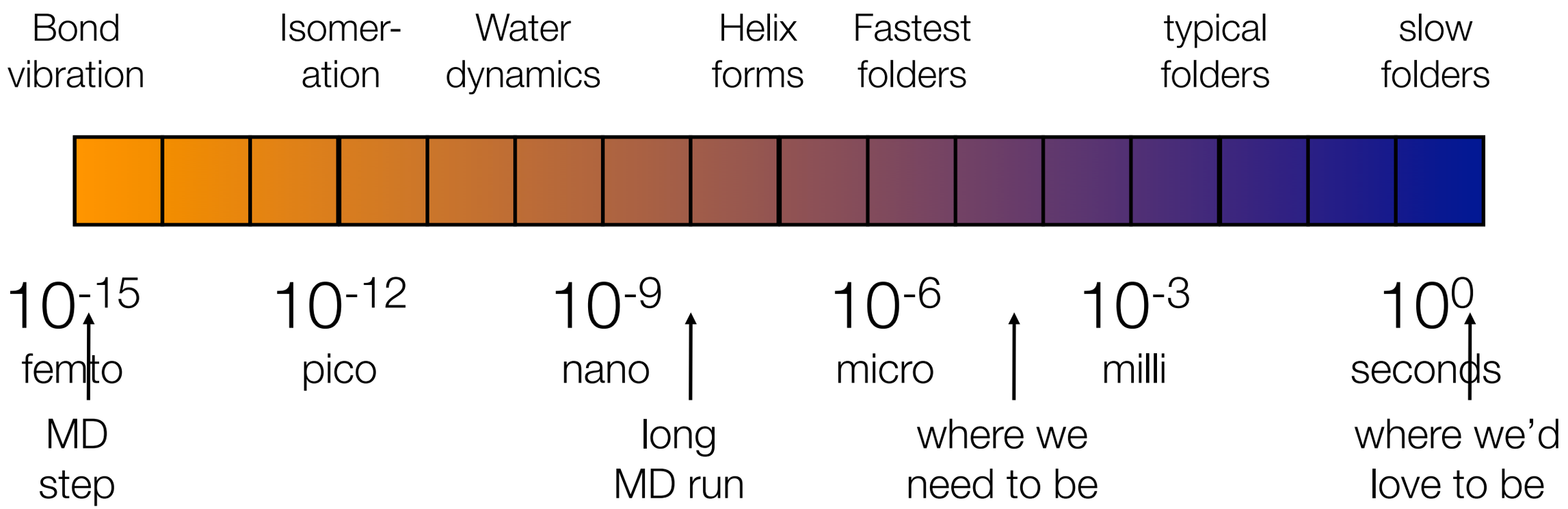
Advanced Topics in Molecular Dynamics: Sampling & Solvation

Vijay Pande

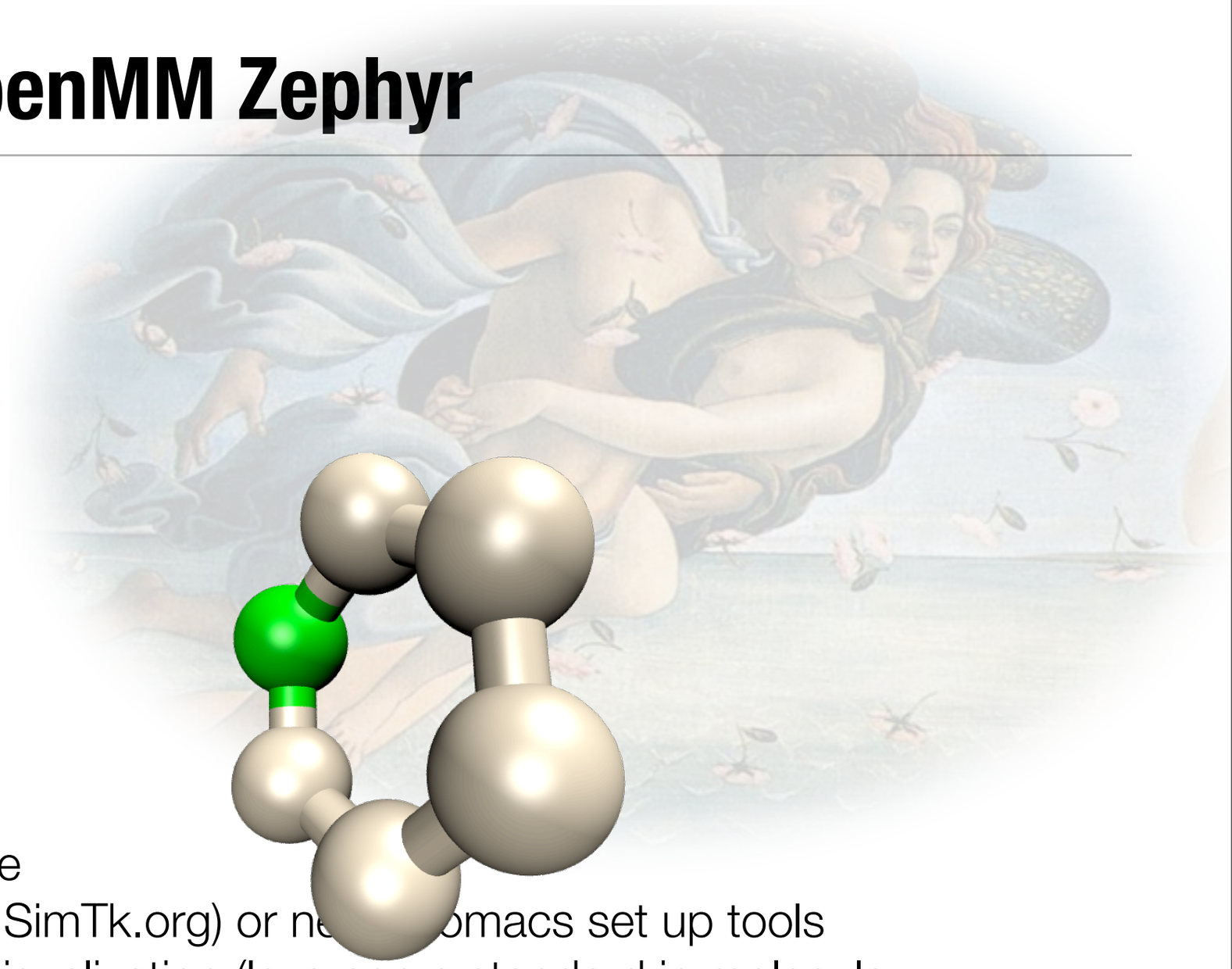
OpenMM Workshop, February 13, 2009

Challenge: timescales



- **Kinetics: reaching experimental timescales**
- **Thermodynamics: convergence**
  - Are the results independent of initial conditions?

New Application: OpenMM Zephyr



- **Goals**
  - make MD easy to run
  - easy but correct setup (not just PDB -> MD, but think about protonation, missing residues, etc)
  - easy to run on GPU's
  - visual feedback
- **Under the hood**
  - Wrap GPU enabled MD code
  - use MMtools (Pande group, SimTk.org) or ne omacs set up tools
  - Use VMD IMD interface for visualization (leverage a standard in molecular visualization)
- **Use of real time visualization**
  - immediate feedback is not just fun, but can be useful
  - key to correct setup, etc

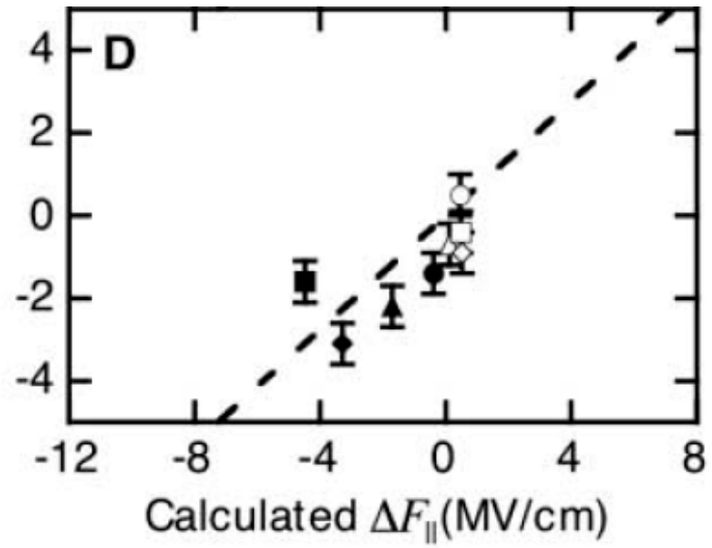
Large speed increases seen using GPU

Molecule	# atoms	ns/day	speedup*	GFLOPS (GPU)	GFLOPS (x86)
fip35	544	576	128	311	657
villin	582	529	136	328	692
lambda	1254	202	255	547	1153
a-spectrin	5078	17	735	805	1702

(*comparing a GTX280 to a single core of a 3GHz core 2 duo using the AMBER code)

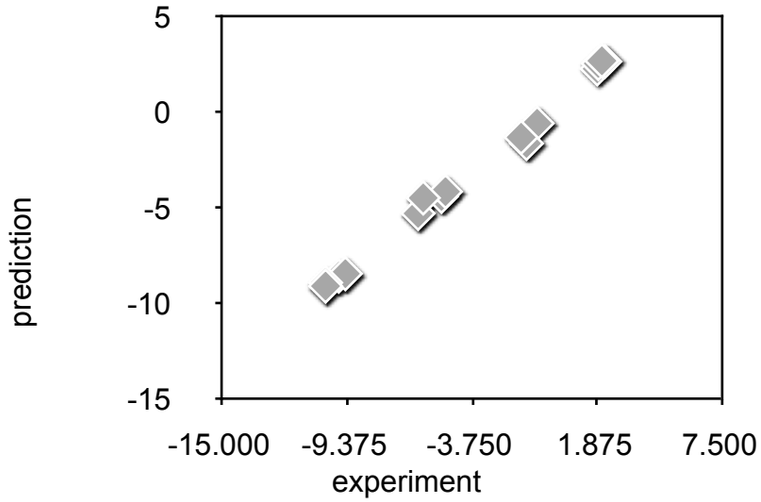
(Shirts, Snow, Zagrovic, et al)

How accurate are atomistic physical models?



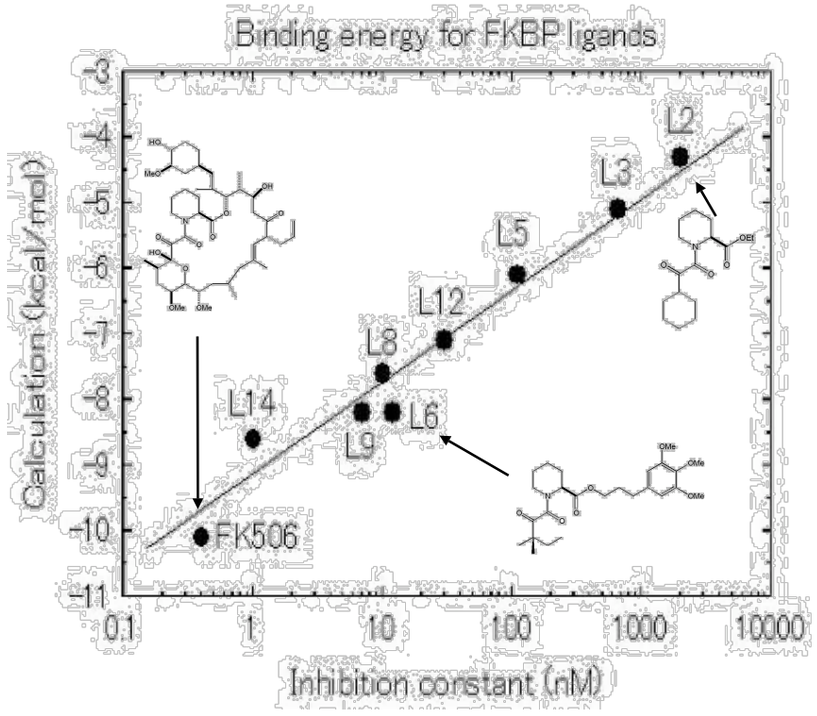
Science, 313 200-4 (2006)

ELECTROSTATICS



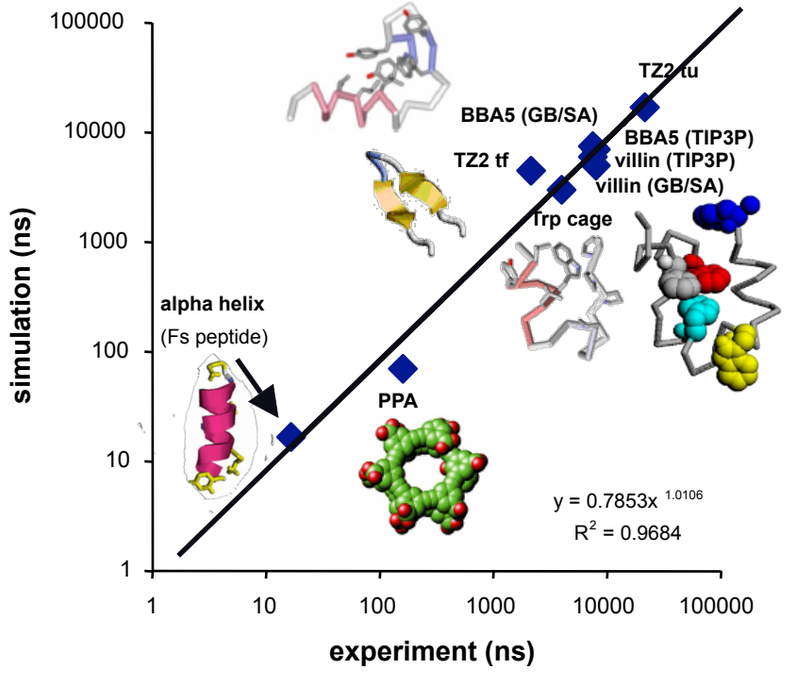
Journal of Chemical Physics, 119 5740-5761 (2003)

SOLVATION FREE ENERGY



Journal of Chemical Physics, 123 084108 (2005)

THERMODYNAMICS



Annual Reviews of Biophysics 34 43-69 (2005)

KINETICS

Case study: implicit solvent

http://pande.stanford.edu

How does Generalized Born (GB) work?

- **Break down water into dielectric and hydrophobicity**

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel},$$

- **Make an ansatz for the form of the dielectric**

$$\Delta G_{el} \approx \Delta G_{gb} = -\tfrac{1}{2}\sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)}$$

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)\right]^{\frac{1}{2}}$$

- **Must calculate the Born Radii (R_i)**

$$R_i^{-1} = \rho_i^{-1} - \frac{1}{4\pi} \int_{solute} \theta(|\mathbf{r}| - \rho_i) \frac{1}{r^4}\, d^3\mathbf{r}.$$

- **Can include salt effects**

$$\Delta G_{el} \approx \Delta G_{gb} = -\tfrac{1}{2}\sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)} \left(1 - \frac{e^{-\kappa f_{gb_{ij}}}}{\varepsilon_w}\right)$$

Can this be put on a more formal ground?

- **Limiting case: single ion of radius r yields the Born eq**

$$\Delta G_{el} \approx -\tfrac{1}{2}\left(1-\frac{1}{\varepsilon_w}\right)\frac{q^2}{r}$$

- **For the linearized PB equation, one can derive the exact result in a spherical geometry (Kirkwood equation), which yields**

$$\Delta G_{el} \approx \Delta G_{gb} = -\tfrac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)}$$

$$f_{Kirkwood} = [\, r_{ij}^2 + R_i\, R_j]^{1/2}$$

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)\right]^{\frac{1}{2}}$$

- **The exp term can be considered an empirical fix for non-spherical geometries**

Different Generalized Born models

- **All have the same general form**

$$\Delta G_{el} \approx \Delta G_{gb} = -\tfrac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)}$$

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)\right]^{\frac{1}{2}}$$

- **But differ in the calculation of the Born radii**
  - Still (Original)
  - Hawkins, Cramer, Truhlar ("HCT")
  - Onufriev, Bashford, Case ("OBC")
  - Mongan, Simmerling, McCammon, Onufriev, Case ("GBneck")

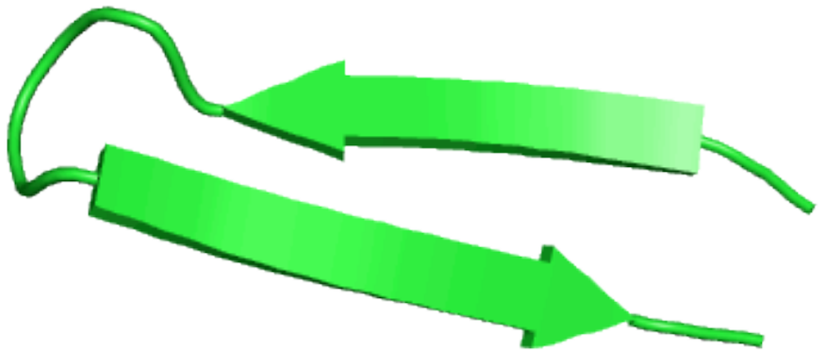
- **Goal is to best model the nature of the dielectric region**

Large number of force fields to choose

- **AMBER**
  - ff94: too helical (explicit solvent)
  - ff96: too beta sheet (explicit solvent)
  - ff99: not helical enough (explicit solvent)
  - ff99sb: modifications to improve torsions
  - ff03: latest, intended to be balanced
- **OPLS**
  - OPLS-ua (unified atom)
  - OPLS-aa: classic all atom force field
  - OPLS-aa/L: new torsions
- **CHARMM**
  - CHARMM19 (unifed atom)
  - CHARMM27 (latest)
- **CHARMM**
  - CHARMM19 (unifed atom)
  - CHARMM27 (latest)
  - CMAP (new torsions for use with CHARMM27 or other CHARMM ff's)
- **Other**
  - GROMOS (van G.)
  - GROMACS
  - Encad (Levitt)
- **Polarizable force fields**

Test systems

(Scott Shell, UCSB; Ken Dill, UCSF)



protein G hairpin
16 residues

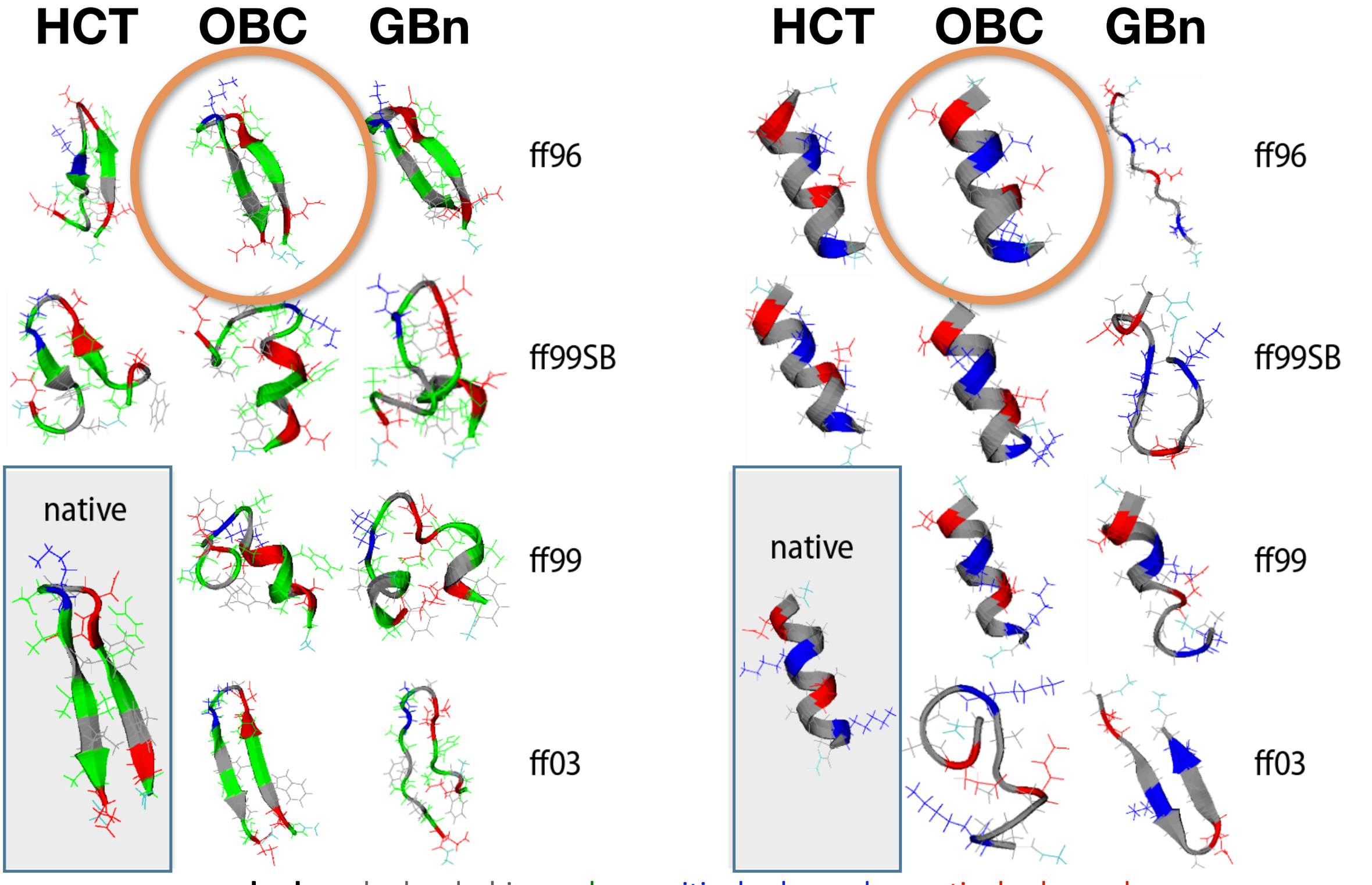
EK peptide
14 residues

multiple force fields
x multiple solvent models
x two test peptides
x three runs each
x 10 ns REMD runs
≈ 12 μs aggregate simulation time

≈ 60 CPU-years of compute time

Results for AMBER ff's

(Scott Shell, UCSB; Ken Dill, UCSF)



color key: hydrophobics, polar, positively charged, negatively charged

Case study: protein thermodynamics

http://pande.stanford.edu

Sampling methods

- **NVE MD: Constant energy**
  - often used to be most faithful to kinetics
  - important test of an MD code (no bugs or numerical issues)

- **NVT MD**
  - uses a thermostat
  - Vijay's opinion (w/data): can be used for kinetics, if a thermostat is used carefully

- **<u>Other thermodynamics methods</u>**
  - **ST**: Serial Tempering
  - **REMD**: Replica Exchange Molecular Dynamics (aka parallel tempering)
  - **MSM**: Markov State Models

Next steps: Generalized Ensemble (GE) methods

- **Basic idea: define new form of kinetics to overcome long timescale behavior**
  - must be able to recover Boltzmann weighted configurations
  - but we can define what ever form of dynamics we want
  - and we can create new potential forms, as we can transform back
  - generalization of methods like "parallel tempering" or REMD

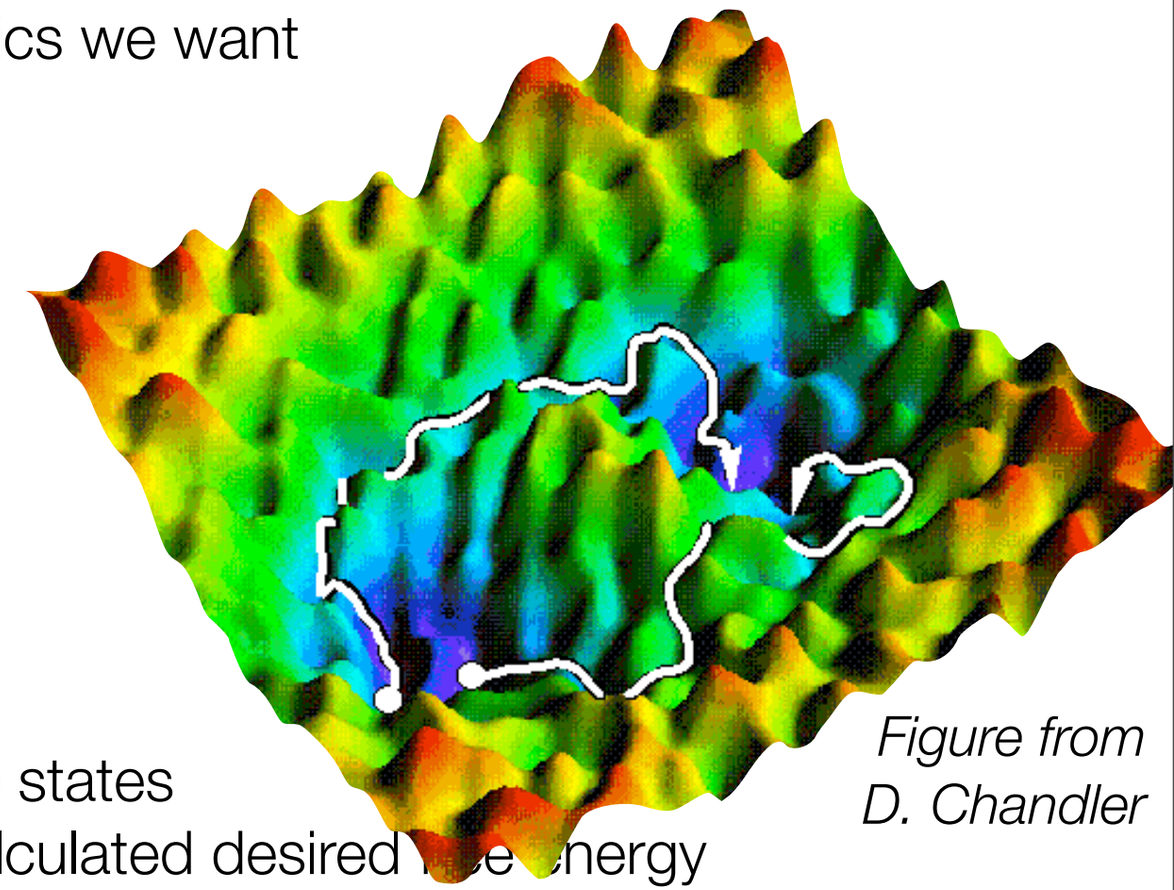


Figure from D. Chandler

- **Game plan**
  - identify what are the factors limiting kinetics (high energy barriers? diffusion?)
  - pick states that drive against these factors
  - Define a new, reduced potential along these states
  - recover original Boltzmann weightings to calculated desired free energy

  (this process will have similarities to the BAR step mentioned in the previous cases -- the question will be given a set of data, what's the best prediction of free energies and Boltzmann weights)

The reduced potential

We define the reduced potential for a state k as a combination of terms

$$u_k(\mathbf{x}) = \beta_k[U_k(\mathbf{x}) + p_k V(\mathbf{x}) + \mu_k^{\mathrm{T}} \mathbf{N}(\mathbf{x})]$$

with thermodynamic parameters for each state

β_k	inverse temperature
U_k	potential energy function
p_k	external pressure
μ_k	chemical potential of exchangeable species

Lyubartsev et al. New approach to Monte Carlo calculations of the free energy: Method of expanded ensembles. JCP 96:1776, 1992.

where

$\mathbf{x}$	microstate or configuration
$V(\mathbf{x})$	volume of simulation box
$\mathbf{N}(\mathbf{x})$	number of each chemical species in system

The distribution function is given by

$$p_k(\mathbf{x}) = Z_k^{-1} \exp[-u_k(\mathbf{x})] \qquad Z_k = \int d\mathbf{x} \exp[-u_k(\mathbf{x})]$$

Covers many common thermodynamic ensembles: NVT, NPT, μVT, μPT

The method of expanded ensembles

Form an expanded ensemble by allowing transitions between thermodynamic states:

$$p(\mathbf{x}, k) = Z^{-1} \exp[-u_k(\mathbf{x}) + g_k]$$

with partition function

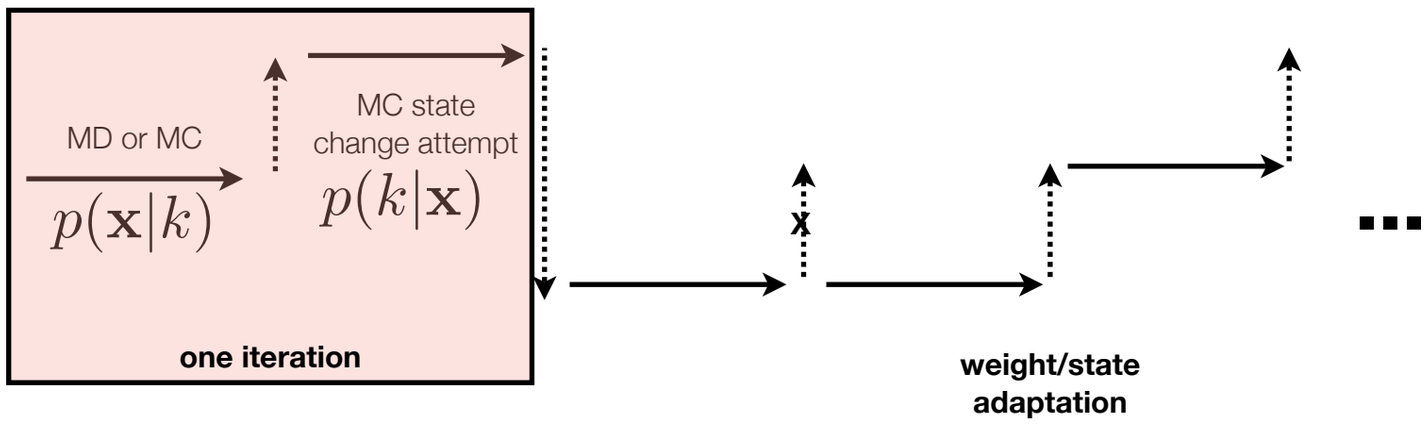
$$Z = \sum_{k=1}^{K} Z_k \exp[g_k]$$

Marinari and Parisi. Europhys. Lett. 19:451, 1992
Mitsutake and Okamoto. Chem. Phys. Lett. 332:131, 2000.
Lyubartsev et al. JCP 96:1776, 1992.

where we have introduced log weights g_k to bias sampling of states.

Current configuration now consists of ($\boldsymbol{x}$, k) pair.

How do we conduct the simulations?



MD or MC moves can be used, or HMC if exact sampling is required.

Sampling could include grand-canonical moves for constant pH (and/or salt concentration).

Multiple ways to conduct MC state change move.

How do we choose states?

- **High temperature barriers?**
  - use high temperature replicas to overcome energy barriers: choose states as different temperatures ("simulated tempering")
  - use umbrella sampling to drive the system throughout configuration space: choose states to be different anchor points for umbrella sampling calculations

- **Alchemical transformations**
  - Goal: calculate free energy difference between two Hamiltonians, via a scaling factor $\mathcal{H}(\lambda)$ $= \lambda \mathcal{H}_1 + (1-\lambda) \mathcal{H}_0$

- **Convenient side effect**
  - in many cases, we actually want the free energy as a function of the state, and this is obtained directly from the GE weights

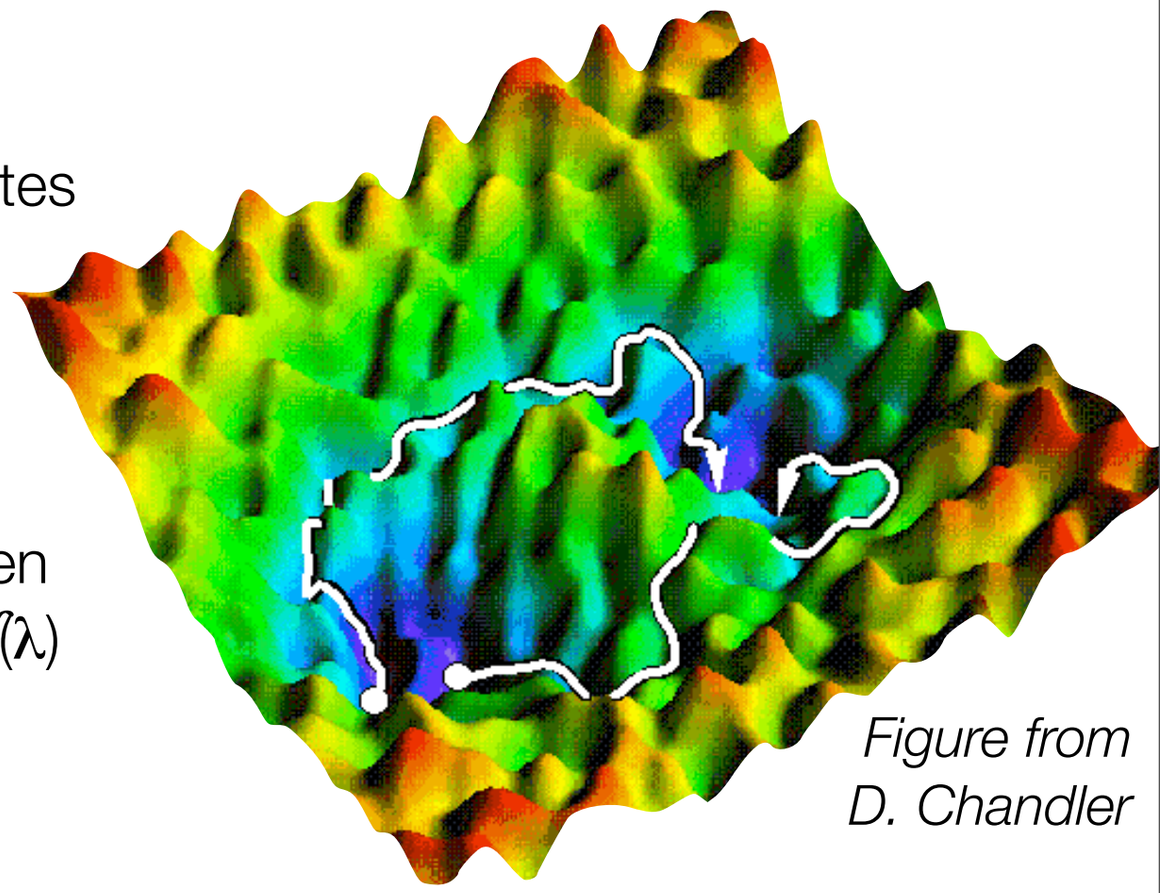


Figure from D. Chandler

How do we choose the weights?

1. Start with a good initial guess

Estimates of $<u_k>_k$ from short simulations of each state can provide an excellent guess.

Sanghyun Park and Vijay S. Pande. Choosing weights for simulated tempering. PRE 76:016703, 2007.

Even initial energies can provide a good initial guess.

2. Several options for automatic updating

(nonequilibrium)

Wang-Landau method

Wang and Landau PRE 65:056101, 2001. DP Landau et al. Am. J. Phys. 72:1294, 2004. Comm Phys Comm 175:36, 2006.
Wei Yang et al. JCP 126:024106, 2007.

(equilibrium)

Bennett acceptance ratio (BAR)

Bennett. J. Comput. Phys. 22:245, 1976. Shirts, Bair, Hooker, and Pande. PRL 91:140601, 2003. Shirts and Pande. JCP 122:144107, 2005.

Weighted histogram analysis method (WHAM) or mulitistate BAR (MBAR)

Kumar, Bouzida, Swendsen, Kollman. J. Comput. Chem. 13:1011, 1992.
Shirts and Chodera. Statistically optimal samples from multiple equilibrium states. *Submitted*, 2007.

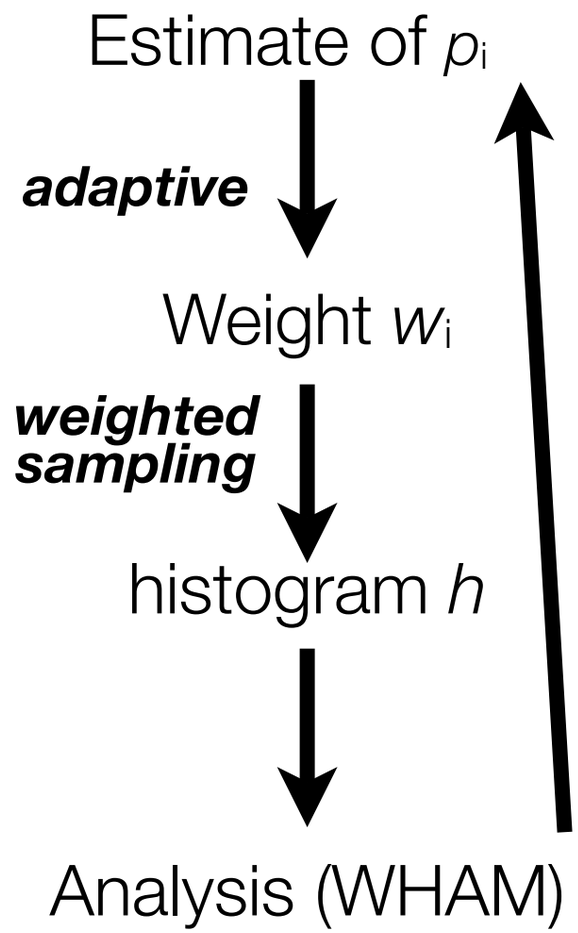
Adaptive Bayesian WHAM (ABWHAM)

Sanghyun Park, Daniel L. Ensign, and Vijay S. Pande. Bayesian update method for adaptive weighted sampling. PRE 74:066703, 2006.

All are extremely simple to implement!

Adaptive Bayesian WHAM

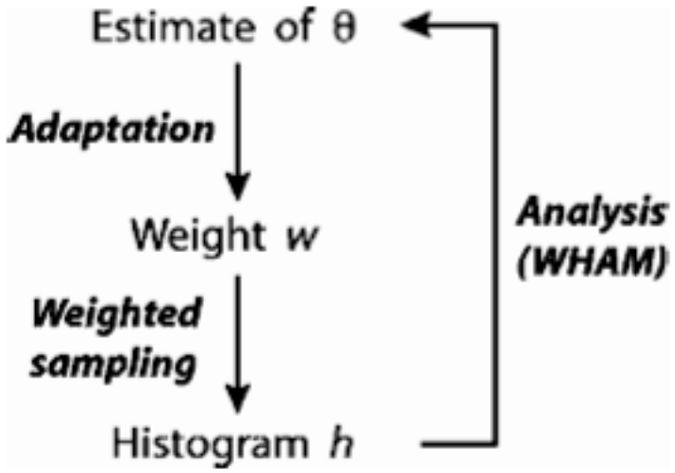
- **Setup**
  - Consider a system that can be in K different states, and let p_i be the probability for the i-th state
  - we want to estimate the parameters p_i by means of weighted sampling
  - We seek an adaptive weighted sampling scheme as outlined on the right
  - Based on the estimates $p_i^{(n-1)}$ from the previous iteration step, new weights $w_i^{(n)}$ are determined in a way that leads to efficient sampling of states

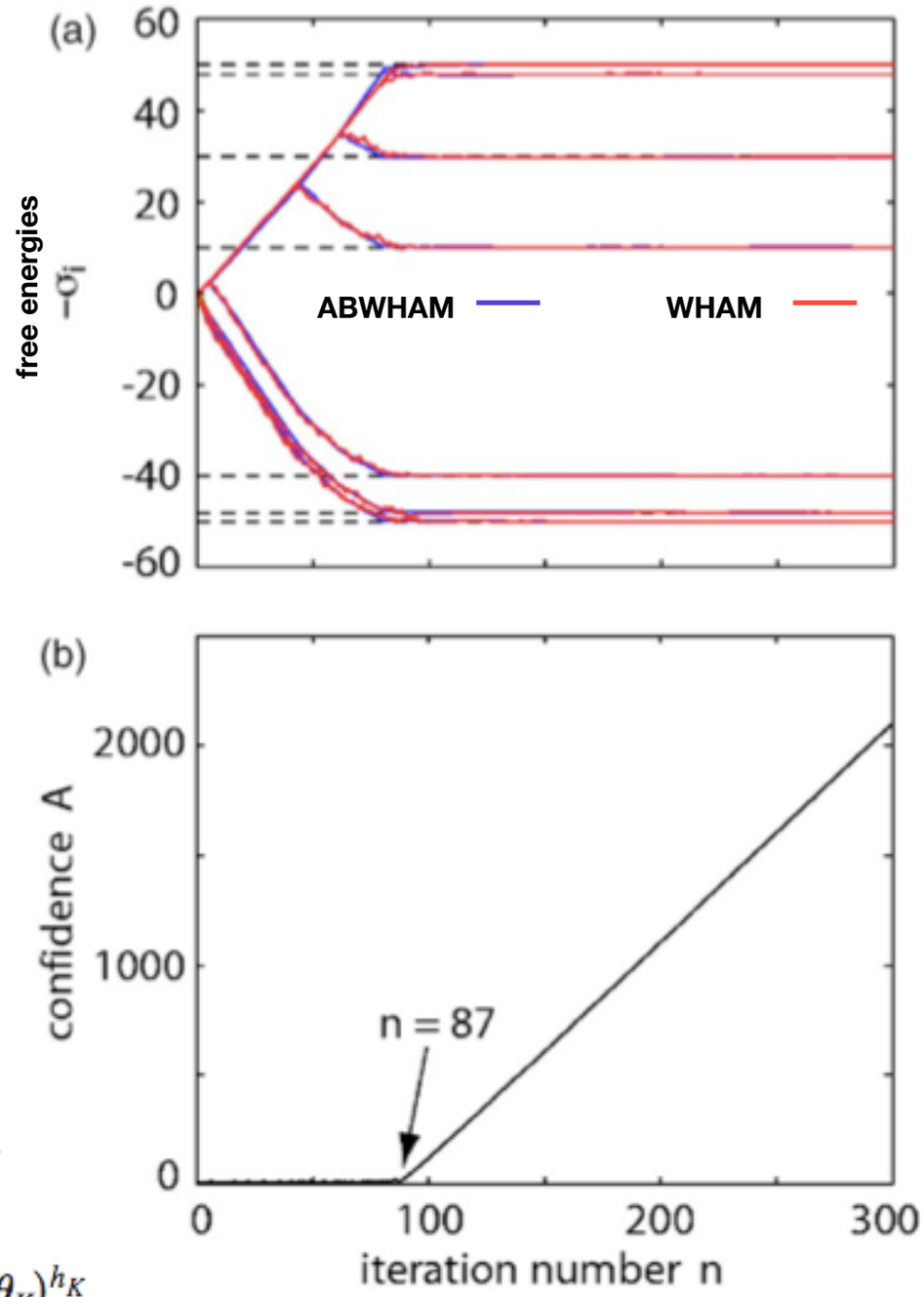


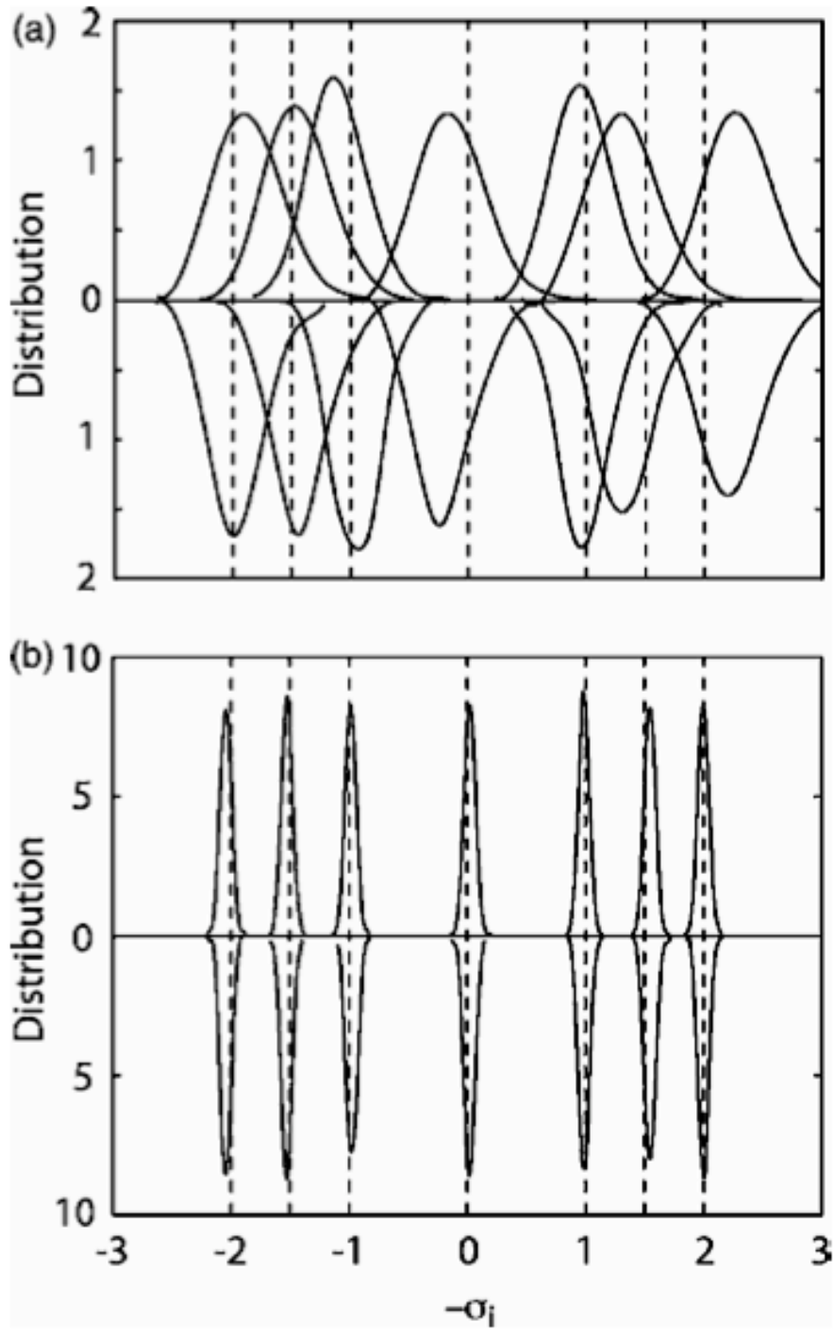
- **Adaptive scheme**
  - Therefore, we attempt to develop a method in which only new data are needed for the update of estimates
  - We want to determine a new estimate $p_i^{(n)}$ from the knowledge of the new histogram $h^{(n)}$, the new weight $w_i^{(n)}$, and the previous estimate $p_i^{(n-1)}$

Sanghyun Park, Daniel L. Ensign, and Vijay S. Pande. Bayesian update method for adaptive weighted sampling. *PRE* **74**:066703, 2006.

ABWHAM converges quickly







$$f^{(n)}(\theta) \propto P(h^{(n)}|\theta, w^{(n)}) f^{(n-1)}(\theta)$$

$$P(h|\theta, w) = \frac{H!}{h_1 ! \cdots h_K!} \phi_1^{h_1} \cdots \phi_K^{h_K}$$

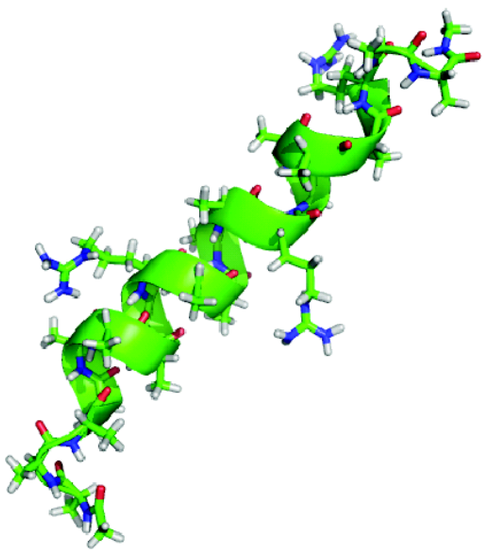
$$P(h|\theta, w) = \frac{H!}{h_1 ! \cdots h_K!} \frac{(w_1\theta_1)^{h_1} \cdots (w_K\theta_K)^{h_K}}{(w_1\theta_1 + \cdots + w_K\theta_K)^H}$$

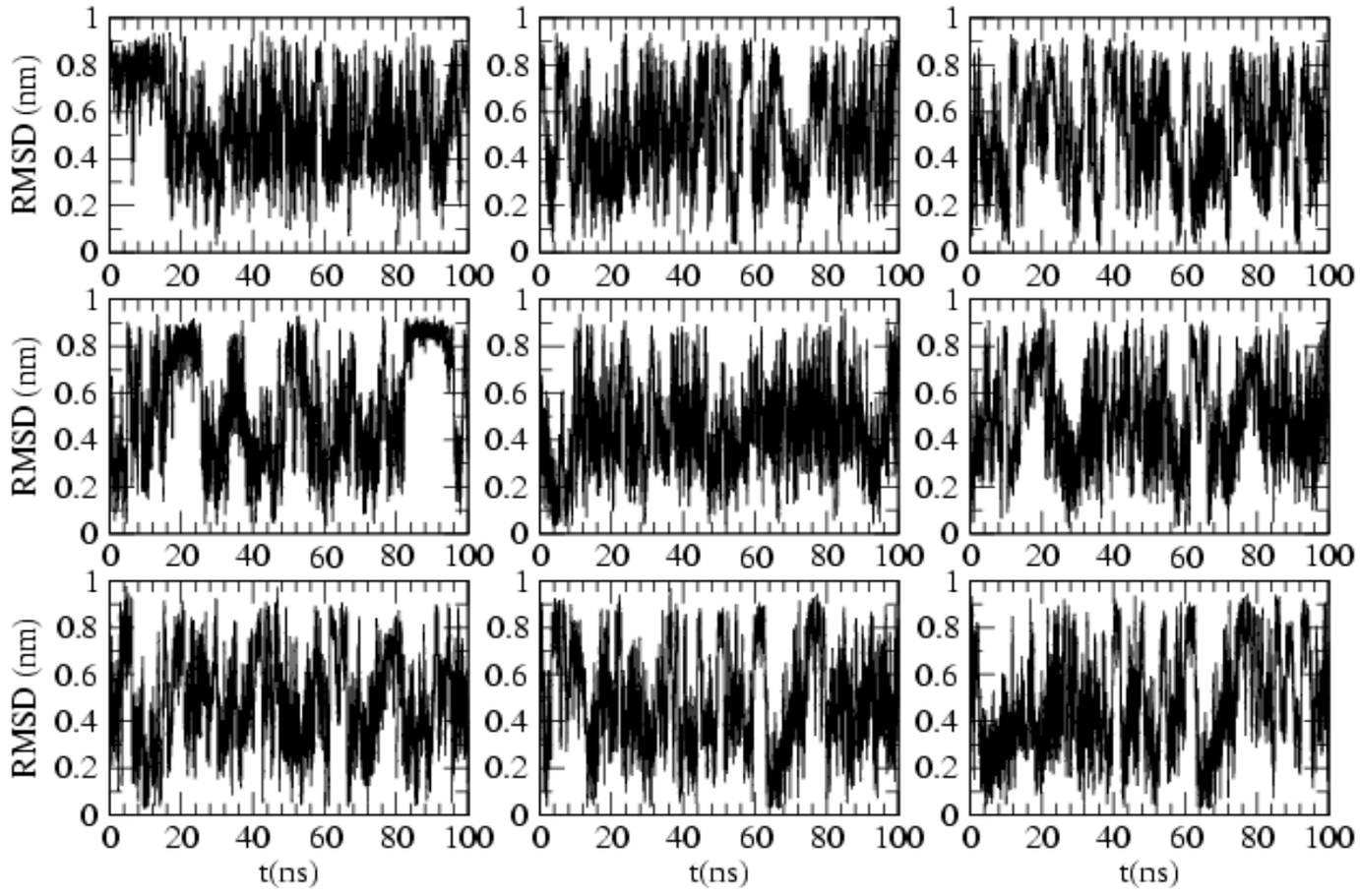
Sanghyun Park, Daniel L. Ensign, and Vijay S. Pande. Bayesian update method for adaptive weighted sampling. PRE 74:066703, 2006.

Application: Fs-Peptide

(Huang & Bowman)

C-α RMSD from ideal helix for a few representative replica walkers in SREM simulations.



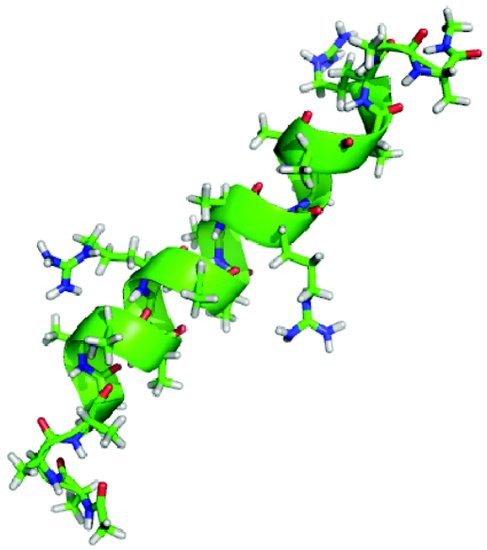


A few transitions between folded and extended states for each replica are observed, indicating there is **Reversible Folding**.

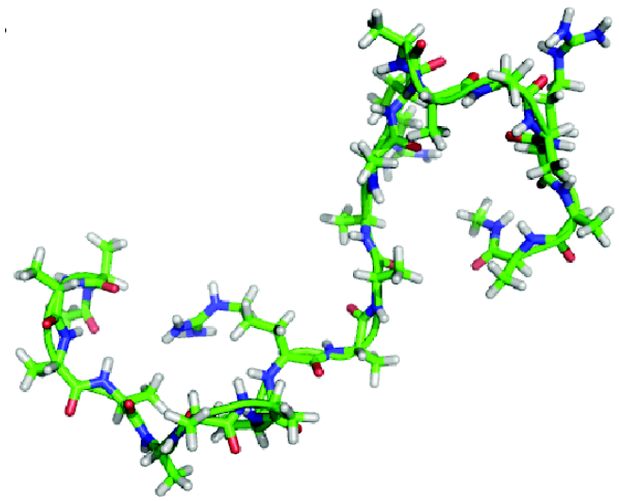
Application: Fs-Peptide

(Huang & Bowman)

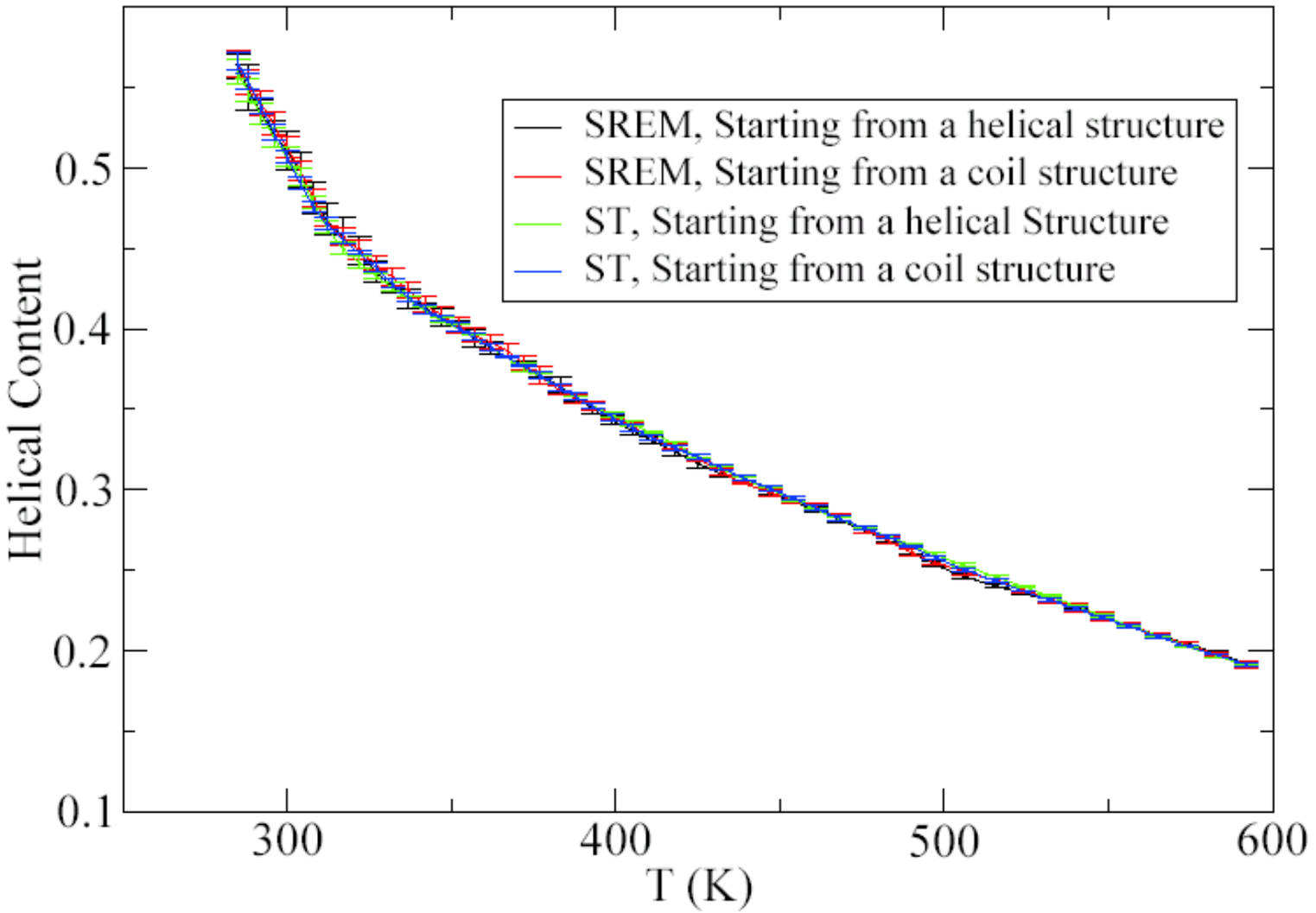
Folded Initial Structure



Extended Initial Structure



Helical Content v.s. T



WHAM is used.

J.D. Chodera, Swope W.C., Pitera J.W., Seok Chaok, K.A. Dill. ***JCTC***, 3, 26--41 (2007)

Case study: small molecule drug design

http://pande.stanford.edu

(Shirts, Bair, Hooker)

Efficient free energy calculation: use forward *and backward* work distributions

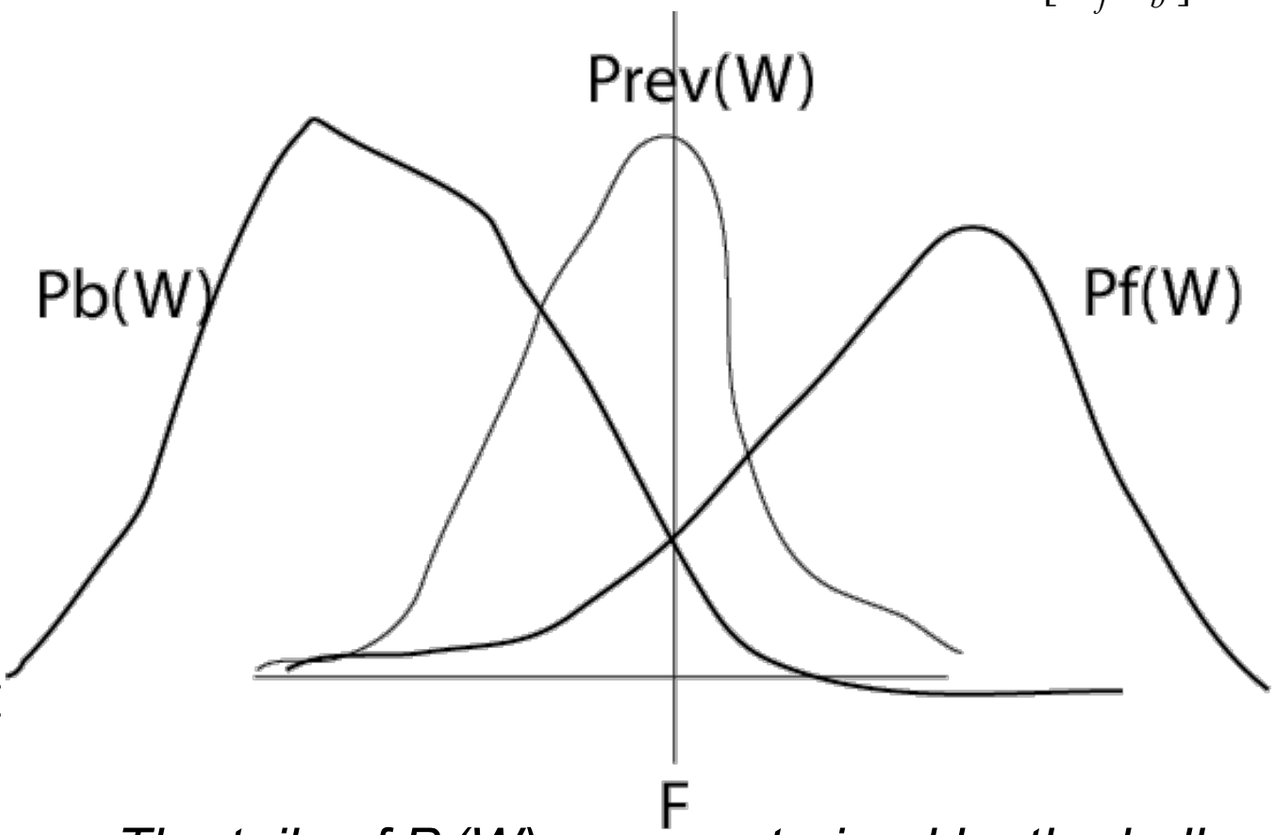
Generalization of Bennett Acceptance Ratio (BAR) Method (Shirts, et al, PRL, 2004)

$$\left\langle \frac{1}{1+\exp[\beta(M+W-\Delta G)]} \right\rangle_f = \left\langle \frac{1}{1+\exp[\beta(M+W-\Delta G)]} \right\rangle_b$$

$M \equiv kT \ln[\, N_f/N_b \,]$

- Plan
  - find $P_f(W)$ and $P_b(W)$
  - average in a new way
  - Find ΔG as the balancing point

- Benefit
  - two distributions are statistically linked
  - use one distribution to help flesh out the tails of the other



The tails of $P_f(W)$ are constrained by the bulk of $P_b(W)$ and vice versa

(Shirts, Bair, Hooker)

How does this work?

1. Start from Crooks' fluctuation theorem

$$\ln\left[\frac{P_F(W)}{P_R(-W)}\right] = \beta(W - \Delta F),$$

2. Use Bayesian method + normalization

$$\frac{P(W|F)}{P(W|R)} = \frac{P(F|W)P(R)}{P(R|W)P(F)} = \frac{P(F|W)}{1-P(F|W)}\frac{P(R)}{P(F)}. \qquad P(F|W) + P(R|W) = 1$$

$$\ln\frac{P(F|W)}{1-P(F|W)} = \beta(M + W - \Delta F) \qquad M \equiv kT \ln[\, P(F|W) / P(R|W) \,]$$

3. This leads to the probabilities

$$P(F|W_i) = \frac{1}{1+\exp[-\beta(M + W_i - \Delta F)]} \qquad P(R|W_i) = \frac{1}{1+\exp[\beta(M + W_i - \Delta F)]}$$

These probabilities hold for all distributions: not parametric!

(Shirts, et al, PRL, 2004)

How does this work?

(Shirts, Bair, Hooker)

4. Define the likelihood

$$L(\Delta F) = \prod_{i=1}^{n_F} P(F|W_i) \prod_{j=1}^{n_R} P(R|W_j)$$

(Shirts, et al, PRL, 2004)

5. Find maximum likelihood

$$\frac{\partial \ln L(\Delta F)}{\partial \Delta F} = \sum_{i=1}^{n_F} \frac{1}{1+\exp[\beta(M+W_i-\Delta F)]} - \sum_{j=1}^{n_R} \frac{1}{1+\exp[-\beta(M+W_j-\Delta F)]} = 0$$

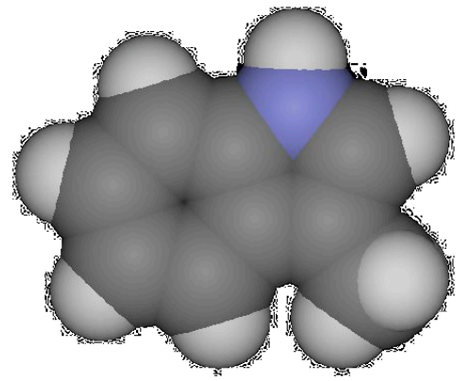
6. Result: new way to average

$$\left\langle \frac{1}{1+\exp[\beta(M+W-\Delta F)]} \right\rangle_f = \left\langle \frac{1}{1+\exp[\beta(M+W-\Delta F)]} \right\rangle_b$$

7. Find the value of ΔF which satisfies the above

Application: solvation free energies

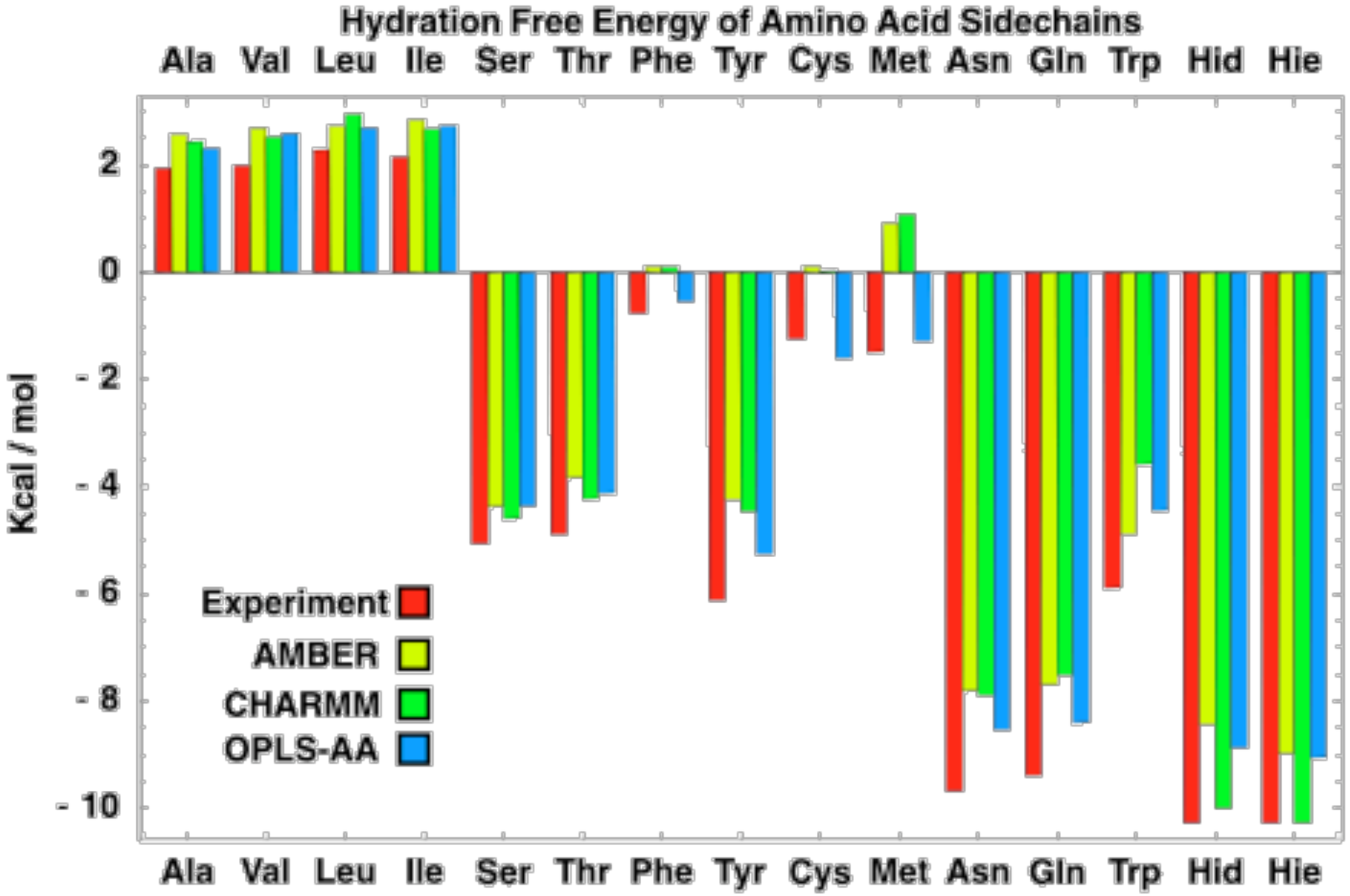
- In our hands, BAR is most efficient
  - FEP is a limiting case of BAR
  - BAR appears to be more efficient than TI too



- Example test:
  - 3-methlyindole (Trp sidechain analog)
  - 1.0 ns at each intermediate
  - **We see BAR yields more precise answers for the same CPU time**

Method	# Intermediates	Value (kcal/mol)	Precision (kcal/mol)
TI	61	3.69	0.05
TI	8	4.41	0.21
BAR	8	3.68	0.05
FEP (0->A)	8	3.43	0.19
FEP (A->0)	8	6.01	0.43
FEP Average	8	4.72	0.24

Comparison with experiment



- RMS deviations from experiment (kcal/mol):

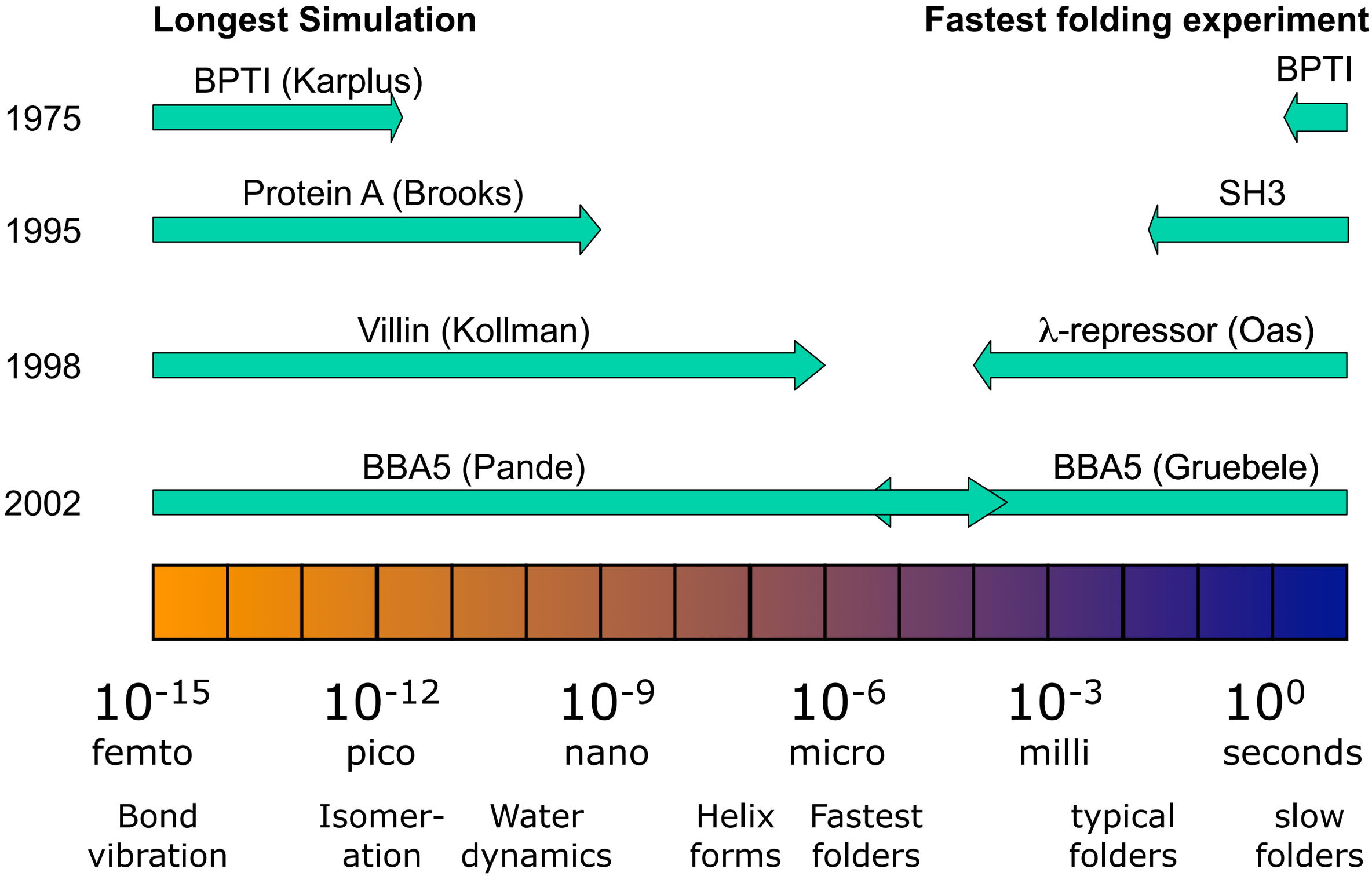
  AMBER 1.35 CHARMM 1.31 OPLS-AA 0.85

29

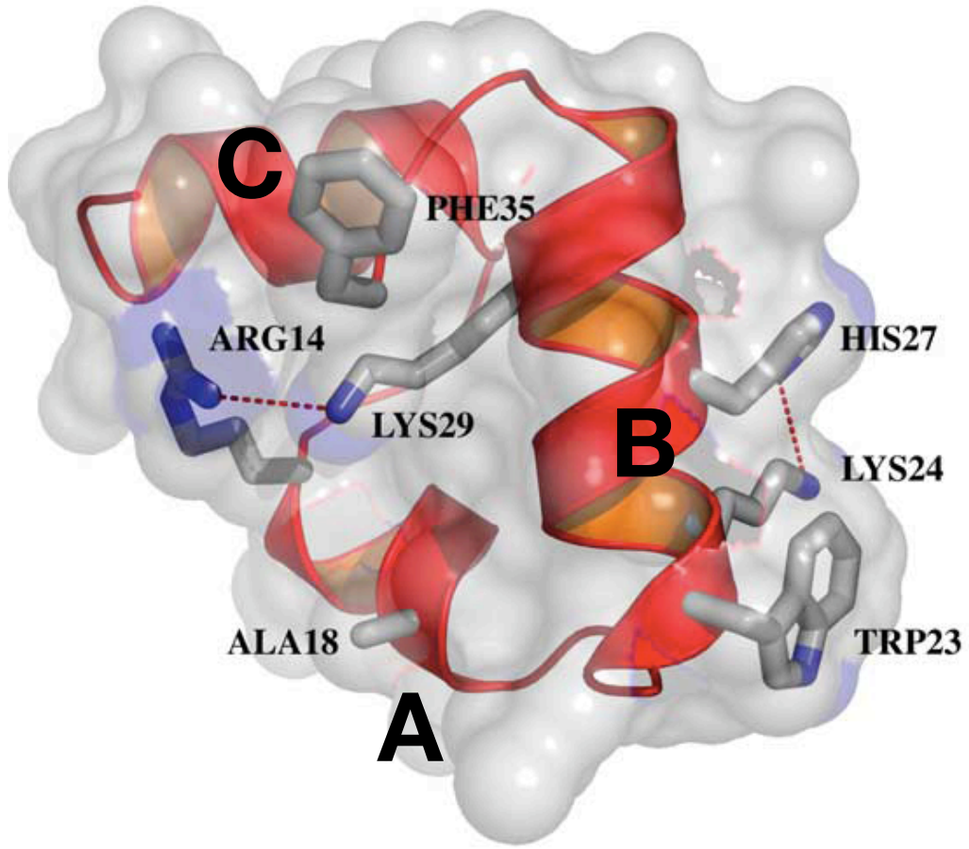
Case study: protein folding kinetics

http://pande.stanford.edu

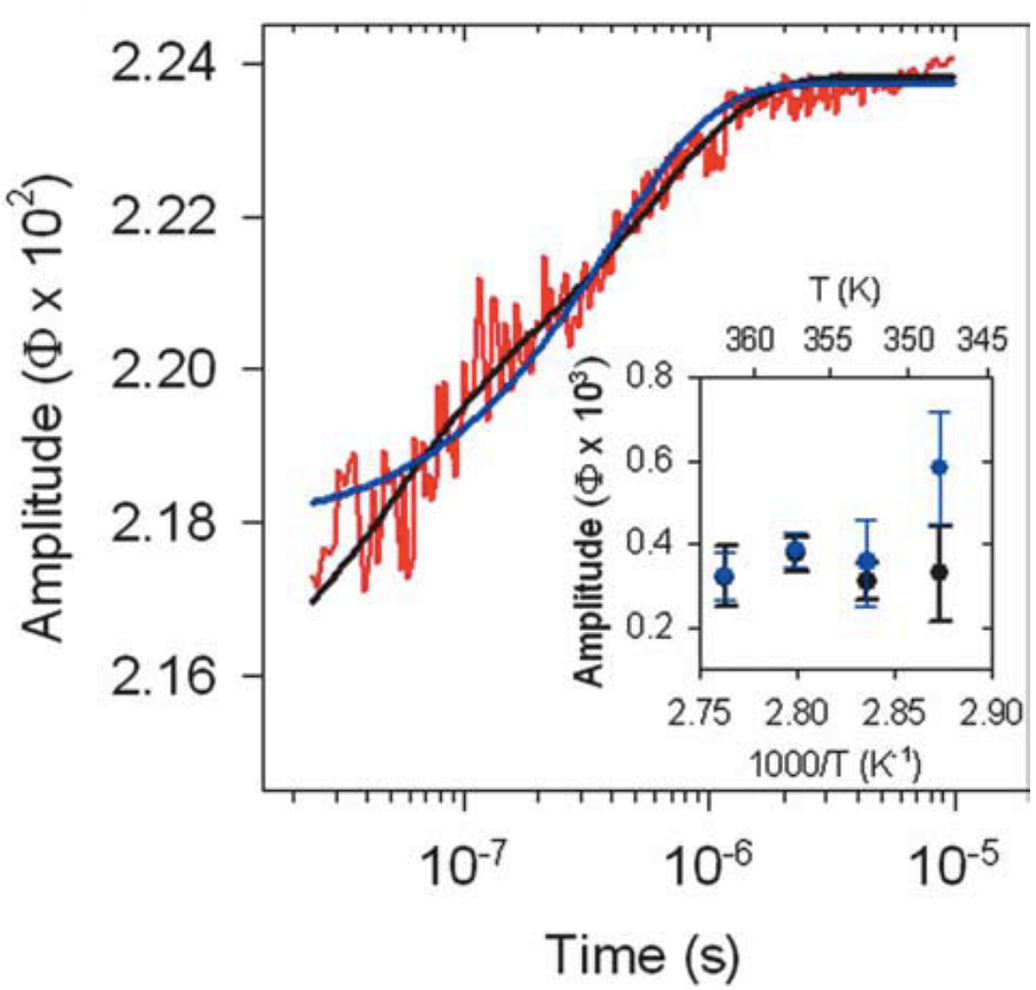
Progress of MD & experiment



A very fast folding protein: k_{fold} ~1/μs



structure



folding kinetics

villin headpiece

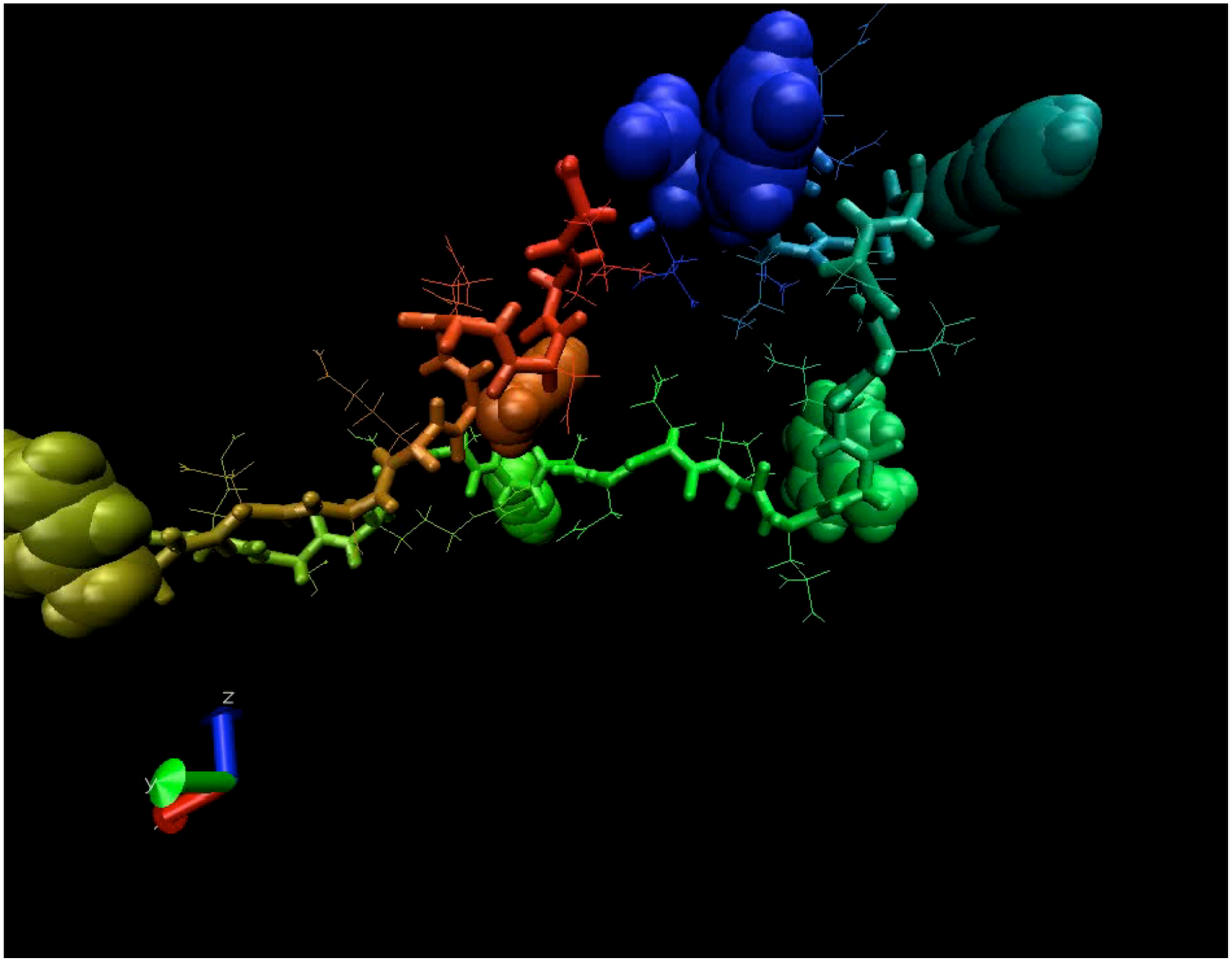
mutant designed by the Eaton Lab

(Kubelka *et al*, *JMB* 2006)

Let's look at a 1μs trajectory for villin: we see stochastic behavior

(Ensign, Kasson)

http://simtk.org



- **Simulation details**
  - villin headpiece (36 residues)
  - Eaton mutant (0.7μs folding time)
  - **explicit solvent**
  - 20,000 atoms total
  - AMBER2003 force field
- **MD Engine**
  - GROMACS 3.3.99 (CVS) code
  - SMP on FAH
- **Visualization (VMD)**
  - spacefill: aromatic resides
  - licorice: backbone
  - rest: sticks
  - color: N-C gradient

One trajectory of thousands, each on the >1 μs timescale

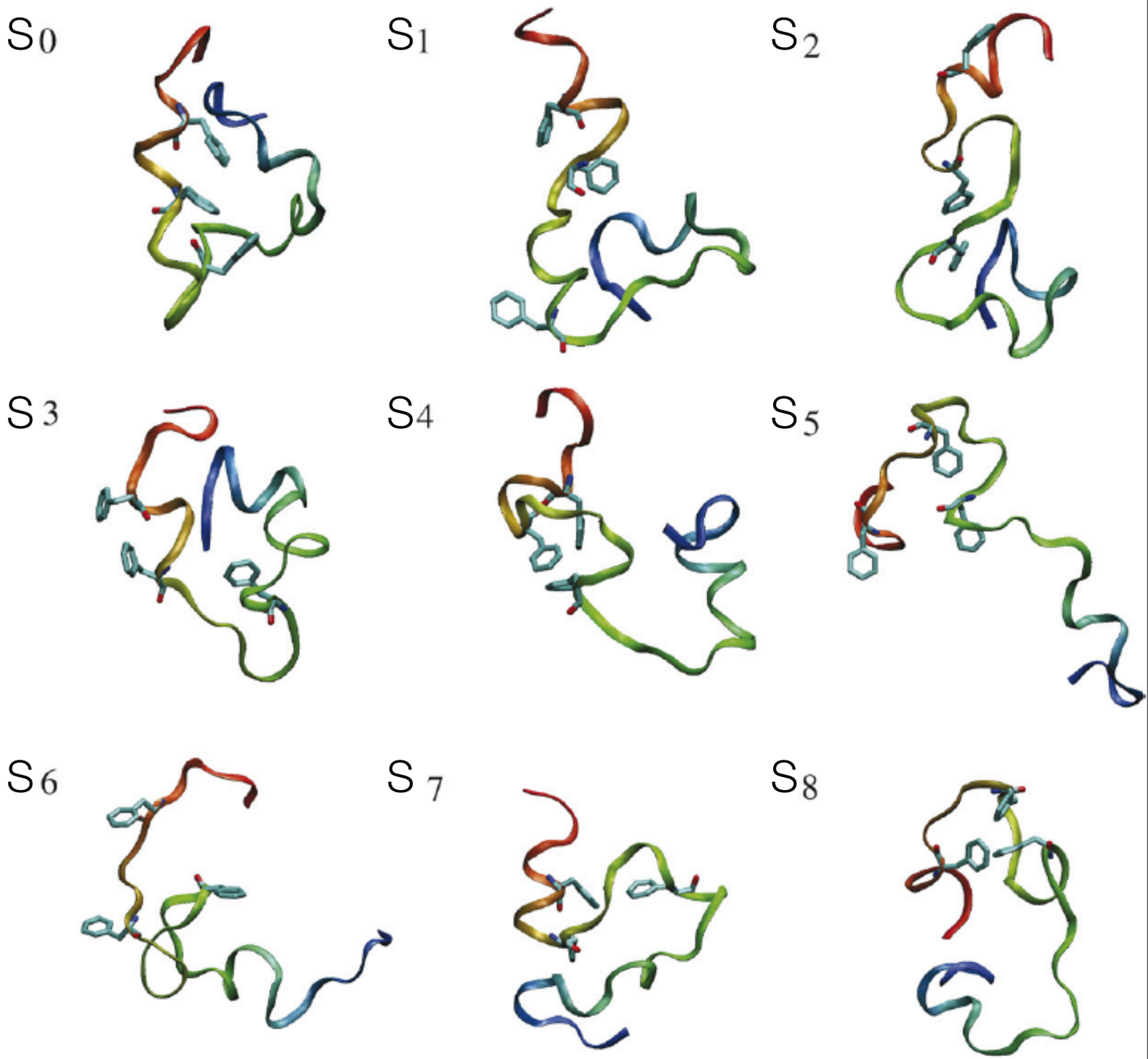
Ensign, Kasson, & Pande. *JMB* (2007)

(Ensign, Kasson)

Looking at ensembles of simulations

- **Starting structures**
  - 9 different structures
  - generated by high temperature unfolding
  - different degrees of native like structure
  - some have helices, other contacts
  - some have no native structure at all

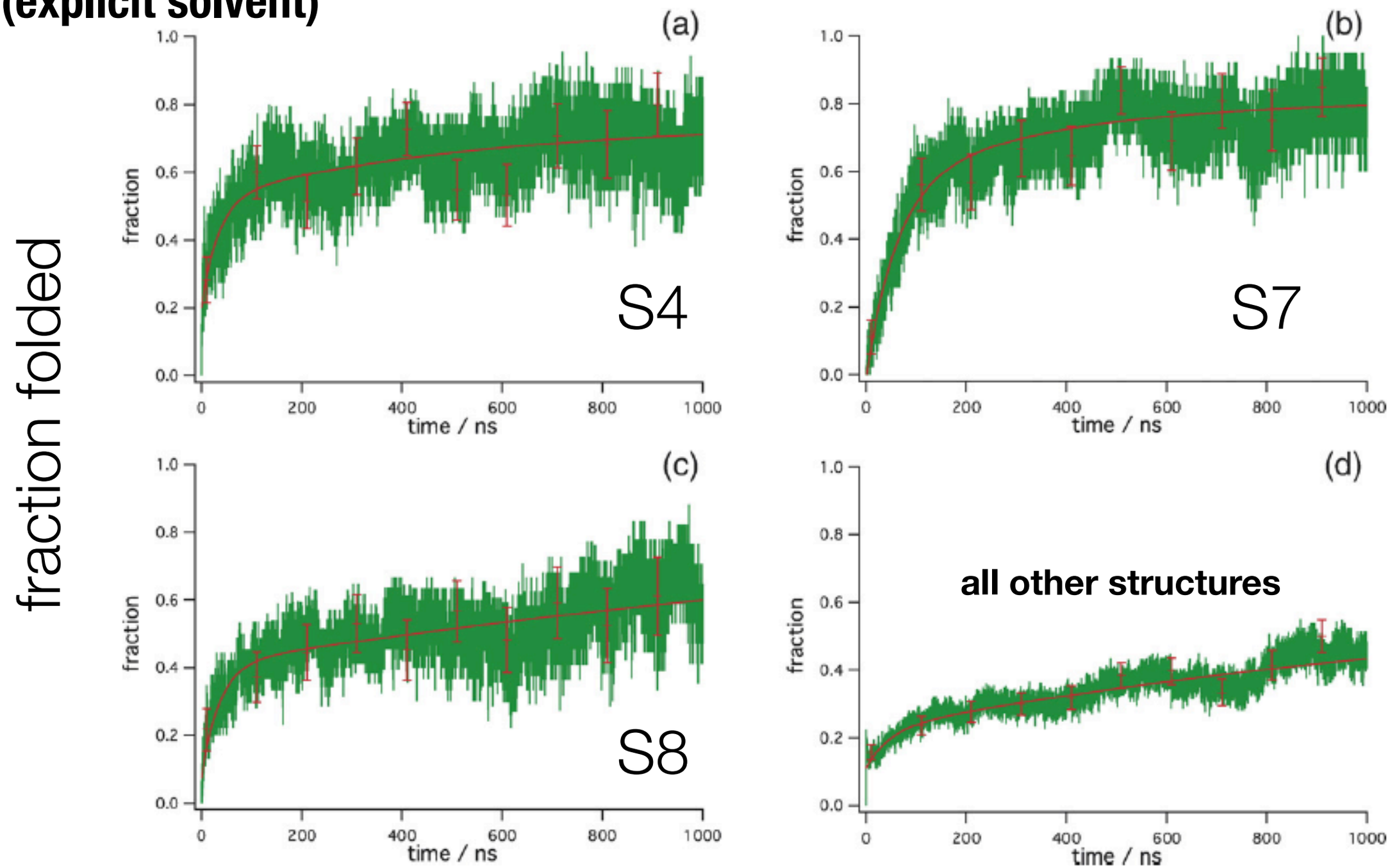
- **Ensemble of trajectories**
  - hundreds to thousands of trajectories per structure
  - each trajectory ~1-2 μs timescale (longer than experimental folding timescale of 0.7μs)



Ensign, Kasson, & Pande. *JMB* (2007)

Ensemble data agrees with experiment

(explicit solvent)



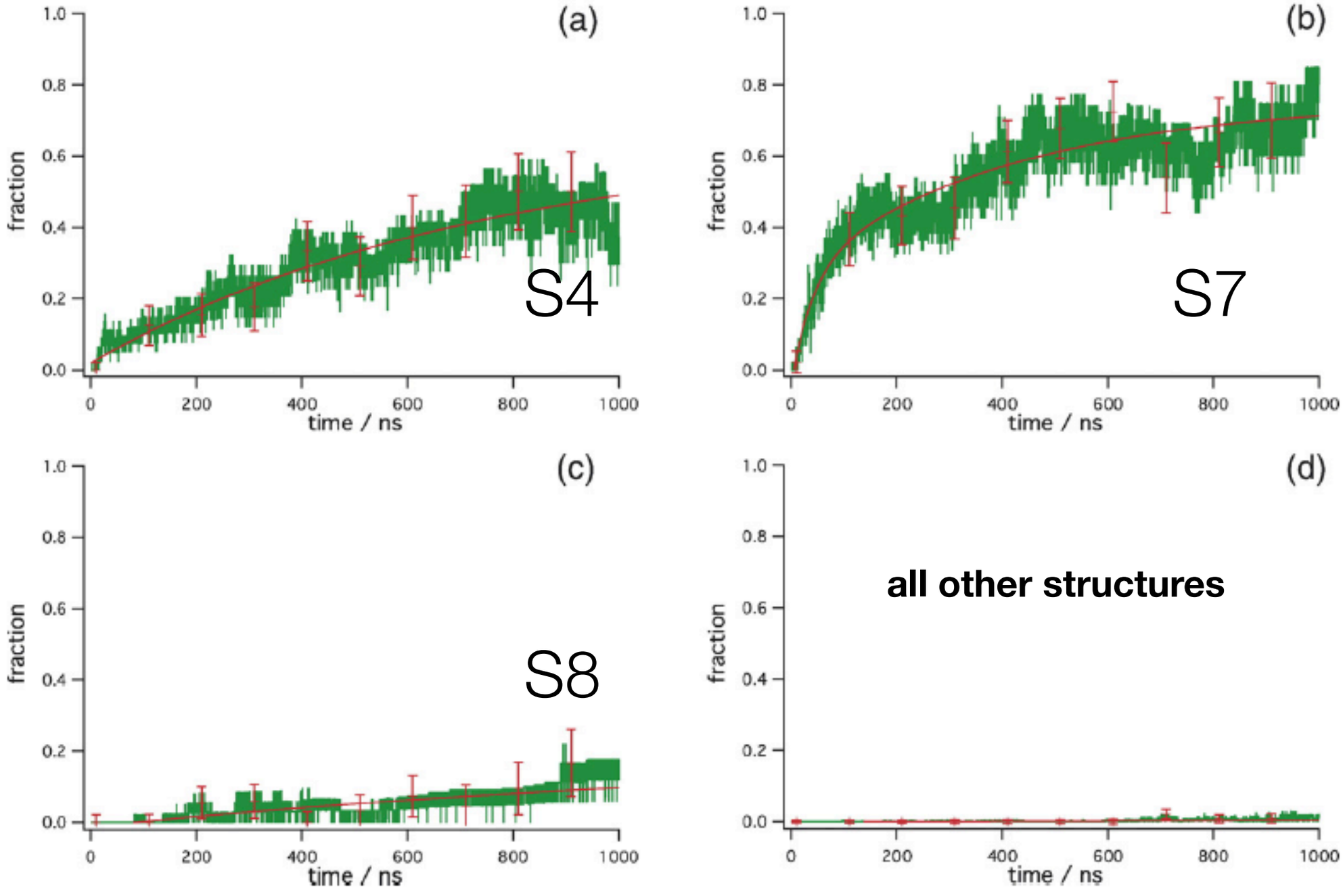
Fraction folded (via Trp-His distance) vs time

Ensign, Kasson, & Pande. *JMB* (2007)

(Ensign, Kasson)

But is the experimental assay looking at folding?

(explicit solvent)

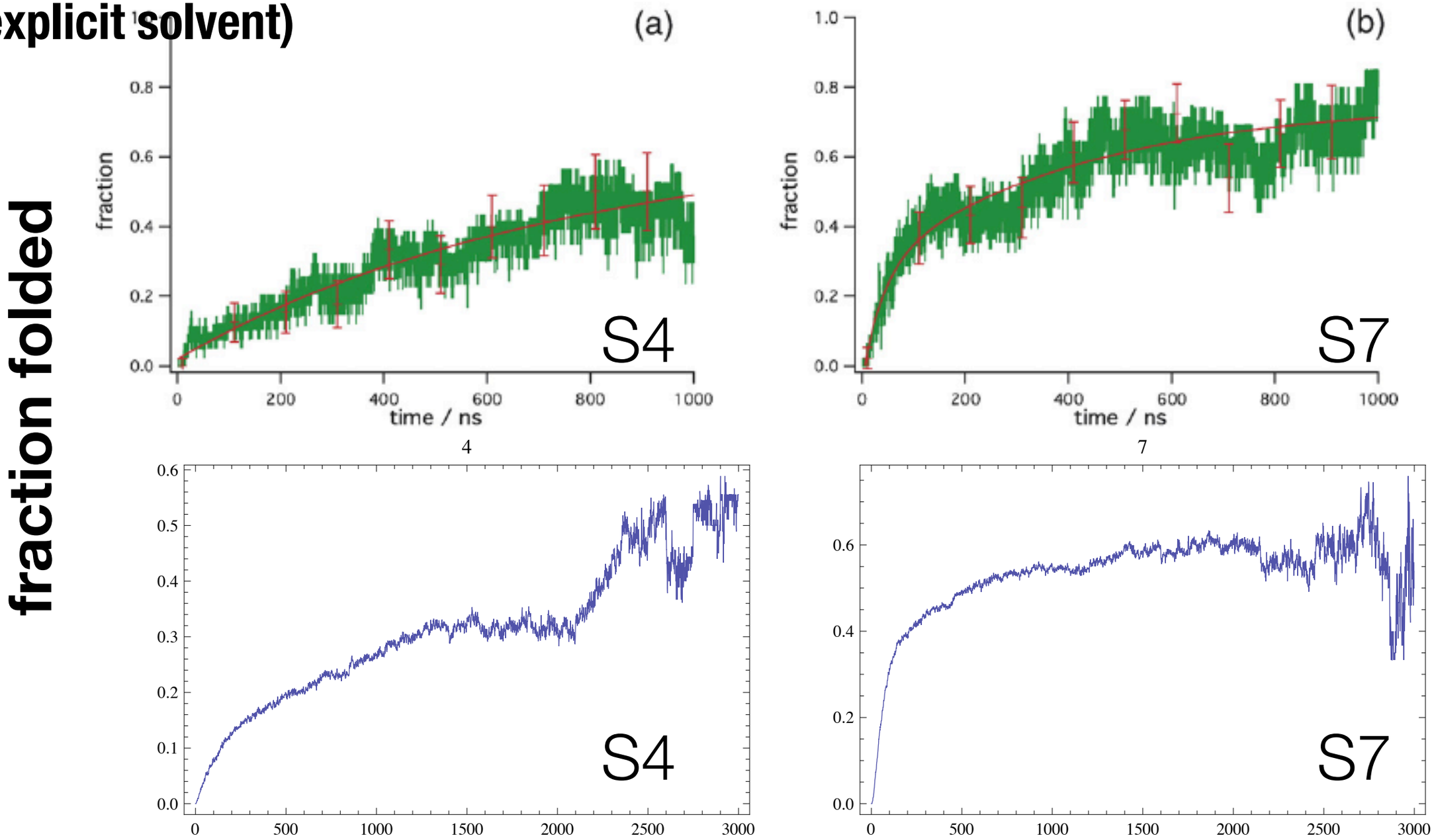


Fraction folded (via comparison to xray structure) vs time

Ensign, Kasson, & Pande. *JMB* (2007)

Comparison between explicit and implicit

(explicit solvent)



(implicit solvent)

Fraction folded (via comparison to xray structure) vs time

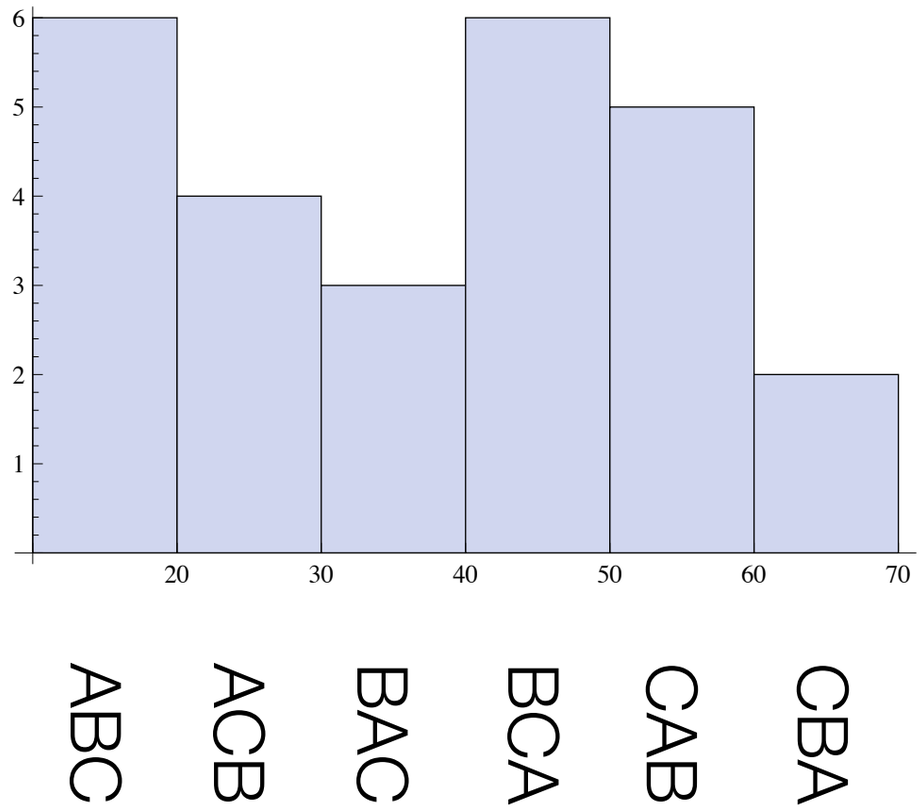
We find a heterogeneous set of folding pathways

- **Do we see a single pathway or many different?**

- **Test this with a simple question: "Is the order of helix formation consistent between simulations?"**
  - for 3 helices (villin), there are 3! = 6 possible orderings
  - histogram shows a very wide variation of pathways seen

- **Other variations possible too**
  - which key core contacts form first?

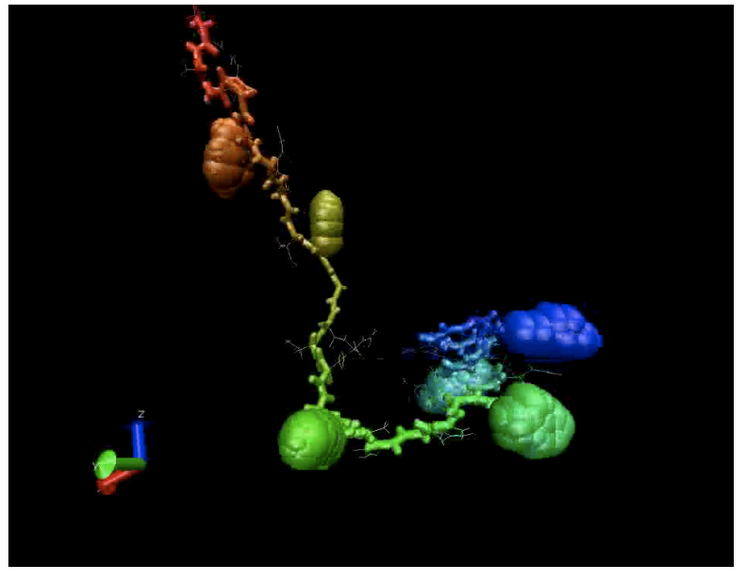
- **A single trajectory (or even a few) would give a misleading picture of the folding dynamics**

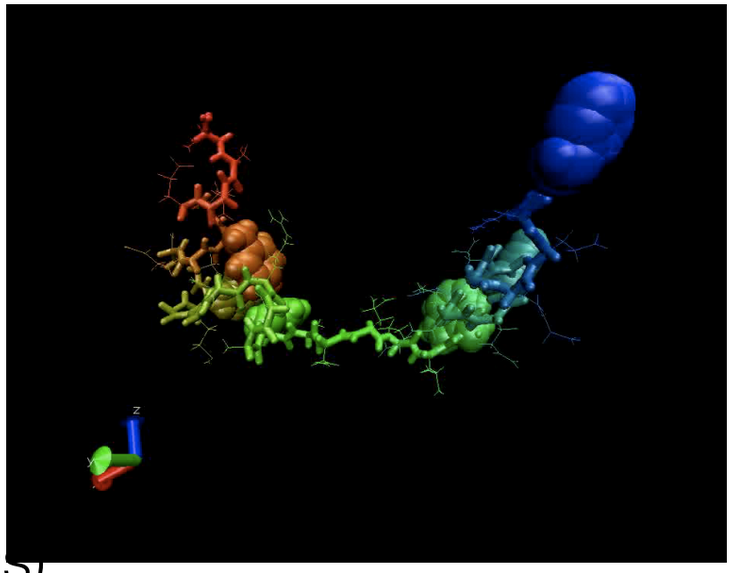


Histogram of folding kinetics: what is the order of formation of each helix A, B, C?

What have we learned about how proteins fold?

- **What did we see in that trajectory?**
  - starts with non-specific hydrophobic collapse
  - unfolds, breaks most contacts
  - refolds, with little native structure
  - some native persist over numerous folding/refolding cycles
  - eventually gets everything right





- **What about other trajectories?**
  - similar behavior in general, but different details
  - great heterogeneity in folding paths

- **General lessons?**
  - Folding is a stochastic process
    (if the folding time is 1ms, then it's not ½ folded at 0.5 ms)
  - Dynamics of even small molecules can be complex & **very heterogeneous**
  - **Even a few long trajectories aren't enough to inform us about the true nature of the complex phase space -- we need a statistical picture**

Case study: long timescale dynamics

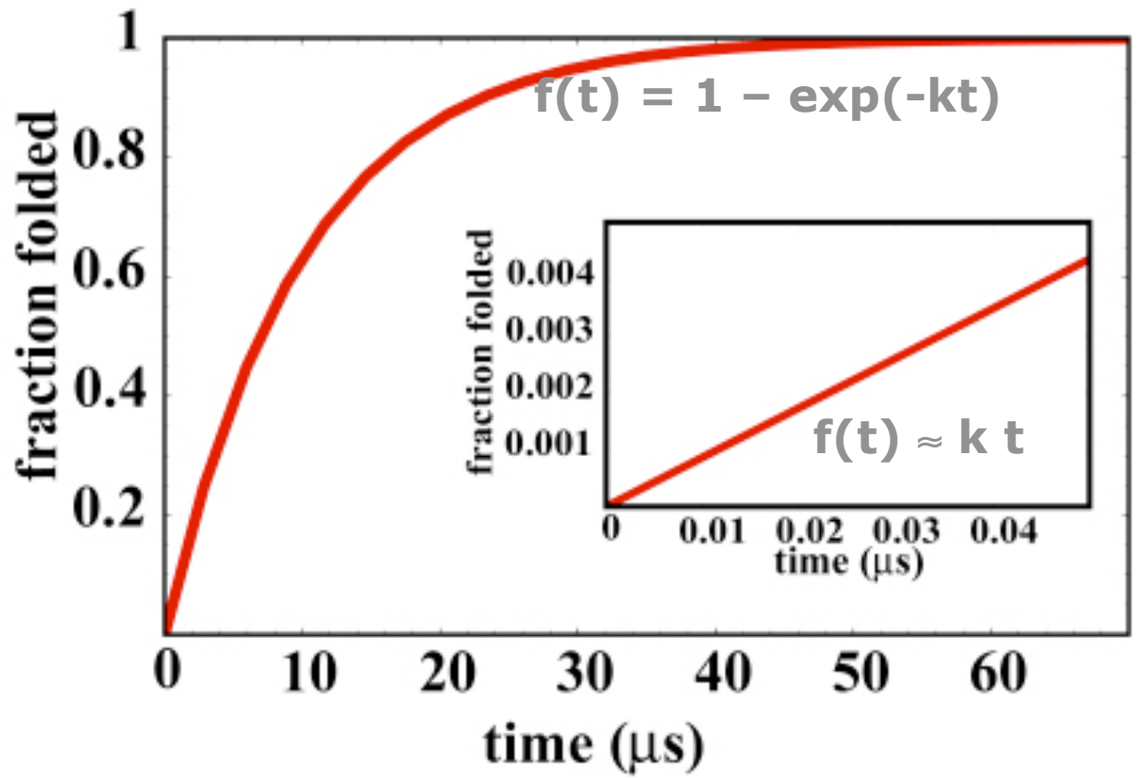
How to overcome long timescales: stochastic kinetic sampling methods

Folding is a stochastic process with exponential kinetics

Fraction that fold:
$f(t) = 1 - \exp(-kt)$

At short times, we get
$f(t) \approx k\,t$

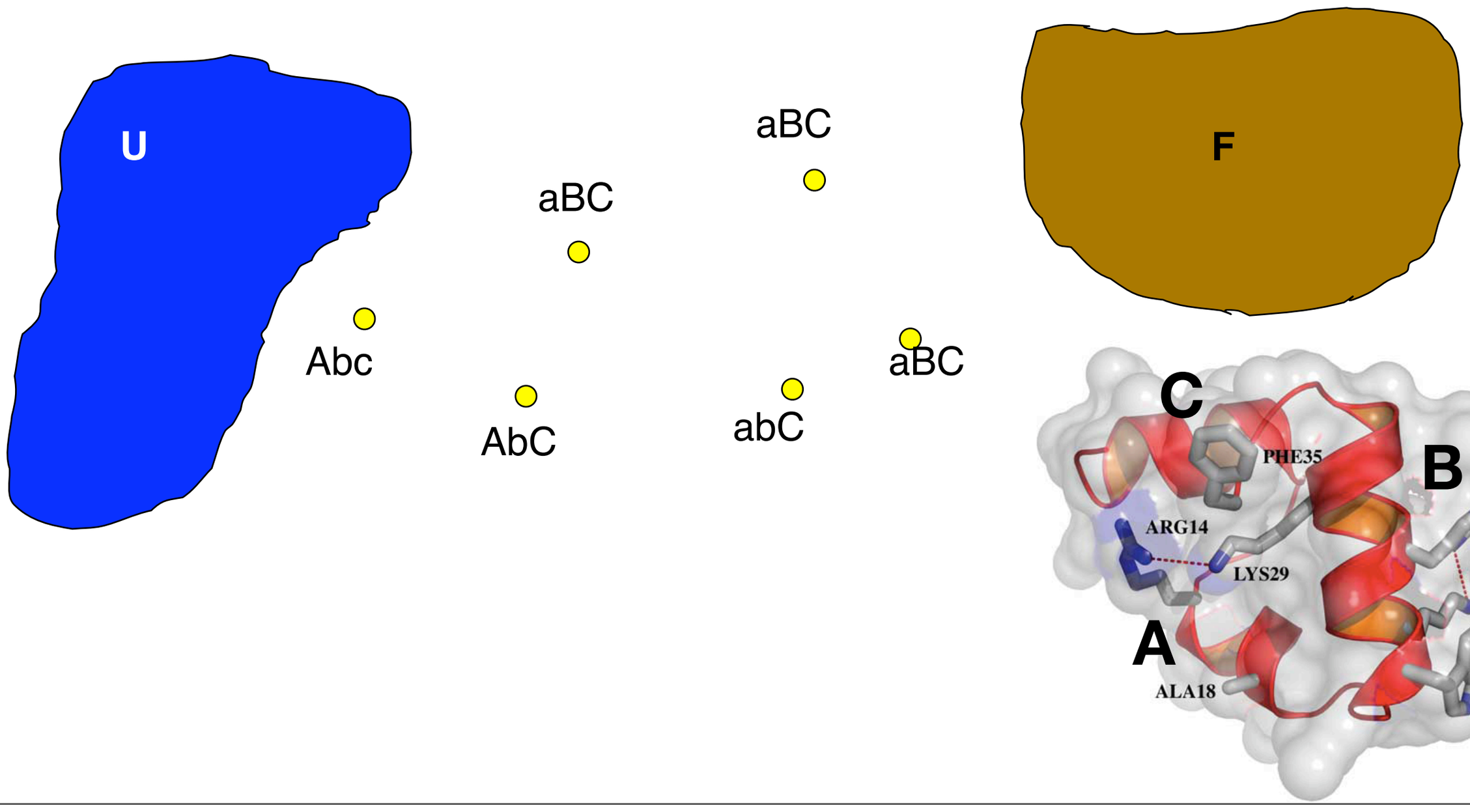
What if we run M Simulations in parallel each of time t?
Mkt will fold



Putting in real numbers: number that fold = *Mkt* =
10,000 simulations x 10,000ns^{-1} x 100ns = 100 events!

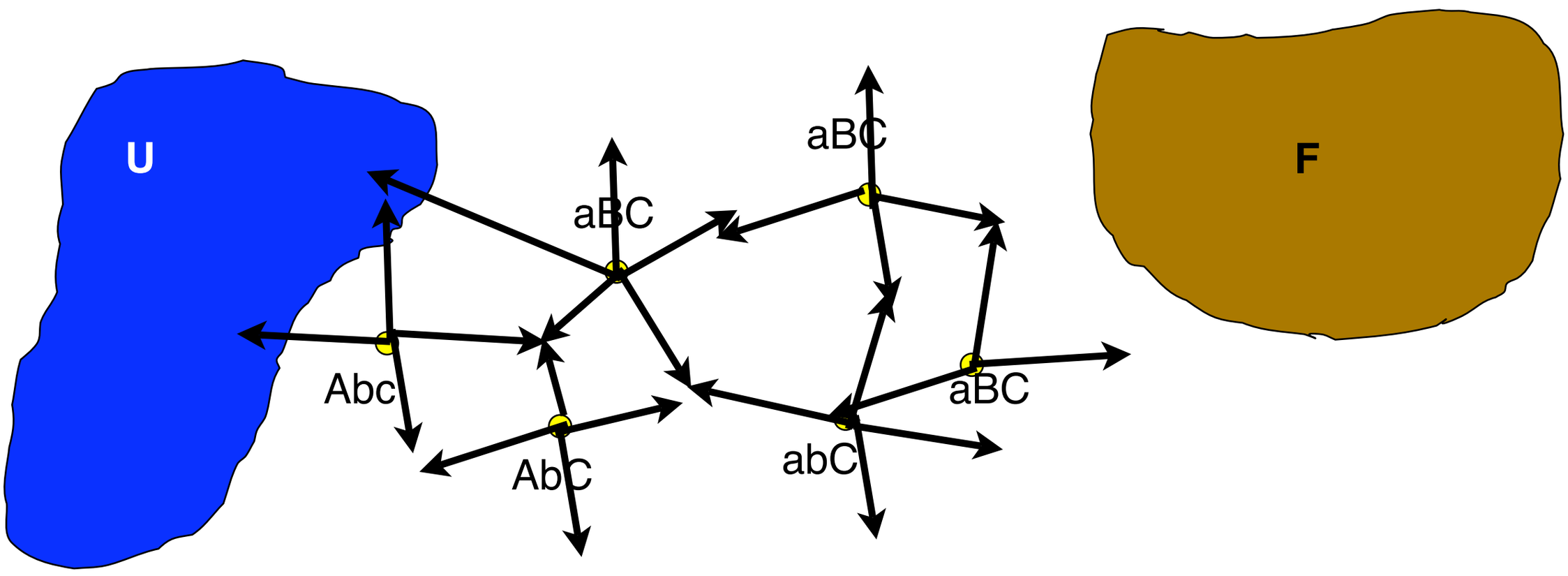
How about a new model based on states & rates

For example, consider villin with 3 helices A, B, & C



Run MD simulations from these initial states

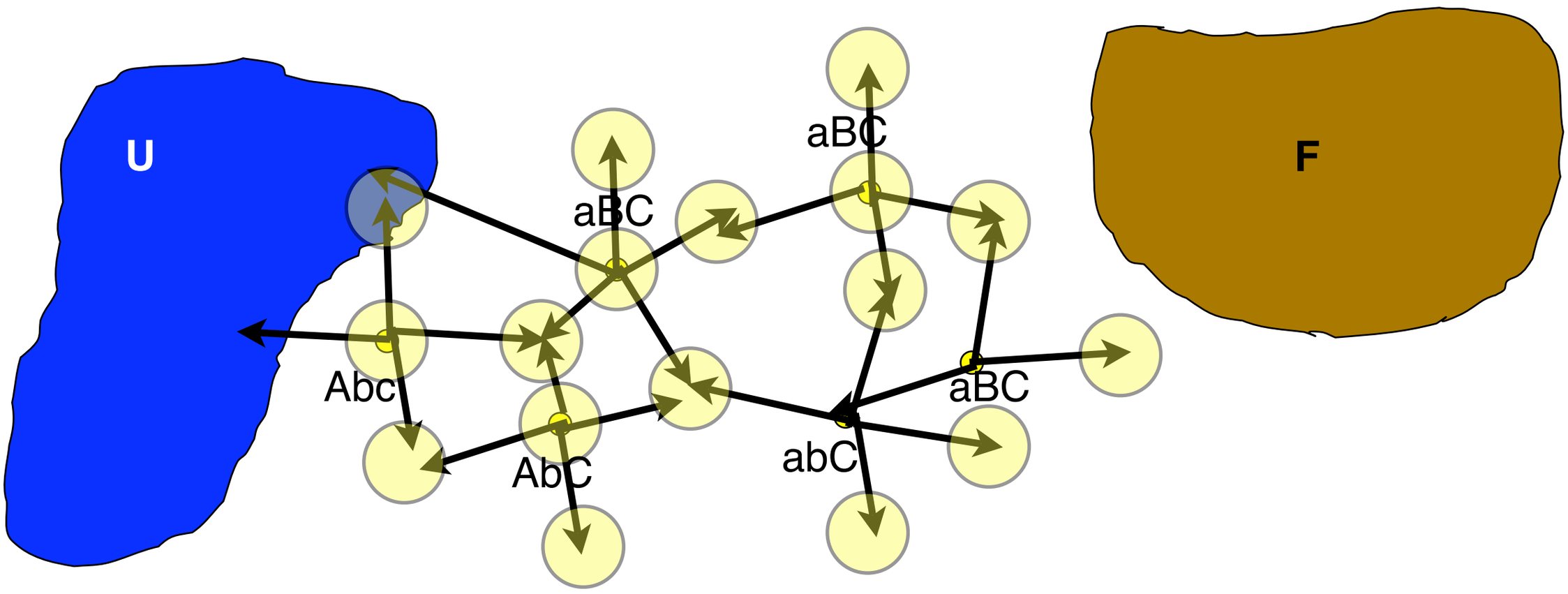
Run ~100 trajectories from each state



(one can use many different means to generate initial conformations for MD; this is just an example)

Find new states and then repeat ...

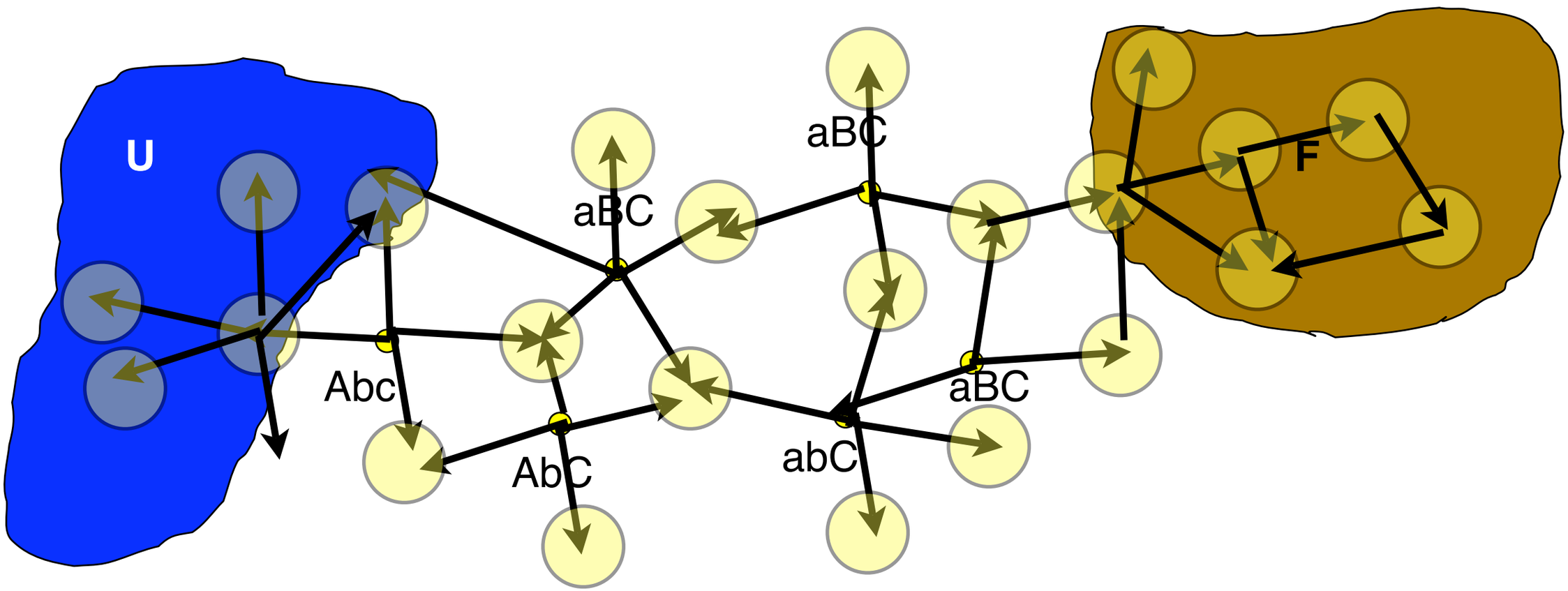
Run a state decomposition algorithm to find new states, and then repeat



***Key concept**: timescales between states (small circles) are **much** faster than between U -> F*

... until convergence

May take several (~5) rounds to converge

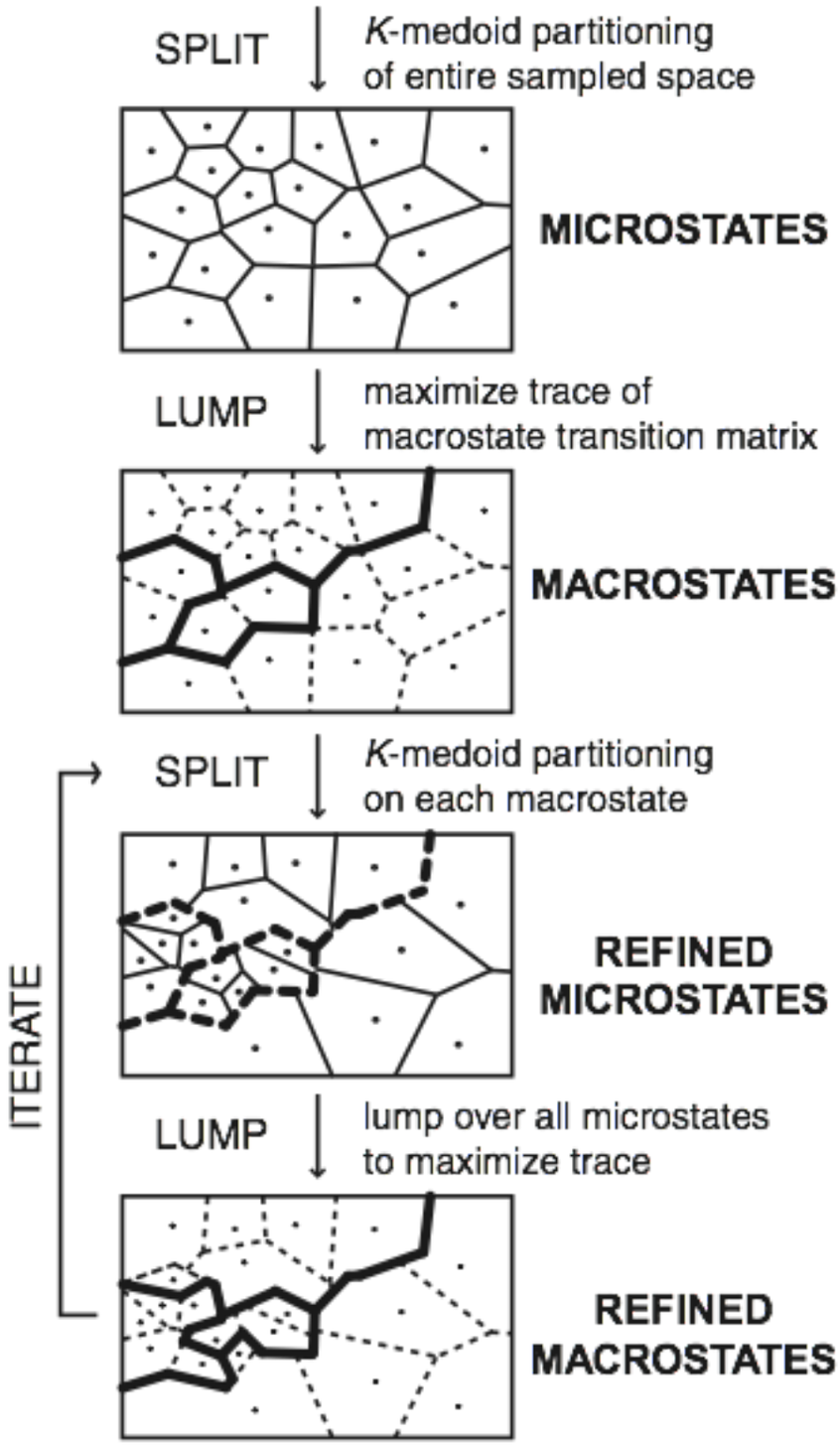


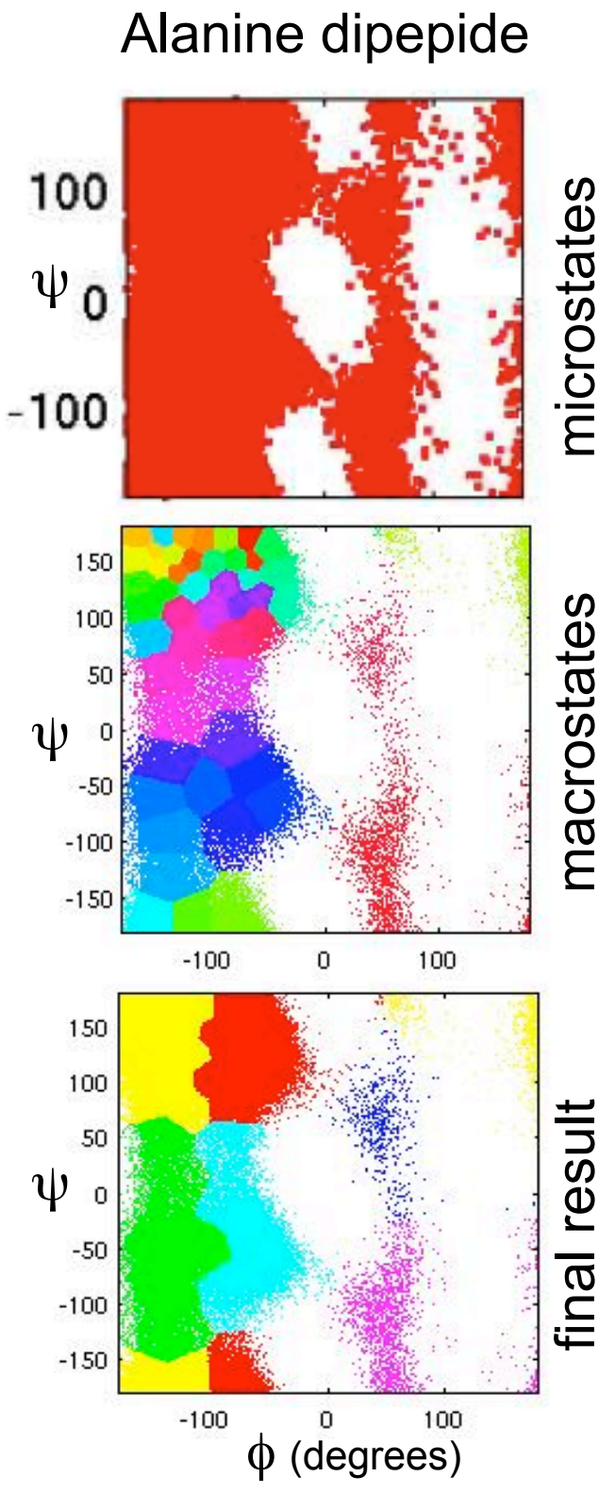
yields a complete description of long timescale kinetics & thermodynamics (predict rates, thermodynamics, & structure)

Automatic State Decomposition: An interative algorithm

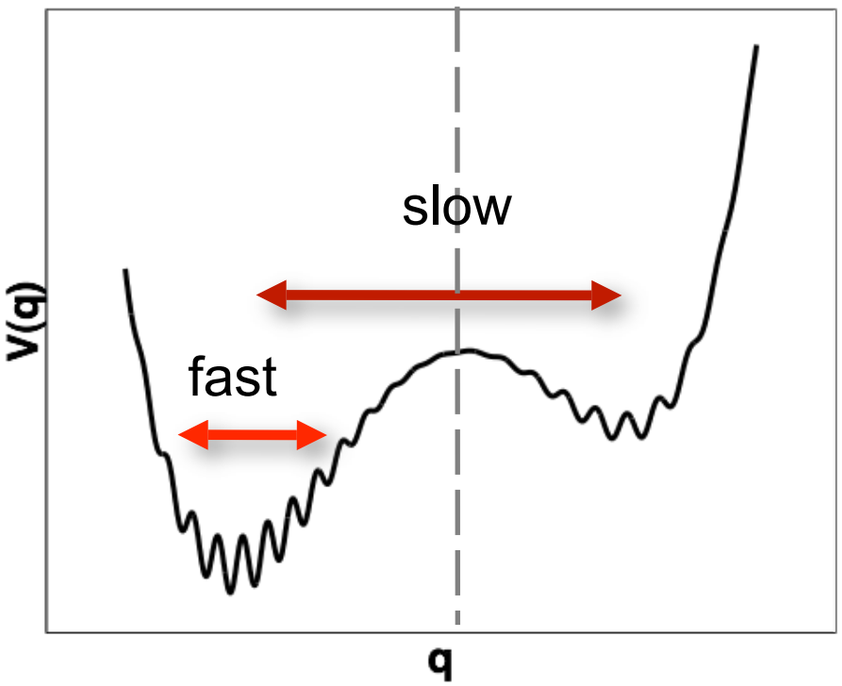
(Chodera & Singhal)

Collaboration between Swope, Dill, and Pande labs



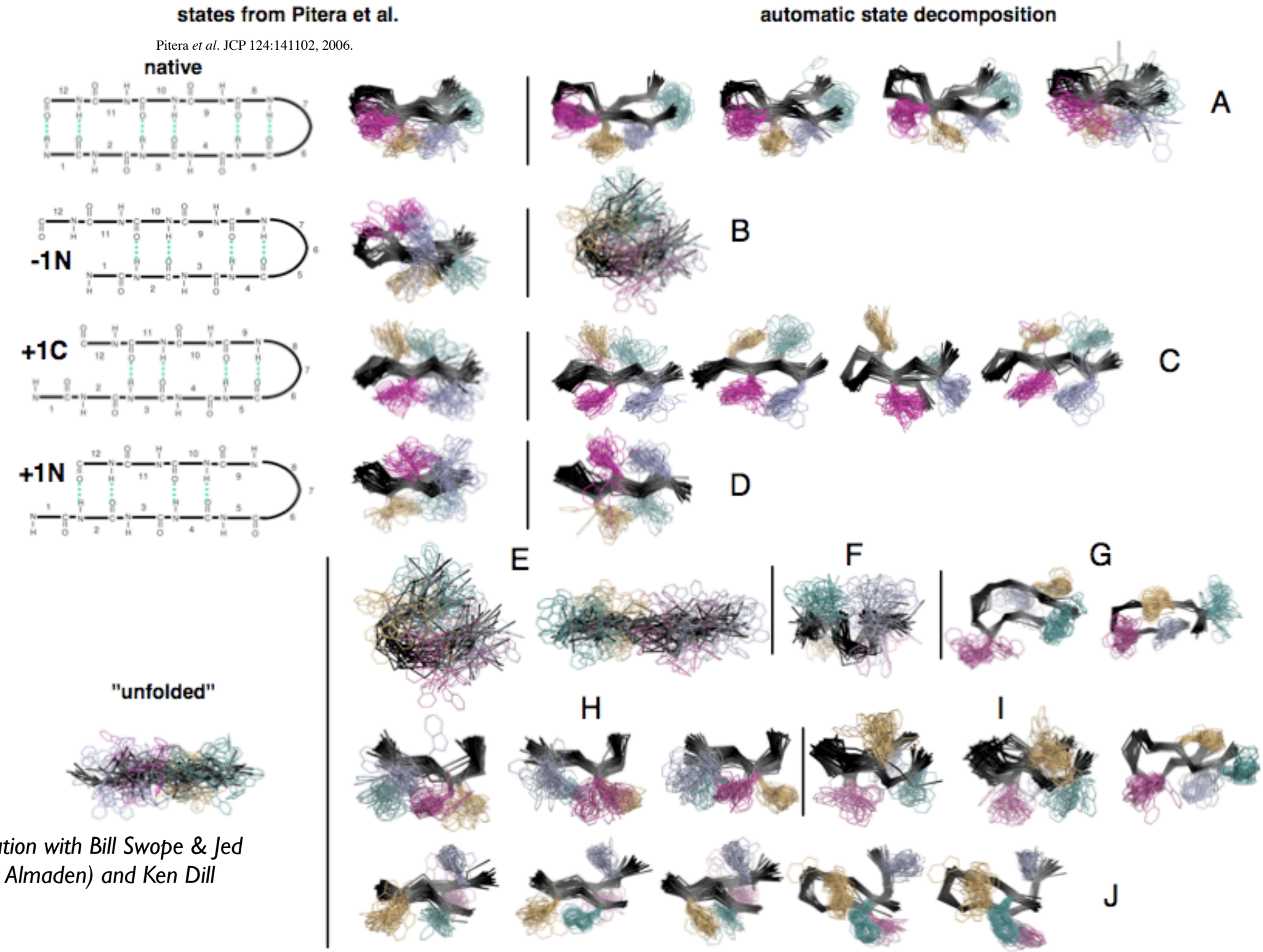


Iterative refinement attempts to locate states for which there is a separation of timescales between fast intrastate dynamics and slow interstate dynamics.



Macrostates reveal a richer decomposition of configuration space than hypothesis-driven study

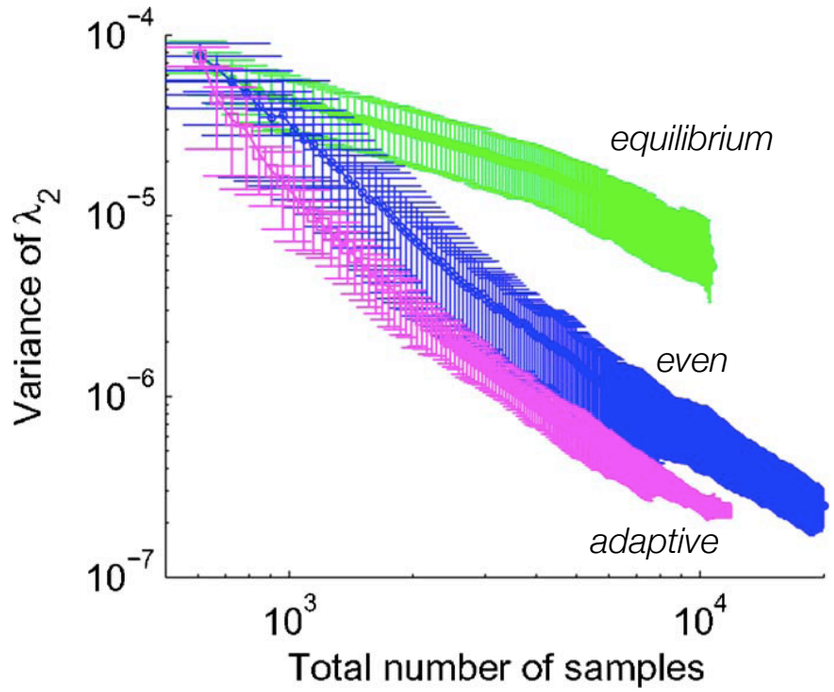
(Chodera & Singhal)



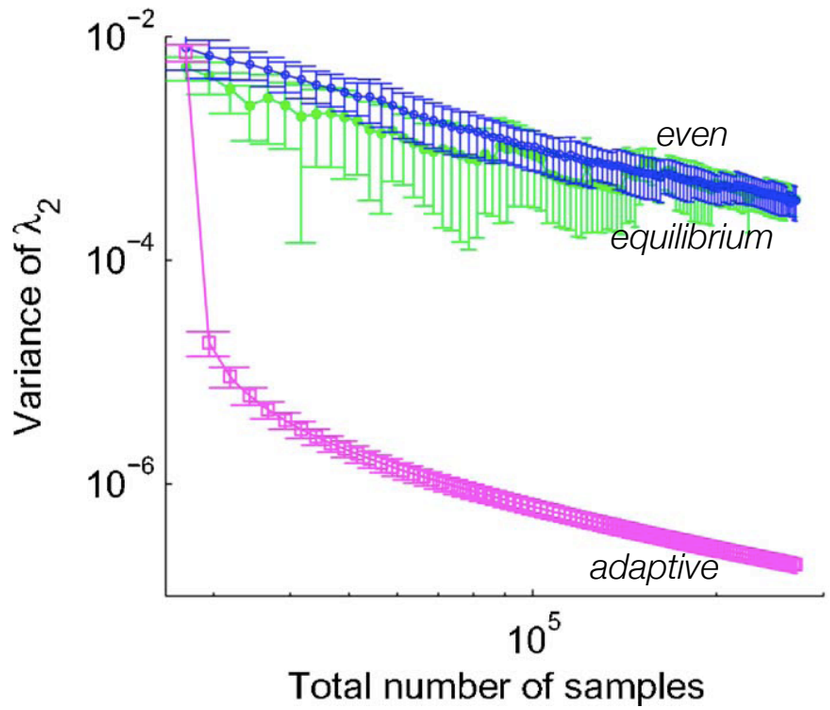
In collaboration with Bill Swope & Jed Pitera (IBM Almaden) and Ken Dill (UCSF).

Adaptive sampling: a big step forward in efficiency

- **Molecular simulation as a statistical problem**
- **Bayesian error analysis methods**
  - allows for on-the-fly adaptive methods
  - add simulations only where needed (to improve uncertaintity)
- **Impact**
  - Optimize trajectory choice based on uncertainty
  - 100x to 1000x speed up -- calculate just what you need, not any more



simple system (alanine dipeptide)



complex system (protein folding of a 36-residue protein)

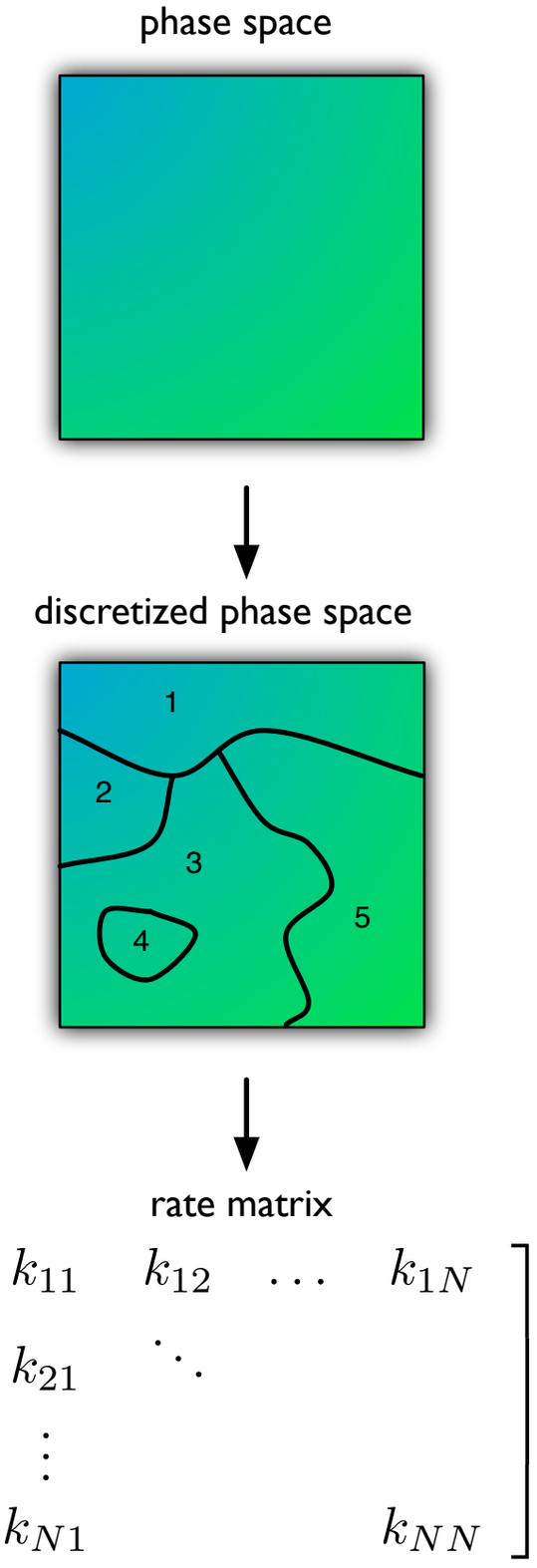
Singhal and Pande, *JCP* (2007)

Conclusions: a paradigm shift in simulation?

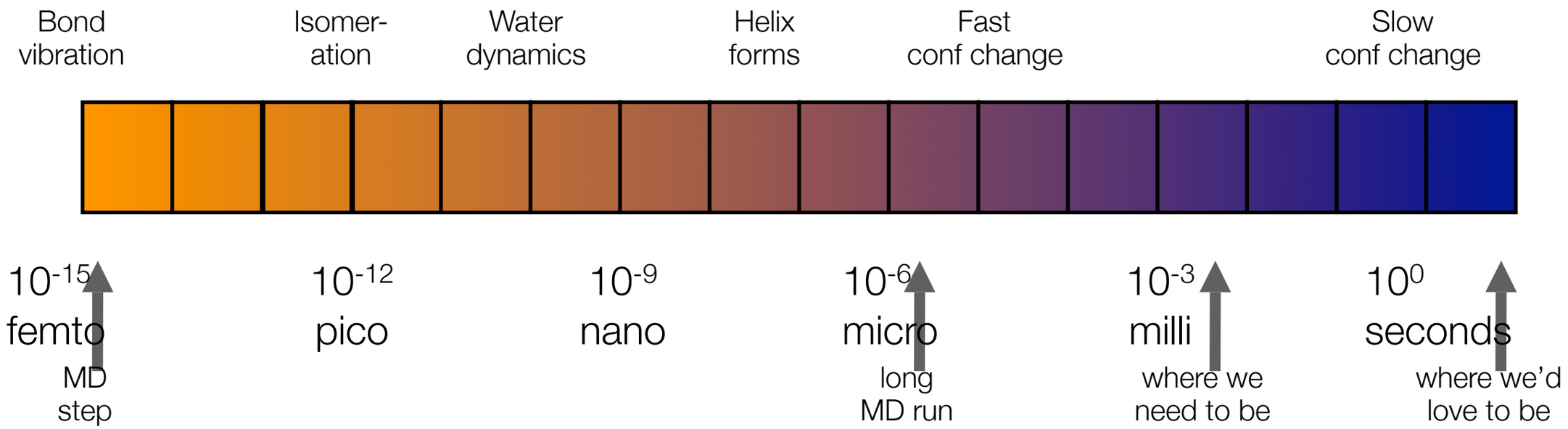
- **Simulations are typically viewed as computational experiments**
  - run, probe, then analyze
  - however, typically done anecdotally (<10 trajectories) due to computational expense

- **New perspective**
  - use simulations to build statistical models of the underlying phenomena
  - Bayesian inferential view of simulations -- simulations are used to parameterize our model

- **Benefits**
  - more powerful methods -- much longer timescales
  - a statistical view of the phenomena of interest (uncertainties, etc)
  - more much scalable than traditional MD
  - **much more efficient (only simulate what you need to simulate)**



A solution to the long timescale challenge?



- **Use a series of complementary methods**
  - default: single CPU does ~ 1ns/day (10^9x gap)
  - Distributed computing (10^4x to 10^5x; cluster: 10^2x)
  - GPU's/streaming (10^2x to 10^3x)
  - MSMs/adaptive sampling (10^2x to 10^3x)
  - total: (10^8x to 10^{11}x = 0.1 to 100 seconds **per day**)

Where to learn more

- **Books:**
  - Leach, *Molecular Modeling*: Great first resource
  - Gromacs manual (http://gromacs.org): has full derivations and detailed explanations

- **Wikipedia**
  - believe it or not, it's pretty well written and has lots of information

- **Folding@Home: http://folding.stanford.edu**