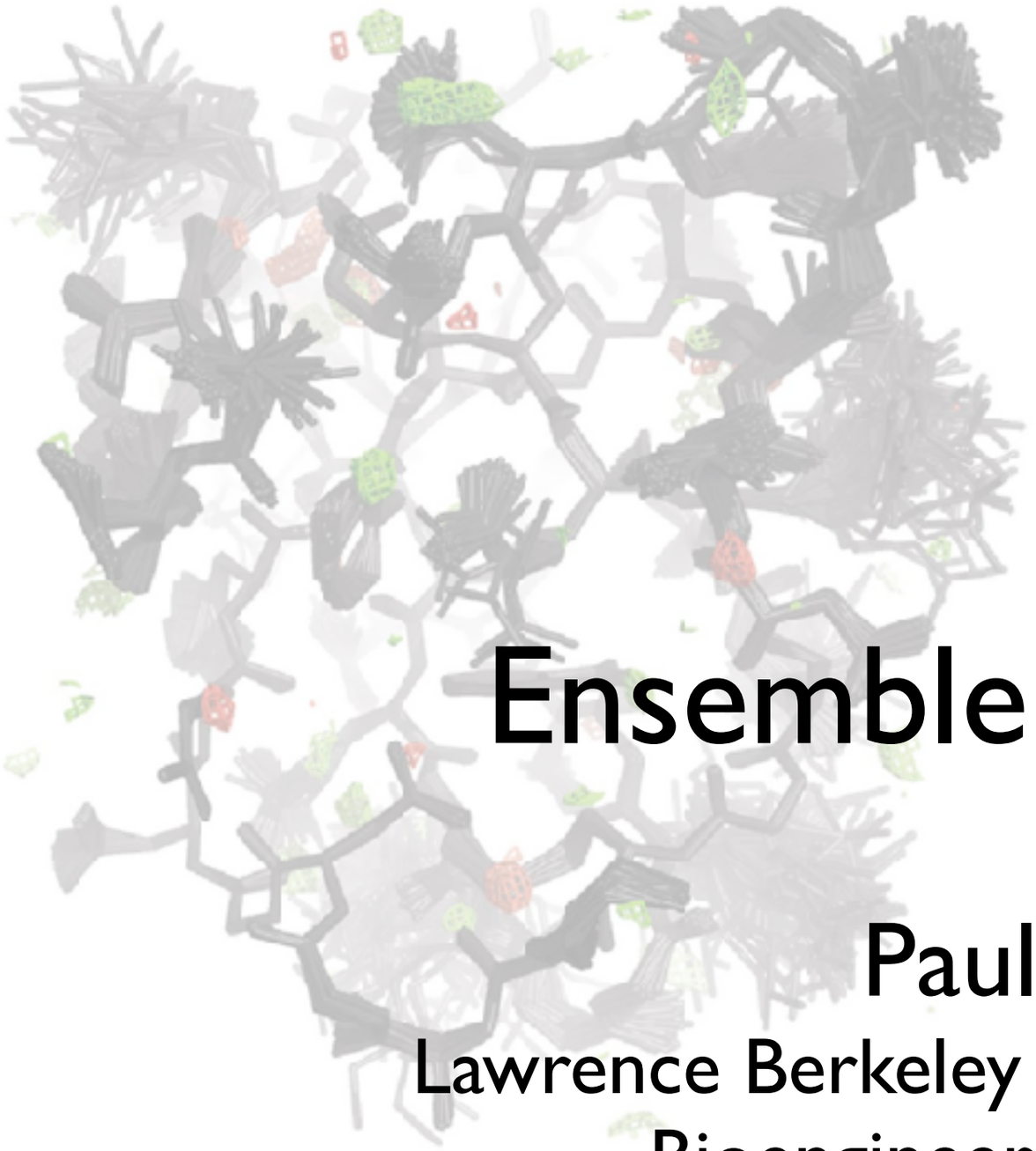


Ensemble Refinement

Paul Adams

Lawrence Berkeley Lab and Department of Bioengineering UC Berkeley

Ensembles in Crystallography

- Use of ensembles as a practical tool:
  - Molecular replacement
  - Phase improvement
  - Refinement
- Acknowledging and representing conformational heterogeneity
  - Building
  - Refinement

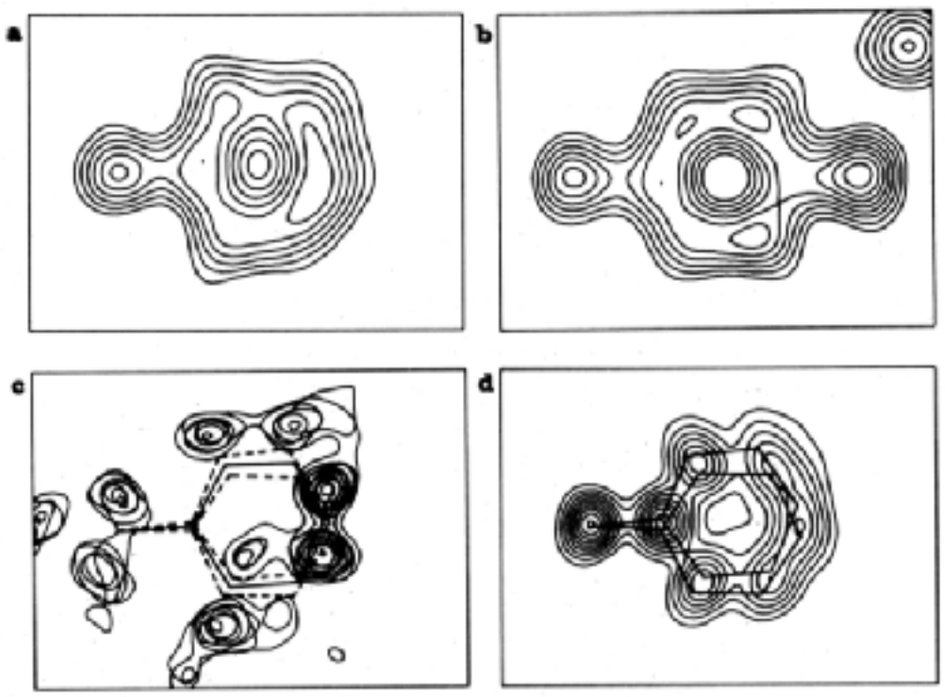
Ensembles

- A set of related but conformationally different models of the same structure

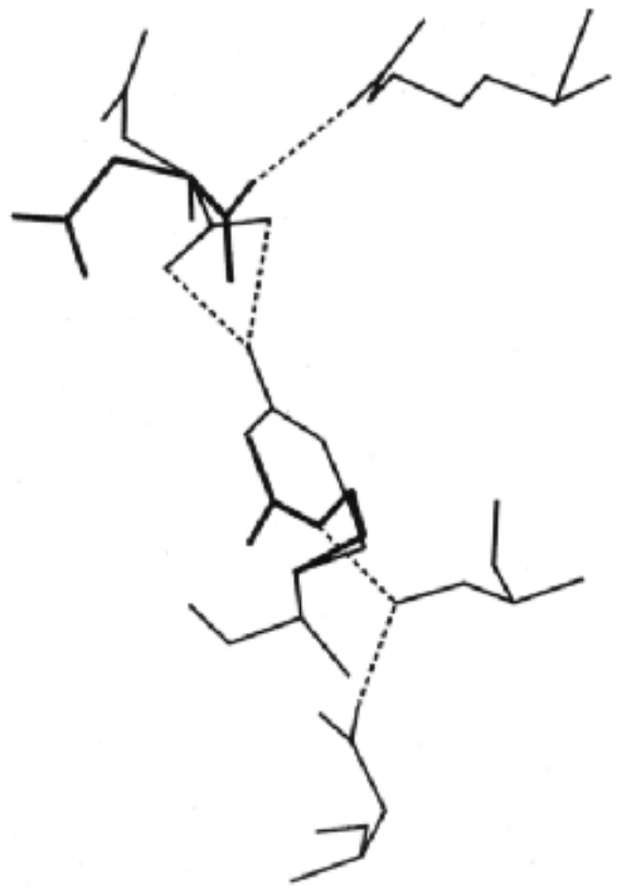
Representing conformational heterogeneity

Conformational Heterogeneity

- Not described in crystallographic models until the 1980s
  - Refinements of high resolution structures



Density supporting alternate conformations



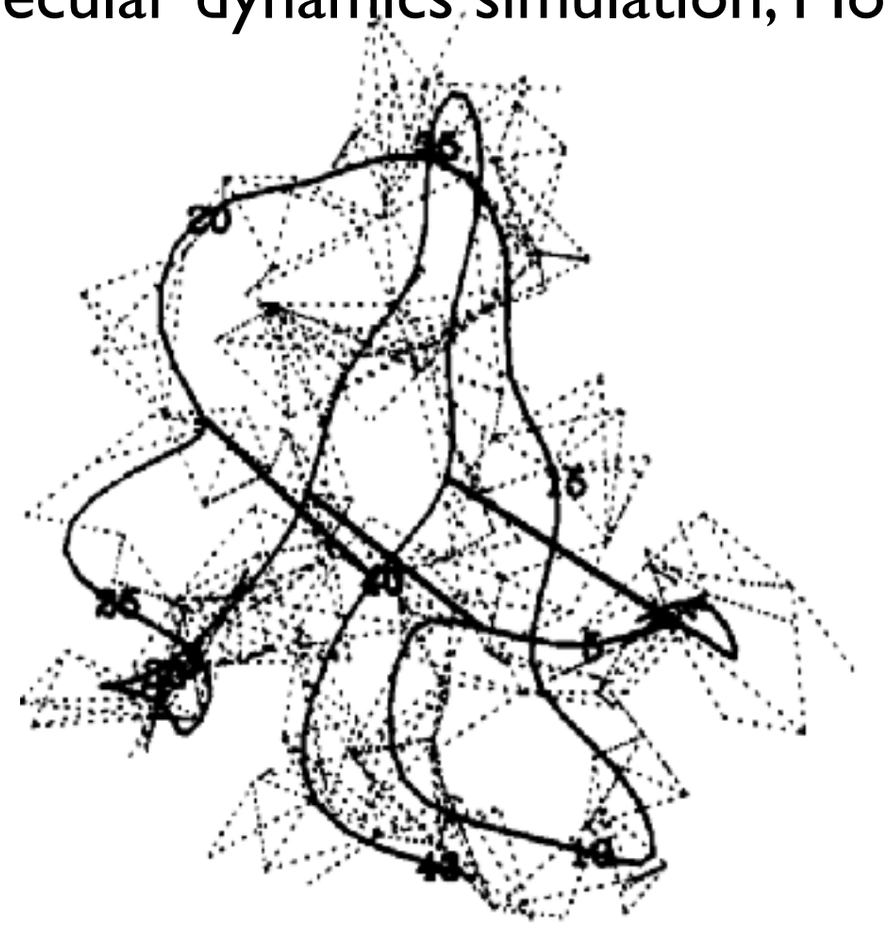
Correlated alternate conformations

Alternate water structures

Smith JL, Hendrickson WA, Honzatko RB, Sheriff S: Structural heterogeneity in protein crystals. *Biochemistry* 1986, 25:5018-27

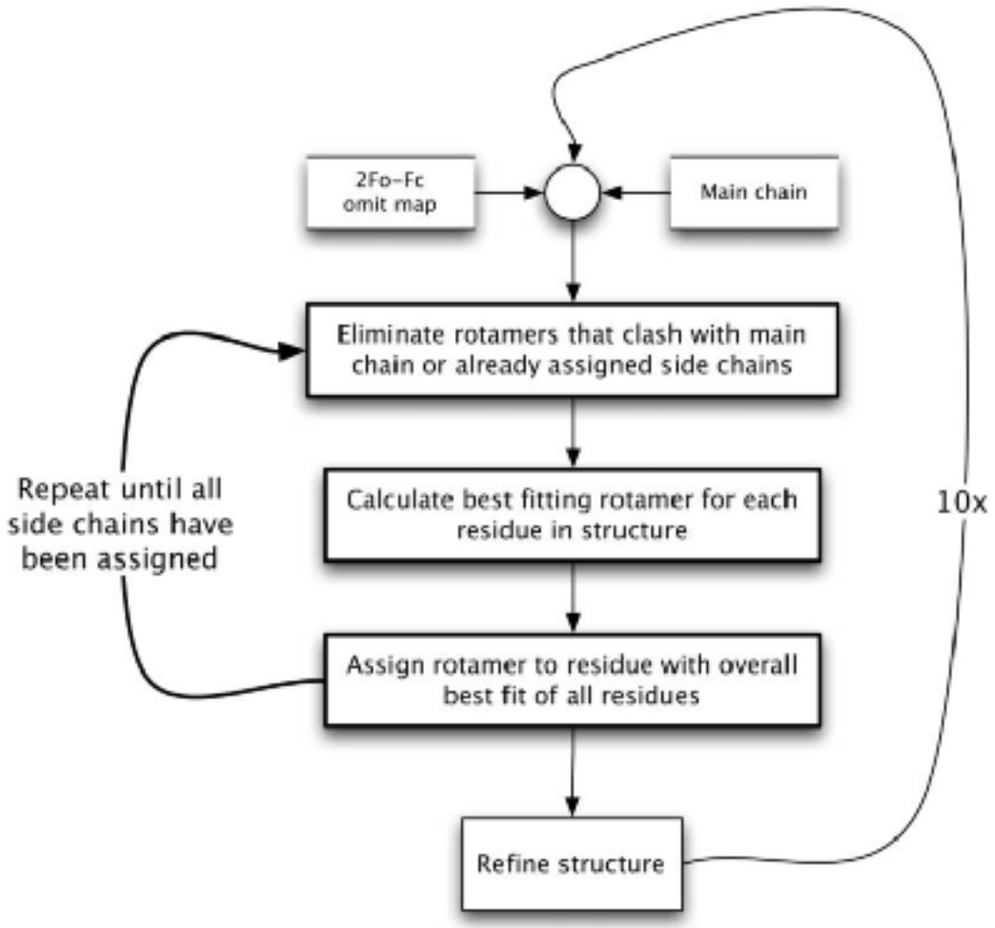
NMR Structure Calculation

- Through-space interactions can define the fold of a molecule (given sufficient distances)
- These interactions can be used as restraints in geometry calculation algorithms:
  - Molecular dynamics simulation, Monte Carlo sampling, energy minimization

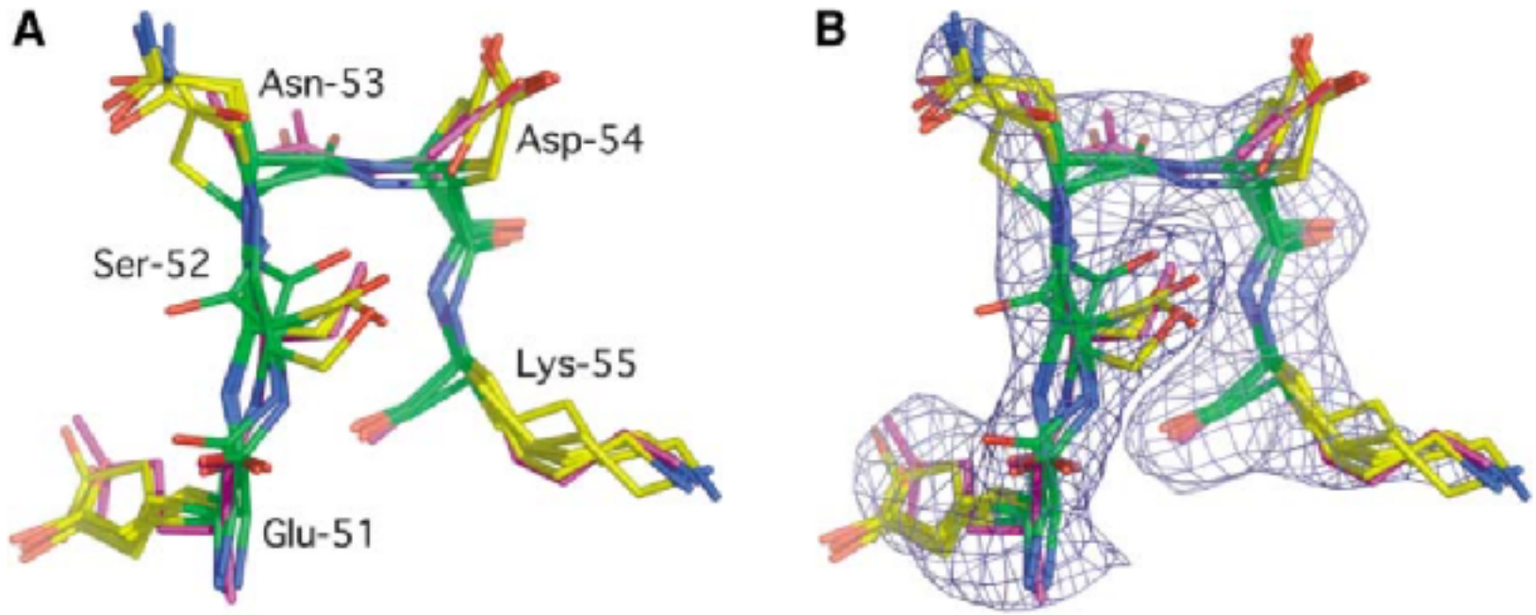


Driscoll PC, Gronenborn AM, Beress L, Clore GM: Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry*. 1989, 28:2188-98.

Ensembles from Model Rebuilding



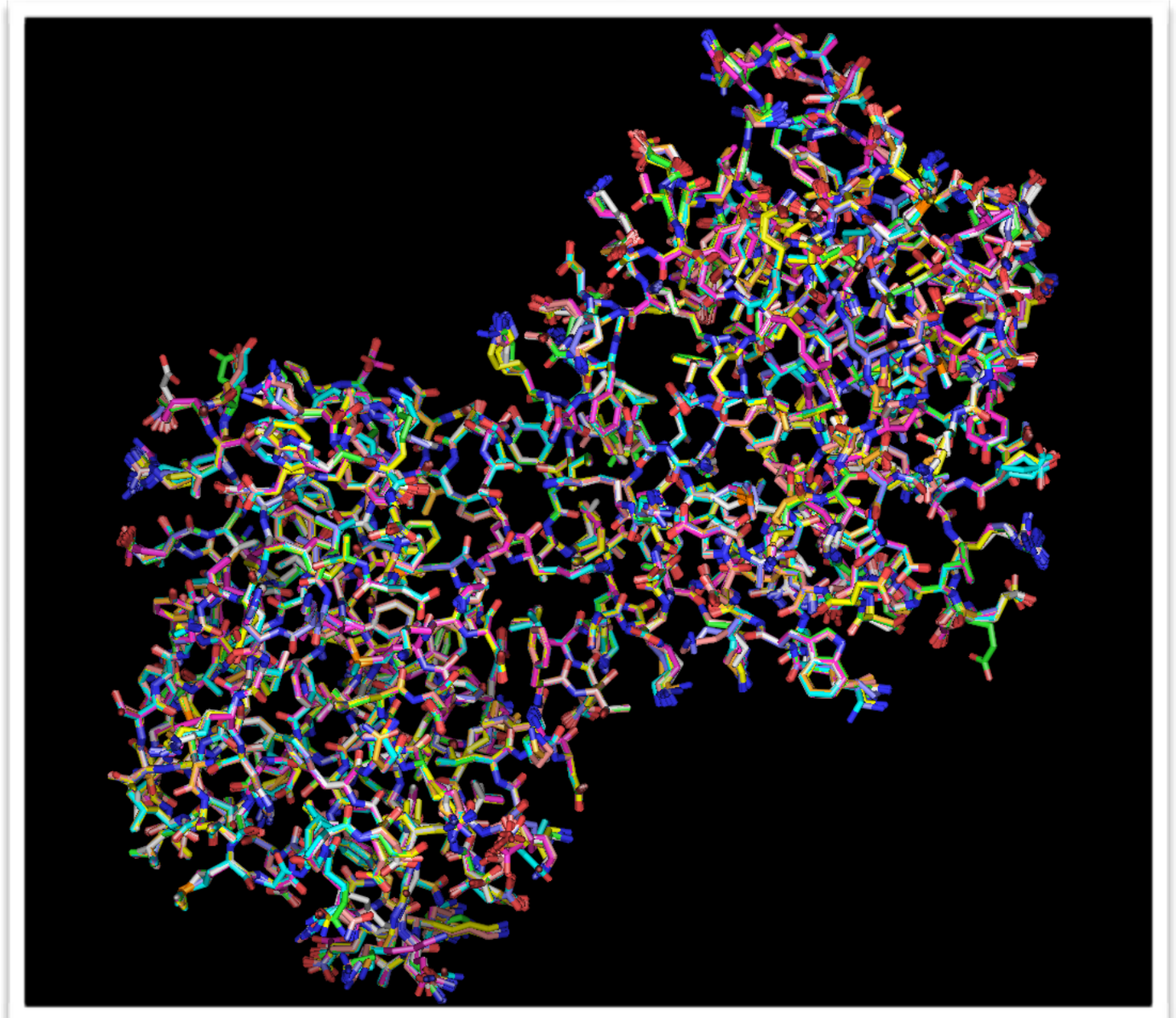
- Automated rebuilding/refinement procedure that creates multiple models consistent with the data
- Average R-free usually better than any individual model



DePristo MA, de Bakker PI, Blundell TL: Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* 2004, 12:831-8

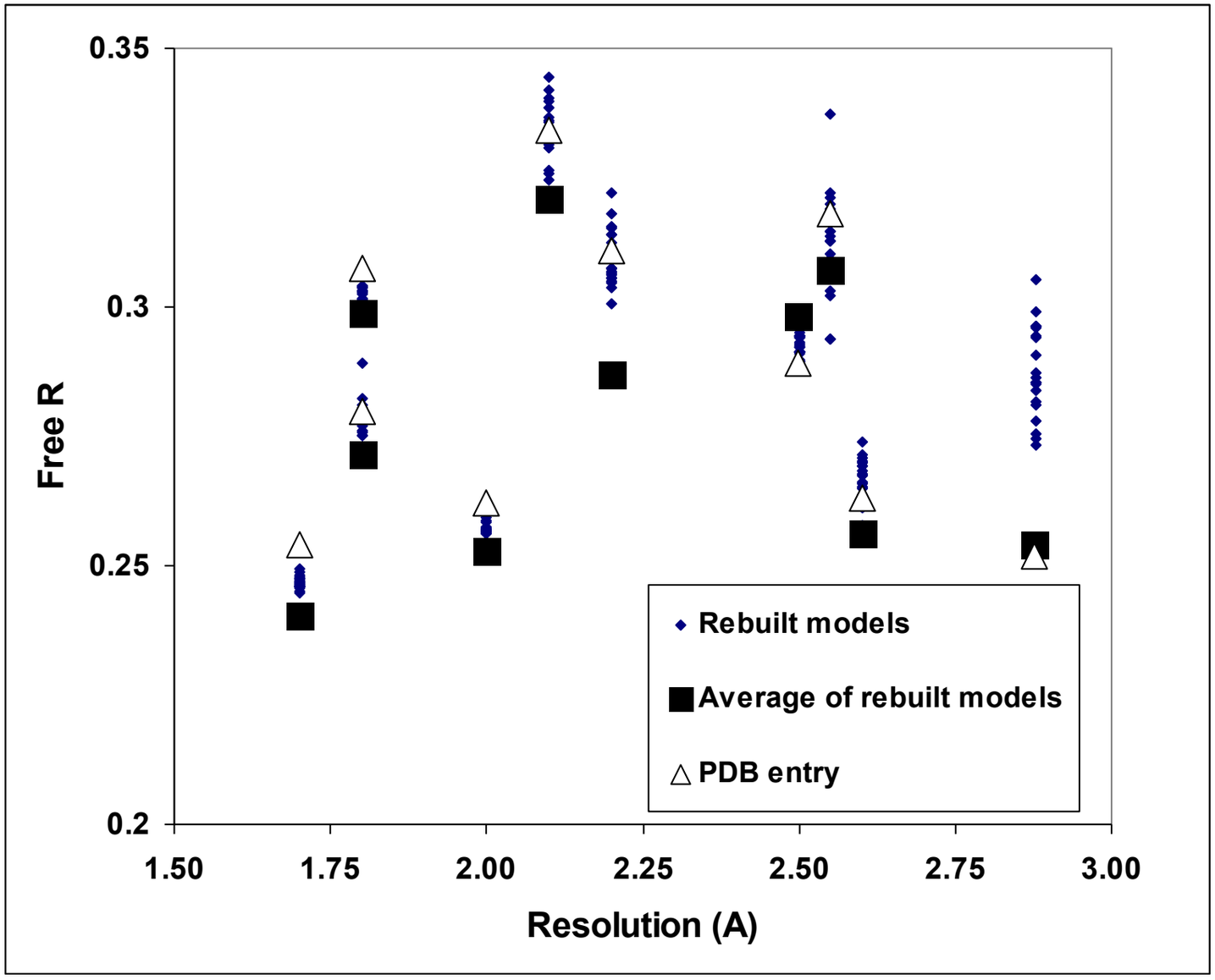
Ensembles from Model Building

- Furnham N, DePristo M, Blundell T, Terwilliger T: PDB Deposits of X-ray structures should be a group of models representing the range of structures compatible with the data. *Nature Struct Mol Biol*, 2006.



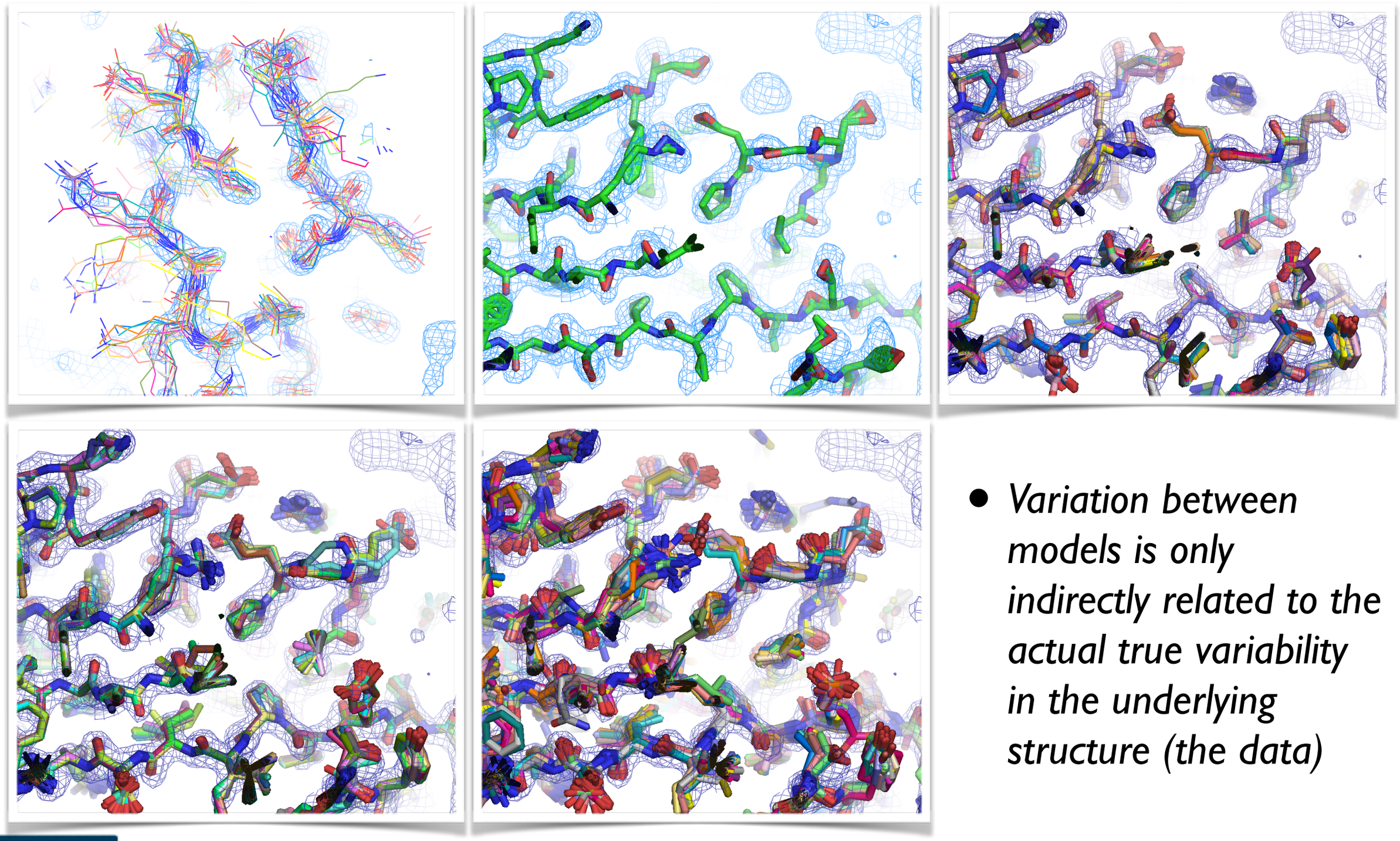
Ensembles from Model Building

- Ensembles of models are a better fit to the data (even when built independently of each other)



Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Adams PD, Moriarty NW, Zwart P, Read RJ, Turk D, Hung LW: Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. *Acta Cryst.* 2007, D63:597-610

Variation in Models Depends on Resolution

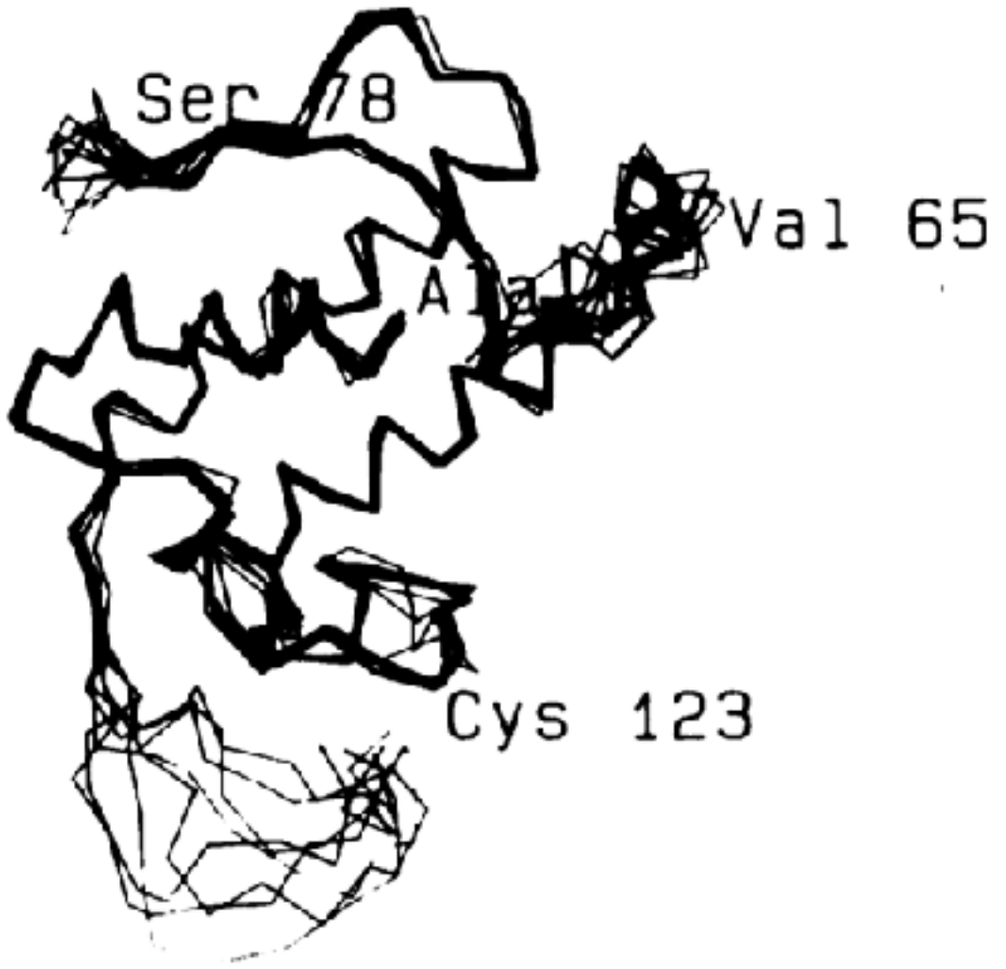


- *Variation between models is only indirectly related to the actual true variability in the underlying structure (the data)*

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Adams PD, Moriarty NW, Zwart P, Read RJ, Turk D, Hung LW: Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. *Acta Cryst.* 2007, D63:597-610

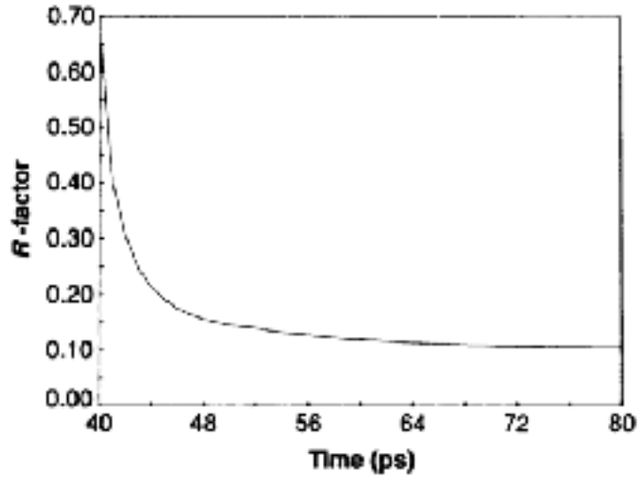
Time-Averaging

- Using MD simulation during refinement to build up an ensemble of models that collectively fit the data
- Atomic displacements modelled by the ensemble
- Captures harmonic and anharmonic displacements



$$E = E_{\text{phys}} + \frac{1}{\sigma_x^2} \sum_{\mathbf{s}} \left(|\mathbf{F_o(s)}| - k|\langle \mathbf{F_c(s)} \rangle|\right)^2$$

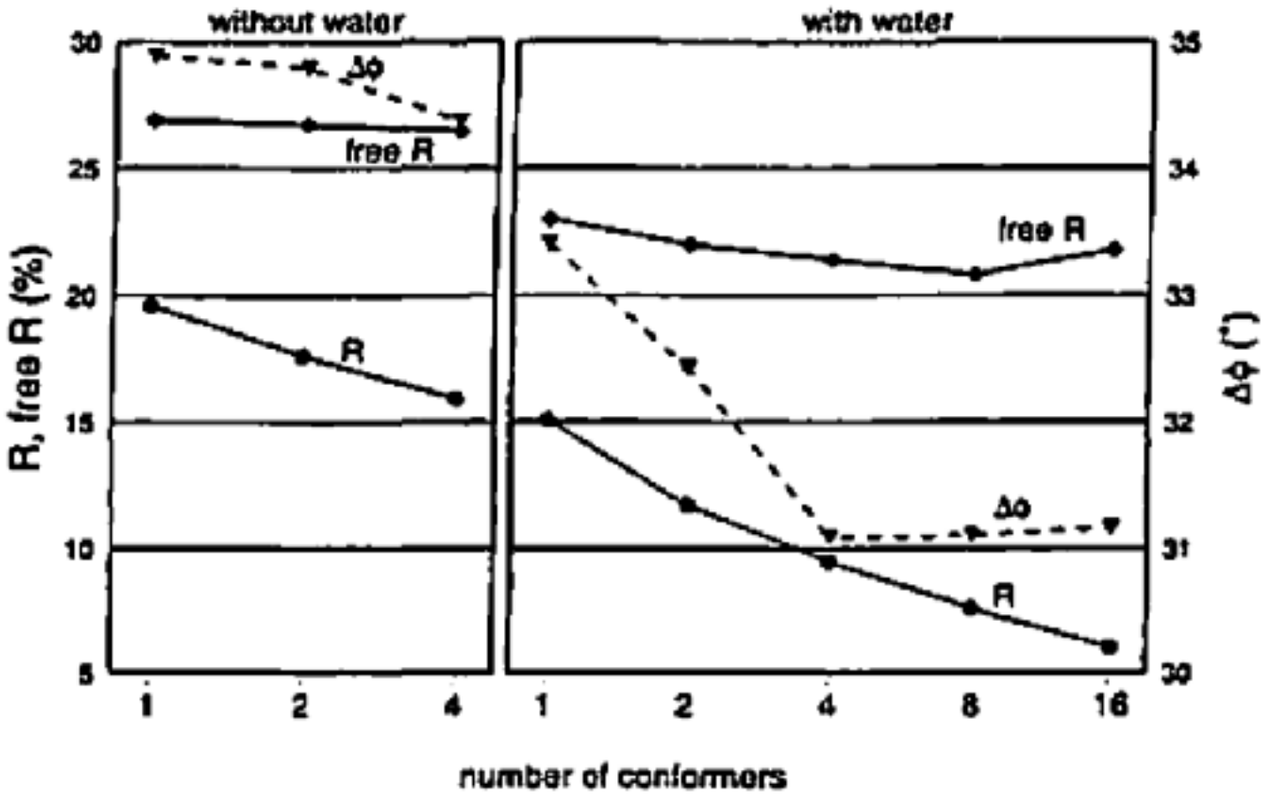
$$\langle \mathbf{F_c(s)} \rangle_{t'} = \frac{1}{\tau_x(1 - e^{-t'/\tau_x})} \int_0^{t'} e^{-(t' - t)/\tau_x} \mathbf{F_c^t(s)} dt$$

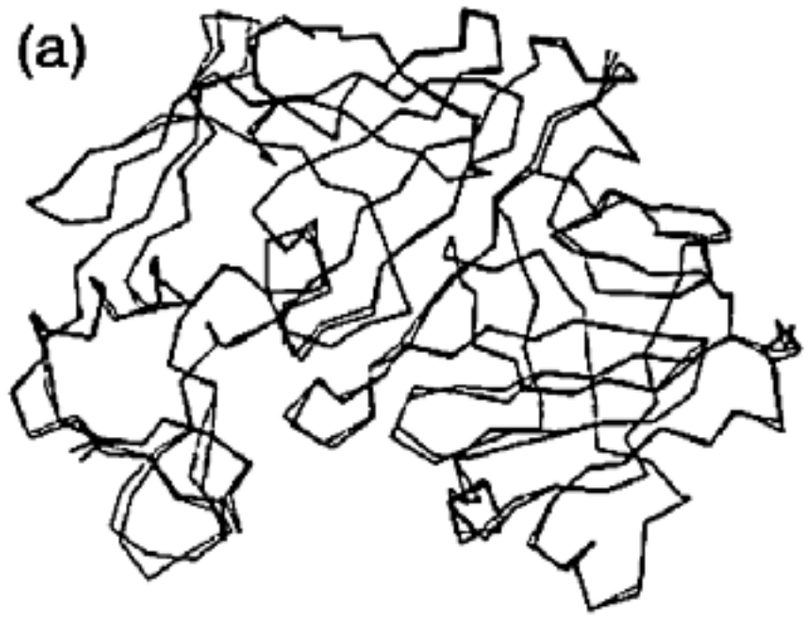


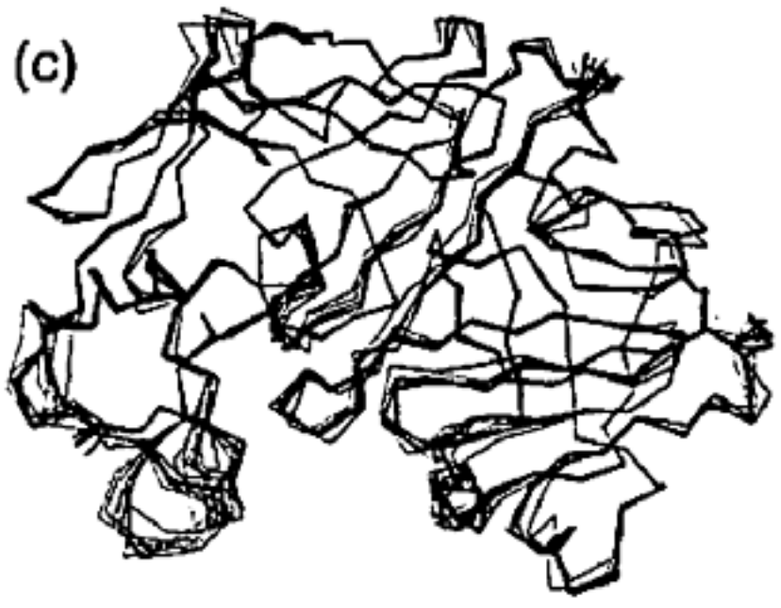
Gros P, van Gunsteren WF, Hol WG: Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. *Science* 1990, 249:1149-52

Multi-copy Refinement

- Simultaneous refinement of multiple copies of the model using simulated annealing
- Captures structural variability without the need for long simulations



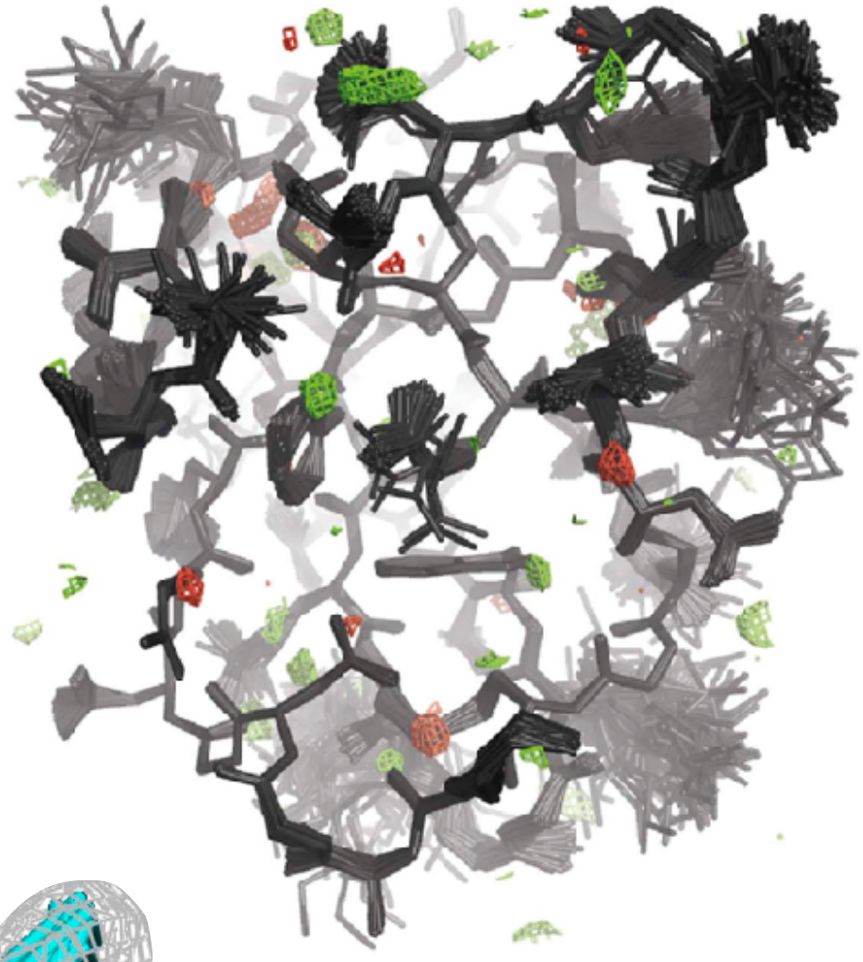


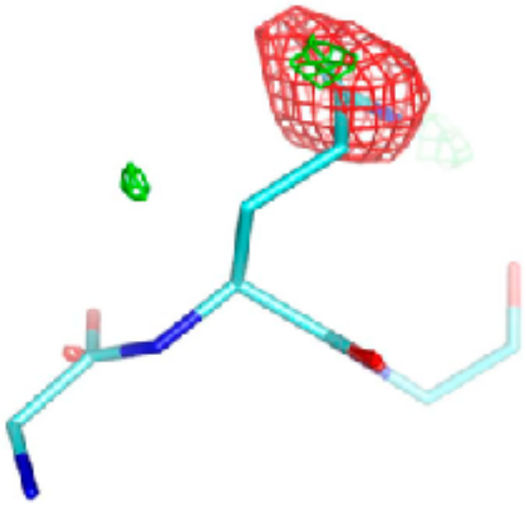


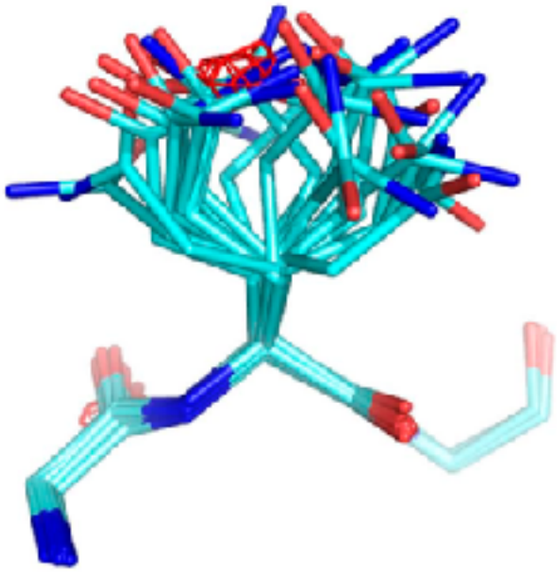
Burling FT, Brunger AT: Thermal motion and conformational disorder in protein crystal structures: Comparison of multi-conformer and time-averaging models. *Israel Journal of Chemistry* 1994, 34:165-175

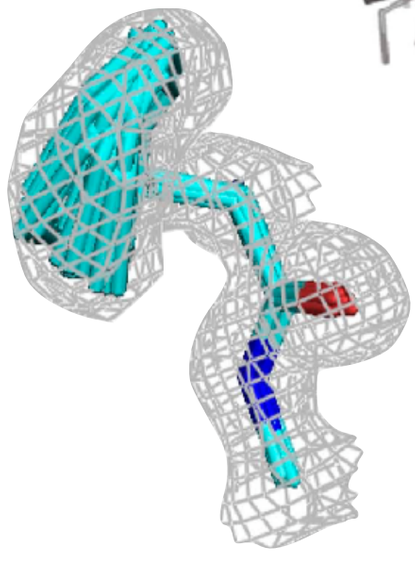
Time-Averaging 2.0

- Original time-averaging suffered from overfitting
- Application of newer refinement algorithms restricts the number of structures modelled to prevent over-fitting of the data
  - TLS refinement, maximum likelihood



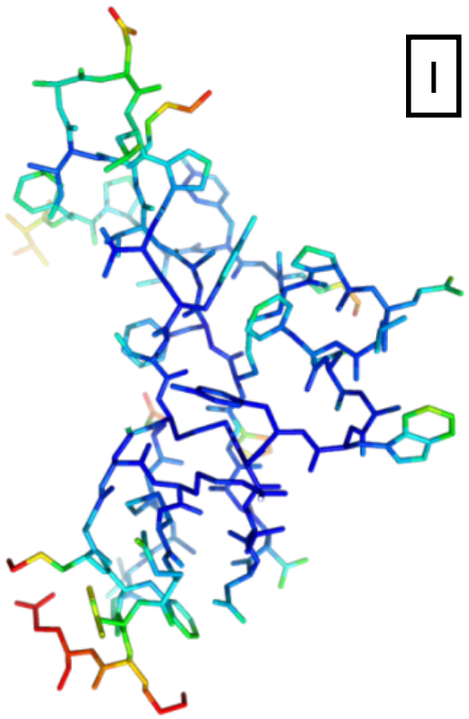


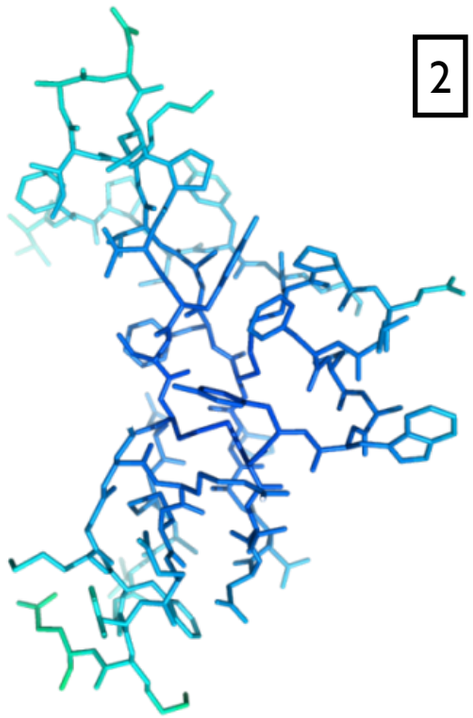




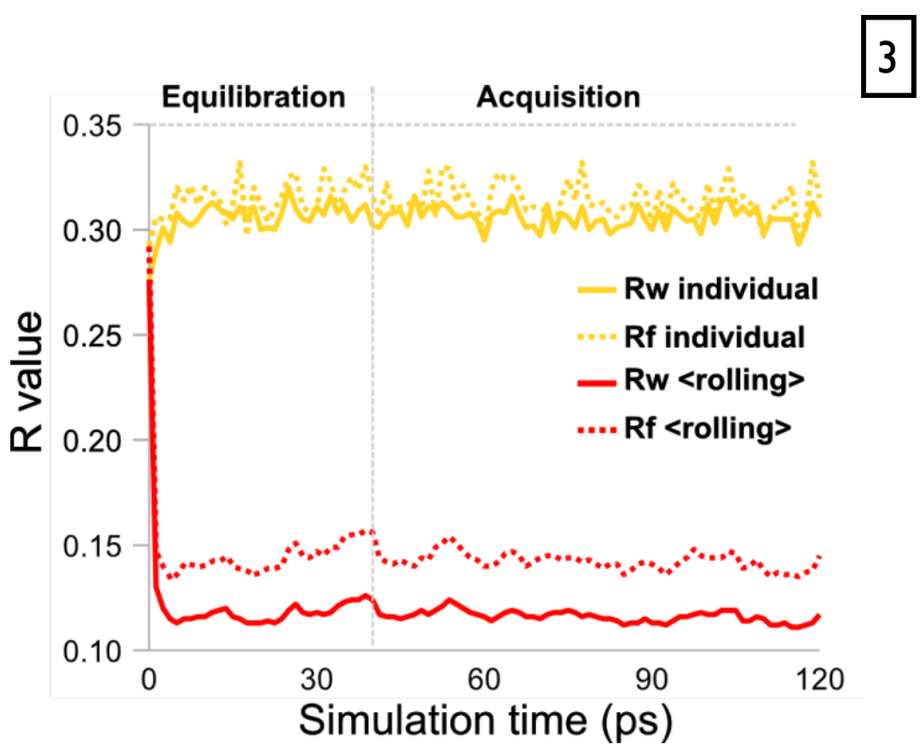
Burnley BT, Afonine PV, Adams PD, Gros P: Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* 2012, 1:e00311

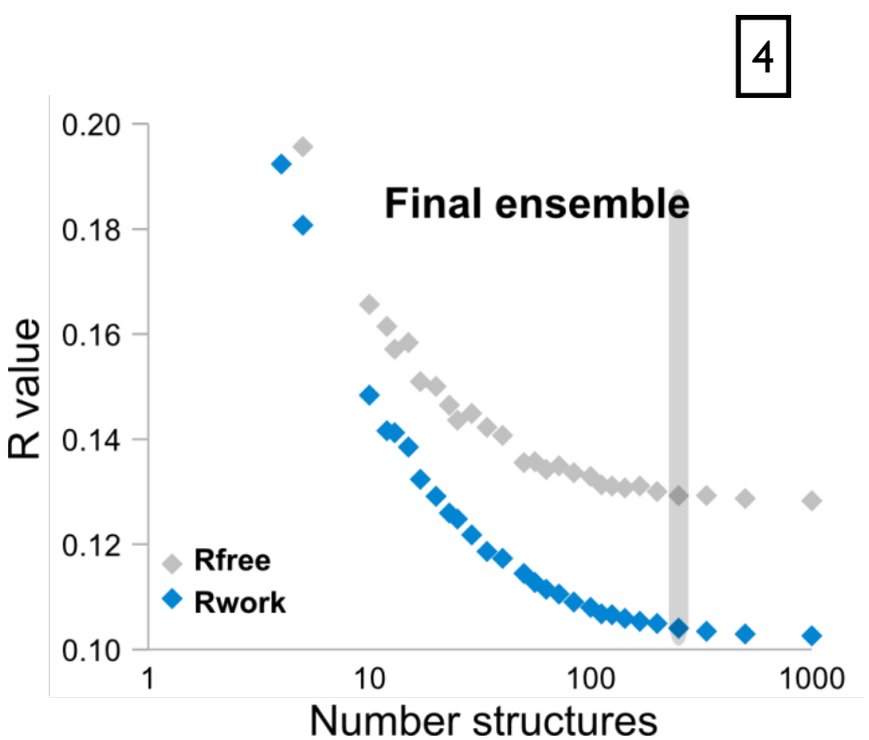
Procedure



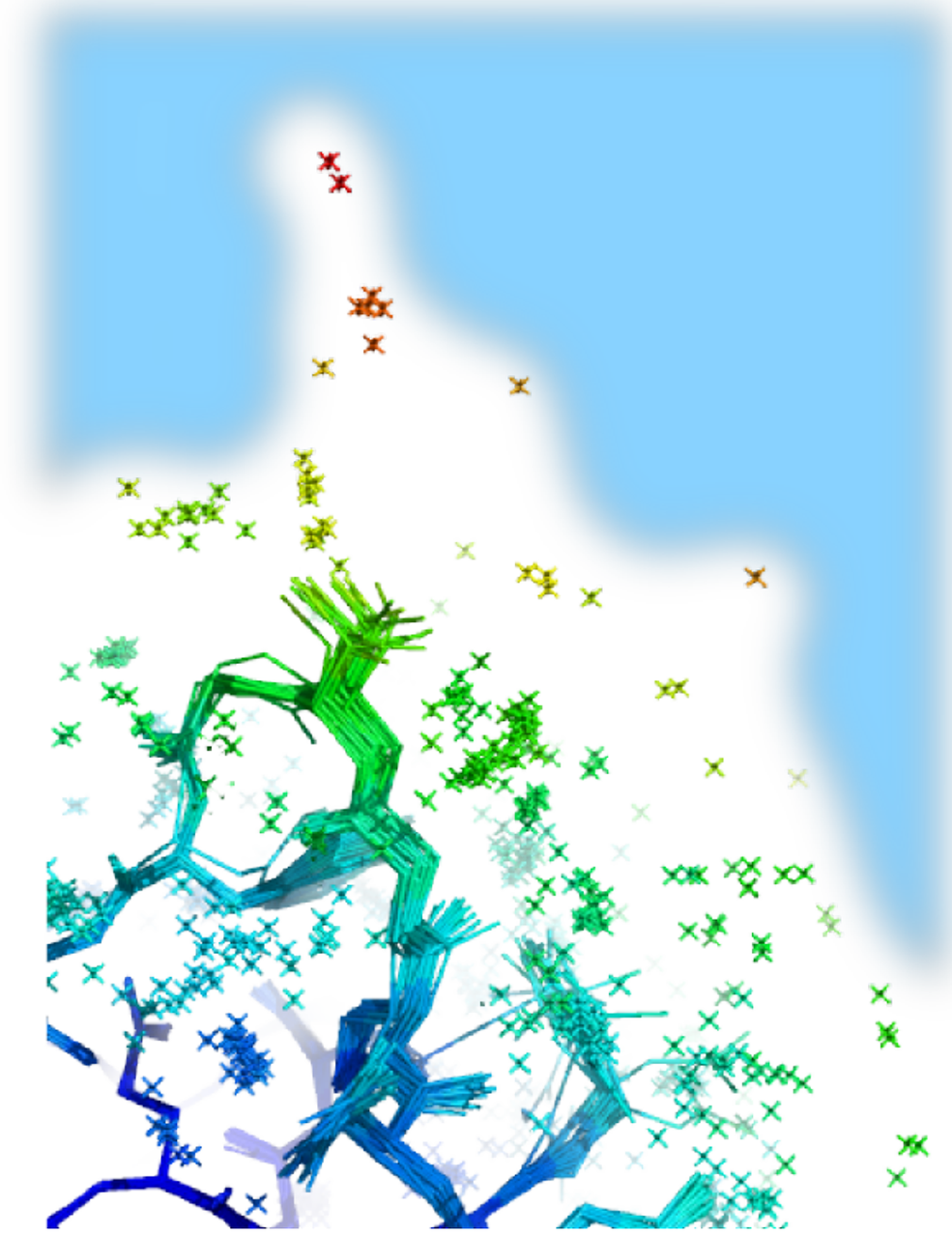


1. Initial refined model
2. Fit TLS model, remove alternate conformations
3. MD simulation with time averaged crystallographic restraints
4. Selection of models for the final ensemble





Dual explicit-bulk solvent model



- Explicit solvent
  - Model with explicit atoms
  - Water picked every 250 steps
    - "standard" rules:
      - > 3 σ in difference map
      - < 3 Å distances
  - B-factor from nearest TLS group

- Bulk solvent
  - Model with 'density mask'

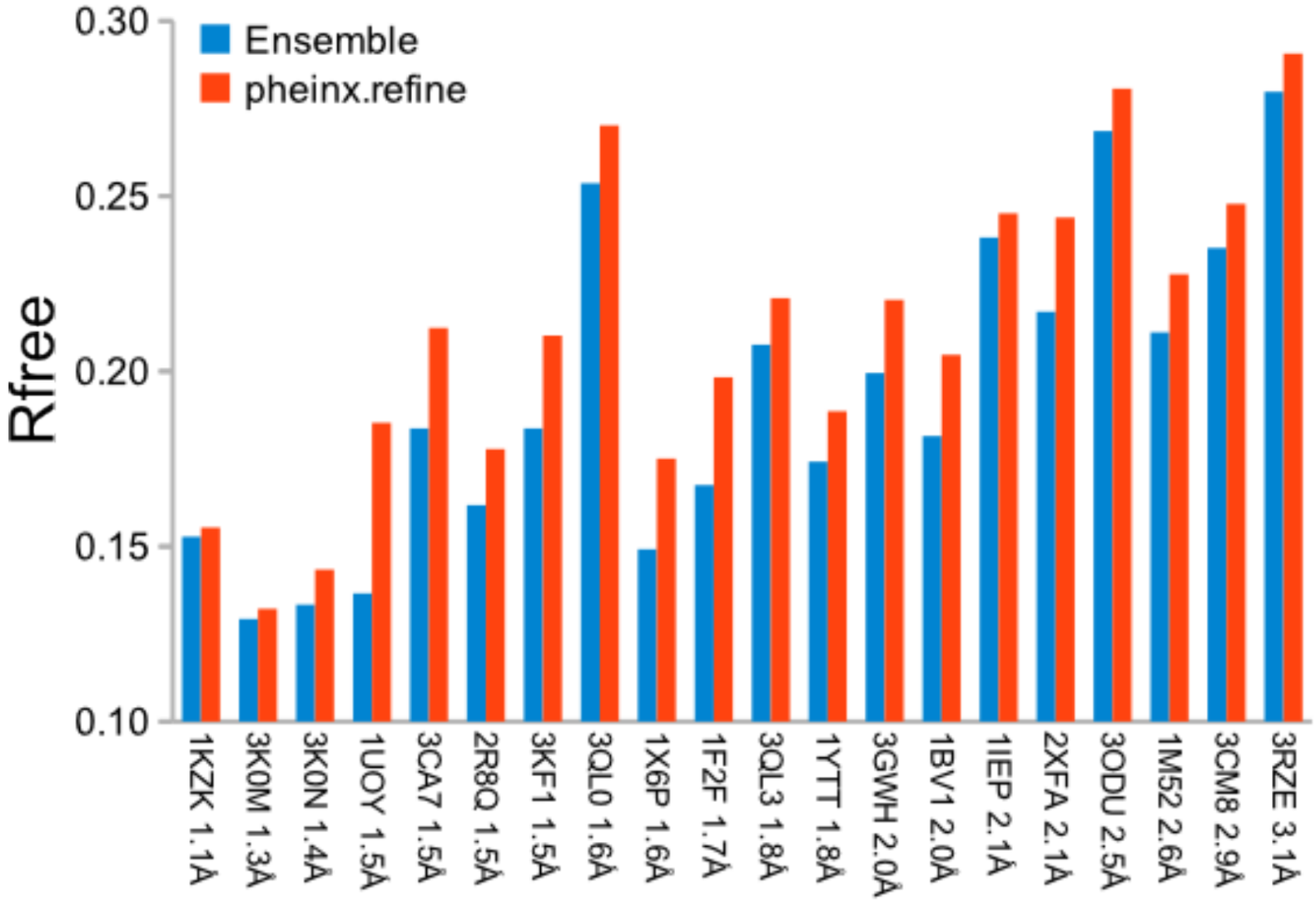
$$\left\langle F_{mask} \right\rangle_t = (1 - e^{-\Delta t/\tau_x}) F^t_{mask} + e^{-\Delta t/\tau_x} \left\langle F_{mask} \right\rangle_{t-\Delta t}$$

BERKELEY LAB
Lawrence Berkeley National Laboratory

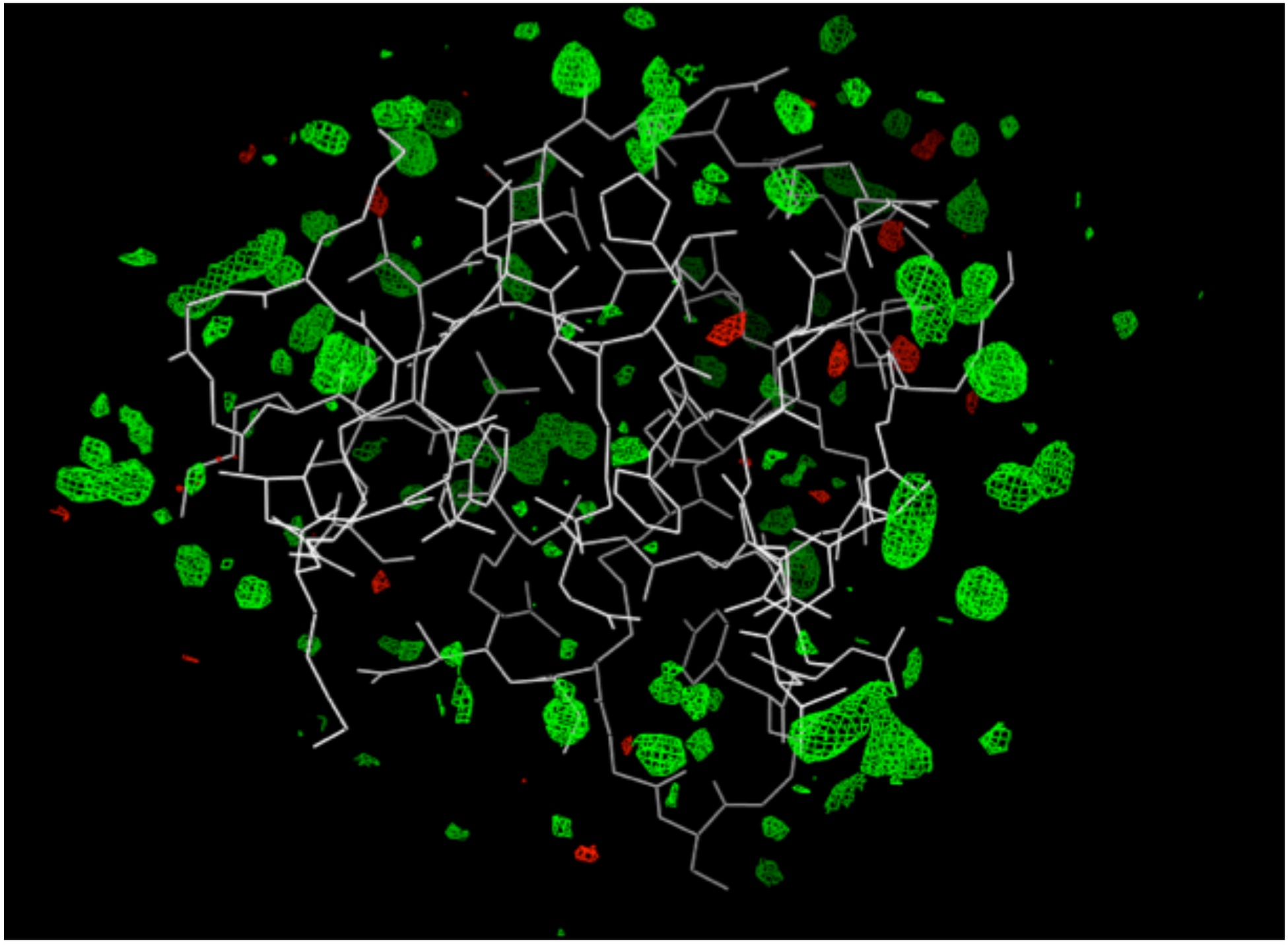
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Model Fit to Data is Improved

- R_{free} reduced in all cases
  - -4.9% (max)
  - -0.3% (min)
  - -1.8% (mean)
- Rf/Rw ratio (mean):
  - 1.23 phenix.refine
  - 1.25 ensemble

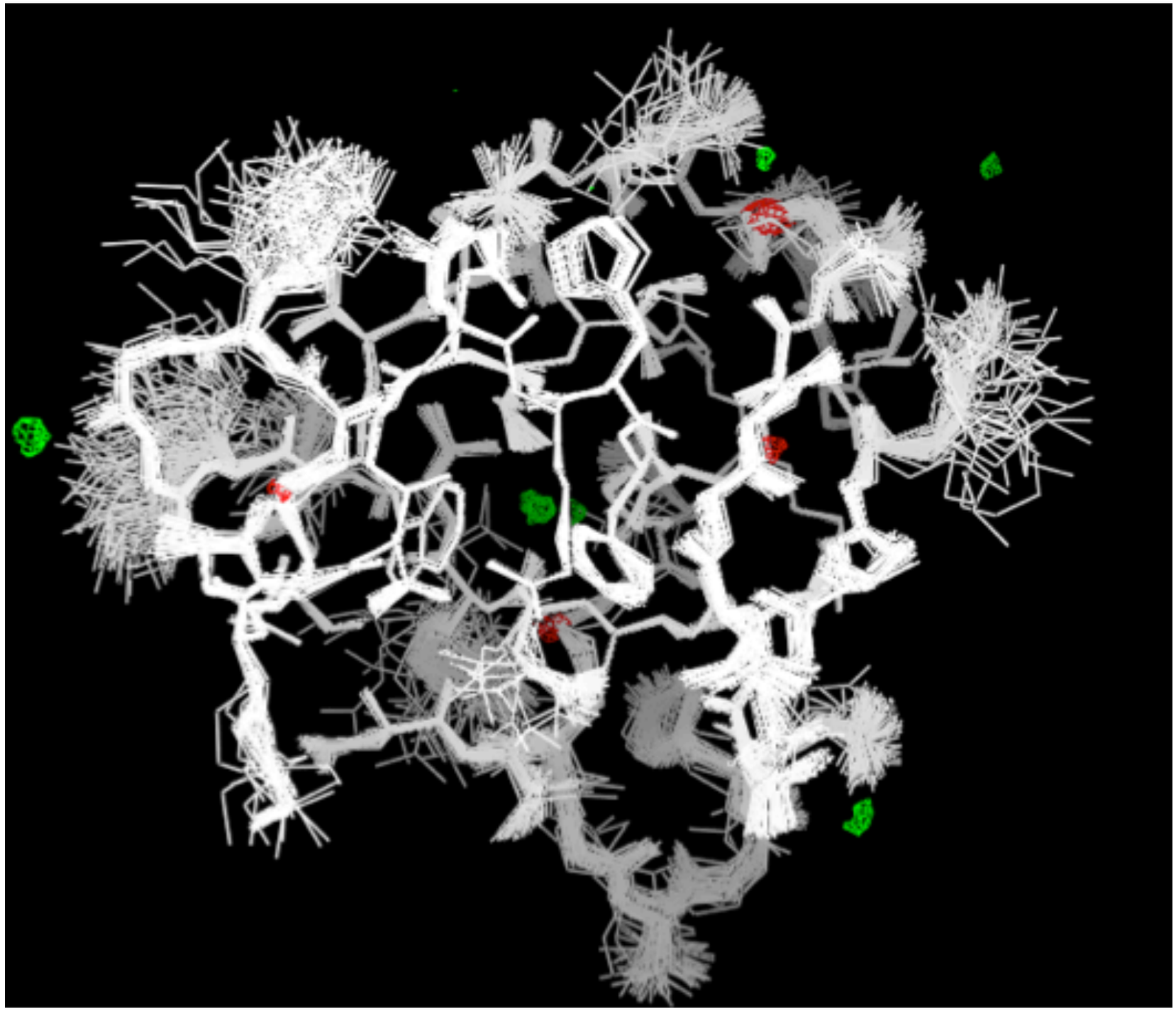


Results



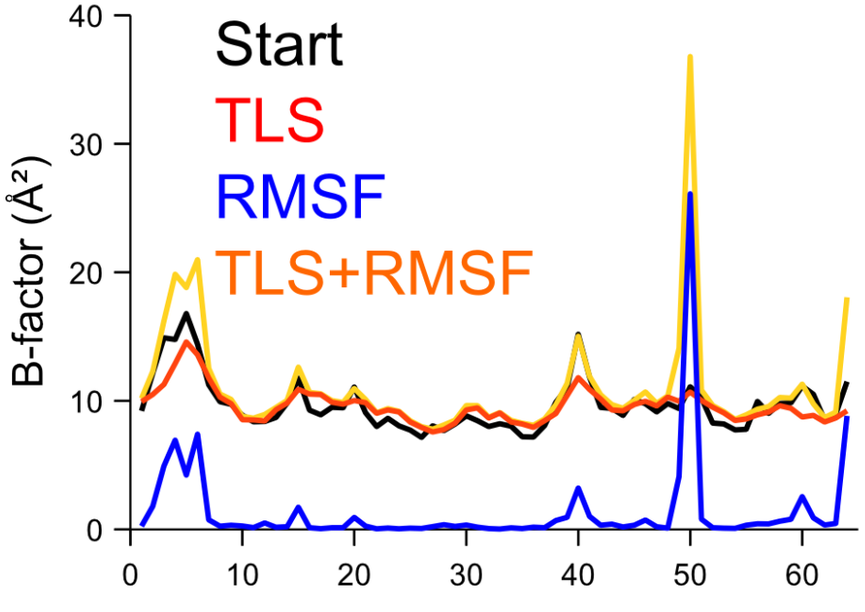
1uoy.pdb | phenix.refine | 1 tls group | mFo-DFc ±0.49 e/Å^3 (3.00 σ)

Results

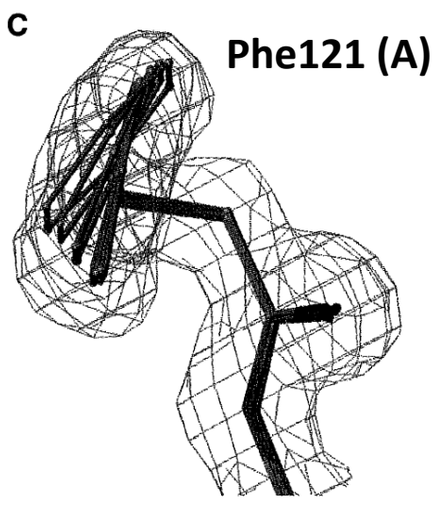


1uoy.pdb | 188 ensemble | 1 tls group | mFo-DFc ±0.49 e/Å³ (4.27 σ)

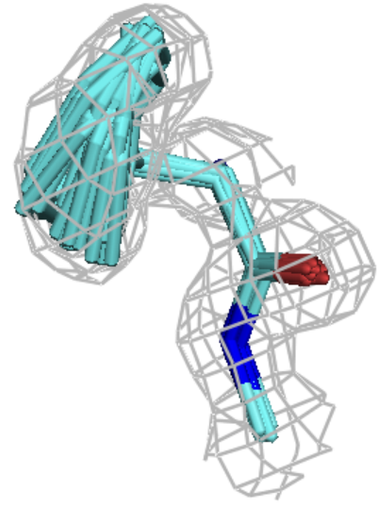
Results



Burling *et al.* (1996)

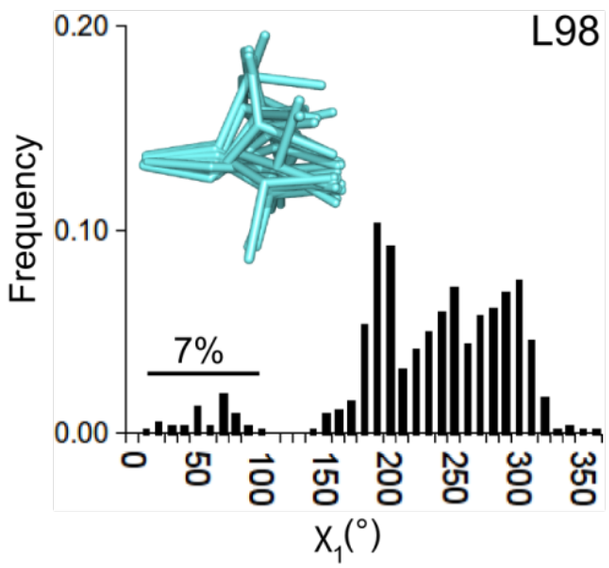


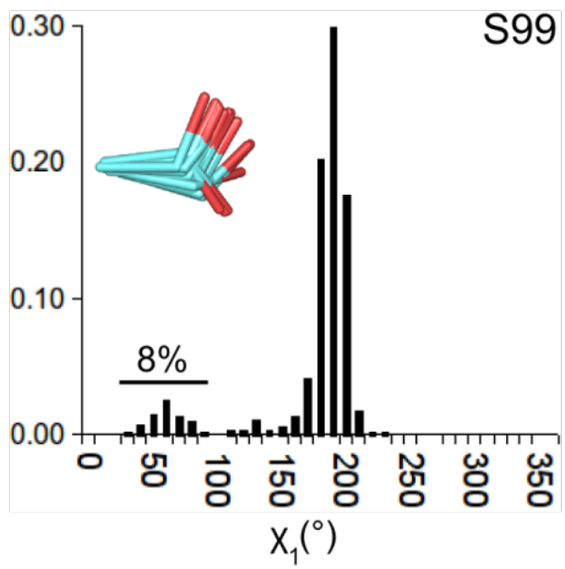
Multi-conformer (Rfree 20.3%)

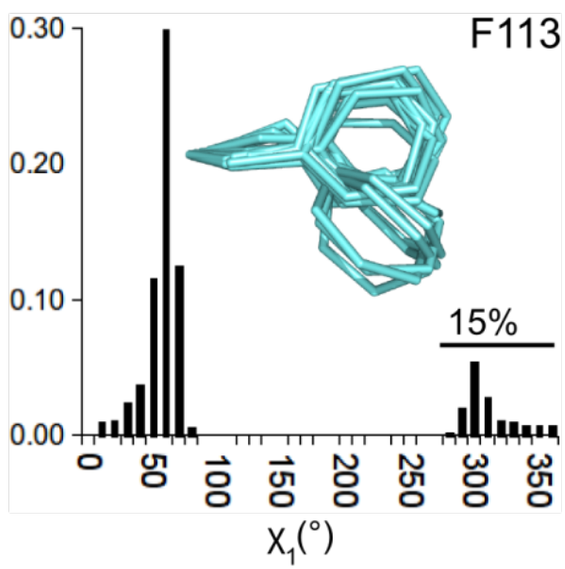


Ensemble (Rfree 17.4%)

Experimental map (1.4σ)





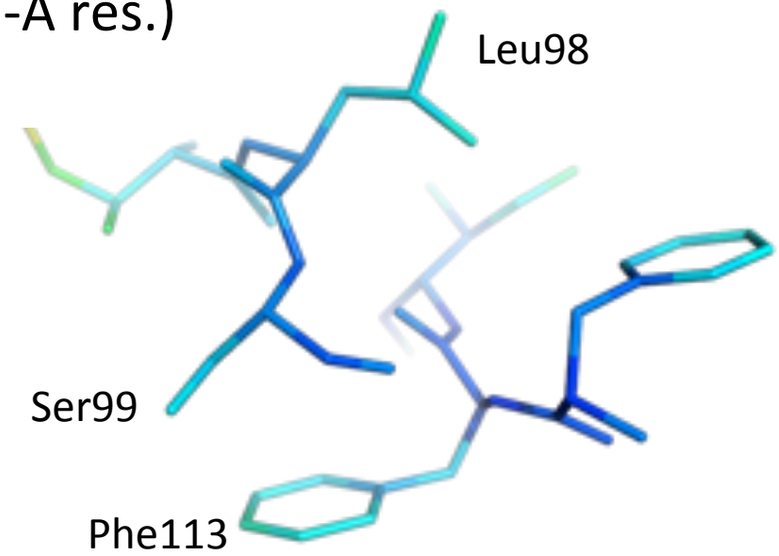


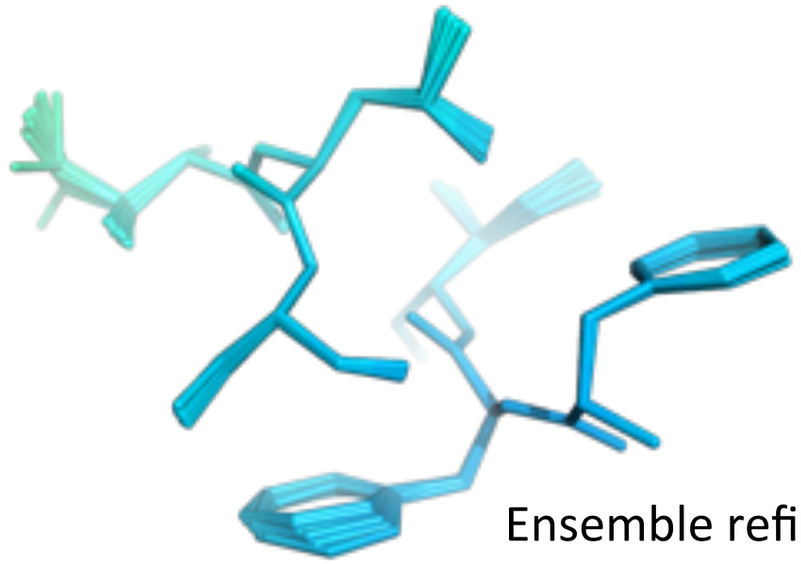
Ensembles consistent with NMR relaxation dispersion data

3K0N & 3K0M: Proline isomerase, Fraser et al. (2009), Eisenmesser et al. (2005)

Ensembles Consistent with Temperature

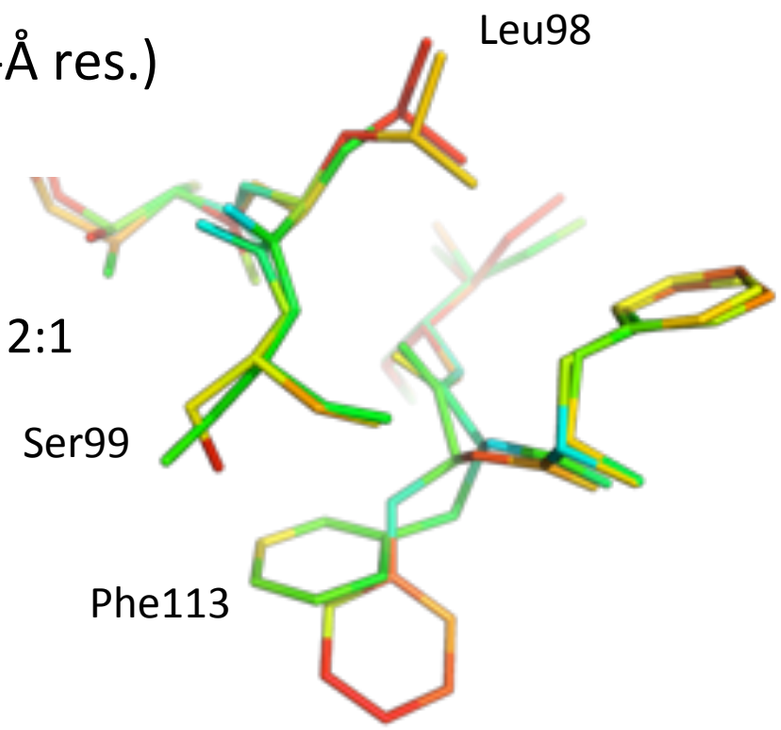
3K0M (1.3-Å res.)
100 K





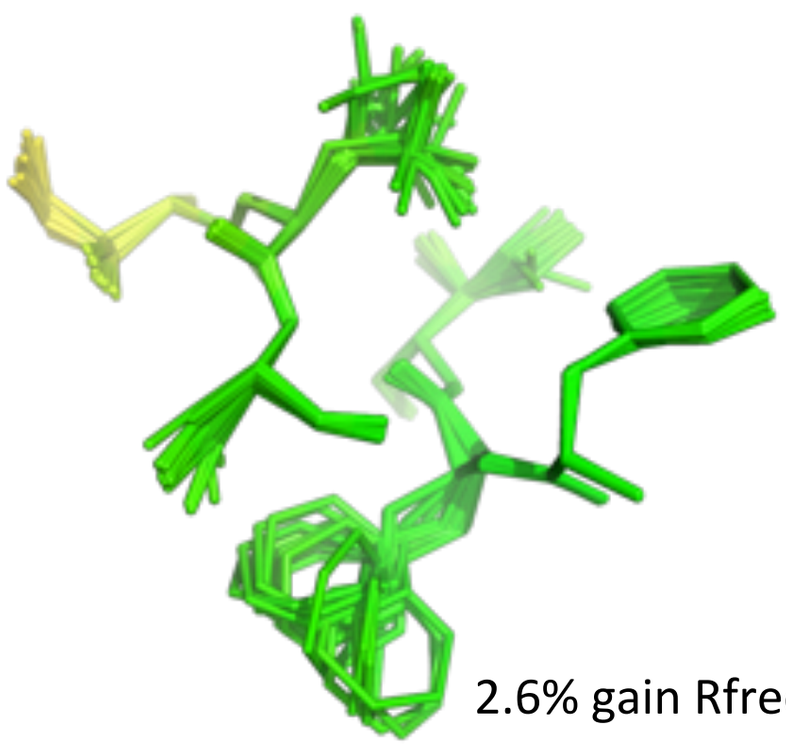
Ensemble refinement:
1.3% gain Rfree

3K0N (1.4-Å res.)
288 K



A:B = 2:1

B-factor
5-25Å²

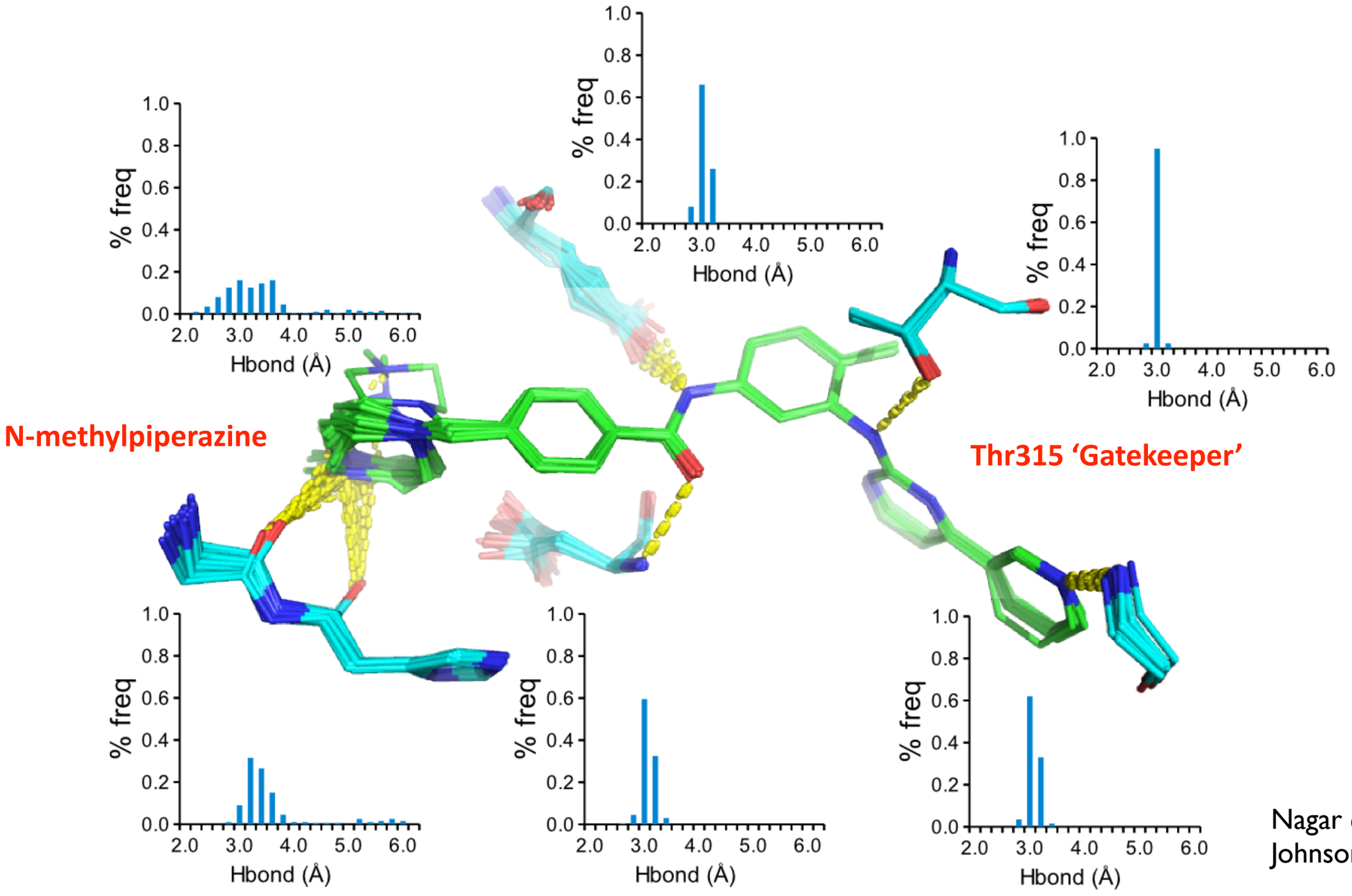


2.6% gain Rfree

BERKELEY LAB
Lawrence Berkeley National Laboratory

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Biological Insight



Nagar *et al* (2002)
Johnson (2009)

Imatinib-ABL Tyrosine Kinase (1IEP)

Conclusions

- Crystallographic data is derived from a time and space average - ensemble models are logical
- Challenging to identify variability arising from the true distribution in the crystal versus uncertainty arising from resolution or computational method
- Molecular dynamics force fields have improved, and should improve ensemble refinement
- Ensembles should routinely be used to represent uncertainty
    - Is the world ready, especially crystallographers?

BERKELEY LAB
Lawrence Berkeley National Laboratory

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Acknowledgments

- **Lawrence Berkeley Laboratory**
  - Pavel Afonine, Yourval Dar, Nat Echols, Jeff Headd, Lida Gifford, Richard Gildea, Ralf Grosse-Kunstleve, Johan Hattne, Nigel Moriarty, Ian Rees, Nicholas Sauter, Oleg Sobolev, Peter Zwart
- **Los Alamos National Laboratory**
  - Tom Terwilliger, Li-Wei Hung
- **Cambridge University**
  - Randy Read, Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Robert Oeffner
- **Duke University**
  - Jane Richardson & David Richardson, Ian Davis, Vincent Chen, Jeff Headd, Chris Williams, Bryan Arendall, Bradley Hintze, Laura Murray
- **Time Averaging**
  - Tom Burnley & Piet Gros (Utrecht)
- **Other Collaborators**
  - Frank DiMaio, David Baker (U Washington)
  - Alexandre Urzhumtsev & Vladimir Lunin
  - Andrew Van Benschoten & Jamie Fraser (UCSF)
  - Pawel Janowski & David Case (Rutgers)
  - Dale Tronrud, Donnie Berholz, Andy Karplus
  - Garib Murshudov & Alexi Vagin
  - Kevin Cowtan, Paul Emsley, Bernhard Lohkamp
  - David Abrahams
  - PHENIX Testers & Users
- **Funding:**
  - NIH/NIGMS:
    - *P01GM063210, P50GM062412, P01GM064692, R01GM071939*
  - PHENIX Industrial Consortium
  - Lawrence Berkeley Laboratory