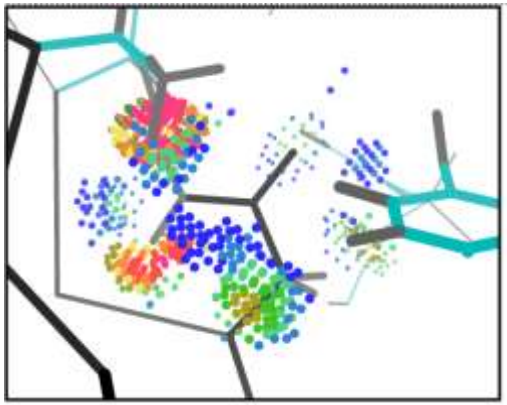
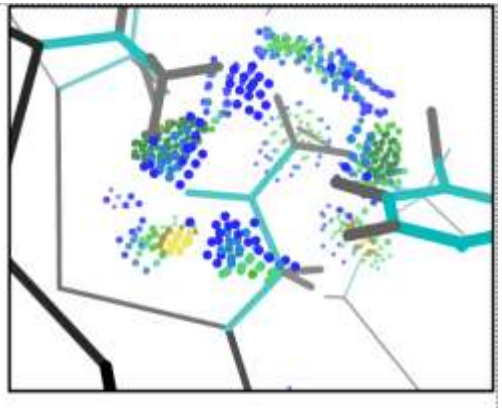


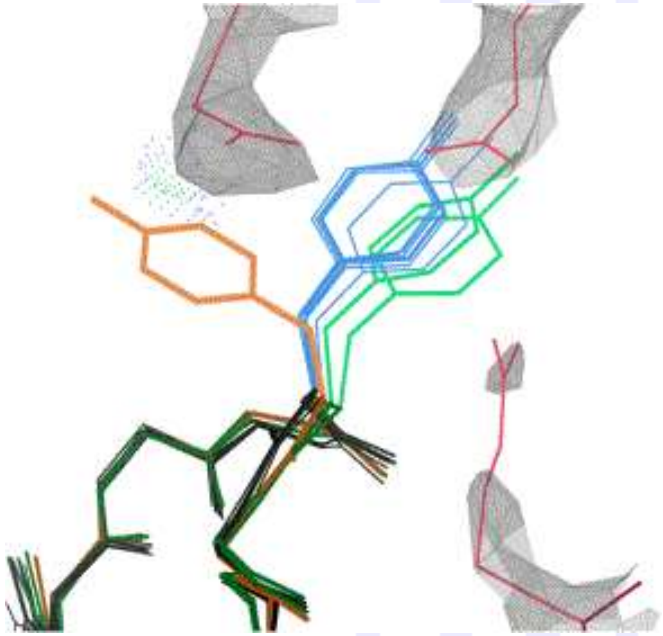


Symmetry-Aware Placement of Hydrogens in Molecules: Reduce & cctbx

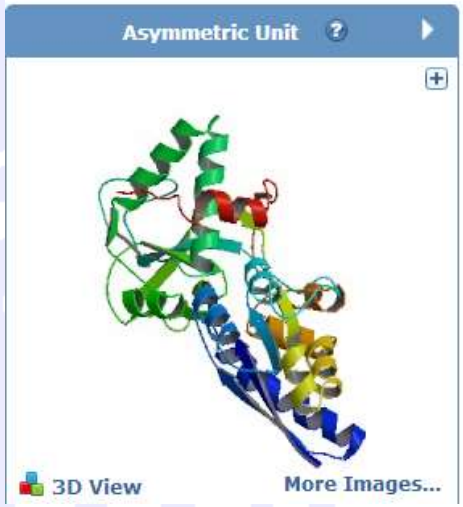
Jack Snoeyink
Auston Sterling
Vishal Verma

Computer Science

UNC Chapel Hill



Outline



Determining Molecular Structure

- Why: "central dogma"
- How: X-ray crystallography
  - Role of symmetry

Structure Validation

- All Atom Contact Analysis: Molprobity
- Hydrogen placement: Reduce

Symmetry-aware Reduce:

- SMP: simple matter of programming

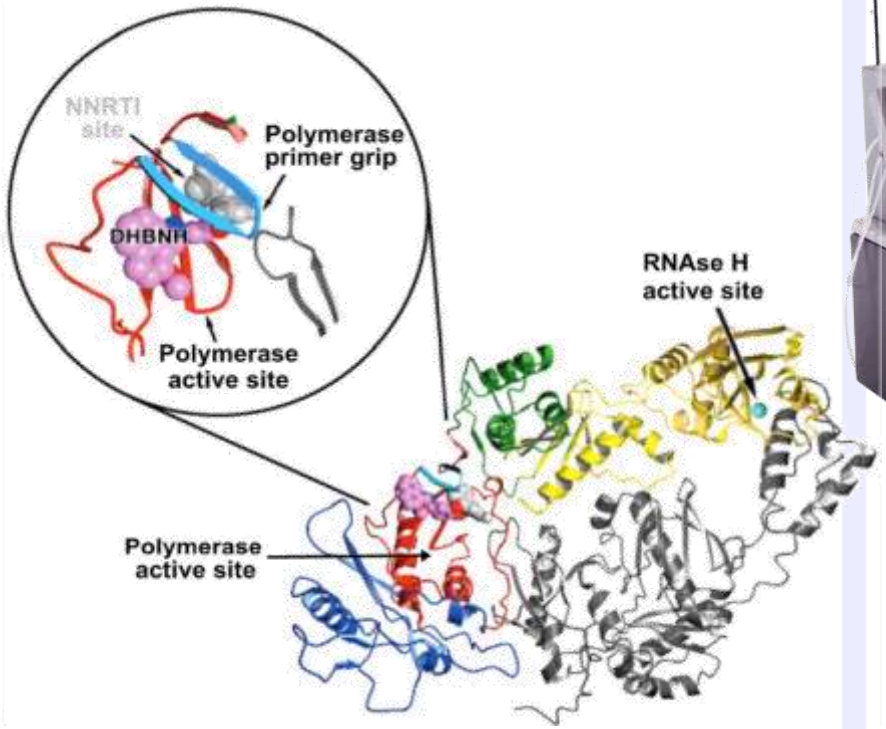
Central Dogma of Biochemistry

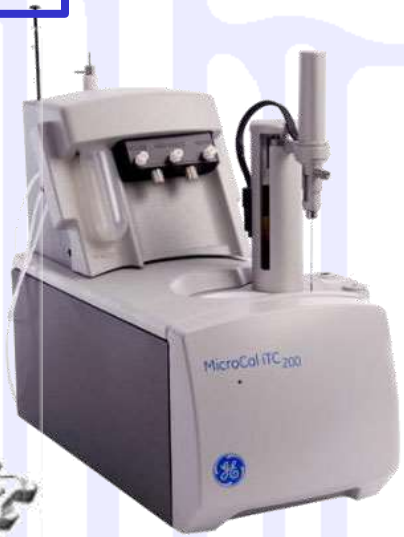
folding

mechanism

Sequence → Structure → Function

VAL LEU SER GLU GLY
LEU HIS VAL TRP ALA
VAL ALA GLY HIS GLY
ARG LEU PHE LYS SER
GLU LYS PHE ASP ARG
THR GLU ALA GLU MET
LEU LYS LYS HIS GLY
ALA LEU GLY ALA ILE
HIS HIS GLU ALA GLU
GLN SER HIS ALA THR
ILE LYS TYR LEU GL
ILE ILE HIS VAL LEU
GLY ASP PHE GLY ALA
MET ASN LYS ALA LEU
ASP ILE ALA ALA LYS
TYR GLN GLY

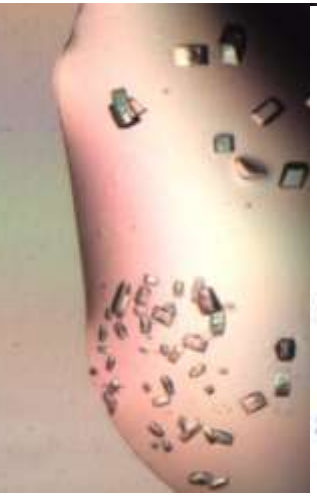


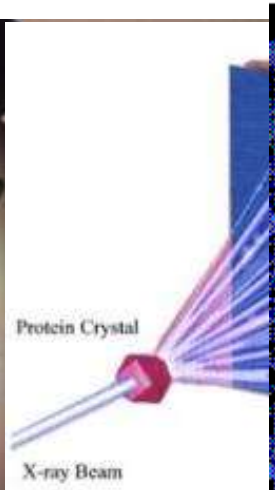


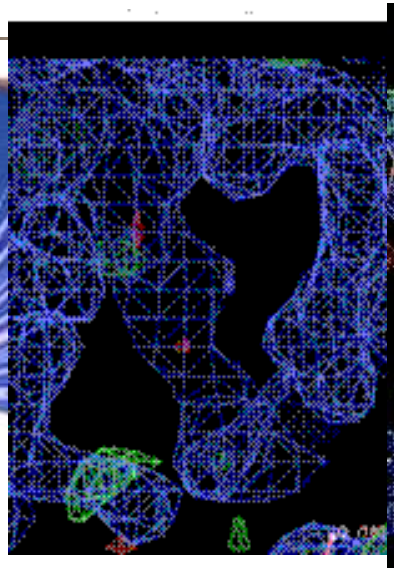
X-ray crystallography to find structure

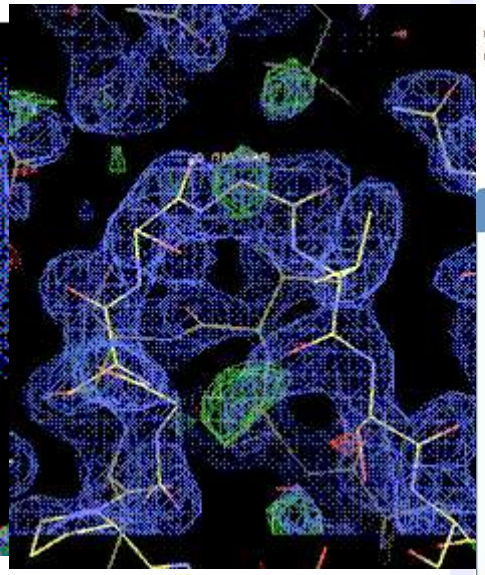
Sequence → Structure → Fu

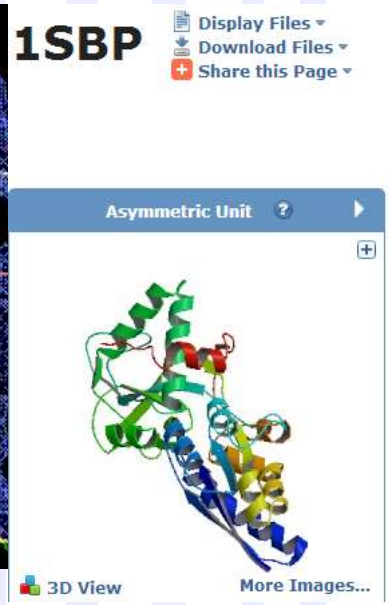
Make crystal	X-ray diffraction	Electron density	Backbone threading	Rotamer placement	Structure Validation







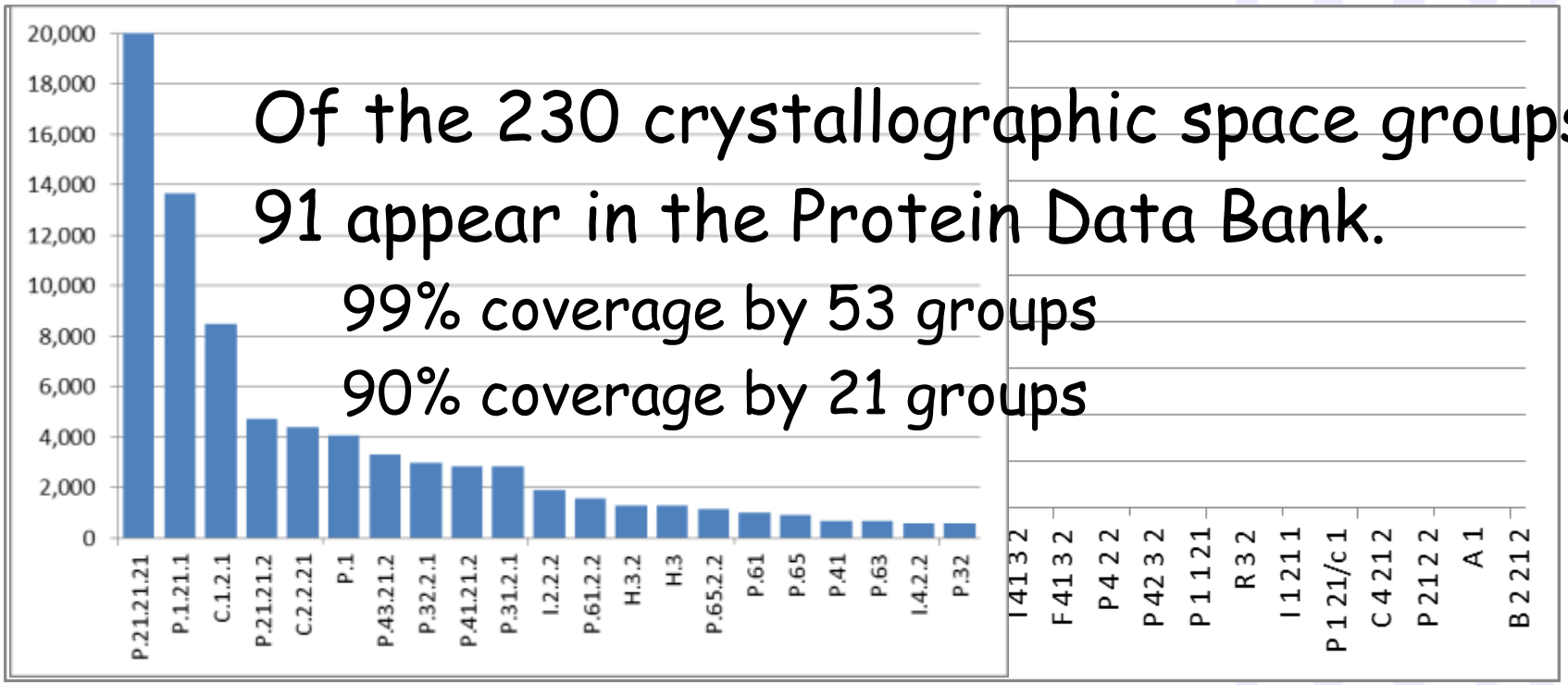




Role of symmetry

Sequence → Structure → Fu

Make crystal	X-ray diffraction	Electron density	Backbone threading	Rotamer placement	Structure Validation

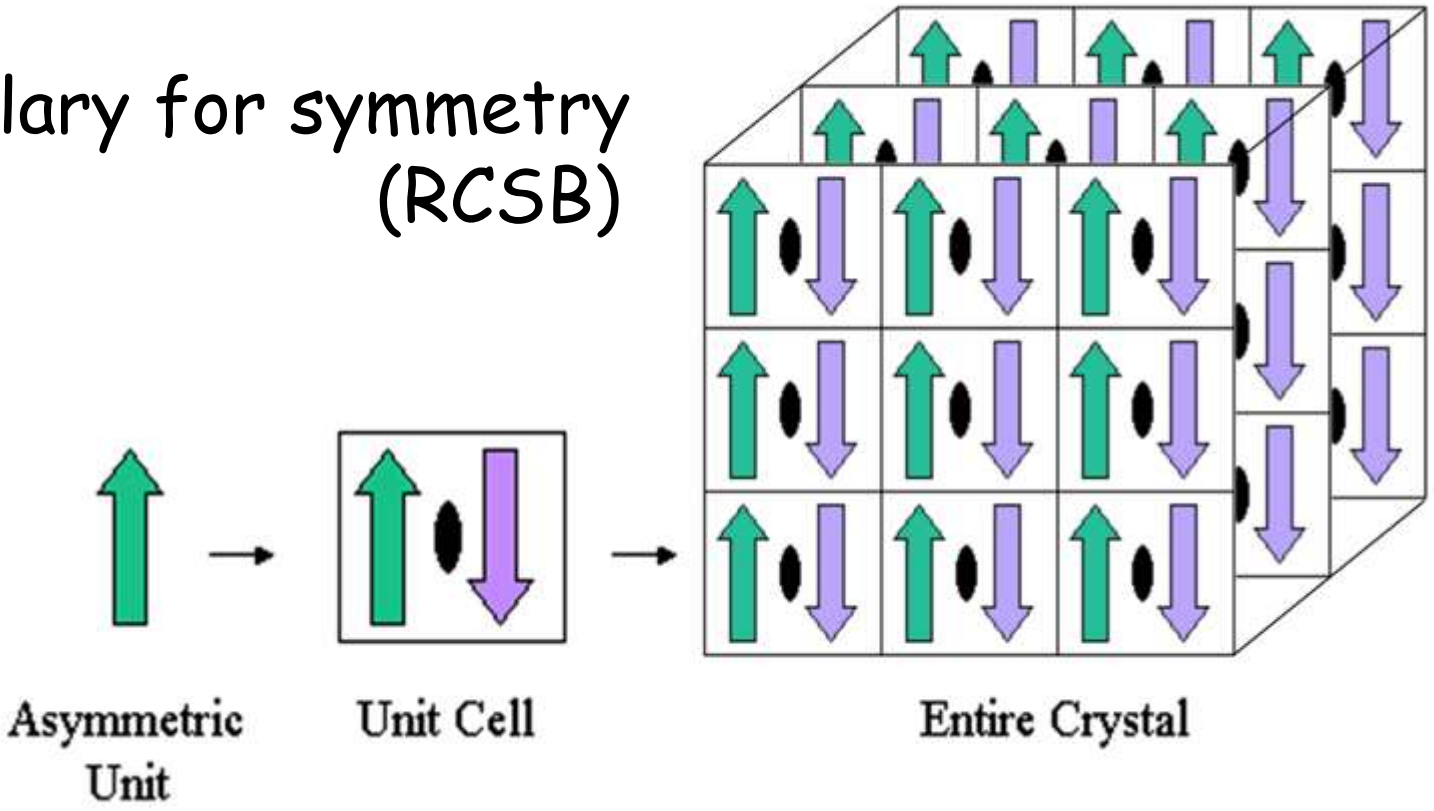


Of the 230 crystallographic space groups,
91 appear in the Protein Data Bank.

99% coverage by 53 groups

90% coverage by 21 groups

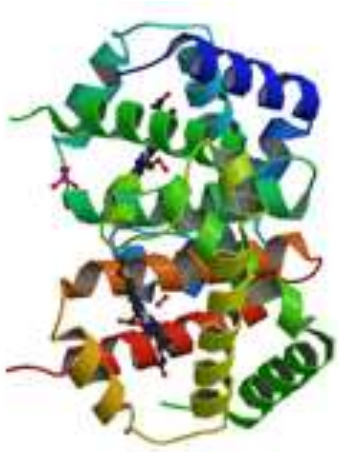
Vocabulary for symmetry (RCSB)



Symmetry group action on Asymmetric unit gives Unit cell.

Note: Asymmetric unit has a model;
Model need not lie inside it, or inside the unit cell.

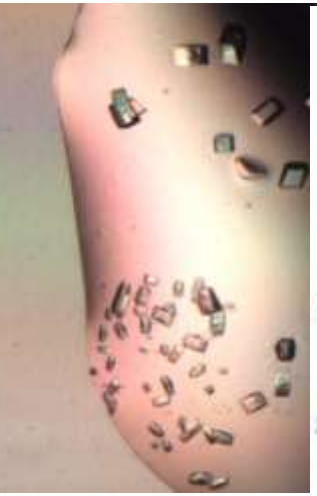
Asymmetric unit vs. Biological assembly

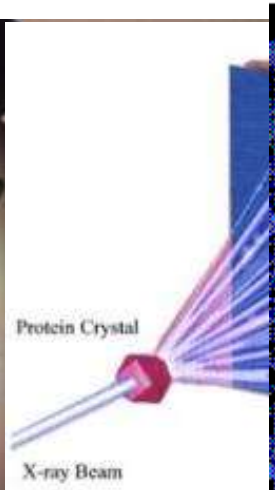
Asymmetric unit with portion of a biological assembly	Asymmetric unit with one biological assembly	Asymmetric unit multiple biological assemblies
		
Entry **1hho** contains **half** a hemoglobin molecule (**2 chains**) in the asymmetric unit. A crystallographic two-fold axis generates the other 2 chains.	Entry **2hhb** contains **one** hemoglobin molecule (**4 chains**) in the asymmetric unit.	Entry **1hv4** contains **two** hemoglobin molecules (**8 chains**) in the asymmetric unit.

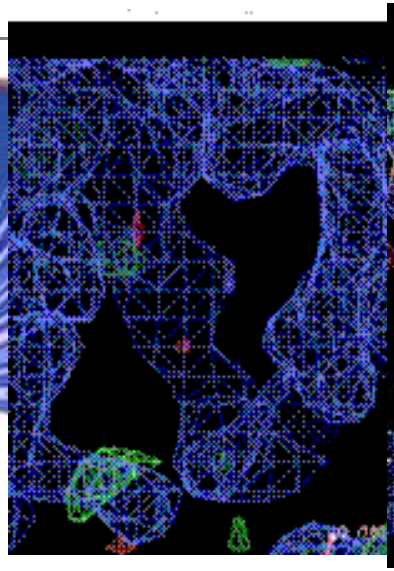
X-ray crystallography to find structure

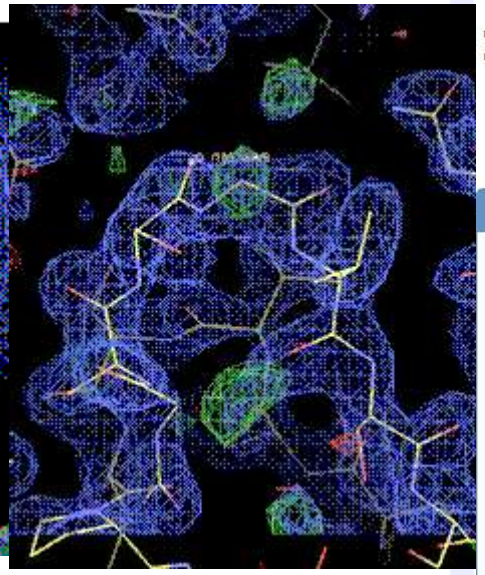
Sequence → Structure → Fu

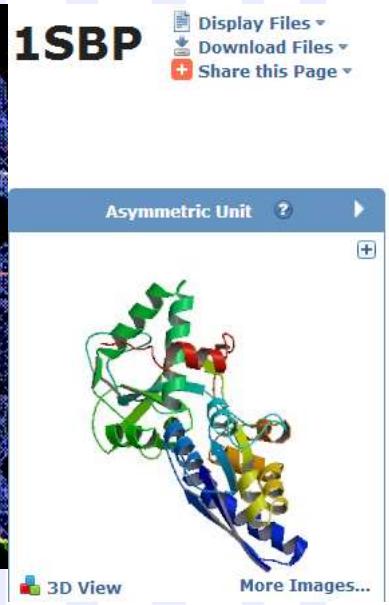
Make crystal	X-ray diffraction	Electron density	Backbone threading	Rotamer placement	Structure Validation







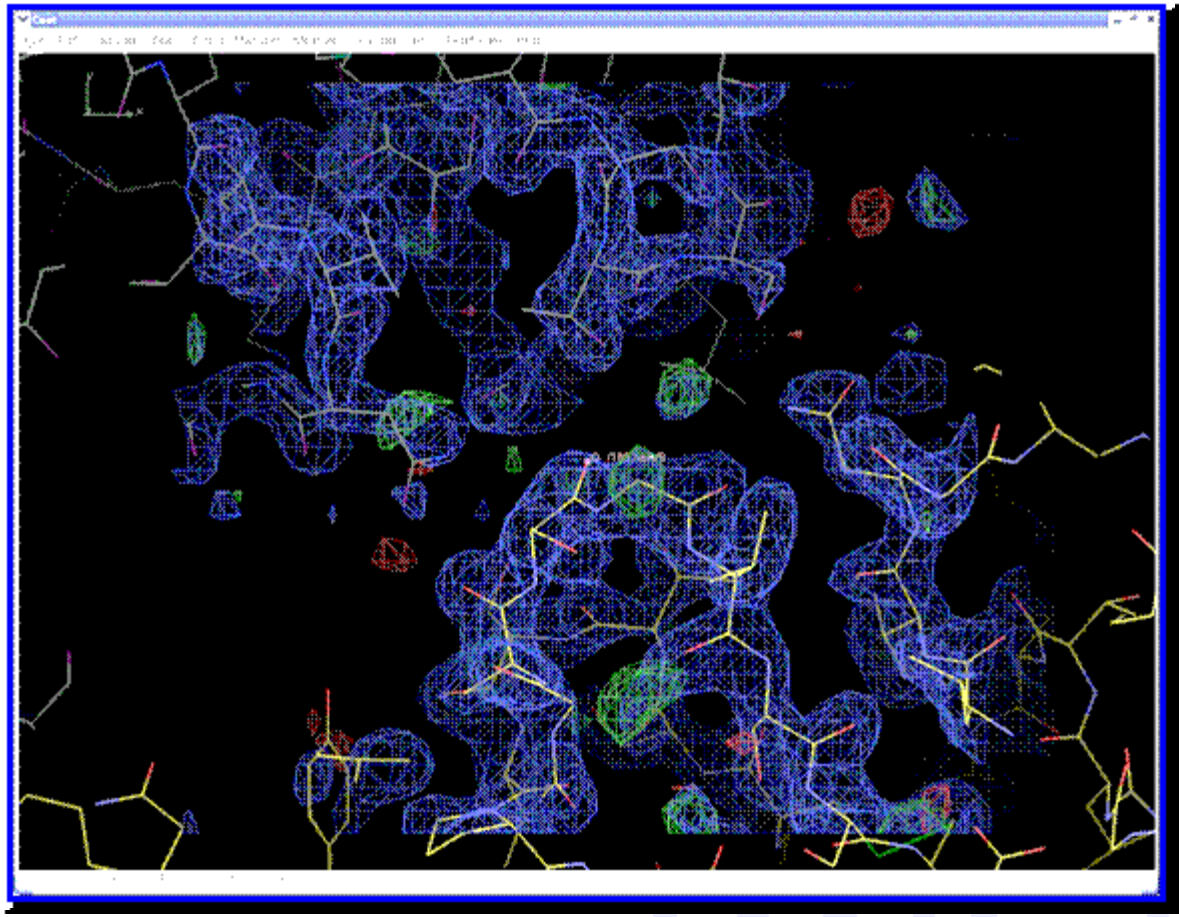




Role of symmetry in threading

Coot tutorial:

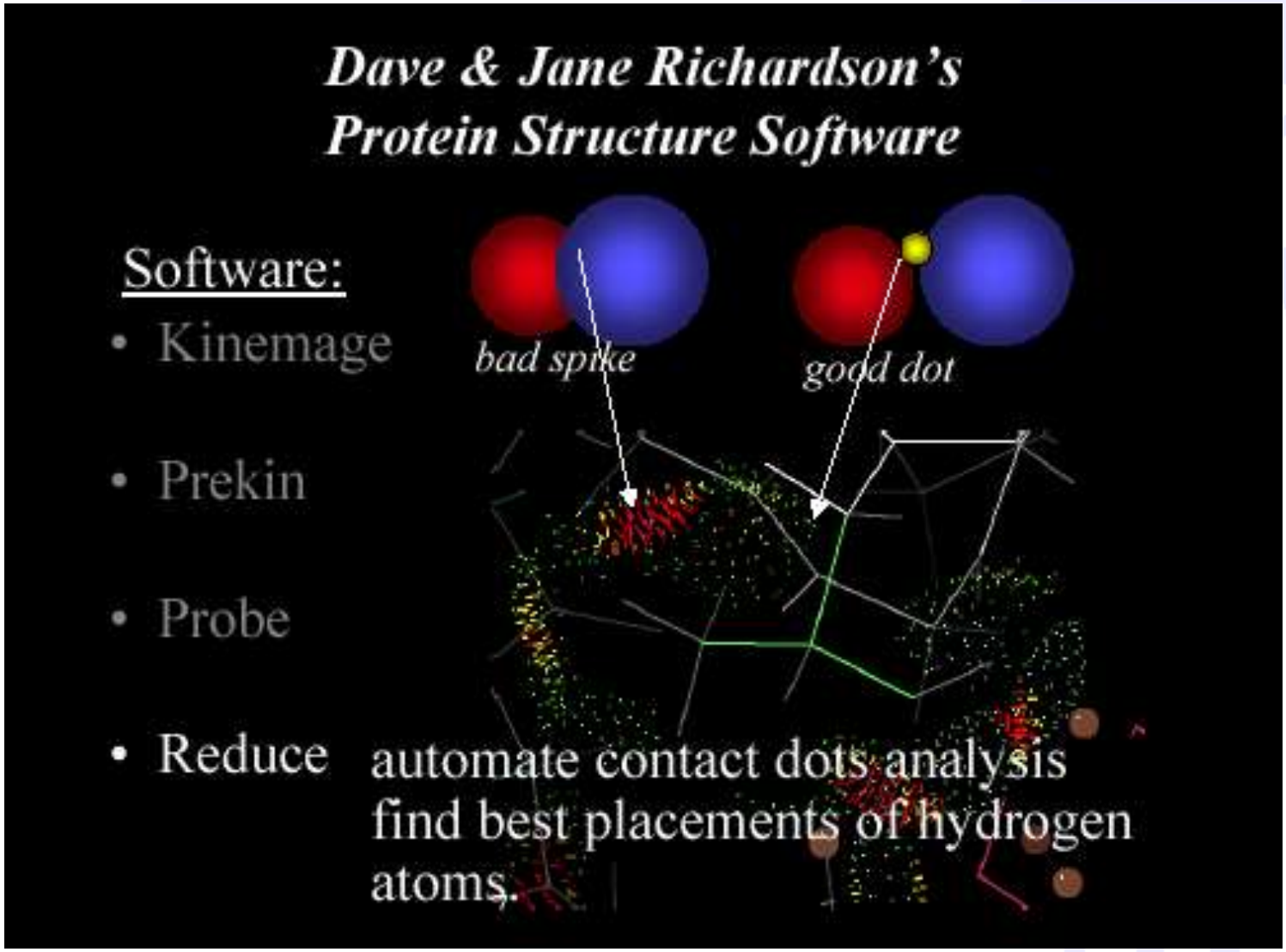
- Density without structure may be symmetric copy
- Check: turn on model symmetry



Key point for me:

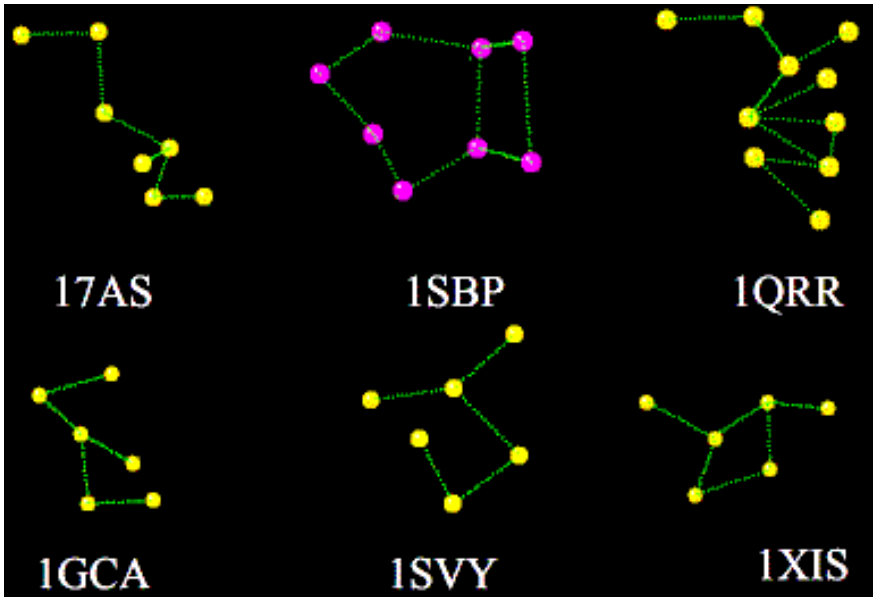
- for consistency use their library: Comp. Cryst. Toolbox (cctbx)

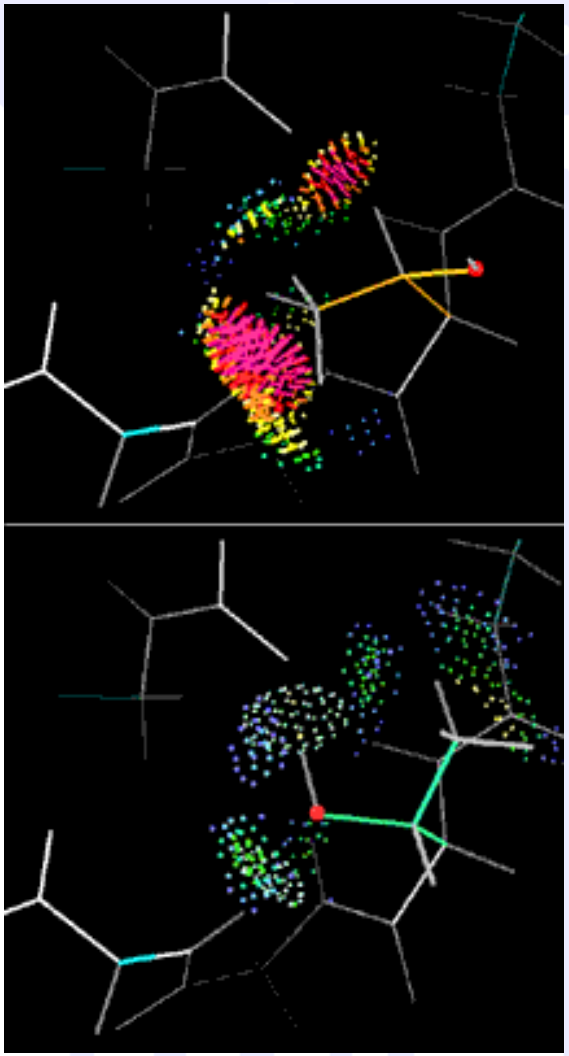
All atom contact analysis: Molprobity demo



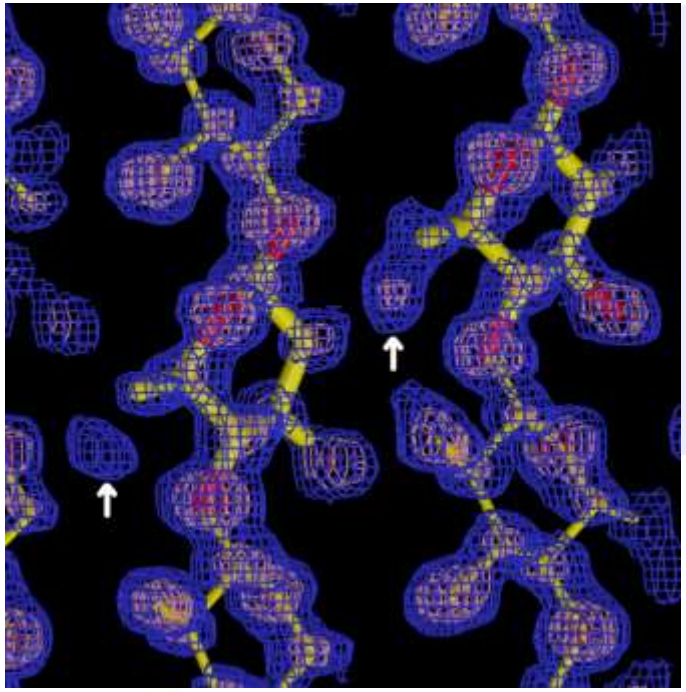
Reduce: Hydrogen placement by dynamic programming on graphs of small treewidth

- Reduce considers flips & rotations, which may interact.
- Interaction graphs' small treewidth allows fast dynamic programming.





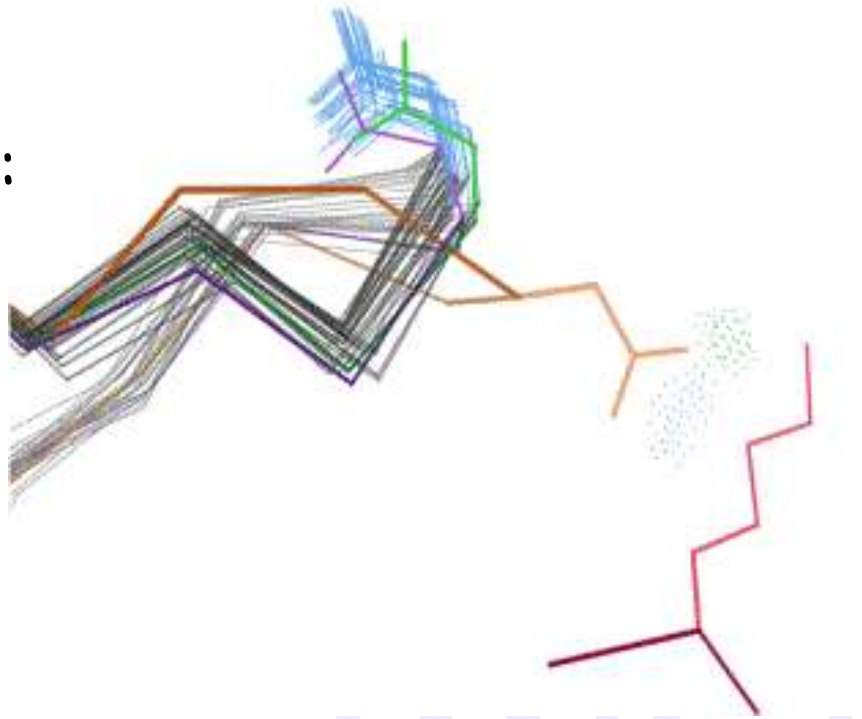
Single model analysis misses "crystal contacts"



- Validation (Reduce)
- Crystallography (PHENIX/Coot)
- Protein folding (Rosetta)

eg. analysis of Rosetta decoys:

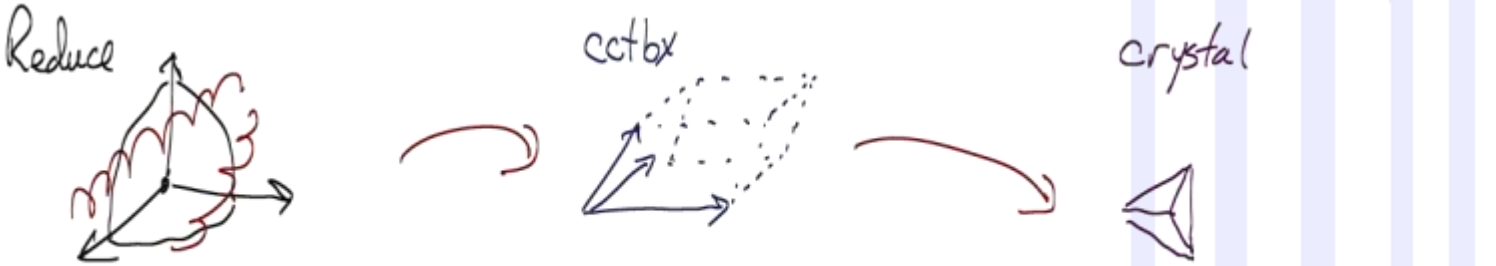
- native
- decoys
- symmetric natives



SymReduce: find neighbors

Q: How should Reduce find its neighbors using the crystallographic symmetries?

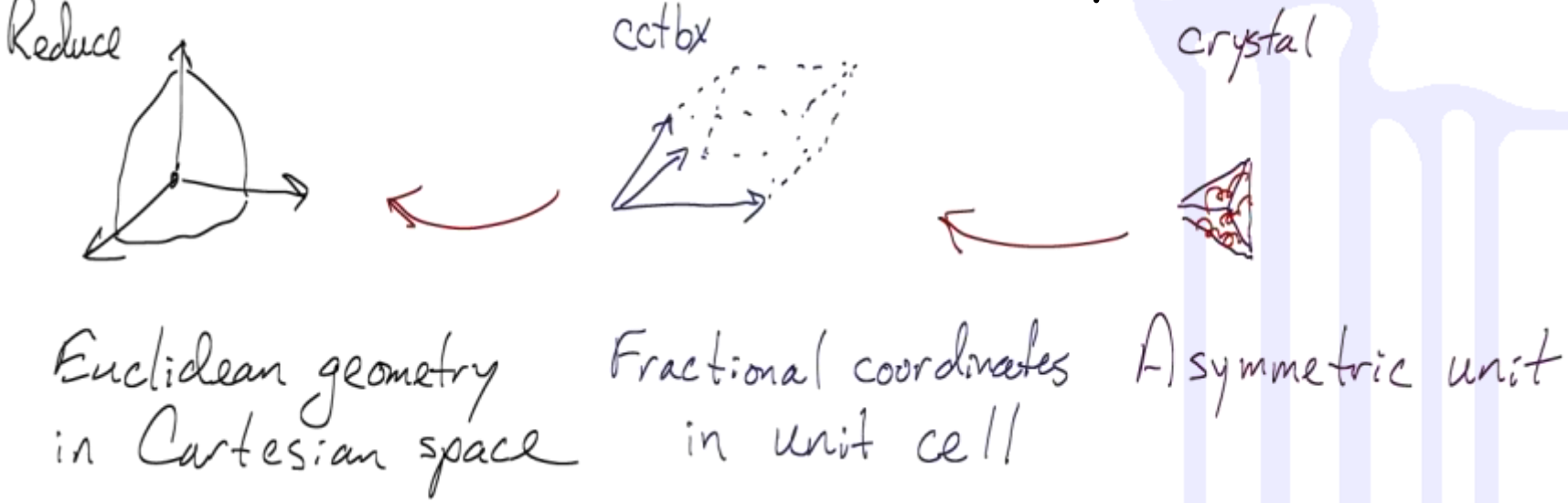
A: Bucketing…Reduce folds lattice into unit cell; cctbx library folds into asymmetric unit



	Euclidean geometry in Cartesian space	Fractional coordinates in unit cell	Asymmetric unit
Data:	vdW spheres	atoms	
Queries:	dots (200-600/atom) or candidate H positions	atoms	

SymReduce: find neighbors

Copy atoms from the asymmetric unit that lie within an interaction distance limit of the asym. unit.



Bucketing must be done in the Euclidean space.

SMP: "simple" matter of programming

- To accommodate candidate H atom positions, we had to add query (x, y, z) functionality to cctbx; otherwise we'd need to add/delete-last.

Remaining to do:

- Testing in PHENIX
- Naming convention for output of symmetric copies of atoms.

Possible extensions

- Speed up, e.g., queries by batching nearby dots.
- Detect if the "right" symmetry has been specified.

Thanks

- Richardson lab (Molprobity)
- Ralf Grosse-Kunstleve (cctbx)
- PHENIX
- Rosetta Commons
- NIH, NSF