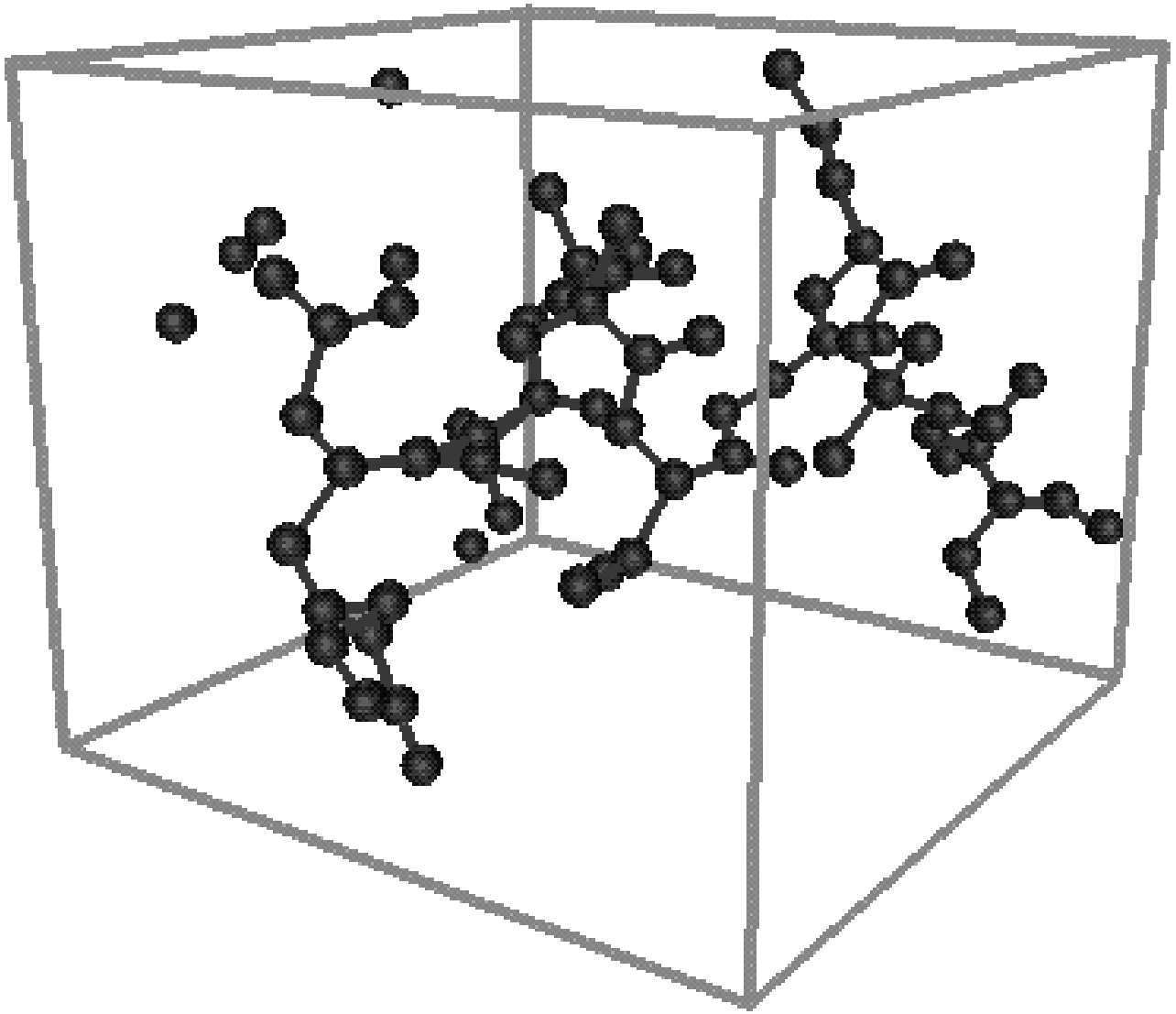


Some *SnB* Applications

Structure	Location	Atoms	Space Group
Riboflavin tetrabutyrate	MSC	180	P1
Vancomycin	Penn	258	$P4_32_12$
I4 Peptide	HWI	289	I4
Gramicidin A	HWI	317	$P2_12_12_1$
Er-1 pheromone	UCLA	328	C2
Crambin	HWI	~400	$P2_1$
Alpha-1 peptide	OCI/U. of T.	471	P1
Rubredoxin	HWI	497	$P2_1$
Scorpion Toxin II	HWI	624	$P2_12_12_1$

Visualization in *SnB* (Ph8755)

Geomview: Geometry Center, U. Minn.



An Interesting I4 Structure

SnB

- **Structure:**
  - Biological Peptide
  - 289 nonH atoms, including 10 sulfurs
  - 1.1Å resolution data
- **Bugs: Special Positions & Refinement**
- **Results (*SnB* 2.0)**
  - PS: 53%
  - PS/Rest: 44%
  - Tan: 25%

Extending Resolution: the I4 Structure

SnB

- **Truncate to 1.2Å - 1.5Å**
- **Solutions at all resolutions**
  - Significant success rates through 1.4Å
  - Detect solutions to 1.4Å with Rmin
- **Need better FOM at lower resolution**
- **Consider Crystallographic R**

Factors Determining Success Rate

SnB

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

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

SnB 2.0: Rationale

- **Improve running time**
- **Build from ground up**
- **Provide additional features**
  - Inverse Fourier
  - Density modification
  - Grid size
  - “Twice Baking”
  - Additional FOMs

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: Varying Peaks

Structure	*non-H Atoms*	*Heavy Atoms*	*Cycles*	*C.E. Peaks*	*Success Rate*
Vancomycin	258	*Cl* 8	200	25	0.8%
I4 Peptide	289	S10	250	50	53.0
Gramicidin A	317	-	275	200	1.1
Crambin	~400	S6	200	100	3.7
Rubredoxin	497	FeS6	400	100	6.2
Scorpion Toxin II	624	S8	500	200	1.0

SnB 2.0: Refinement

Structure	*PS Standard*	*PS Restricted*	*Tangent*
Vancomycin	0.6%	0.4%	0.3%
I4 Peptide	53.0	44.0	25.0
Gramicidin A	0.9	0.5	0.0
Crambin	4.8	3.7	2.2
Rubredoxin	6.0	5.2	3.6
Scorpion Toxin II	1.4	1.0	0.6

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.

Summary

SnB

- ***Shake-and-Bake:* Dual-Space Direct Methods**
- **Targeted at 100-800 atom structures**
- **Publicly available**
- **LEVY / EVAL (Bob Blessing) for |E|'s**
- ***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)


SnB