DETERMINING MINIMAL ENSEMBLE MODELS OF SAXS EXPERIMENTS

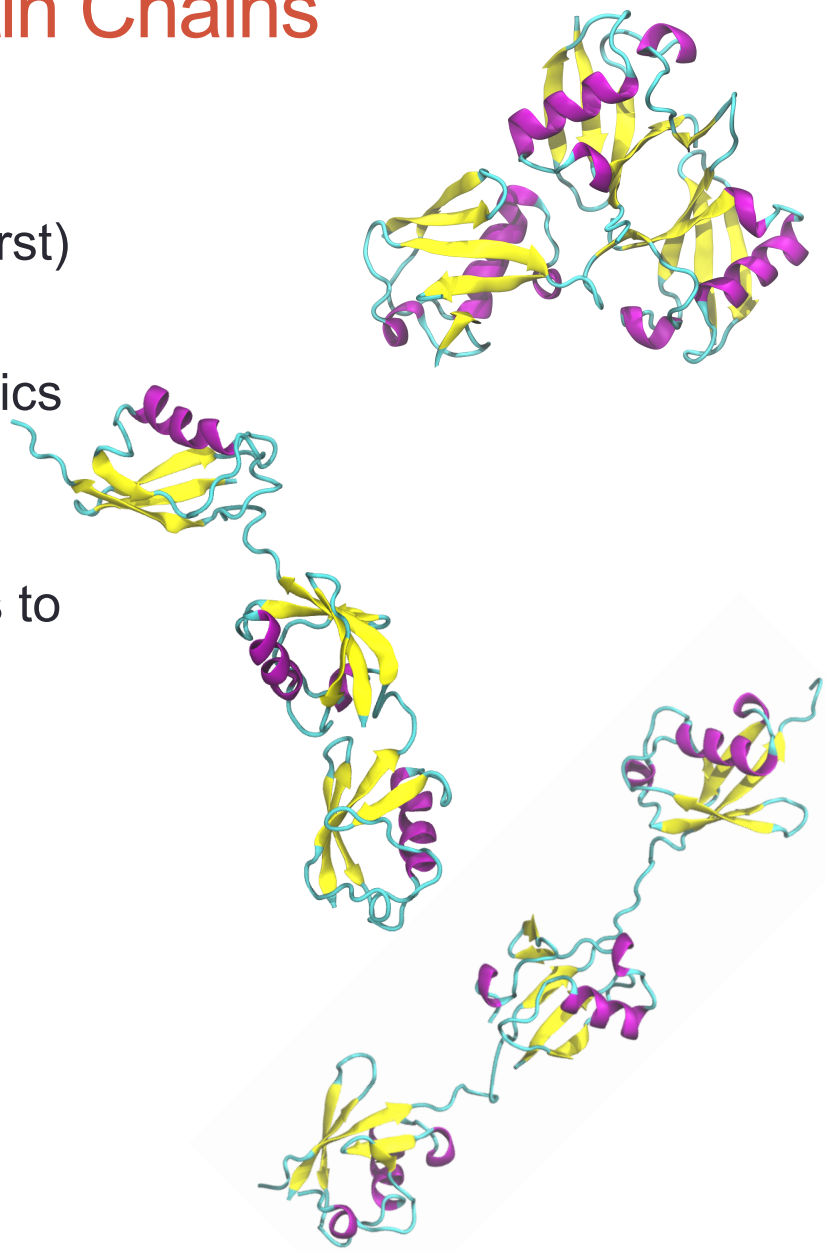
Jeff Wereszczynski
Illinois Institute of Technology, Chicago IL
Biophysics Collaborative Access Team (BioCAT)

Biophysics Collaborative Access Team (BioCAT)

- Located at the Advanced Photon Source (APS), BioCAT is an NIH funded SAXS facility
- Comprises an undulator based beamline, (18-ID) associated laboratory and computational facilities.
- Available to all scientists on basis of peer-reviewed beam time proposals

SAXS + Modelling of Tri-Ubiquitin Chains

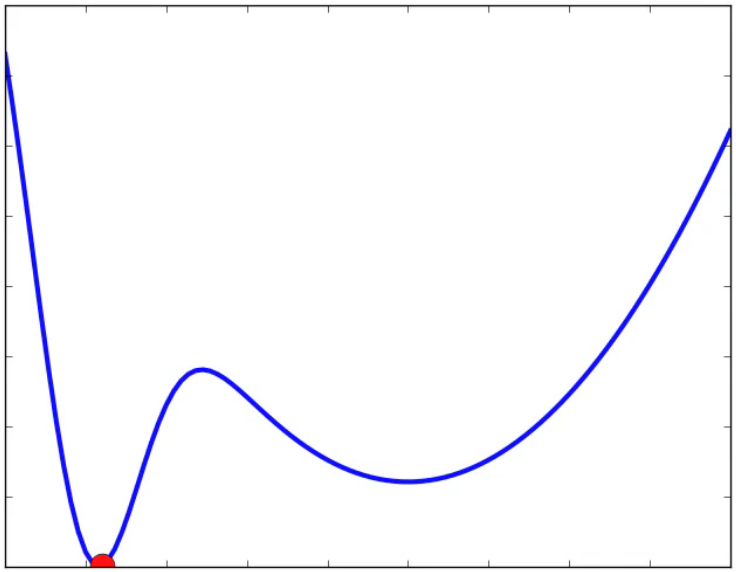
- SAXS experiments performed on diverse tri-ubiquitin systems (Eric Strieter, UMass Amherst)
- Conventional + accelerated molecular dynamics simulations of similar systems (us)
- Bayesian refinement of simulation ensembles to determine the minimal basis set to match experiments



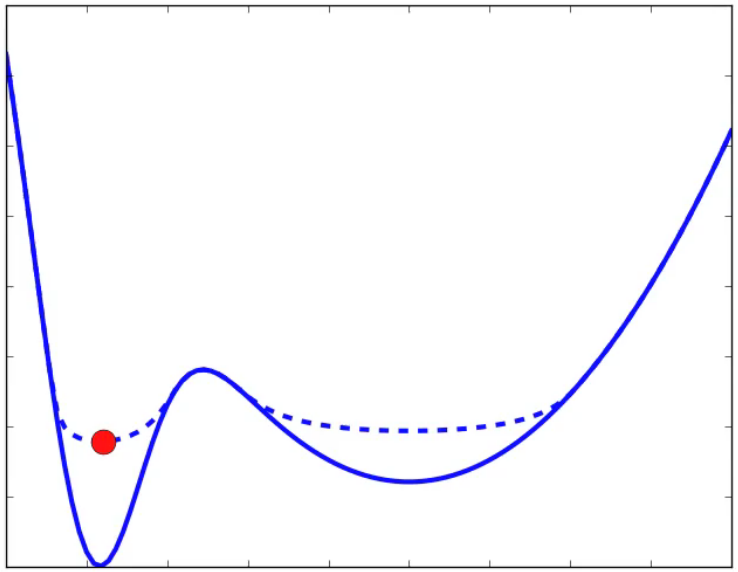
Bowerman *et al.* J. Chem. Theory Comput. (2017) + Poster

Accelerated Molecular Dynamics (aMD) Speeds Sampling (In Theory)

Conventional MD



Accelerated MD



$$V^*(r) = V(r) + \Delta V(r)$$

$$\Delta V(r) = \begin{cases} 0 & V(r) \geq E \\ \dfrac{\left(E - V(r)\right)^2}{\alpha + E - V(r)} & V(r) < E \end{cases}$$

Hamelberg, Morgan, & McCammon *J. Chem. Phys.* (2004)

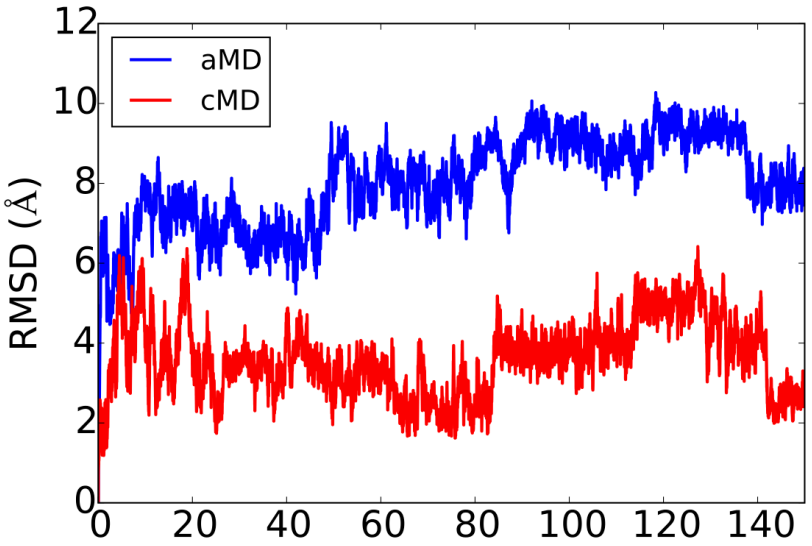
Accelerated Molecular Dynamics (aMD) Speeds Sampling (In Practice)

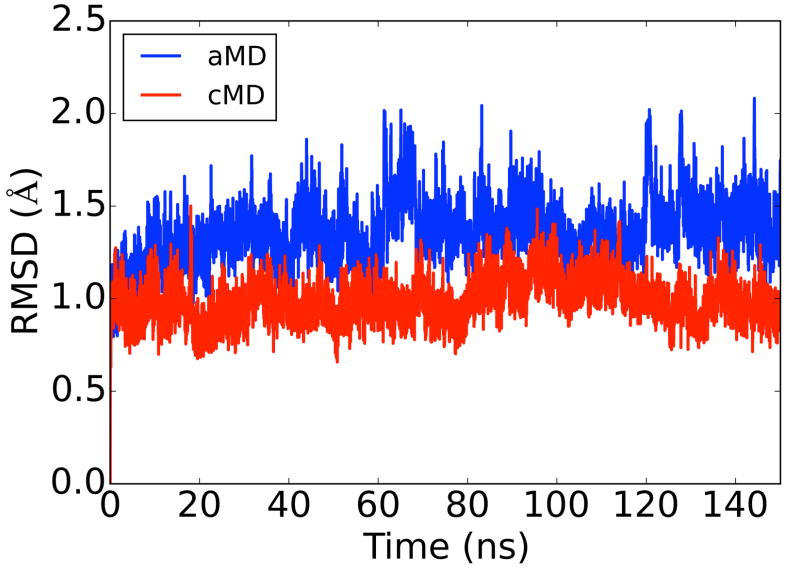
Determining parameters:

- Run a short simulation (~500 ps)
- Extract energies
- Use “back of the envelope” calculations to estimate E and α

Here: 150 ns of aMD vs cMD, explicit solvent

- Trimer RMSDs show increased sampling of large scale motions
- Monomer RMSDs show local secondary structure is maintained





Determining Ensemble of Structures to Fit SAXS Data

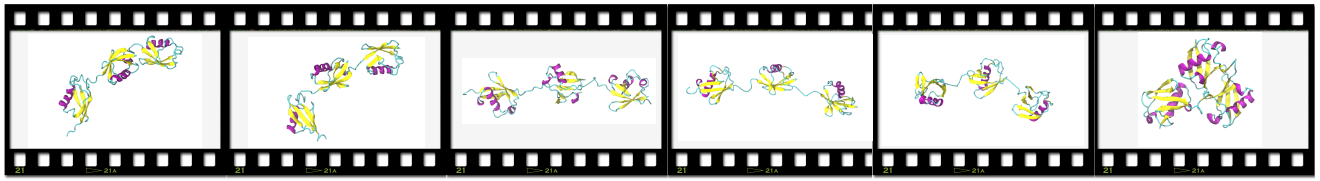
Generate candidate structures
- aMD, cMD, Monte Carlo, TAMD, etc…

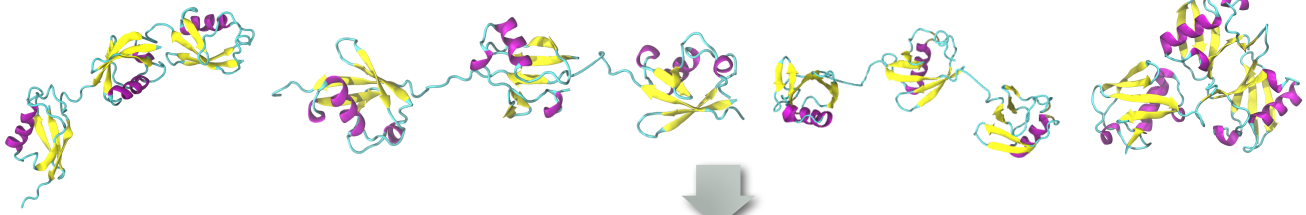
Pare down the structures into a manageable number
- RMSD based clustering

Compute theoretical scattering profiles for each structure
- Crysol (here), SasCalc, Capriqorn, etc.

Cluster scattering profiles
- $\chi 2_{free}$ based hierarchical clustering

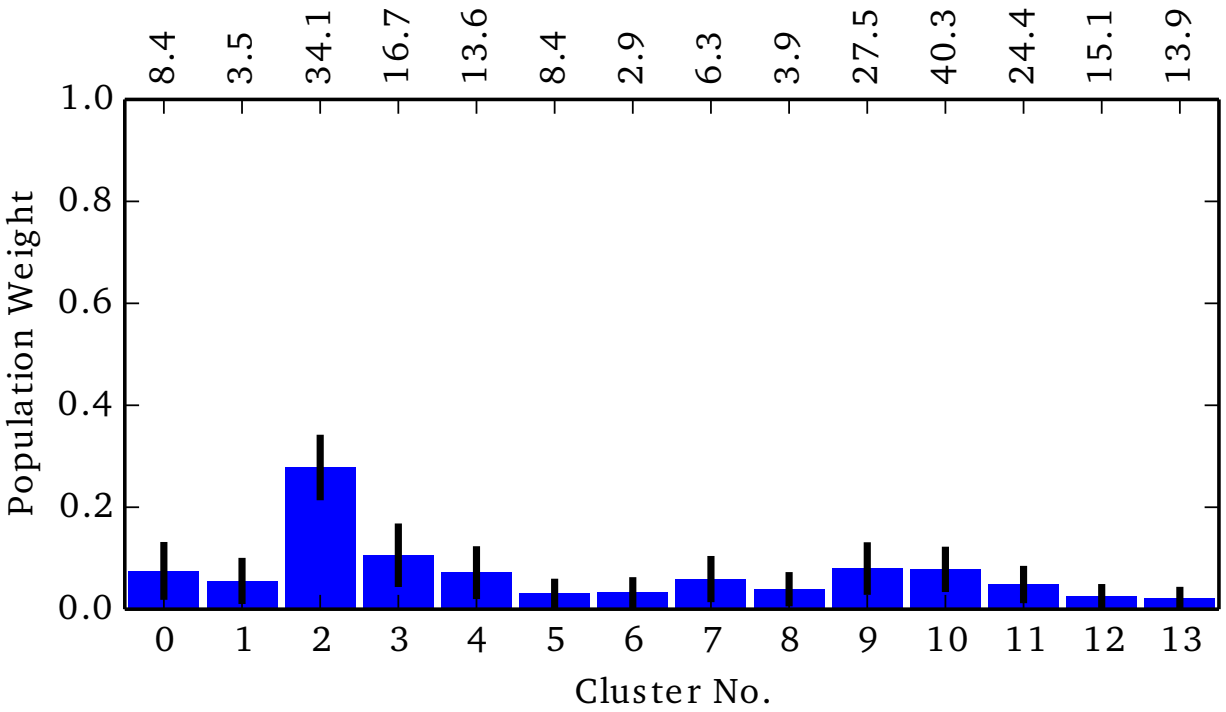
Determine populations of states
- Bayesian Monte Carlo algorithm



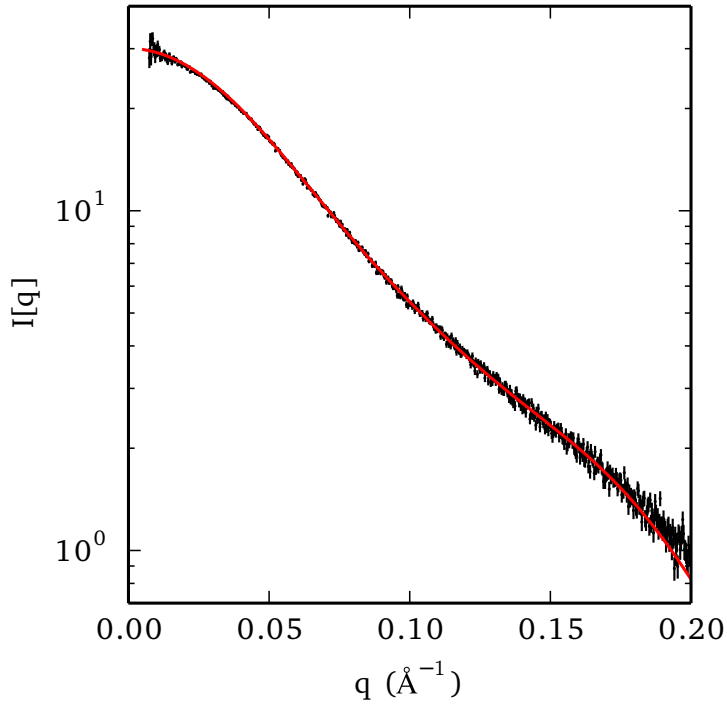


χ^2_{free}: Rambo & Tainer *Nature* (2013)

(Over)Fitting SAXS Data to a Population of States



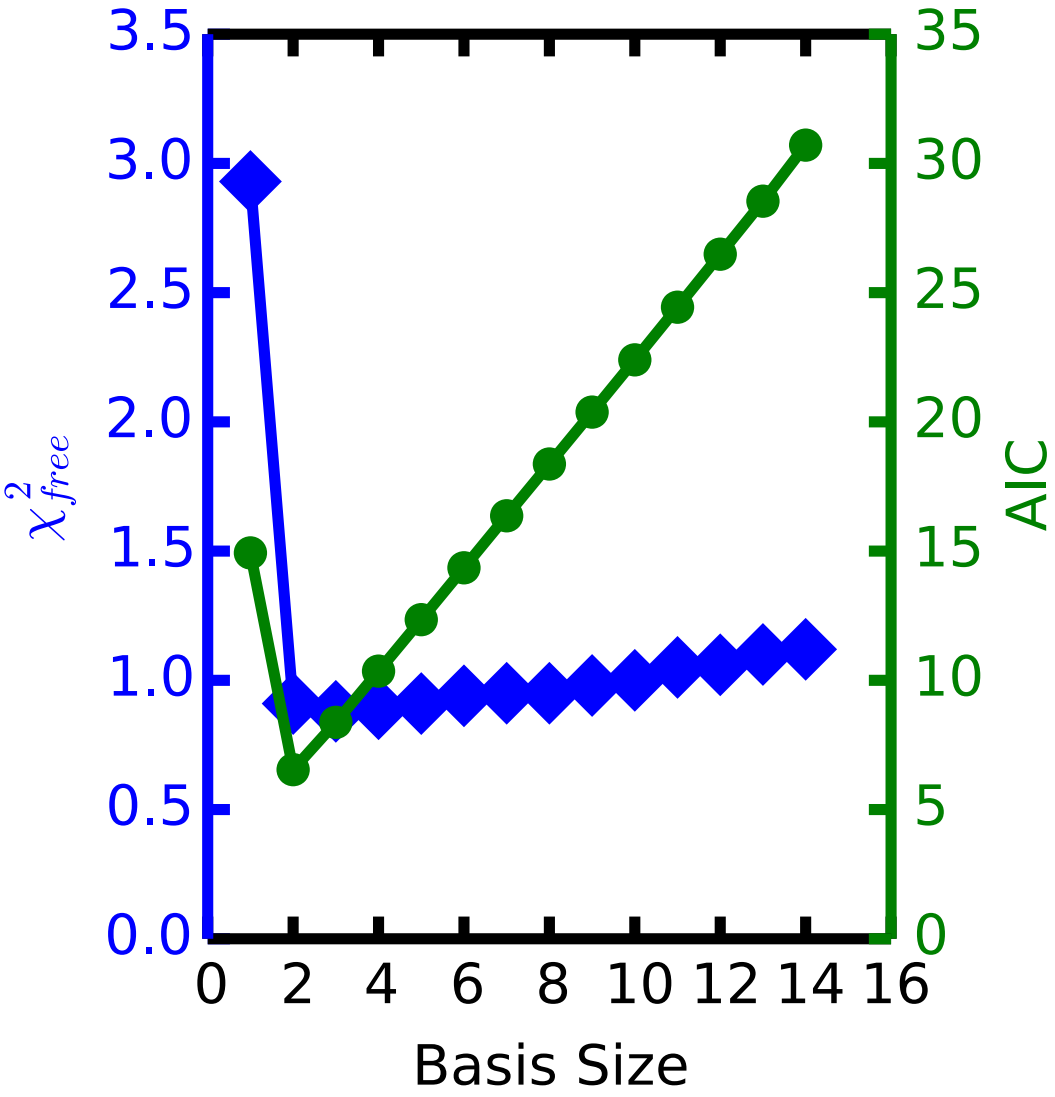
$$\chi^2_{free} = 1.12$$



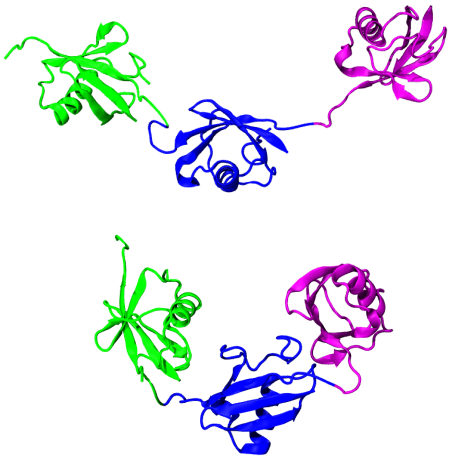
Resisting Overfitting with Iterative Refinement to Find Minimal Basis Set

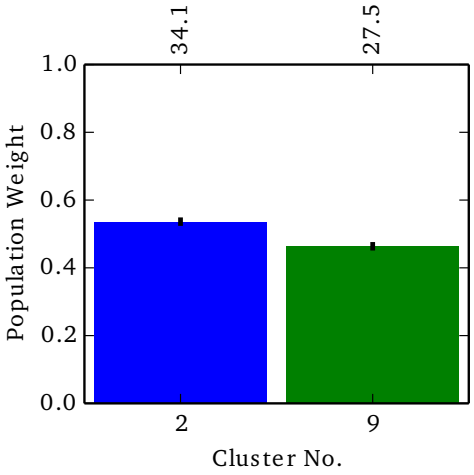
1. Compute populations with single scatterer
2. Compute each permutation of two scatterer basis sets, take the value with minimal χ^2
3. Repeat N times until only all scatterers in basis set
4. Choose ensemble size that minimizes and the Akaike information criterion (AIC)

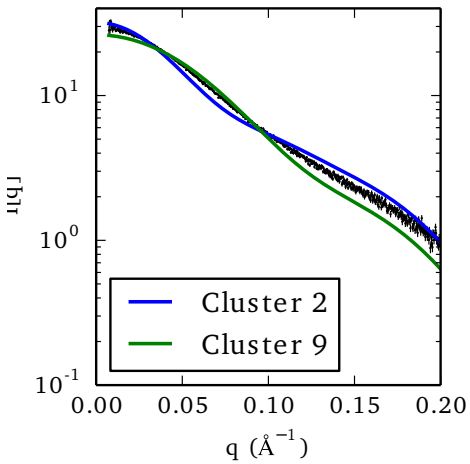
$$AIC = 2k - 2\ln L$$

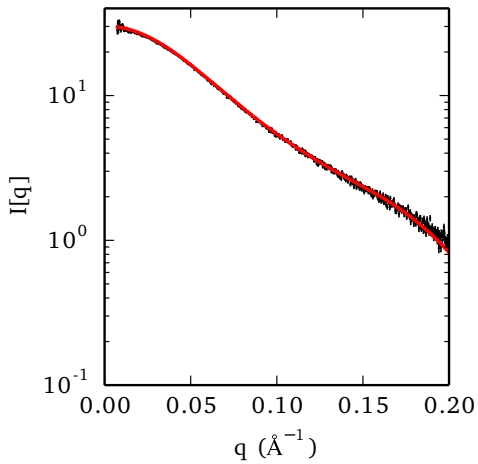


Iterative Refinement to Find Minimal Ensemble







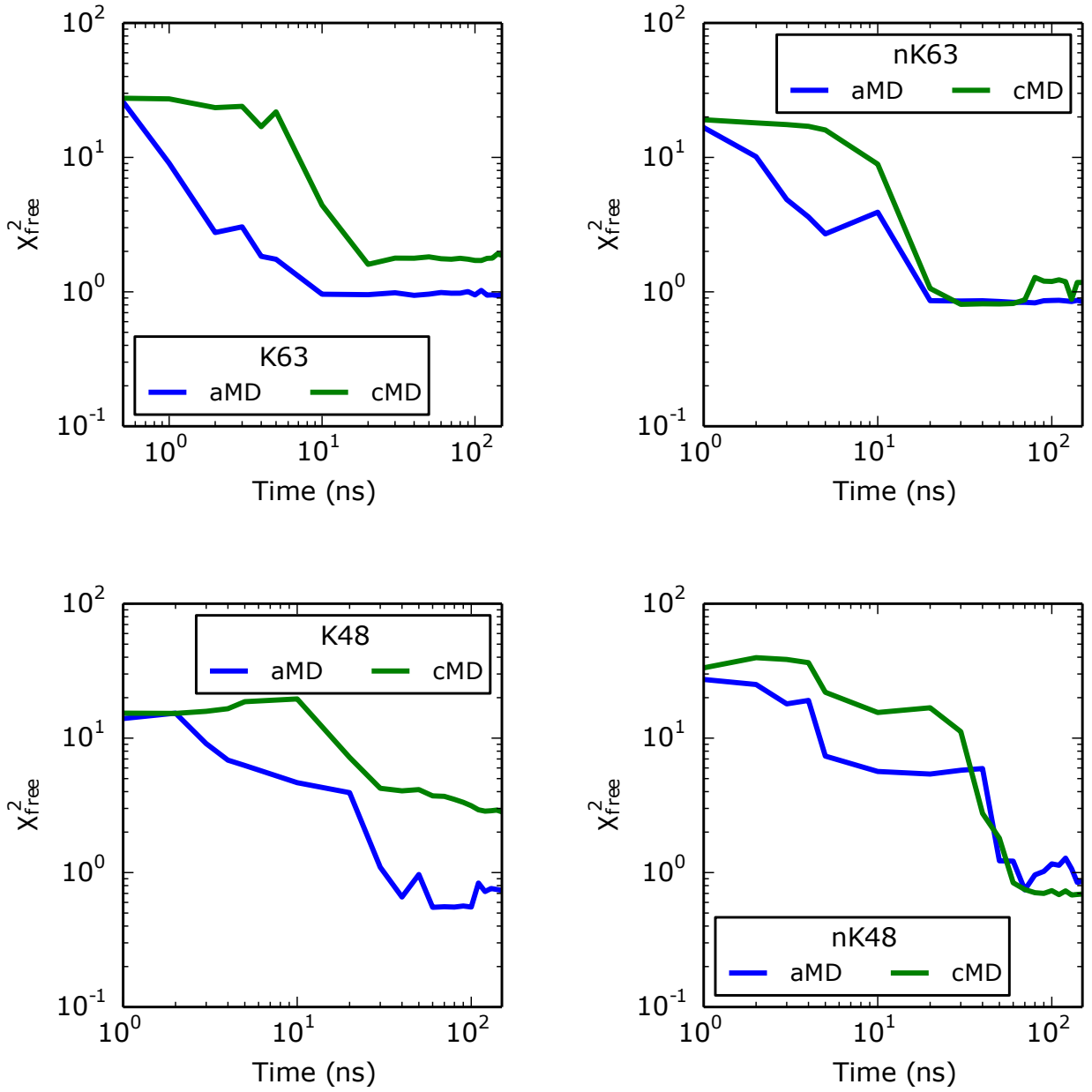


Cluster	R_g (Å)	Distal Group Distance (Å)	Interdomain Angle
2 (54%)	32.6 ± 0.2	68.9 ± 0.4	120°±0.1°
9 (46%)	23.3 ± 0.3	41.0 ± 1.5	97°±8.3°
Combined	28.3 ± 0.3	~	~
Experiment	28.0 ± ~1.0	~	~

Results for All Systems

System	χ^2_{free}	# Final Structures
K6	2.4	1
K11	0.7	2
K29	0.9	1
K48	0.7	1
nK48	0.9	2
K63	1.0	2
nK63	0.9	2

Convergence of accelerated and conventional MD

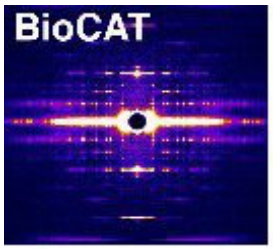


Conclusions & Future Directions

- Accelerated MD can dramatically improve sampling
- SAXS + Bayesian methods can help to “reweight” aMD simulations
- Many tri-ubiquitin systems adopt both compact and extended states in solution
- We are working to apply these methods to other systems.
- aMD + ensemble refinement are coming soon to a SASSIE near you!

Acknowledgments

- Group Members:
  - Samuel Bowerman
  - Amy Rice
- Dr. Srinivas Chakravarthy (ANL)
- Prof. Eric Strieter (UMass Amherst)
- Funding:
  - NIH: R15GM114758, R35GM119647
  - NIH: P41GM103622 (BioCAT)
- XSEDE Computing Resources



XSEDE

Extreme Science and Engineering Discovery Environment