Flexibility Prediction: Preserving Detailed Balance in Side Chain and Backbone Movers

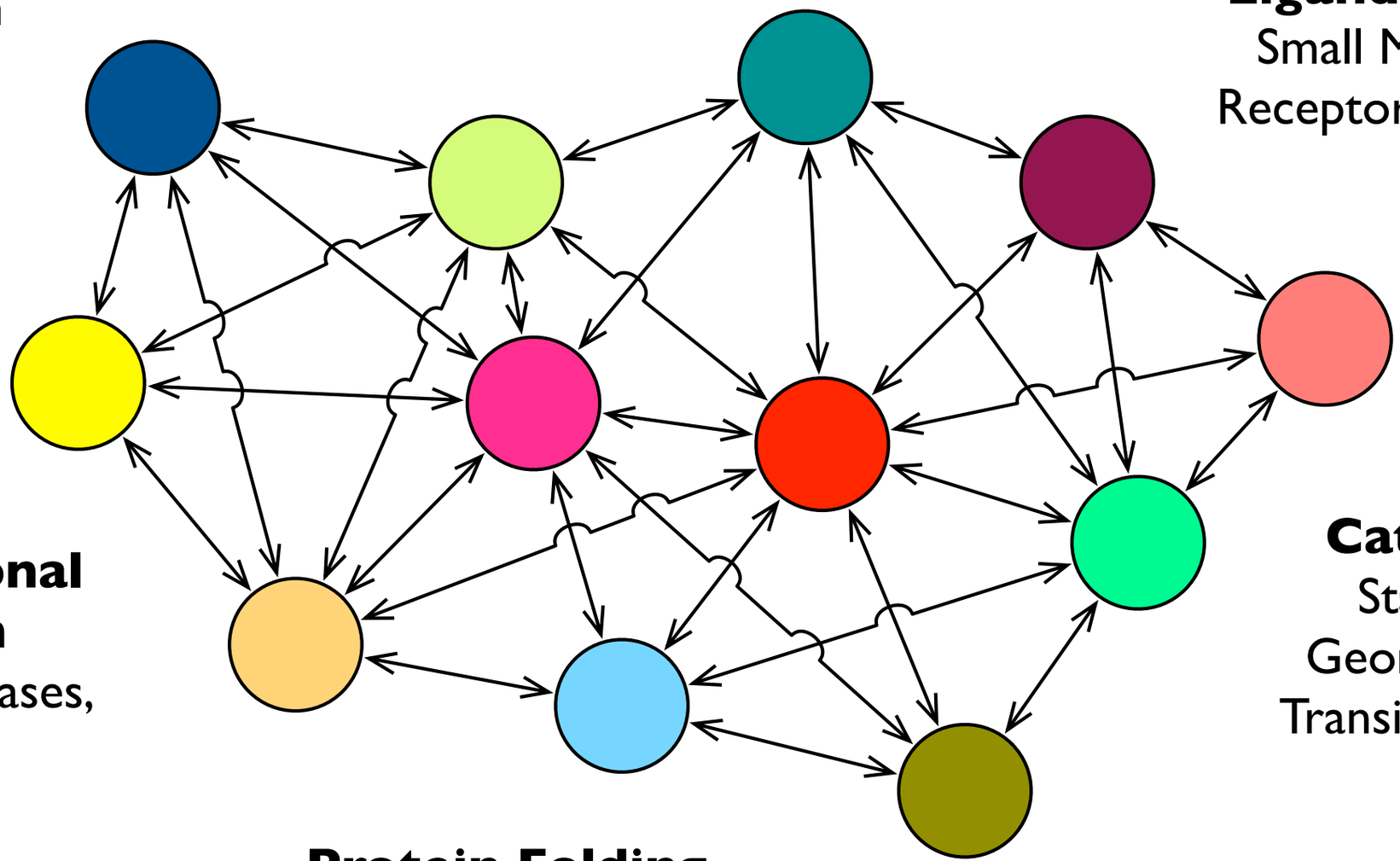
Colin A. Smith - Kortemme Lab

RosettaCON - Wednesday, August 4, 2010

Predicting Flexibility With Monte Carlo Simulations

Protein Protein Interaction
Free Energy
Cost of Binding

Ligand Binding
Small Molecule/
Receptor Flexibility



Post Translational Modification
Accessibility to Kinases,
Proteases, etc.

Catalysis
Stabilize
Geometry of
Transition State

Protein Folding
Free Energy vs.
Potential Energy Basin

Rigorous Monte Carlo Must Preserve Detailed Balance

Analogous to to the Law of Mass Action:

$$A \Leftrightarrow B$$

Population

Transition Probability

$$K_{eq} = \frac{[B]}{[A]} = \frac{k_{forward}}{k_{reverse}}$$

Solution #2:

$$a_{AB} = \frac{1}{q_{AB}} e^{(E_A - E_B)/kT}$$

Mover Proposal Density

$$p_{AB} = q_{AB} a_{AB}$$

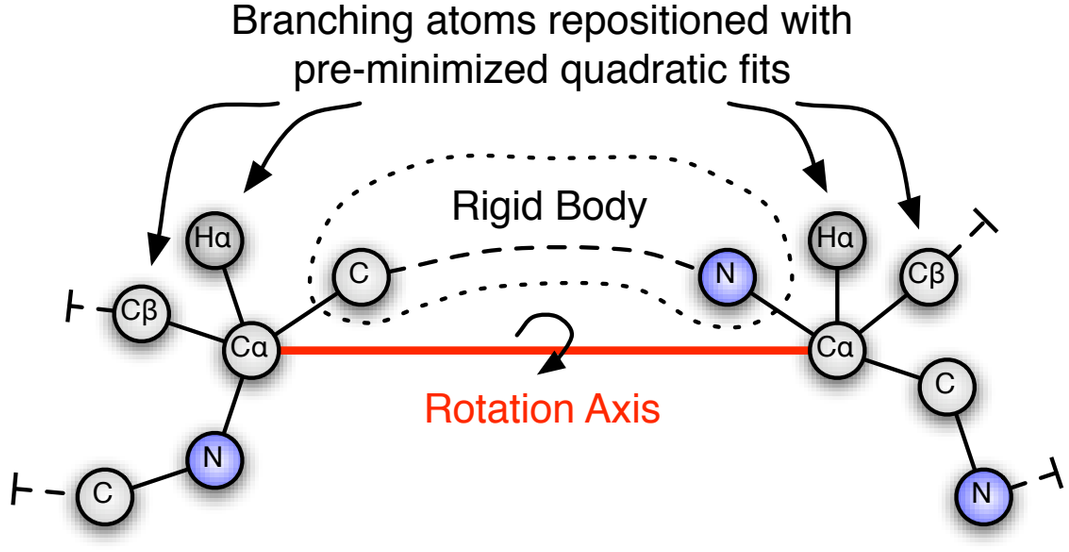
Metropolis Criterion

$$\frac{\pi_B}{\pi_A} = \frac{p_{AB}}{p_{BA}} = \frac{q_{AB} a_{AB}}{q_{BA} a_{BA}}$$

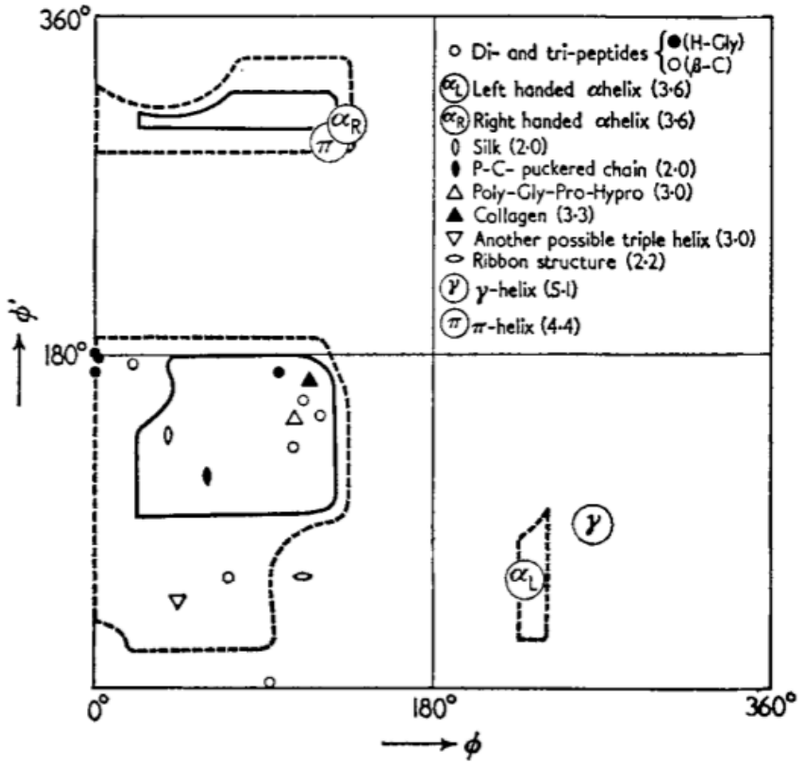
Problem: $q_{AB} \neq q_{BA}$ Solution #1: $q_{AB} = q_{BA}$

Movers Using Solution #1

- BackrubMover
  - Problem: Branching atoms sampled nonuniformly
  - Solution:
    Turn off branching atom sampling



- SmallMover and ShearMover
  - Problem: Phi/Psi angles biased by Ramachandran plot
  - Solution:
    Turn off Ramachandran biasing



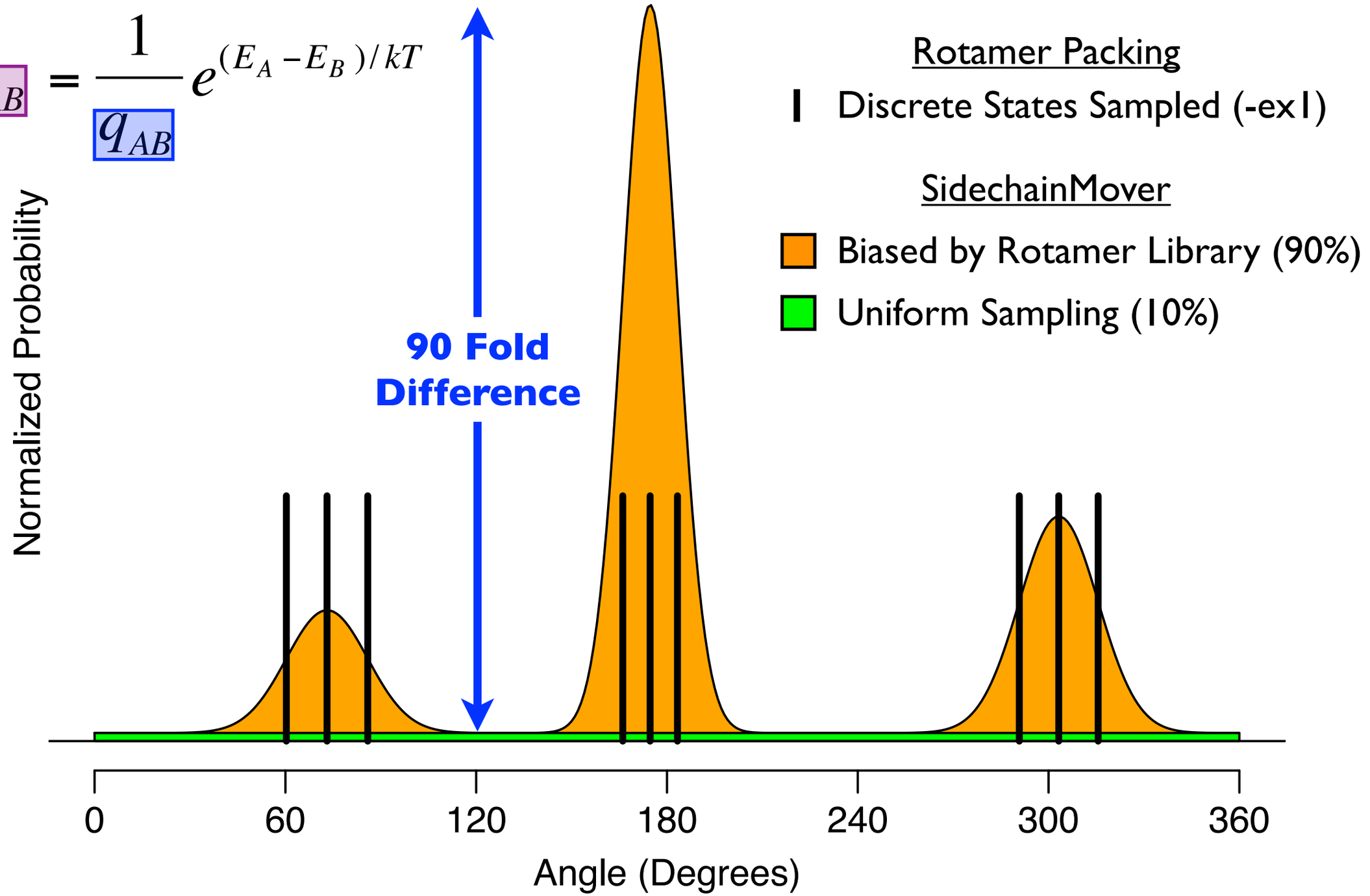
Rotamer Packing vs. SidechainMover

Rotamer Packing	SidechainMover
high probability rotamer wells sampled uniformly	all rotamer wells sampled with PDB probability
discrete chi angles sampled within rotamer wells	chi angles sampled continuously from a Gaussian distribution
always selects a random rotamer configuration	can also sample within current rotamer well or all chi angles uniformly
you have to worry about library resolution (-ex1, -ex2, etc.)	all possible chi angle combinations are sampled

Rotamer Packing vs. SidechainMover

Solution #2:

$$a_{AB} = \frac{1}{q_{AB}} e^{(E_A - E_B)/kT}$$



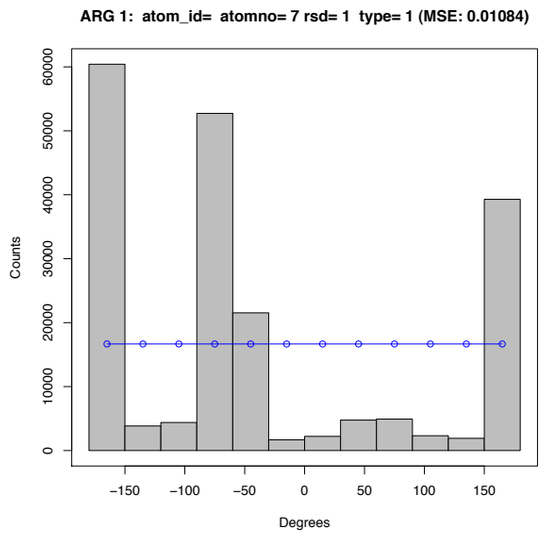
Testing Detailed Balance

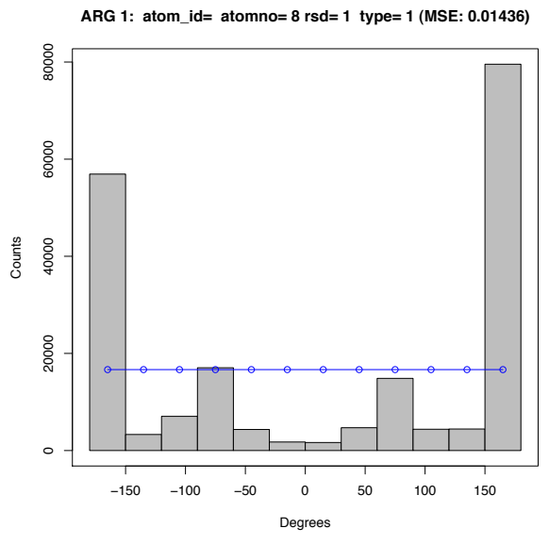
- Strategy: Run simulations and make sure movers generate expected populations.
- Easiest population to predict: Empty scoring function: $a_{AB} = \frac{1}{q_{AB}} e^{(E_A - E_B)/kT}$ (with $E_A - E_B = 0$)
- Use simplified systems for nightly testing:
  - SidechainMover: Single residue poses
  - Other movers: 8 residue polyalanine

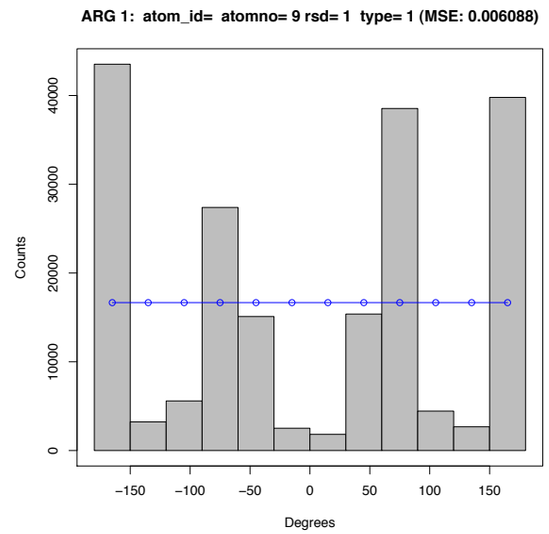
SidechainMover Example: Arginine Angle Histograms

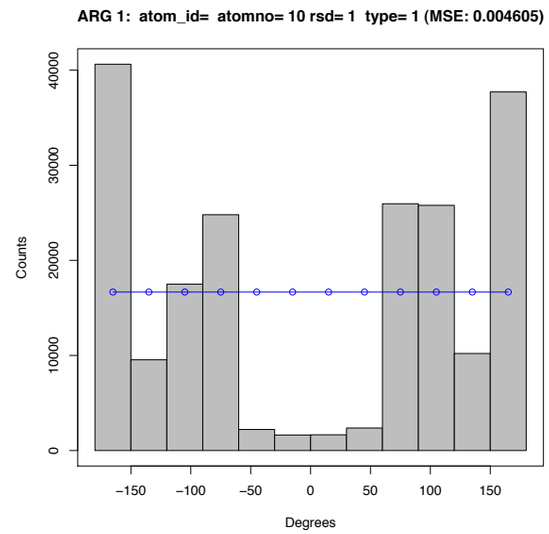
chi 1 | chi 2 | chi 3 | chi 4

Without Detailed Balance

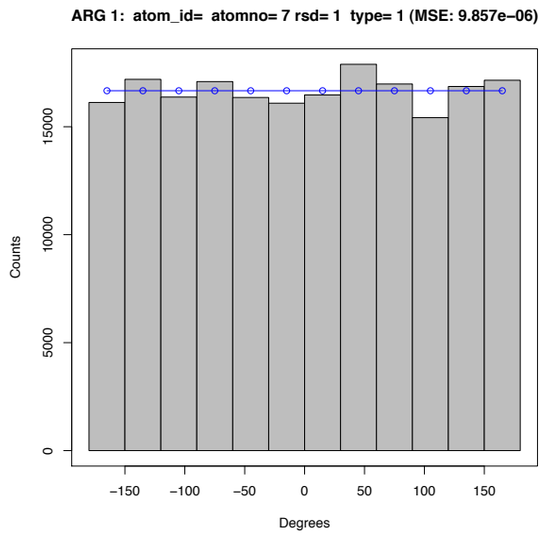


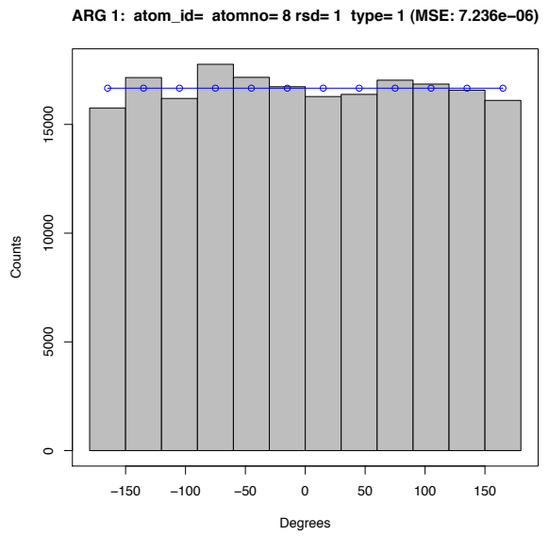


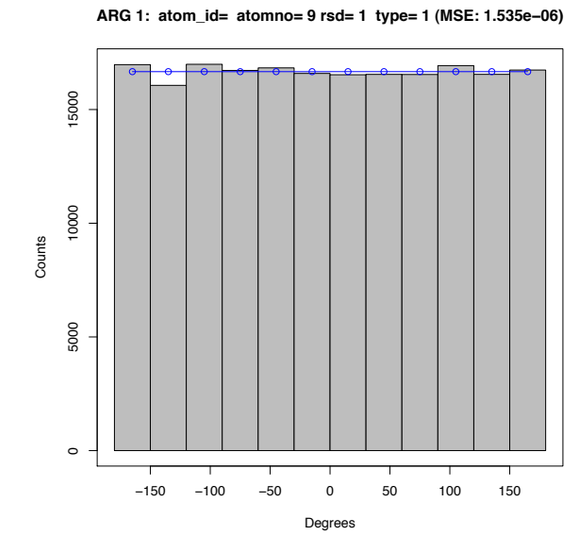


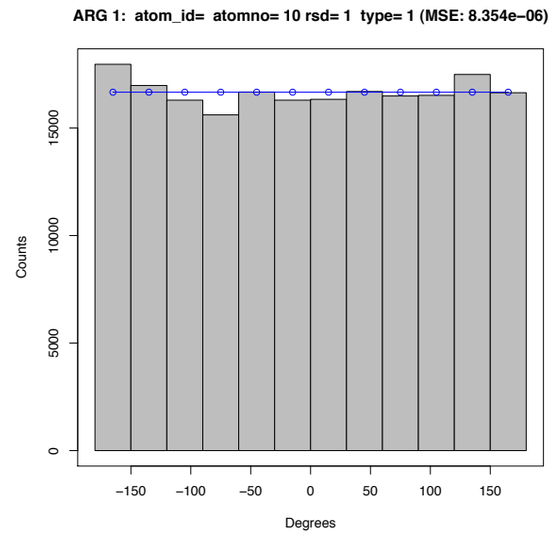


With Detailed Balance

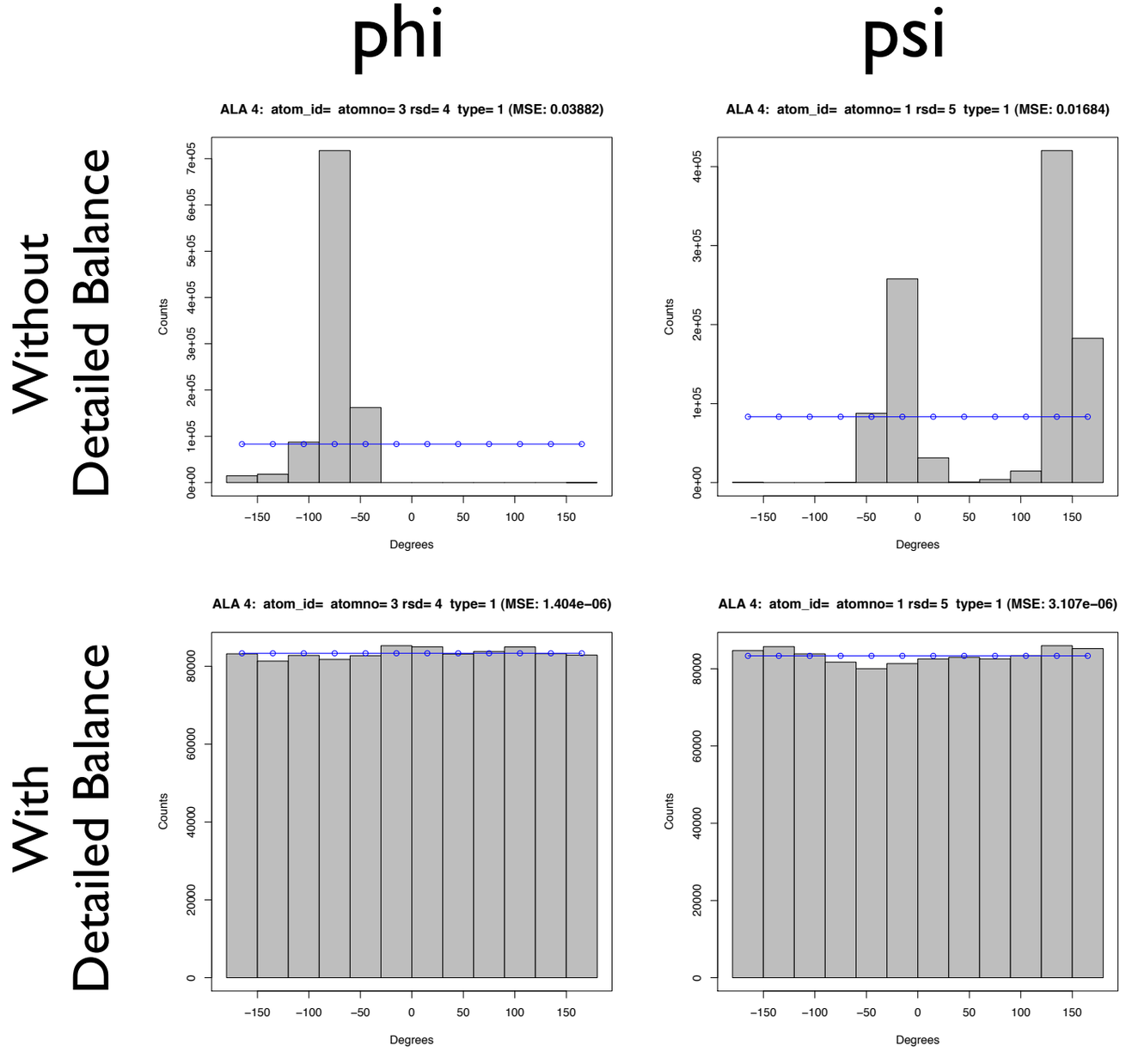








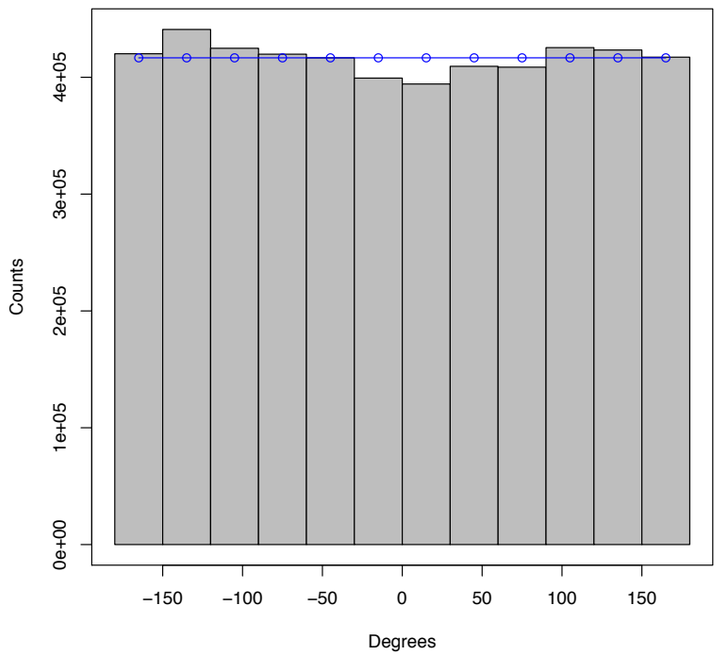
SmallMover Example: Polyalanine Residue 4



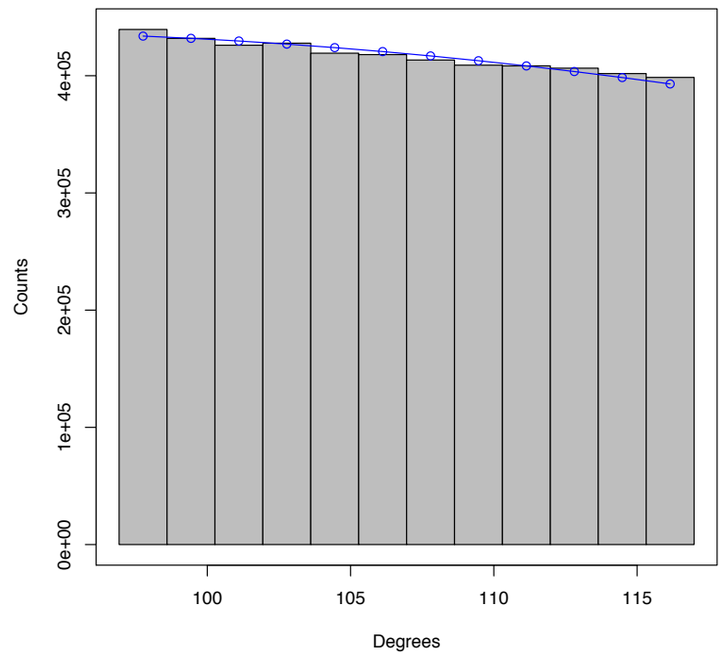
BackrubMover Example: Polyalanine Residue 4

With or Without Detailed Balance

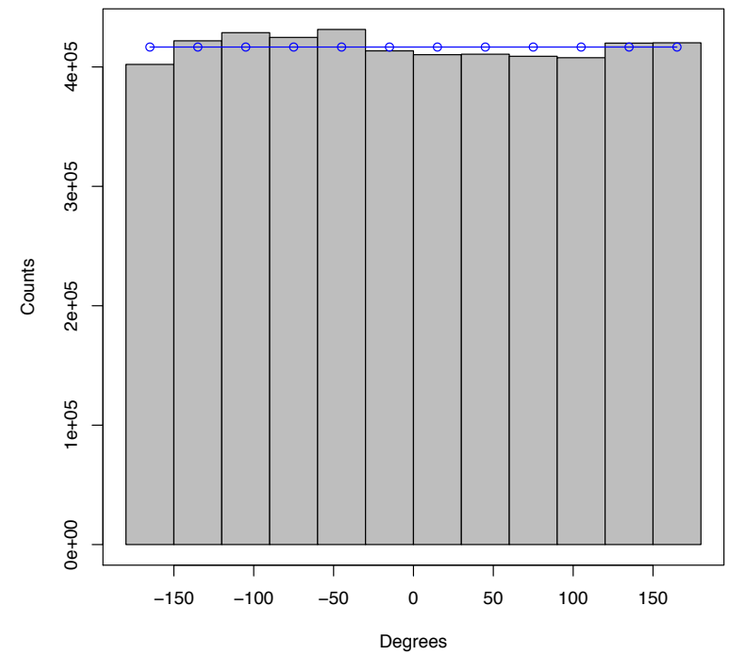
phi



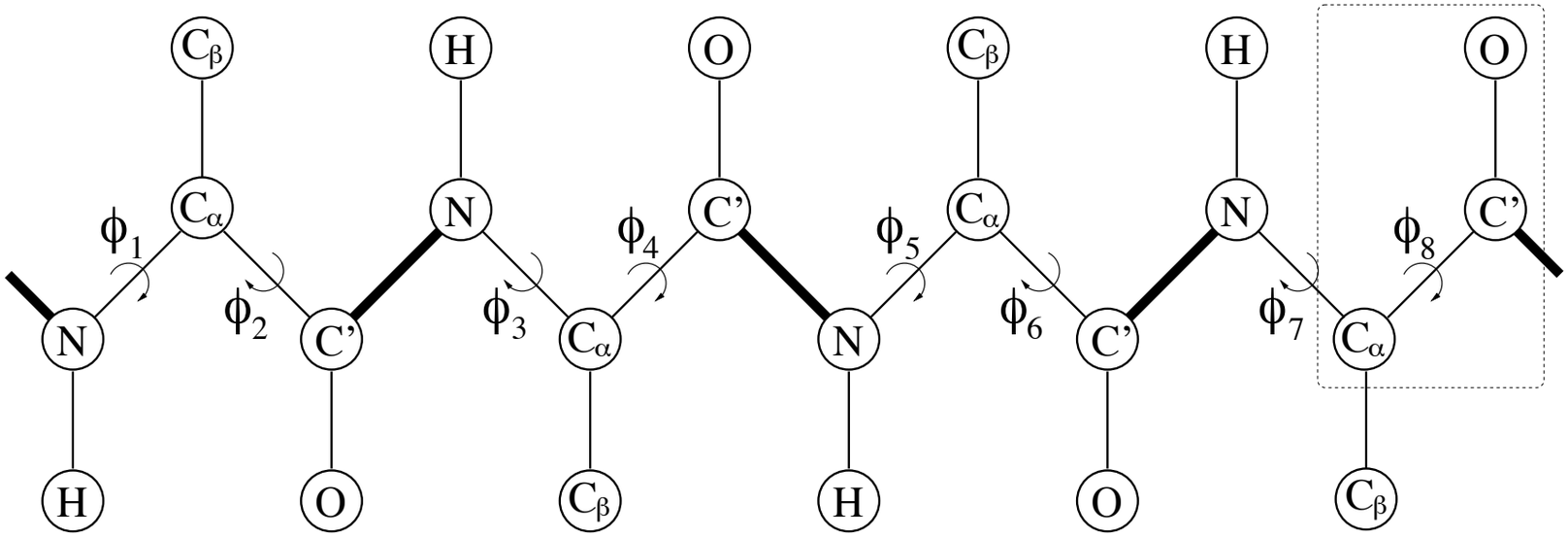
N-CA-C bond angle



psi

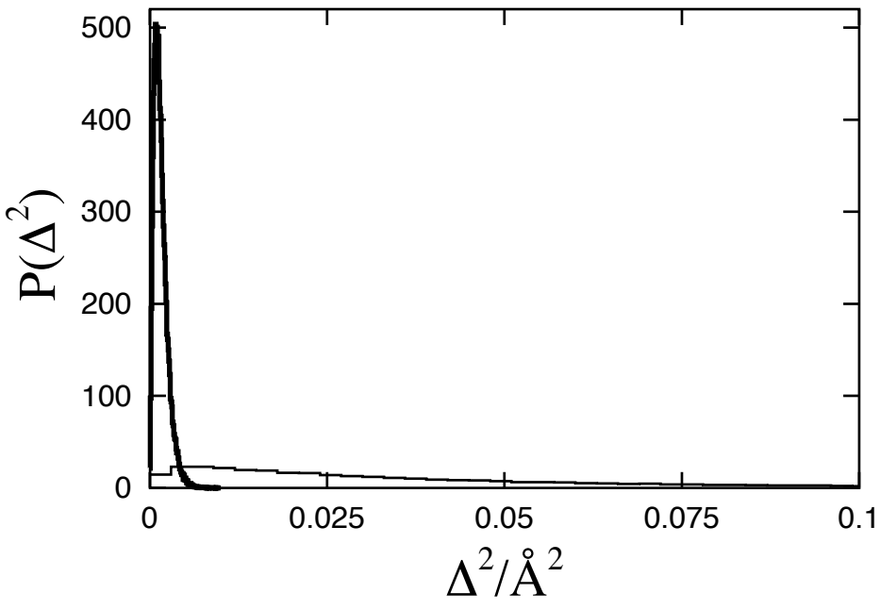


Yuan Liu: Backbone Biased Gaussian Mover

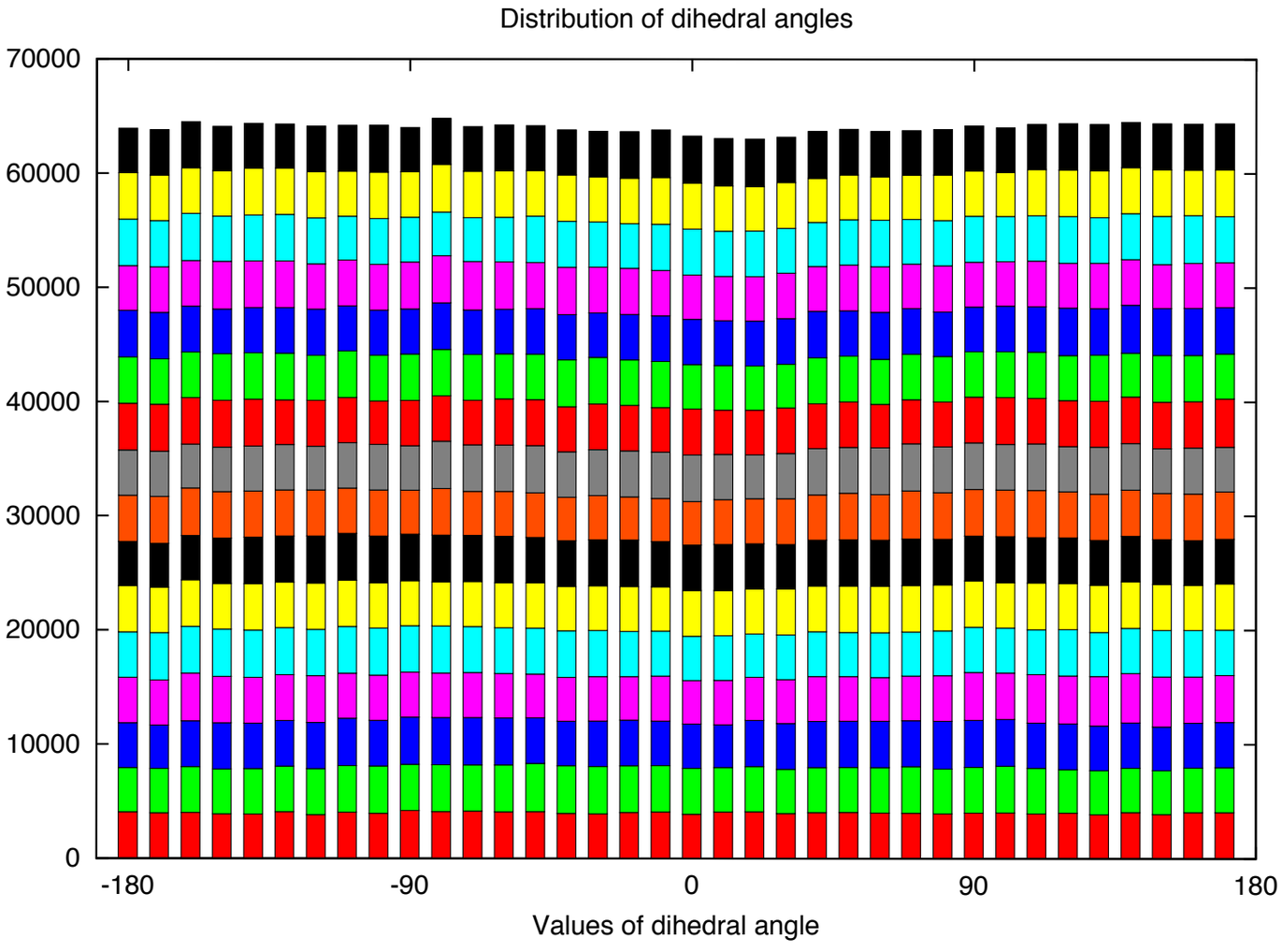


$$\Delta^2 = \sum_{I=1}^{3} (\delta \vec{r}_I)^2$$

Cα, C' and O are the first three atoms of downstream rigid body



Distribution of displacement of the downstream(thick line) is much more local than distribution of randomly perturbed dihedral angles(thin line)



Monte Carlo Update for Chain Molecules: Biased Gaussian Steps in Torsional Space
G. Favrin, A. Irbäck and F. Sjunnesson, J.Chem. Phys. 114, 8154 (2001)

What kT to Approximate Room Temperature Flexibility?

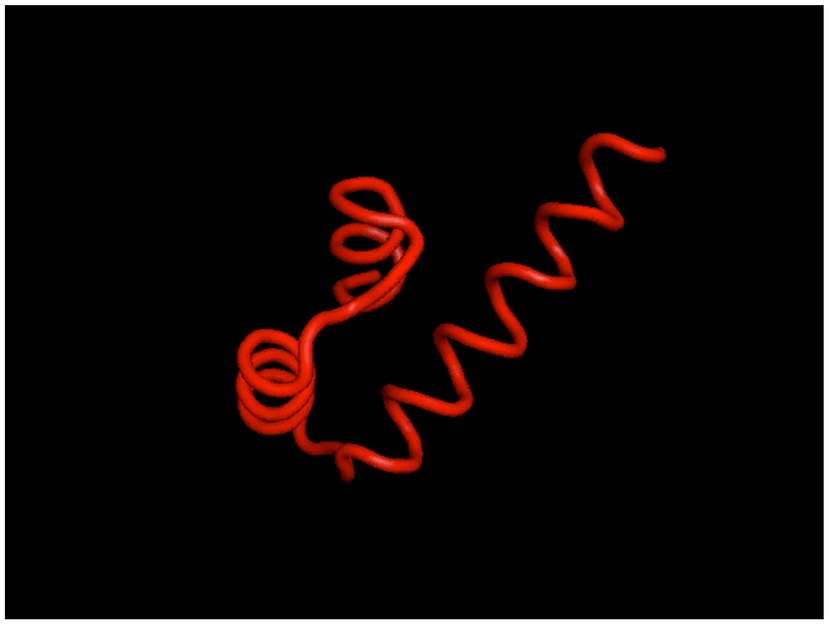
$$a_{AB} = \frac{1}{q_{AB}} e^{(E_A - E_B)/kT}$$?

- Set an upper bound on Rosetta "room temperature" by running extended simulations of proteins
- Pick the temperature where none unfold

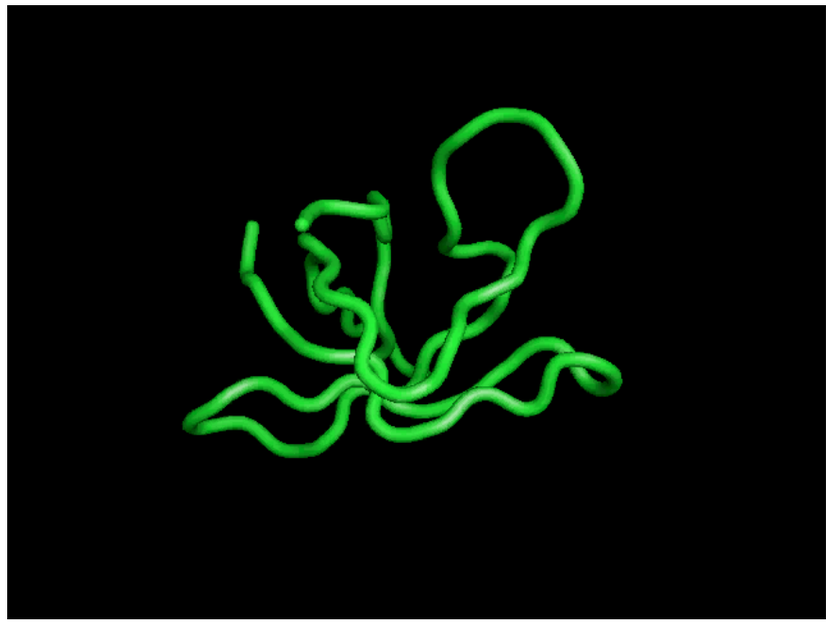
All 6 Proteins Unfold at 0.6 kT

kT = 0.6

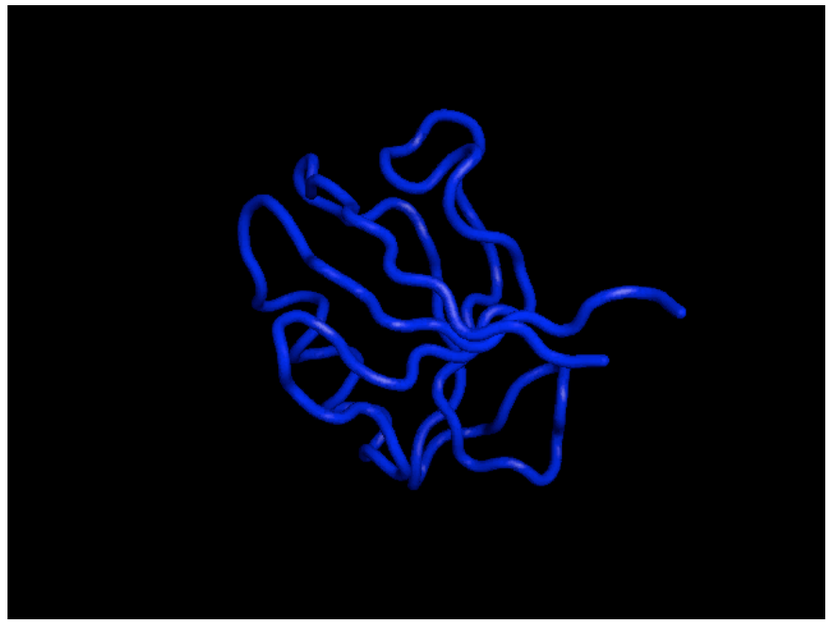
1B72A



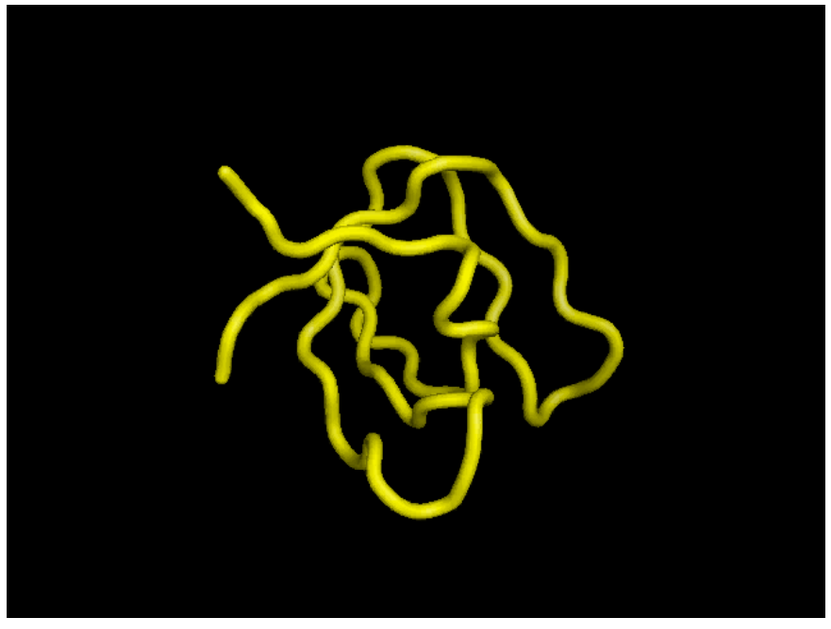
1CSP



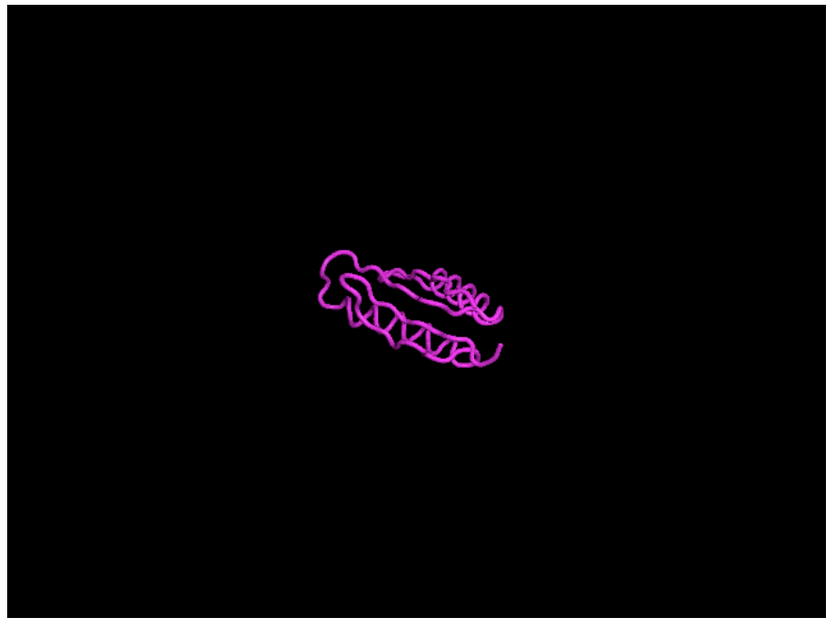
1MFG



1SHFA



1TIG

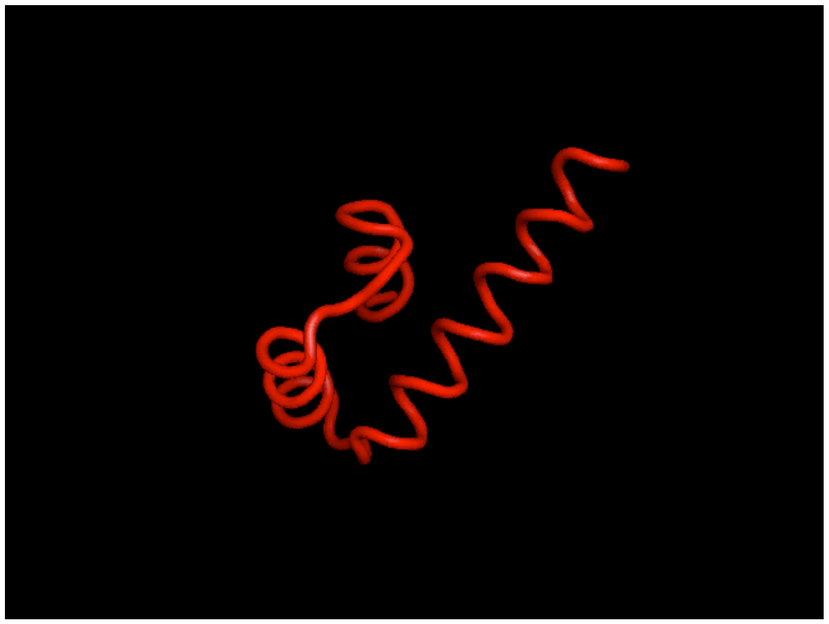


2REB2

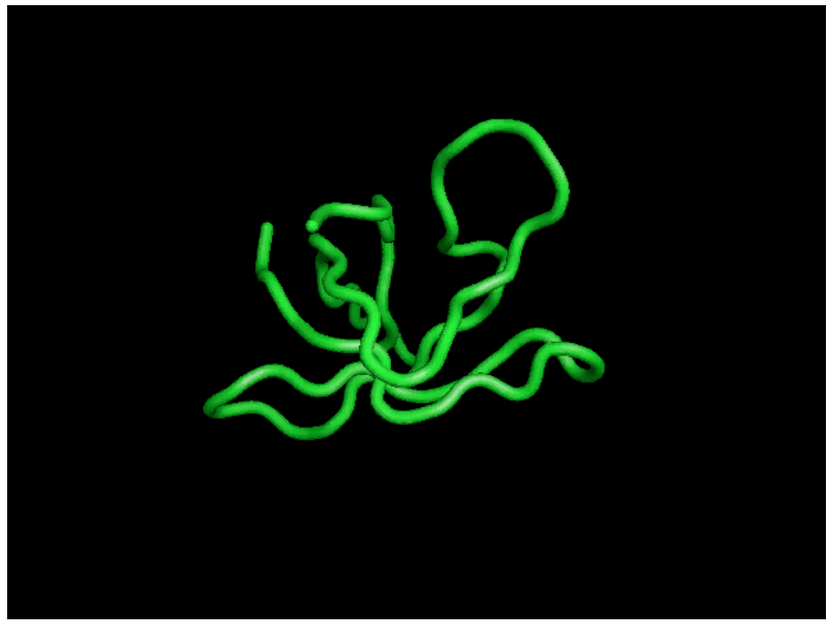
All 6 Proteins Are Stable at 0.3 kT

kT = 0.3

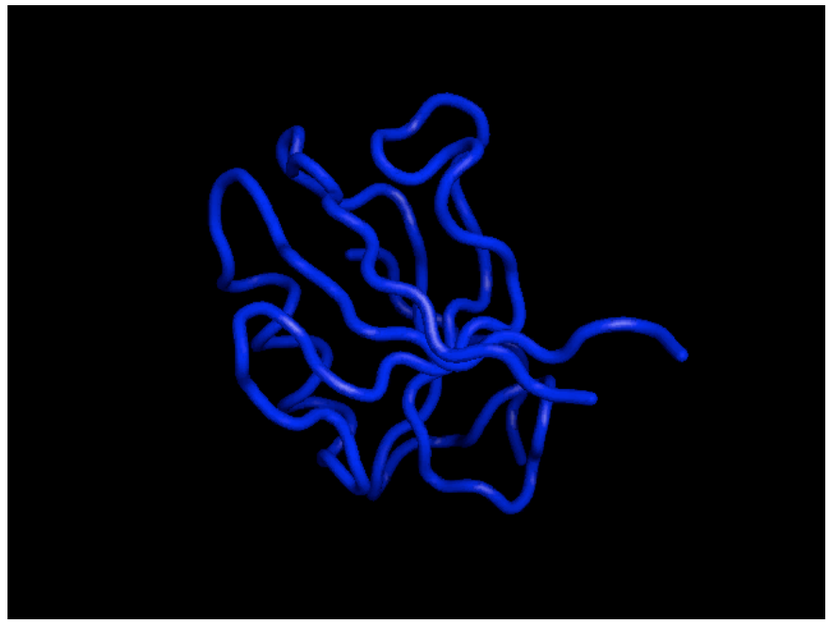
1B72A



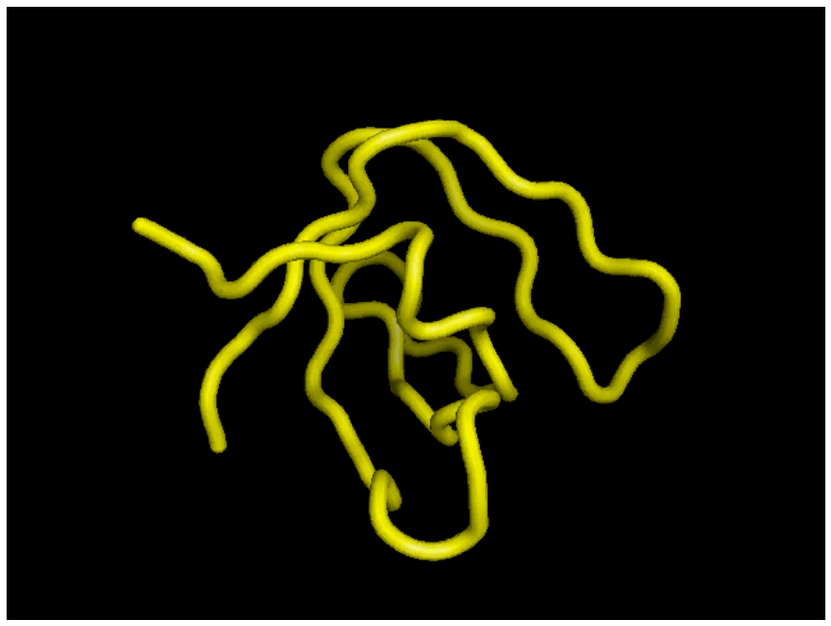
1CSP



1MFG

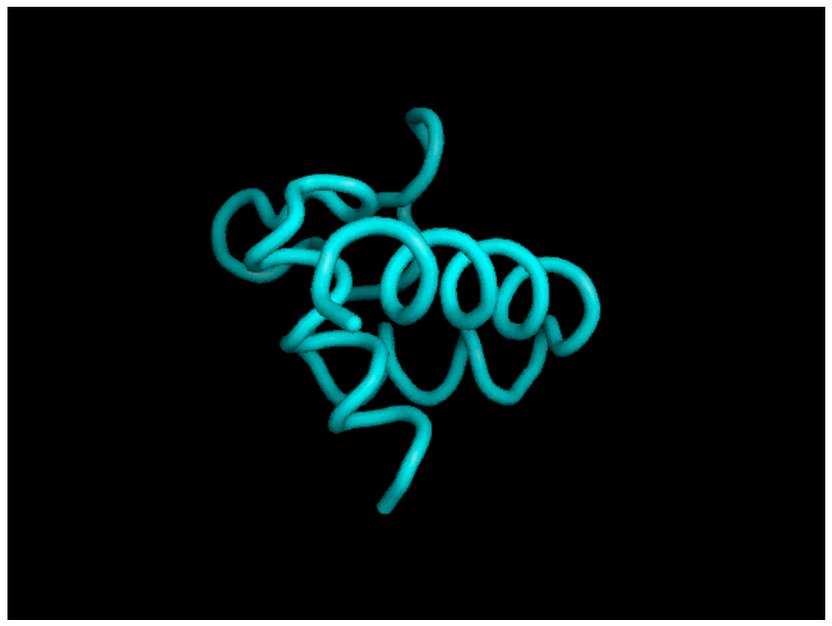


1SHFA



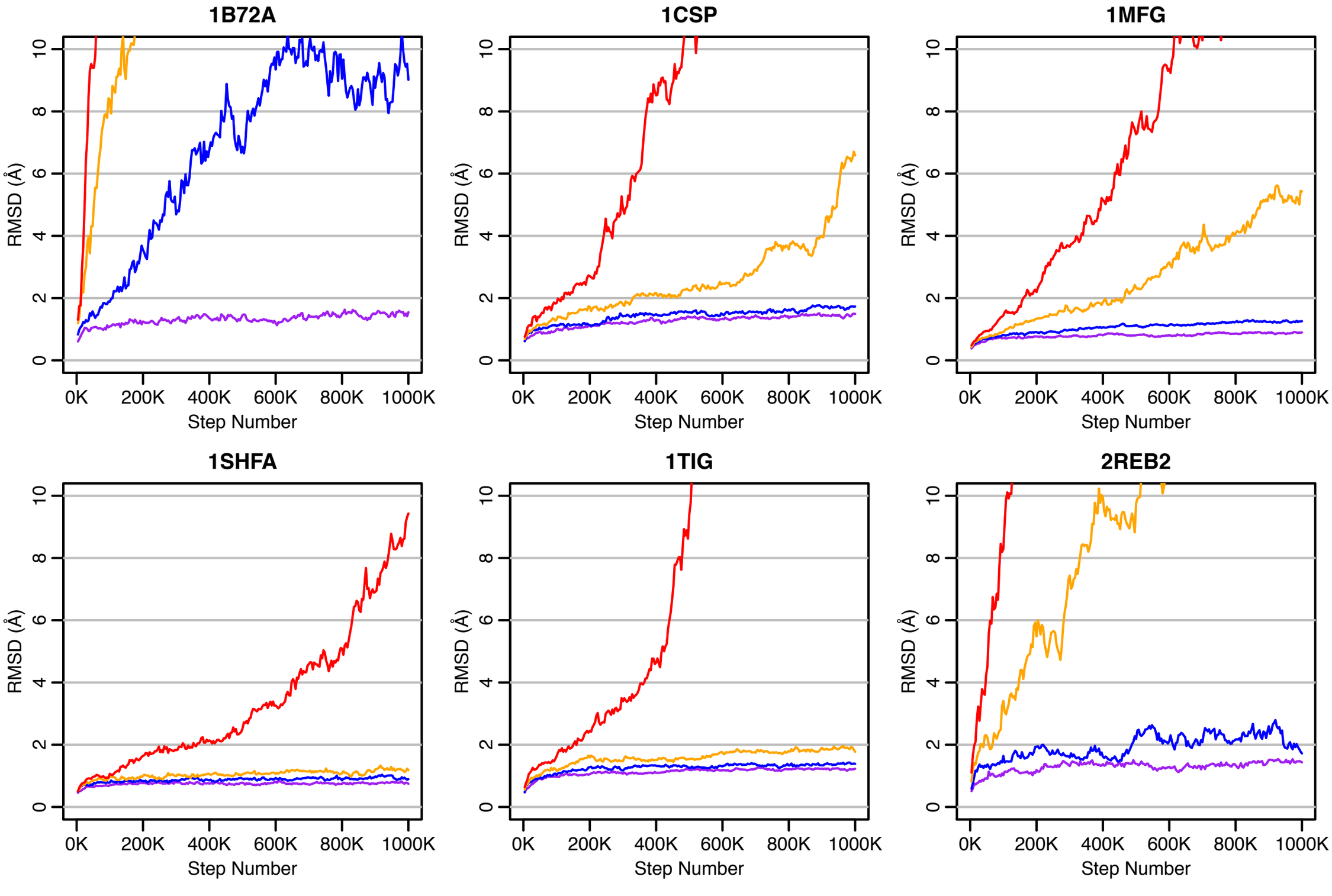
1TIG

2REB2



10 Simulation Mean RMSD

kT = 0.3 **kT = 0.4** **kT = 0.5** **kT = 0.6**



Model System: Cyclophilin A

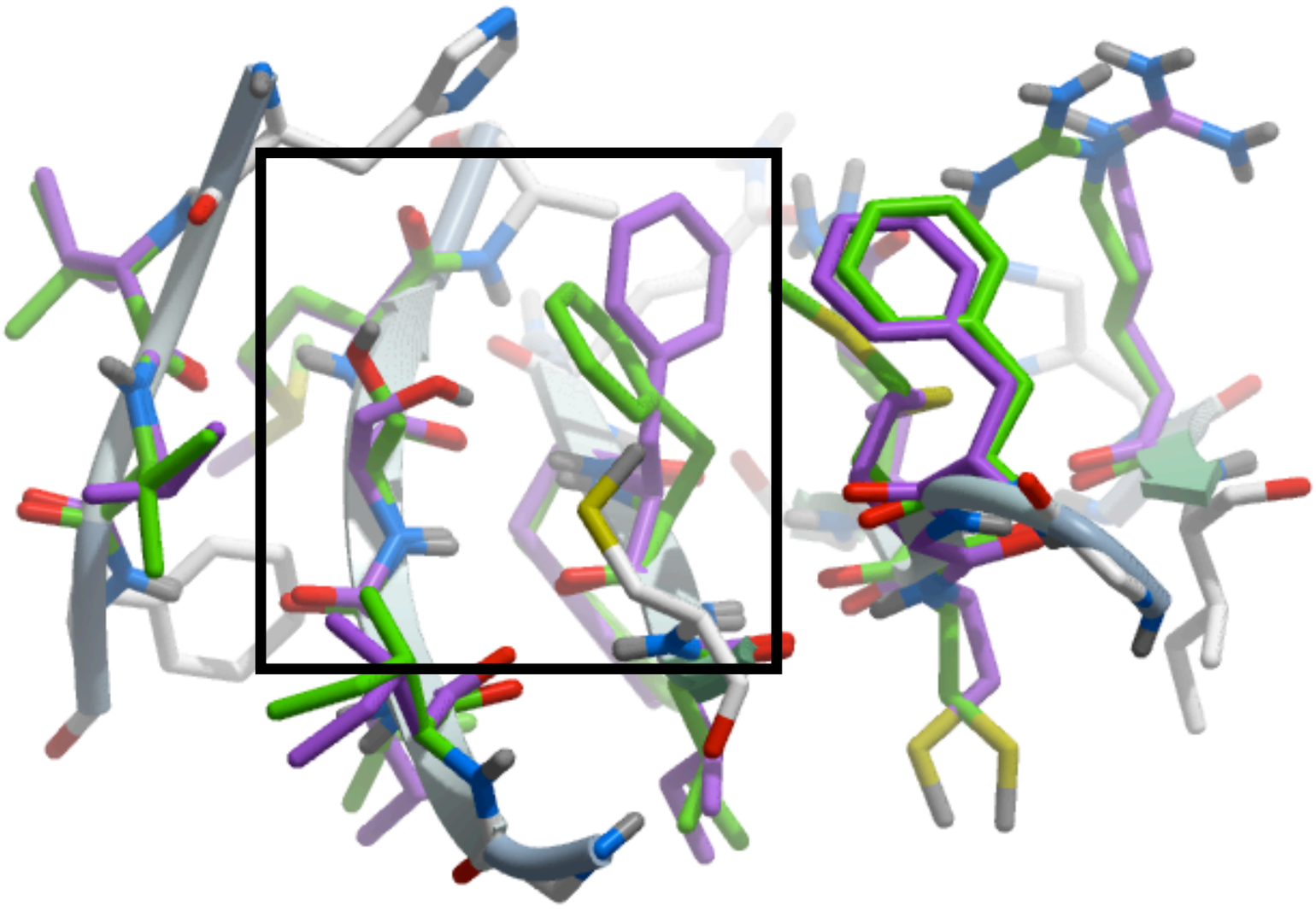
- Catalyzes proline cis-trans isomerization
- Localized side chain/backbone flexibility near the active site has been observed crystallographically and shown to be temperature dependent
- Studies have shown correlations between the rate of protein motion and catalysis
- Even if there is no causation, this is still an interesting system because of the crystallographic motion

Crystallography shows alternate conformations at room temperature

Most of the alternate conformations occur in the 2 beta sheets of the beta sandwich

Fraser et al. *Nature* 2009

Crystallography shows alternate conformations at room temperature

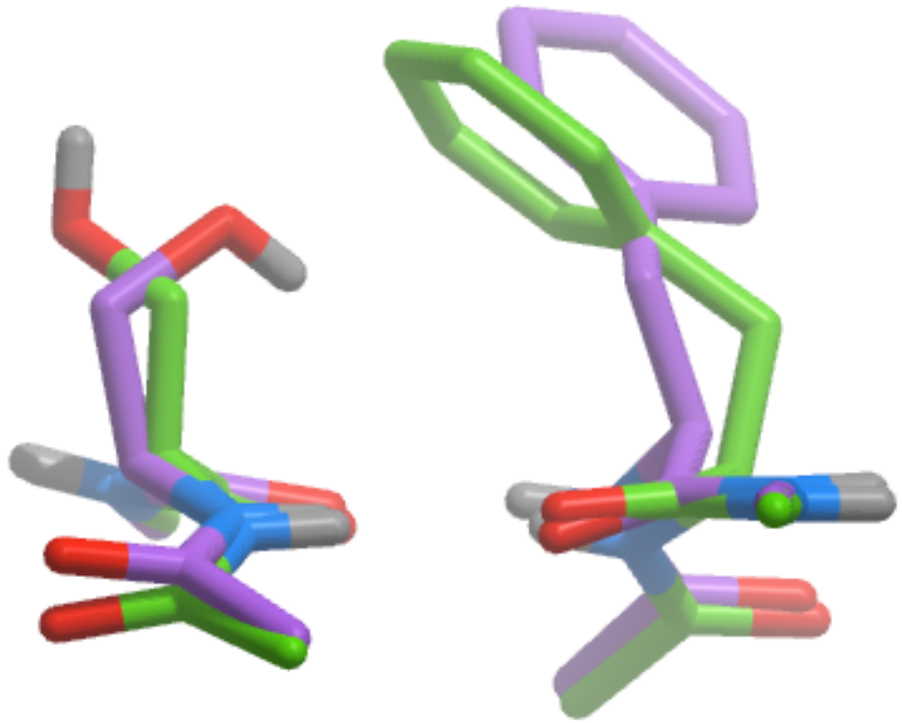


Major Conformation (A)

Minor Conformation (B)

Fraser et al. *Nature* 2009

Crystallography shows alternate conformations at room temperature



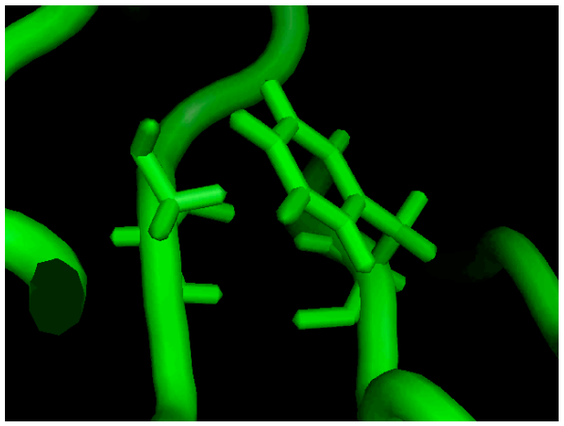
Major Conformation (A)

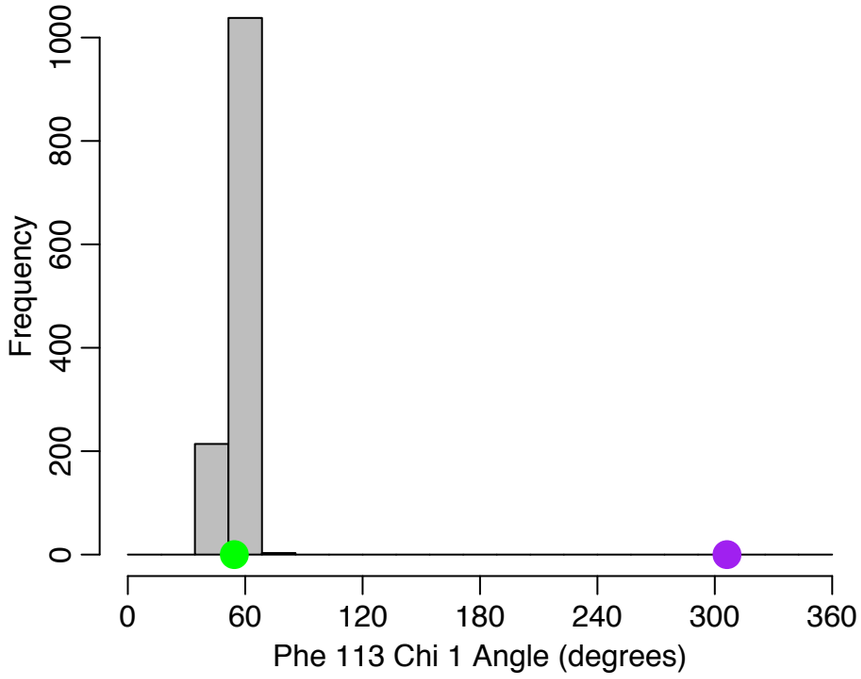
Minor Conformation (B)

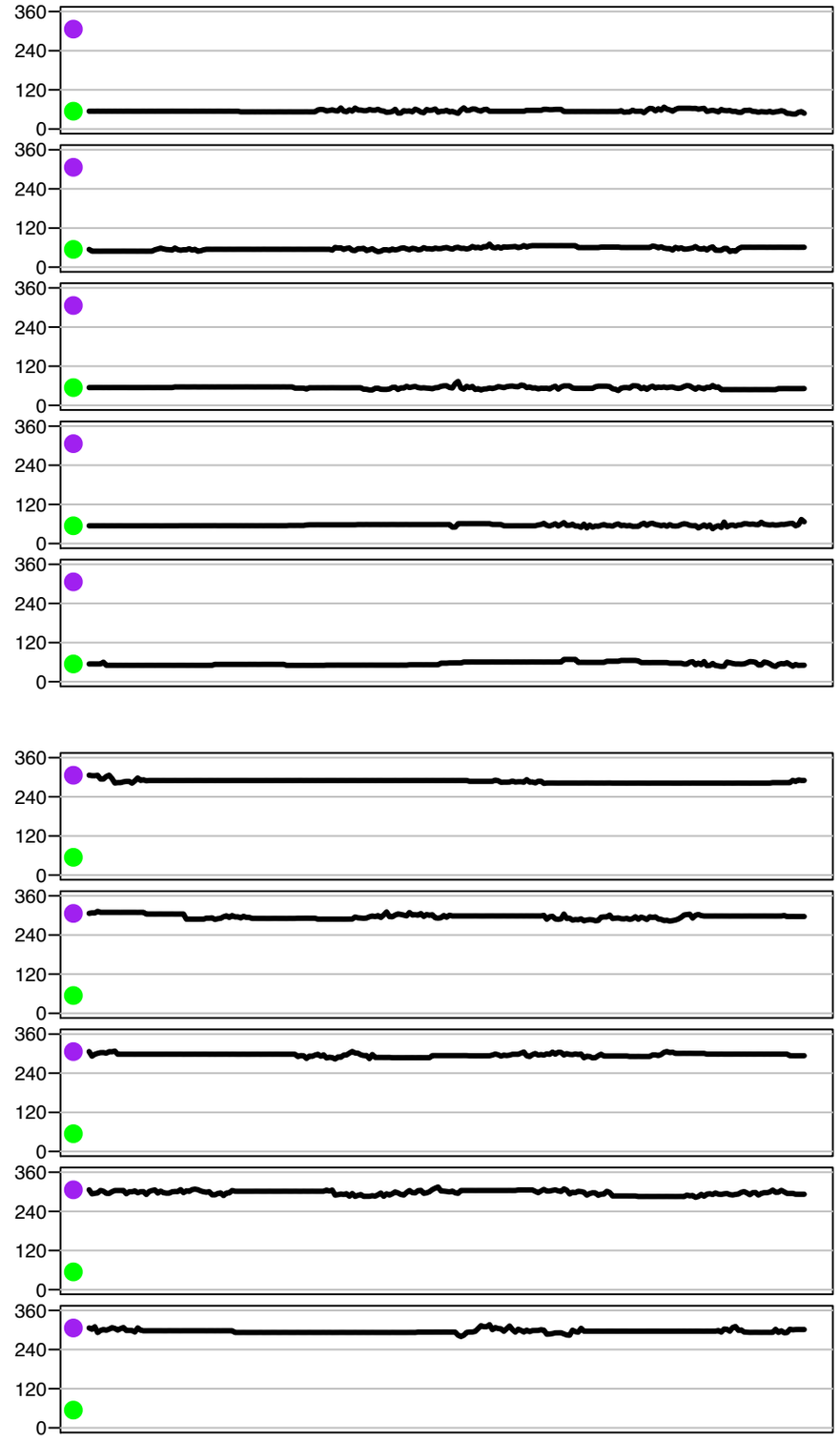
Fraser et al. *Nature* 2009

Little Phenylalanine Motion at 0.3 kT

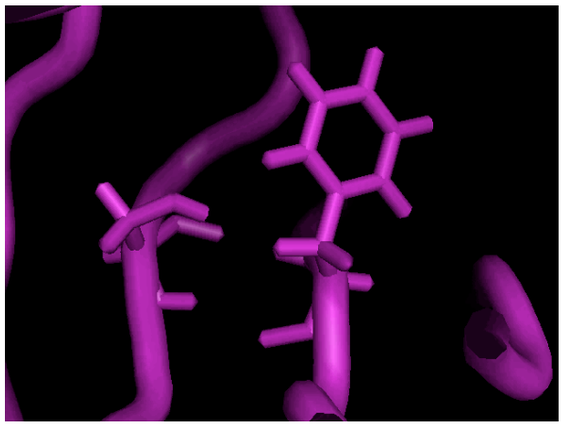
Major (A)

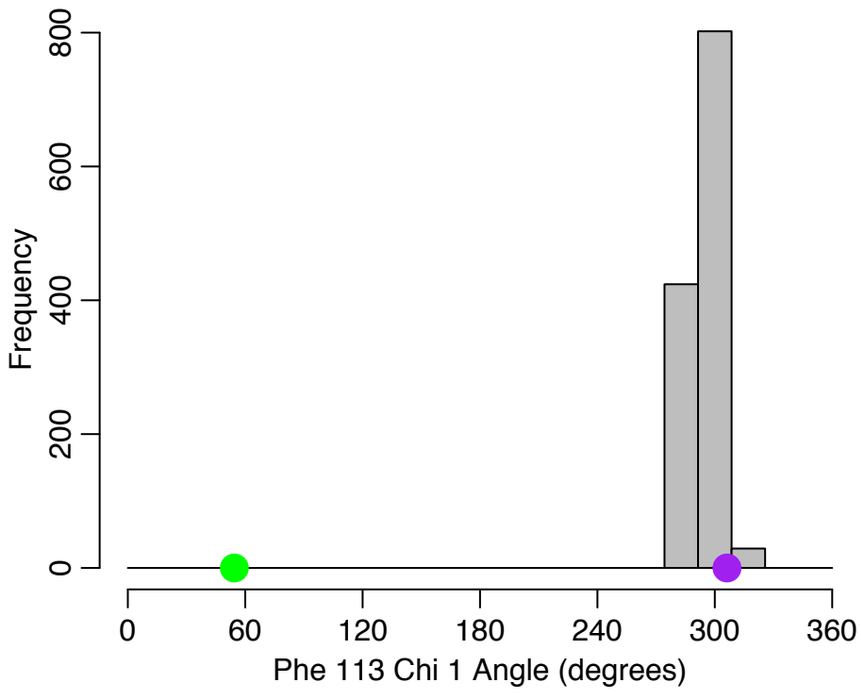






Minor (B)



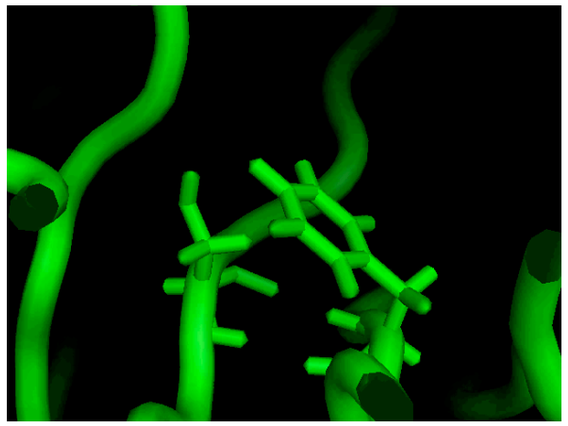


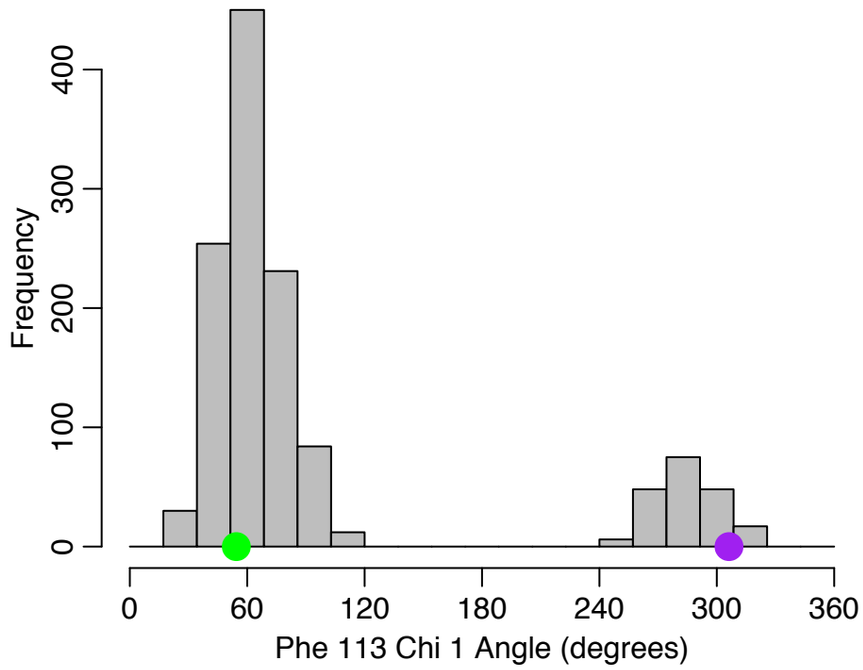
Cyclophilin Protocol Changes

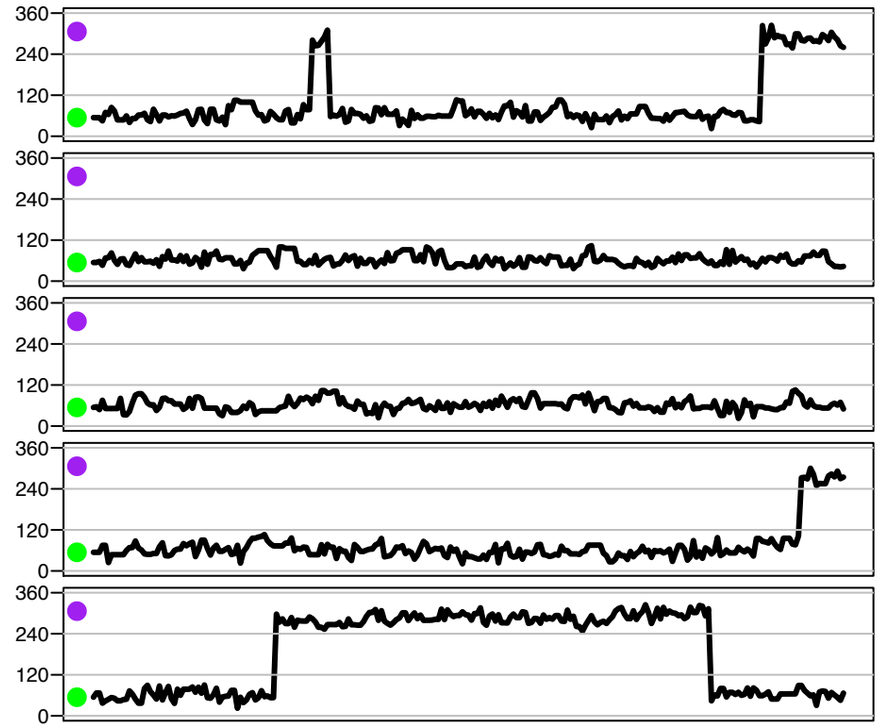
- Restricted sampling to 8 angstroms around Ser 99 & Phe 113
- Raised temperature to 4.8 kT (tried 0.6-2.4 along the way)
- Added simultaneous backrub side chain sampling
- Increased inter-rotamer sampling from 45% to 70%

Phenylalanine Intraconverts at 4.8 kT

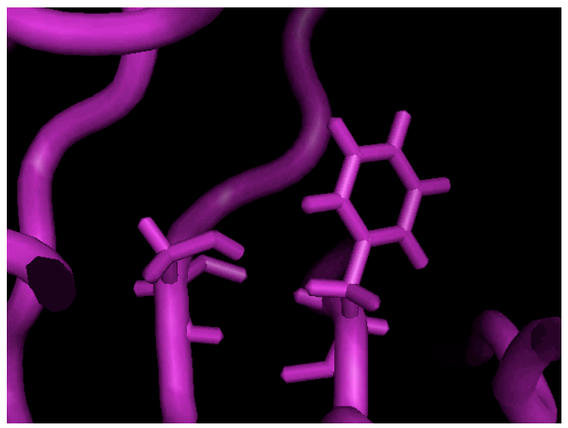
Major (A)

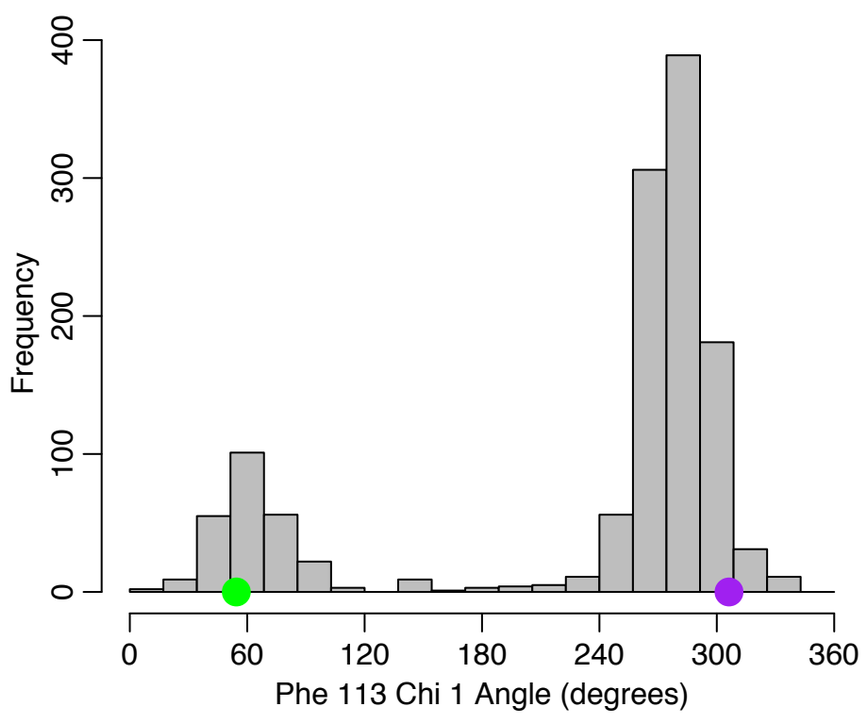


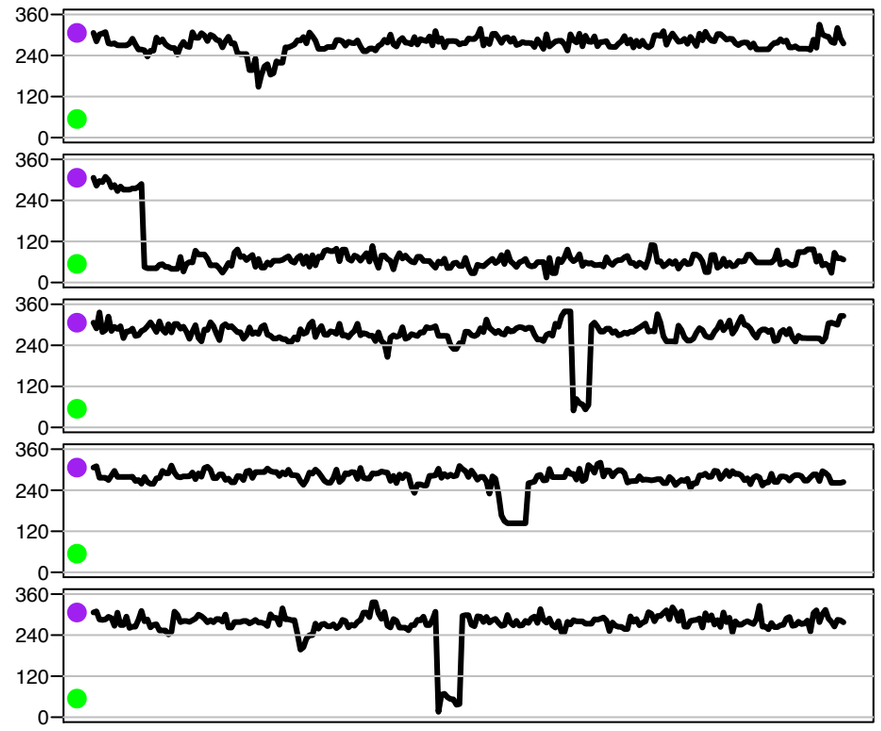




Minor (B)







Thanks

Thomas Bliska (Williams College)
Tanja Kortemme (UCSF)
Jerome Nilmeier (UCB)
Yuan Liu (UW)
Liz Kellogg (UW)
Andrew Leaver-Fay (UNC)
David Baker (UW)

DOD
Genentech