Milestoning: Extending time scales of molecular simulations

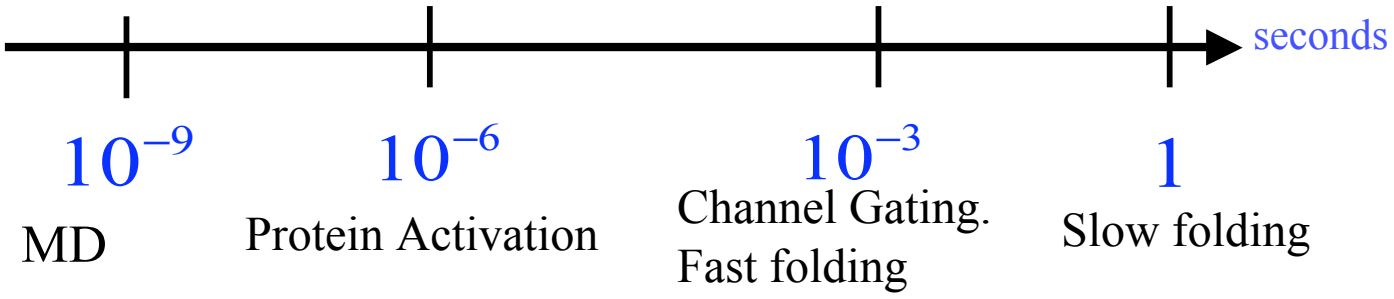
Anthony West, Peter Majek, and RE
Department of Chemistry and Biochemistry and
Institute of Computational Sciences and Engineering (ICES)
University of Texas at Austin

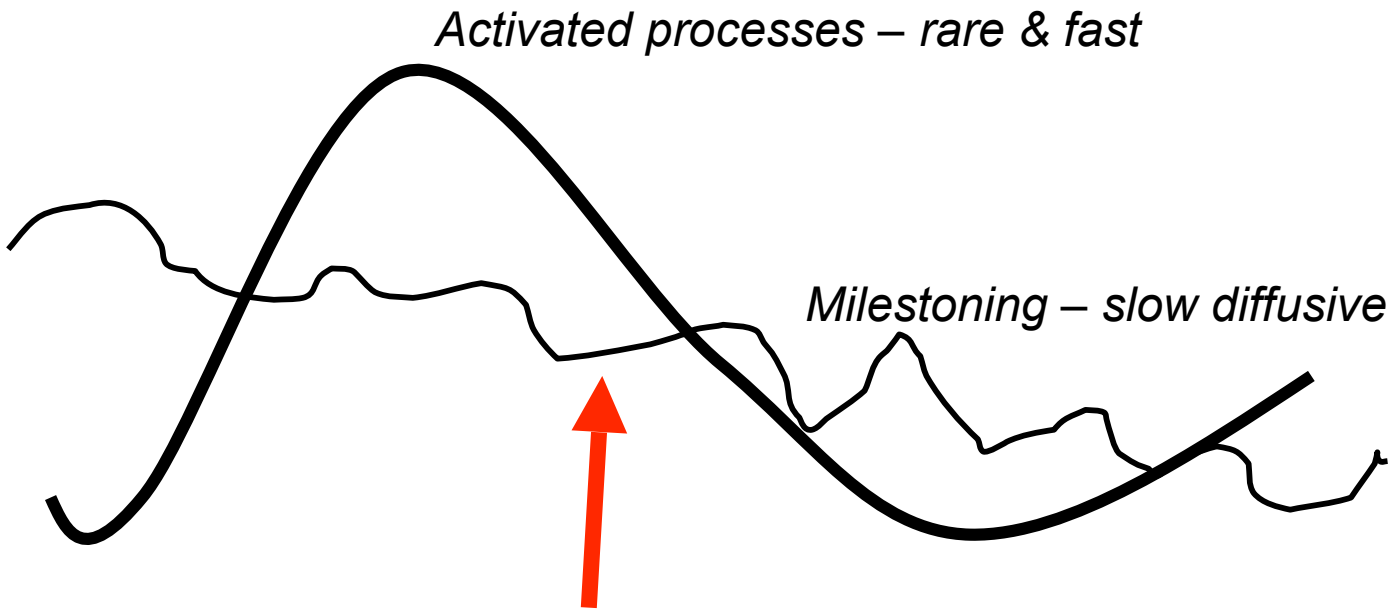
$ NIH

Program

- The problem
- A taste of theory
- Ala dipeptide
- Folding of a helical peptide WH21
- Allosteric transition in Scapharca hemoglobin
- The recovery stroke in myosin

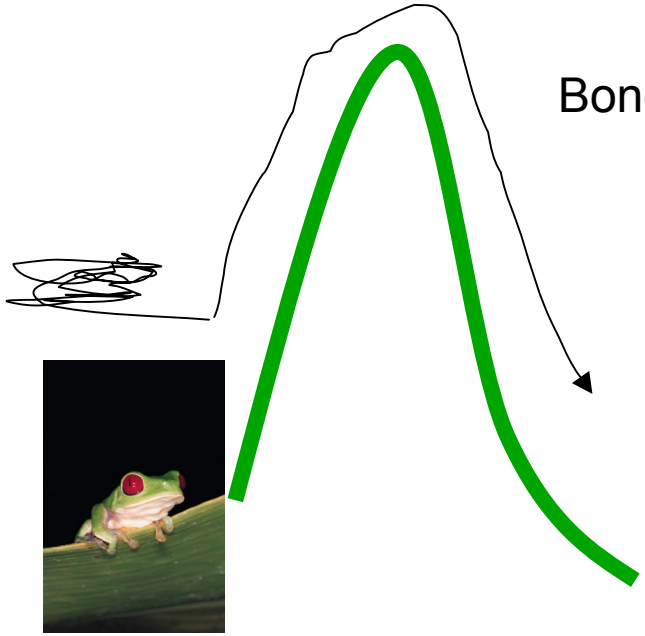
Long time processes in biophysics: activation or long range diffusion





Long time dynamics

- Rare events (<u>short</u> infrequent trajectories)



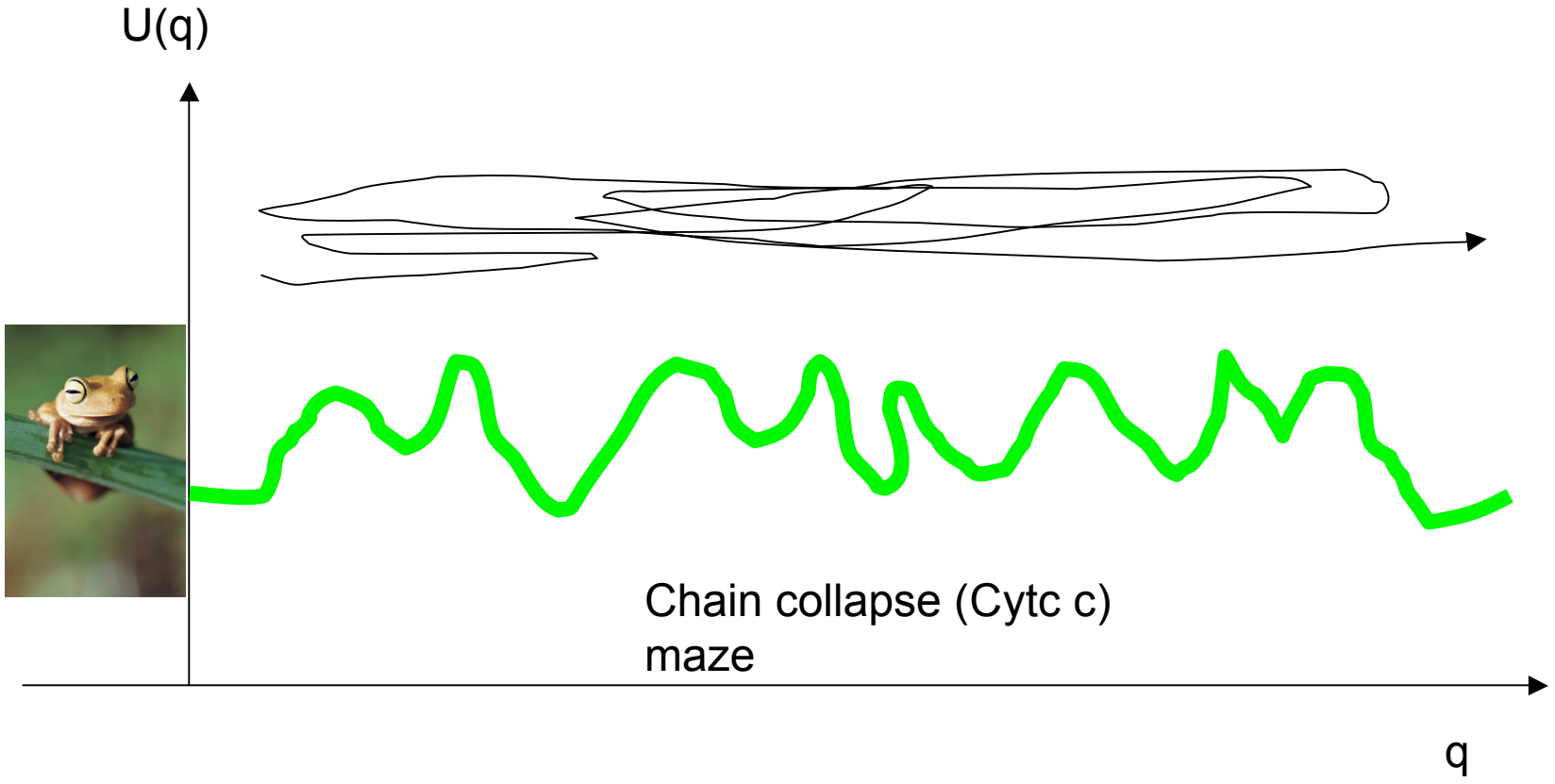
TST (Eyring)

Sampling of complete (rare) traj. :

1. TPS (Chandler, Dellago, Bolhuis)
2. TIS/PPTIS (Moroni, Bolhuis, van Erp)
3. FFS (Allen, Frenkel, ten Wolde)
4. WE (Kim, Huber)

Long time dynamics: Diffusion on rough energy landscape

Milestoning (West, Elber)
Markovian Milestoning (Venturoli and Vanden Eijnden)



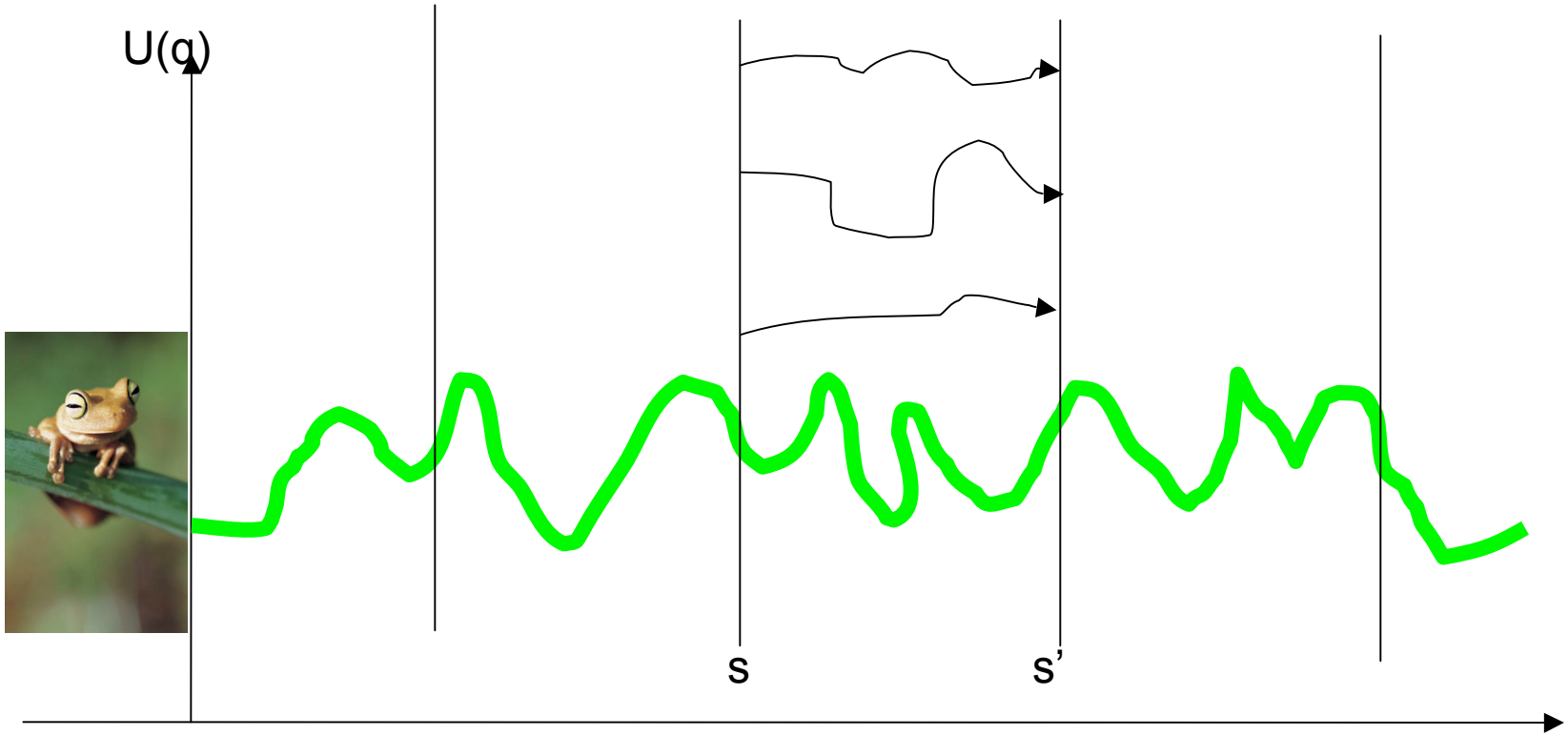
Complexity increases exponentially with system size

- Exponentially larger number of minima and transition states

- Broad distribution of barrier heights and minima depths (studies on myoglobin [Elber, Karplus], peptides [Czerminski, Elber], numerous systems [Wales et al.])

Milestoning:

$K_{s,s'}(t)$ The probability that traj. starting at Milestone s will reach Milestone s' after time t



Assume that we know $K_{s,s'}(t)$

How can we calculate the overall time dependence of the system?

With the matrix $K_{s,s'}(\tau)$ determined, compute kinetics

$$Q_s(t) = \eta_s \delta\left(t - 0^+\right) + \int_0^t \left[\sum_{s'} Q_{s'}(t') K_{s',s}(t - t') \right] dt'$$

$$P_s(t) = \int_0^t Q_s(t') \left[1 - \int_0^{t-t'} \sum_{s'} \left[K_{s,s'}(\tau) \right] d\tau \right] dt'$$

$$\langle t \rangle = \mathbf{1}^t \cdot \int_0^\infty \tau \mathbf{K}(\tau) d\tau \cdot \left[\mathbf{I} - \int_0^\infty \mathbf{K}(\tau) d\tau \right]^{-1} \cdot \Gamma_s$$

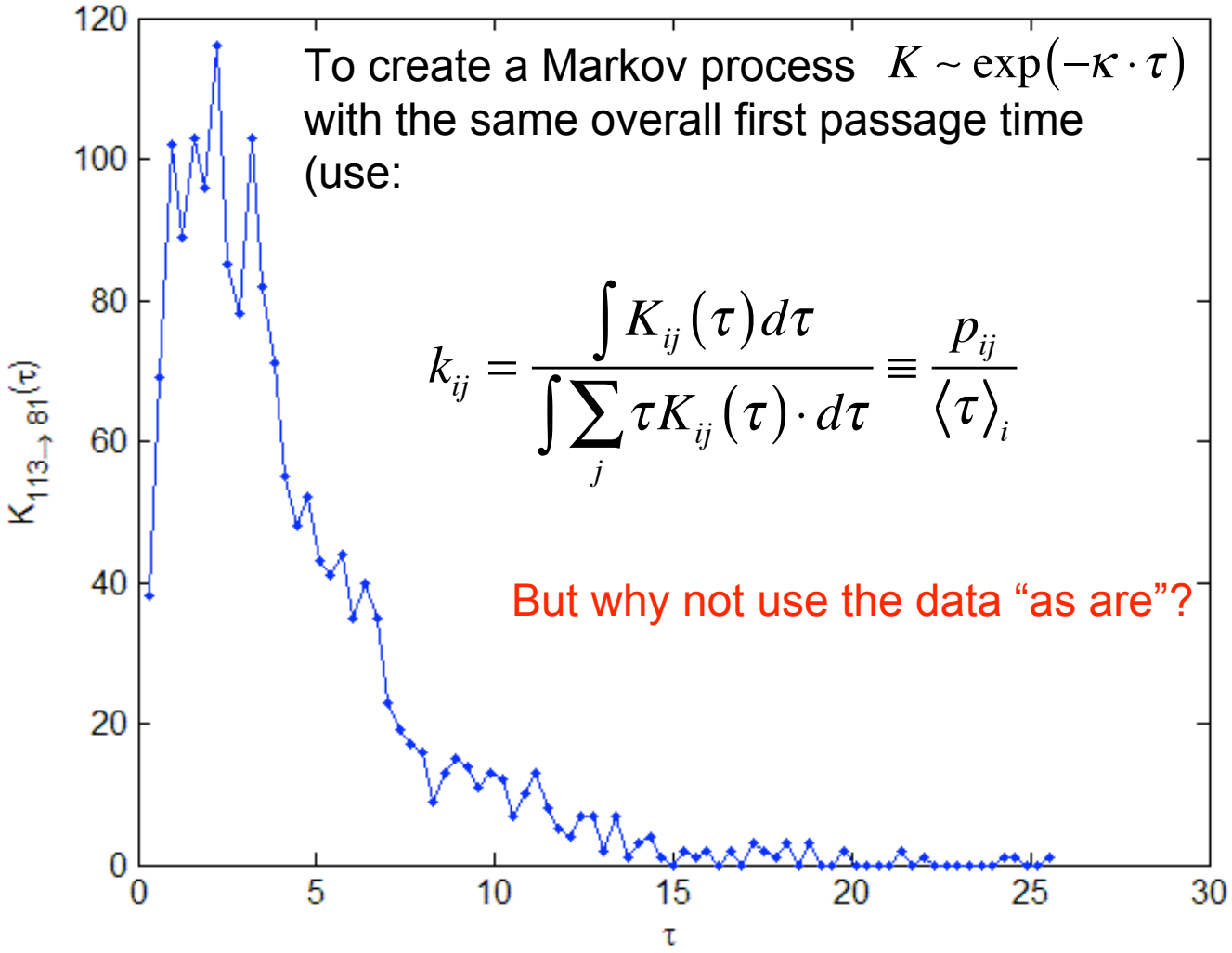
s,s' -- milestones

$Q_s(t)$ -- probability of making a transition to s at time t

$P_s(t)$ -- probability of being at s at time t

- by direct integration
- by Laplace transform and moments of the first passage time (Shalloway)
- by trajectory statistics (Vanden Eijnden)

Example for local first passage time distribution: Alanine dipeptide



To create a Markov process $K \sim \exp(-\kappa \cdot \tau)$ with the same overall first passage time (use:

$$k_{ij} = \frac{\int K_{ij}(\tau) d\tau}{\int \sum_j \tau K_{ij}(\tau) \cdot d\tau} \equiv \frac{p_{ij}}{\langle \tau \rangle_i}$$

But why not use the data "as are"?

Equivalent to Generalized Master Equation

- **The generalized Master equation has time dependent rate coefficients**

$$\frac{dP_s(t)}{dt} = \int_0^t \sum_{s'} \left[-R_{s',s}(\tau) P_s(t-\tau) + R_{s,s'}(\tau) P_{s'}(t-\tau) \right] d\tau$$

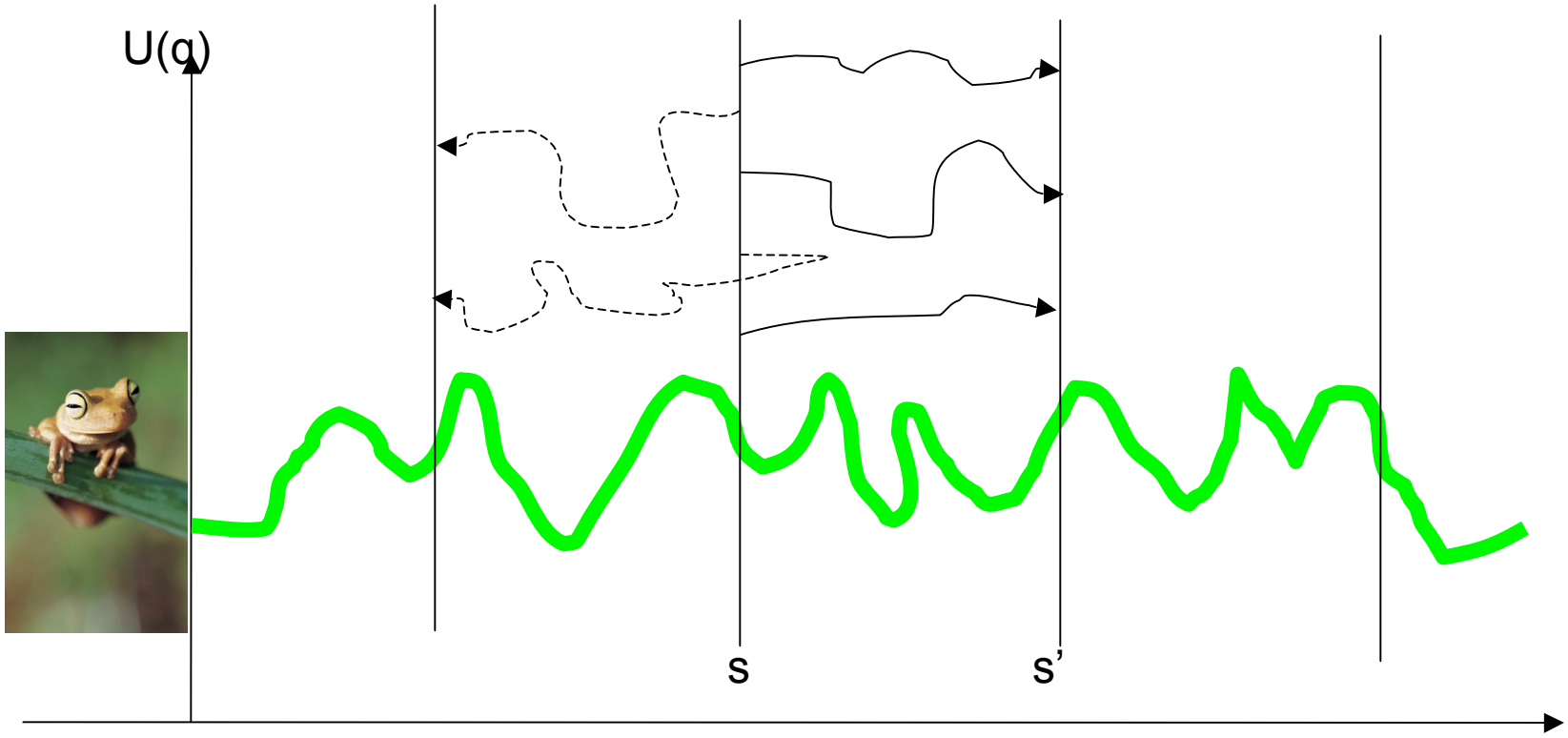
- **K in the QK formulation is easier to compute than R and the Laplace transforms are related by**

$$\tilde{R}_{s,s'} = \frac{u\tilde{K}_{s,s'}(u)}{\left(1 - \sum_{s'} \tilde{K}_{s',s}(u)\right)}$$

How to compute K?

$K_{s,s'}(X,t)$ Sample from the “appropriate” distribution at s and run Short traj. to s’

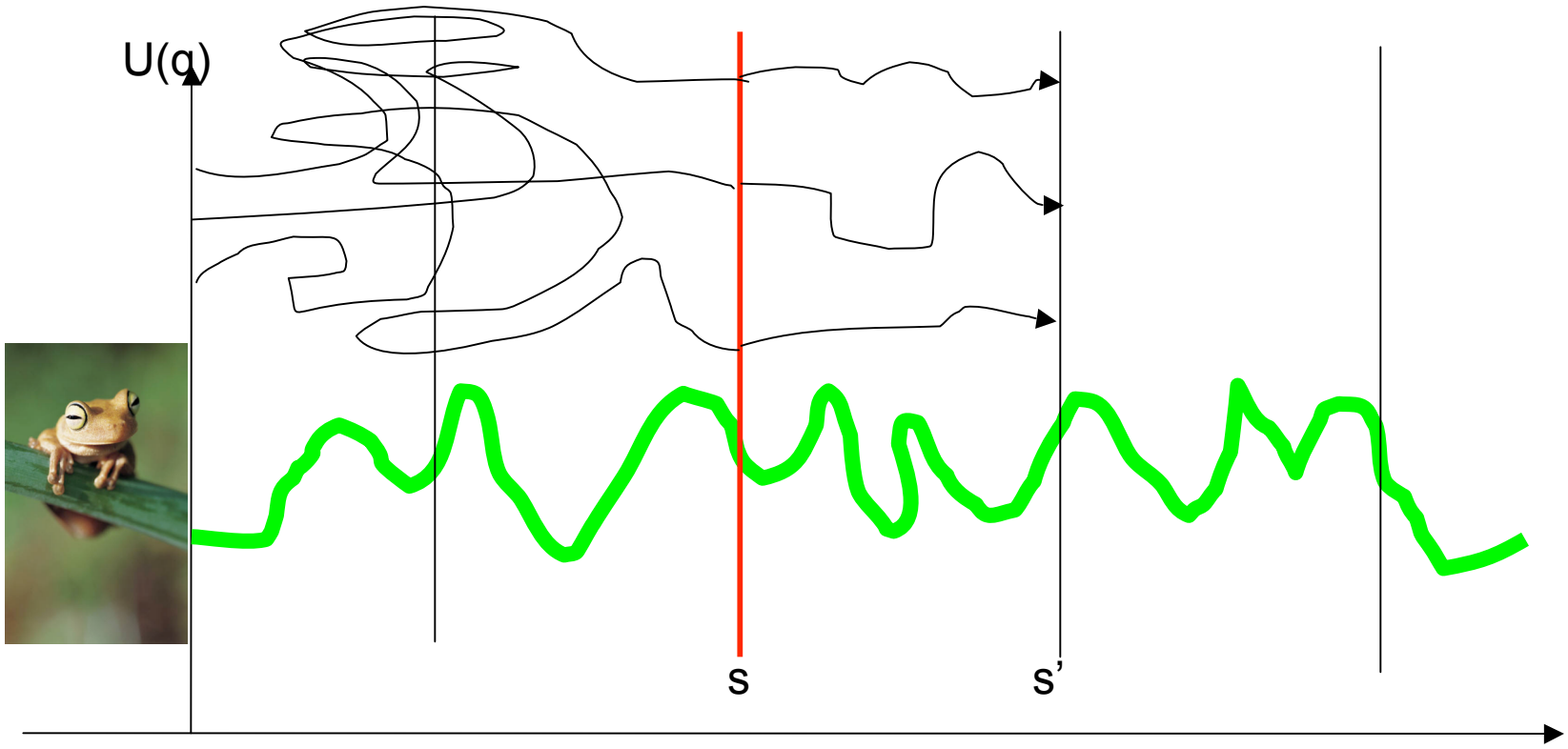
How to obtain the “appropriate” initial distribution?



How to compute K?

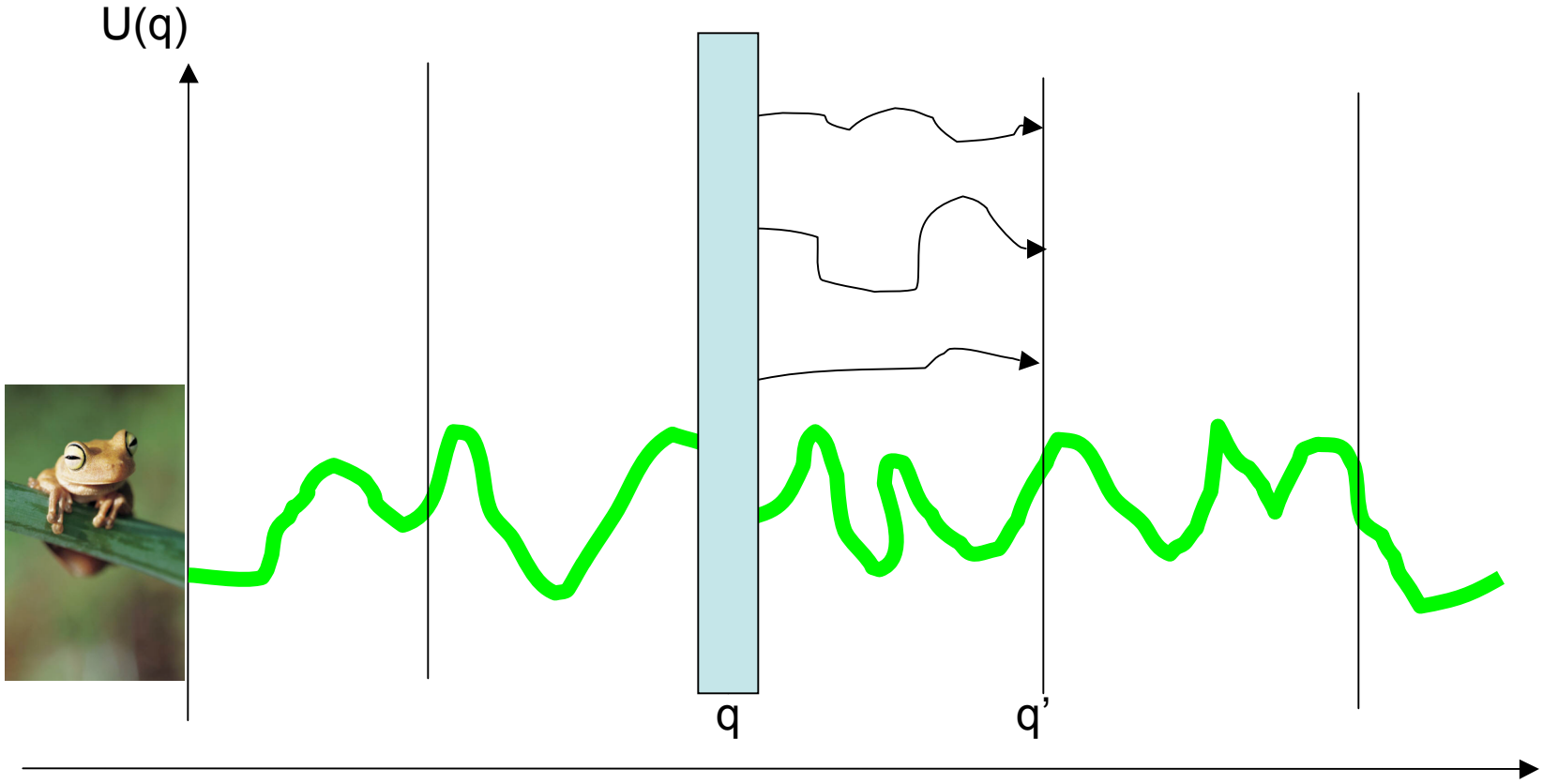
$K_{s,s'}(X,t)$ How to obtain the “appropriate” initial distribution?

Run exact trajectories to s (TIS FFS)



Initiate traj. at the Milestone from Equilibrium

$$\sim c(x)\exp\left(-\beta U(x,q)\right)$$



Stationary Flux

How to initiate at q,X in Milestoning? we use the thermal distribution

$$\sim c(x)\exp\left(-\beta U\left(X;q\right)\right)$$

Milestoning assumes that the initial distribution at q is sampled from equilibrium and compute K from short trajectories between Milestones (interfaces/Voronoi cells) sampled from equilibrium at the initial Milestone. The procedure is approximate but very efficient.

Other approaches: Sampling within the Milestoning cell, Matching fluxes (Venturoli, Vanden Eijnden). More accurate but costly.

If q is a committer the results are exact for Brownian dynamics (Vanden-Eijnden, Ciccotti, Venturoli, Elber, JCP)

Efficiency

- *Diffusive speedup*: $t \sim L^2 \rightarrow t \sim M(L/M)^2 = L^2/M$
- *Parallelization speedup*: $t \sim L^2/M^2$
- Exponential bootstrapping at large barrier: $\frac{1}{qq'} \rightarrow \frac{1}{q} + \frac{1}{q'}$
- E.g.: *microsecond* allosteric transition rate predicted for *Scapharca* (in accord with experiment) based on an ensemble of picosecond trajs totaling 10 ns
- Results on myosin for the recovery stroke predict *submillisecond* timescale (similar to experiment) using nanosecond simulations

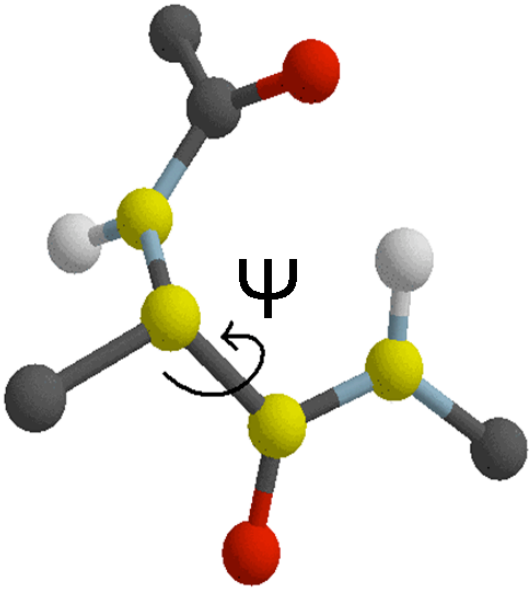
The more milestones the better?

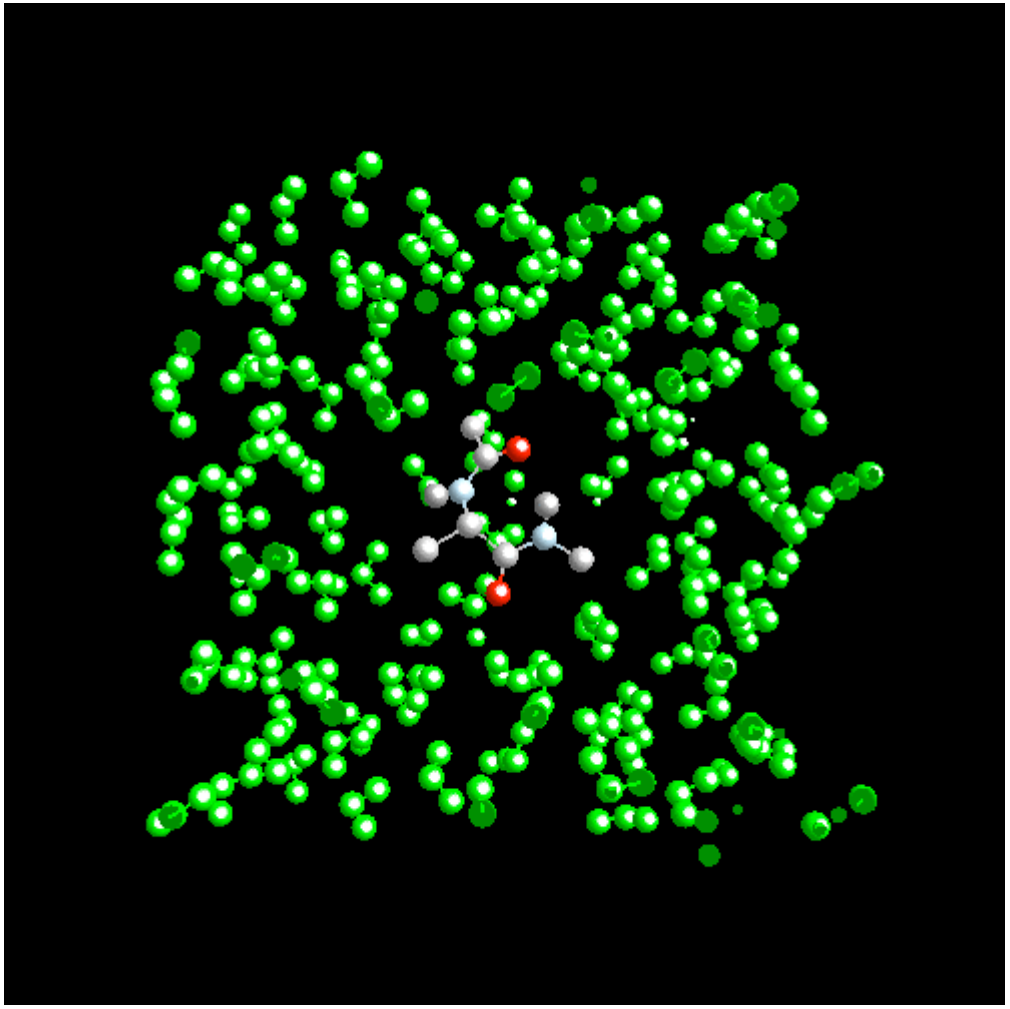
The equilibrium assumption may be violated!

Simple checks of Milestoning using eq. dist. to initiate traj.

- Double (or half) the number of Milestones
- Check typical time of velocity correlation, must be shorter than local transition times.
- Compared equilibrium distribution at the Milestone and arriving FPT distribution.

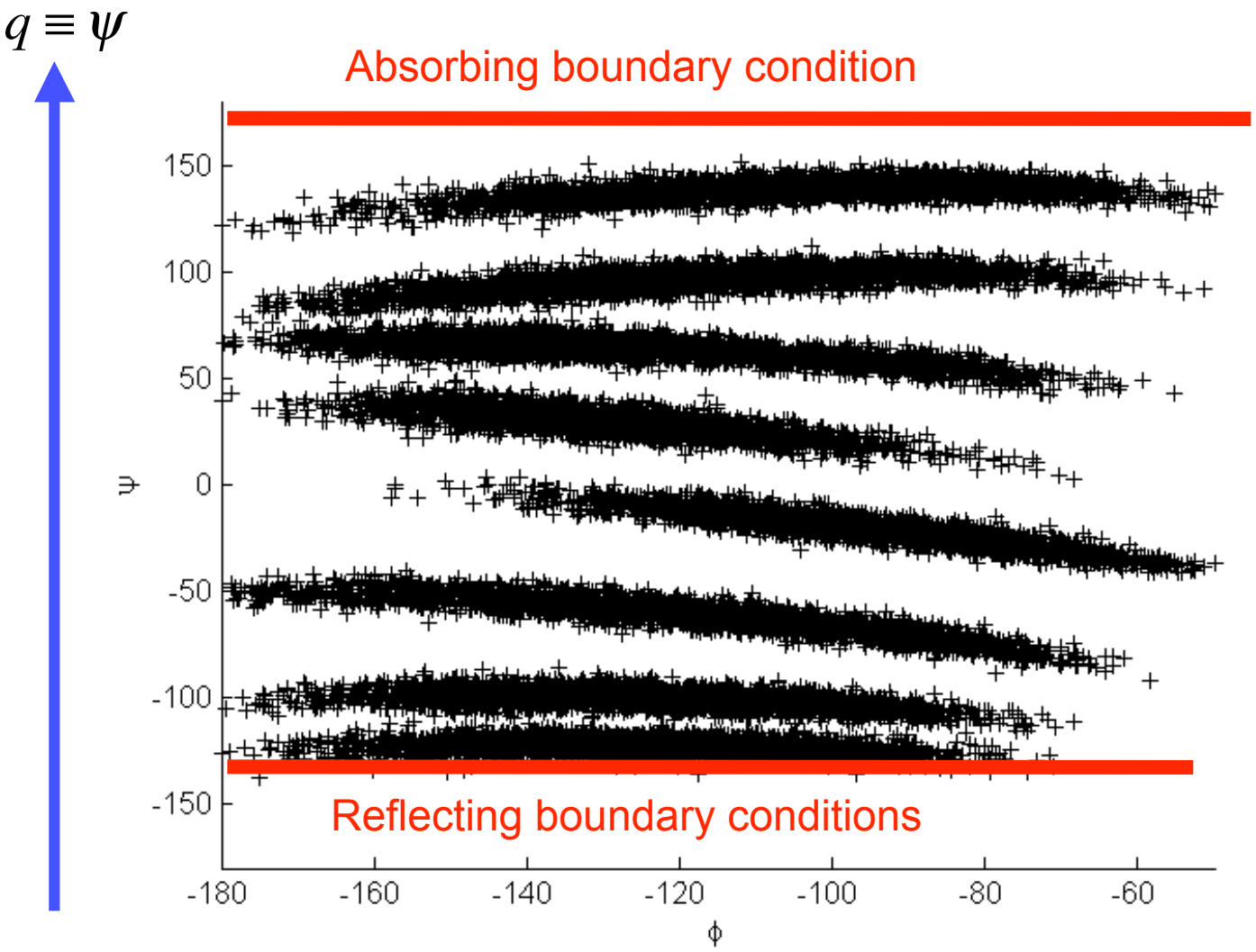
Alanine Dipeptide



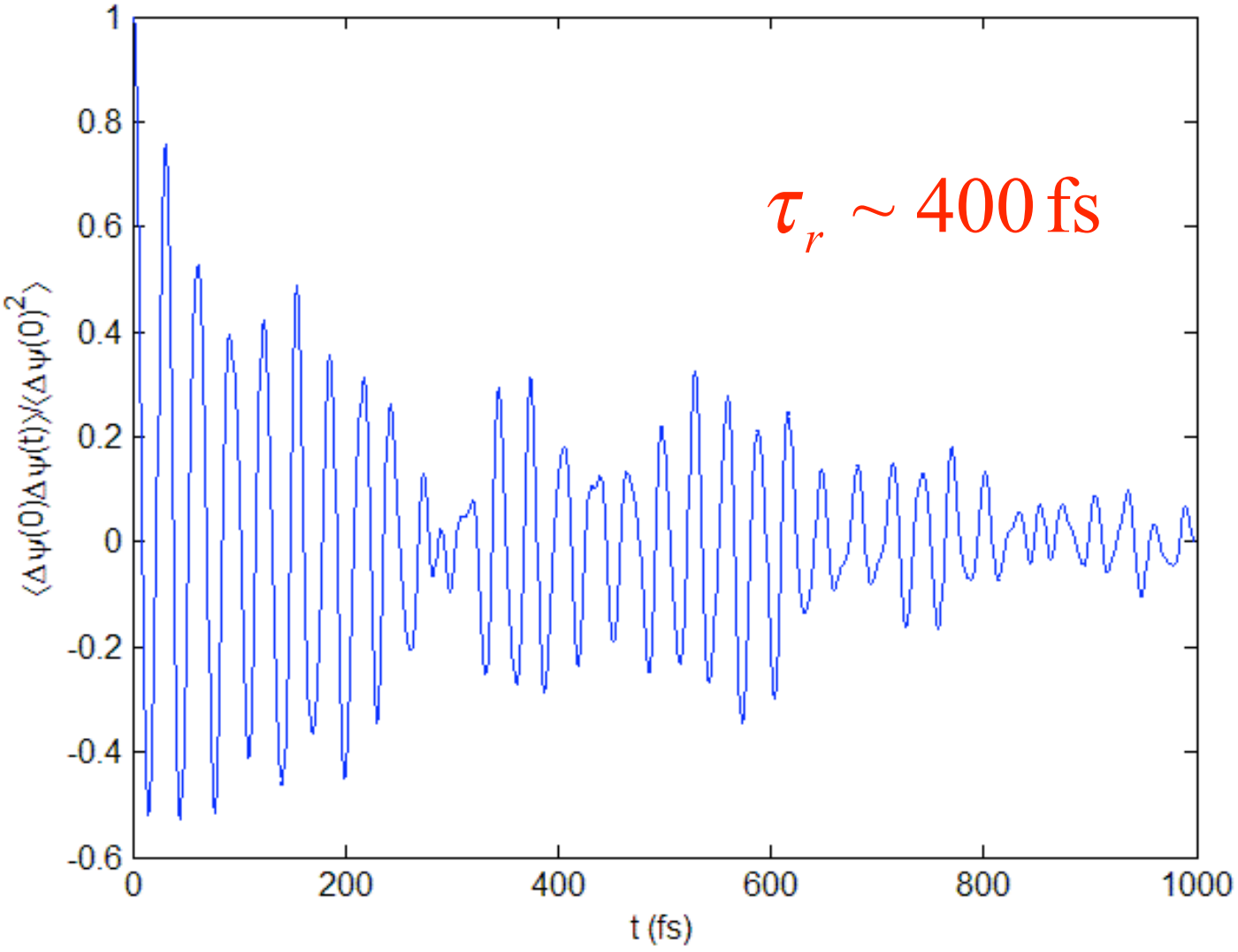


JCP, **126**, 145104 (2007)

Some OEQ ensembles



Torsion velocity auto-correlation indicates when milestoning assumption is being violated



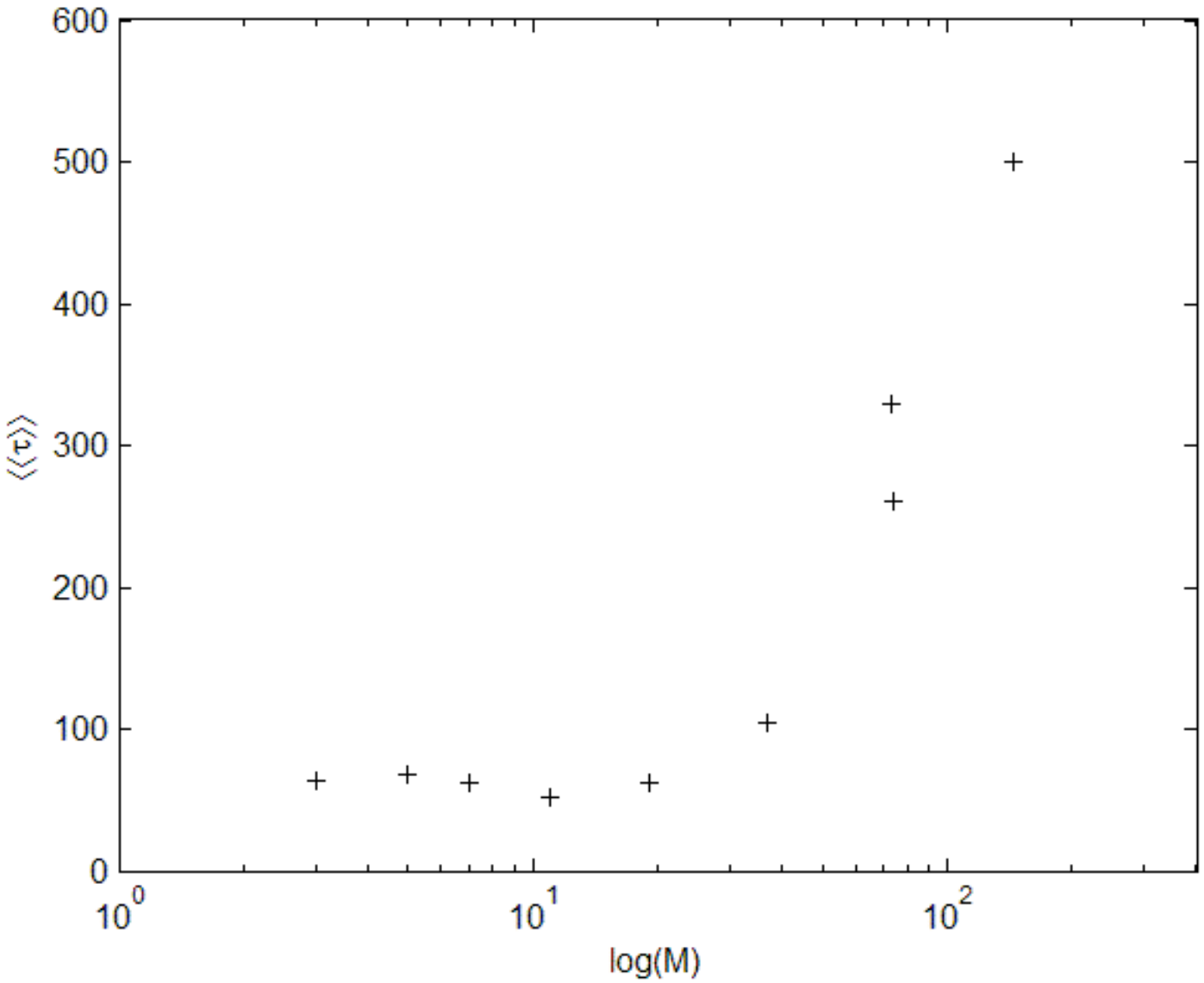
Average Incubation Times vs. Velocity Relaxation Time

$$\langle \bar{\tau} \rangle \equiv \frac{1}{M} \sum_{i=1}^{M} \int_0^\infty \tau K_{s_i}(\tau) d\tau$$

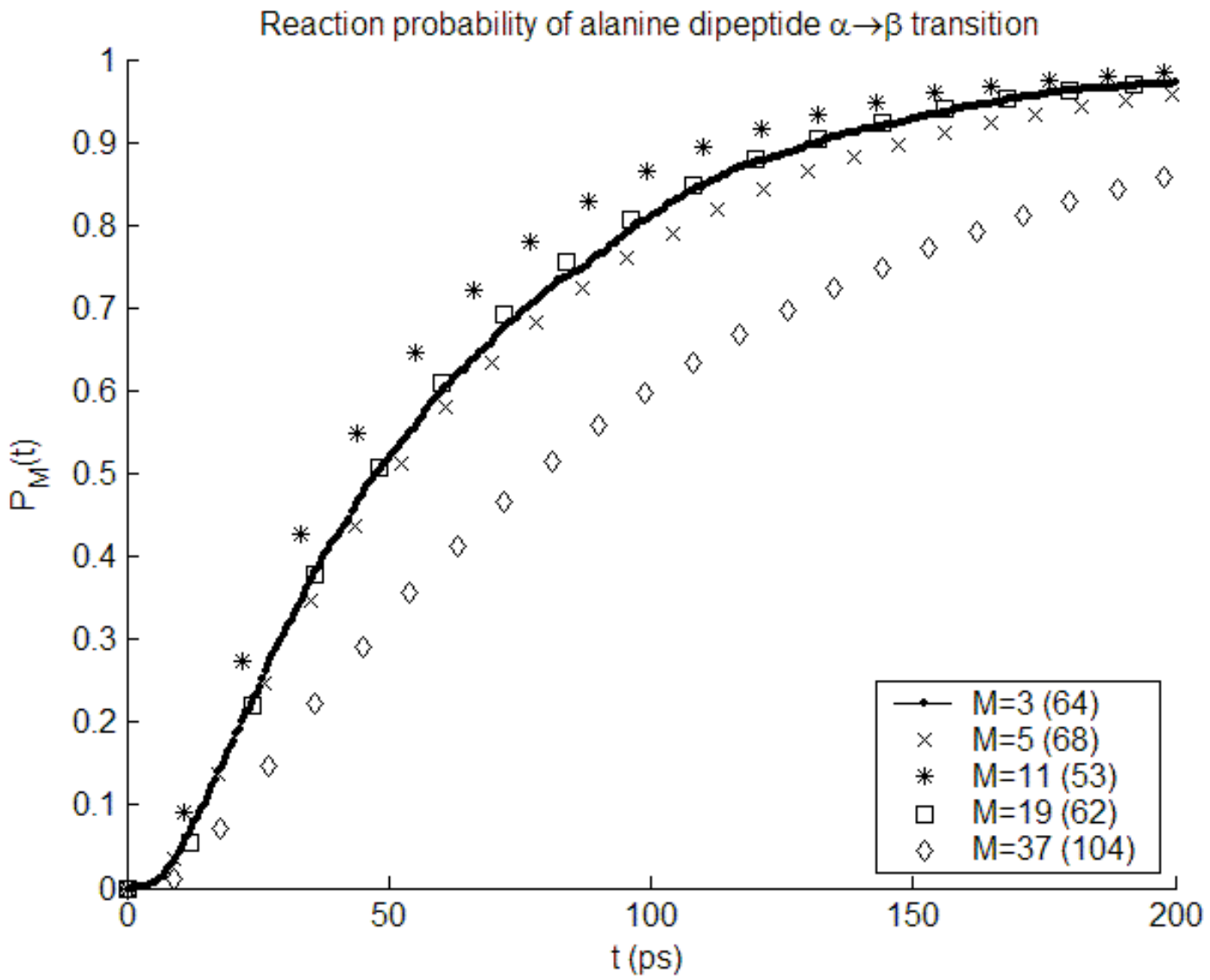
M	$\langle \bar{\tau} \rangle$ (fs)
144	31
74	58
73	58
37	129
19	373
11	1305
7	3581
5	10902

$\uparrow \tau_r > \langle \bar{\tau} \rangle$

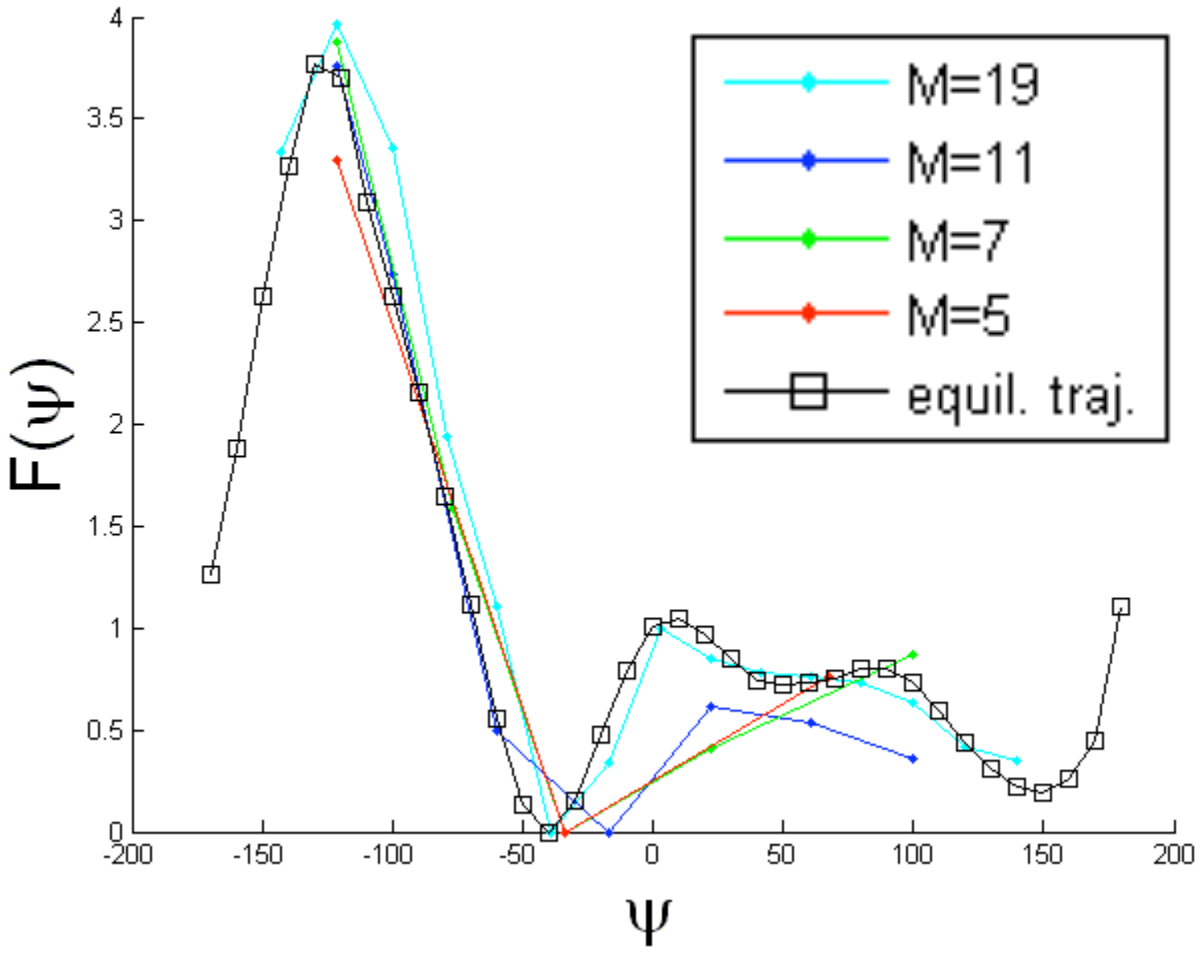
Rate Results



Reaction curves



Free energy “for free”



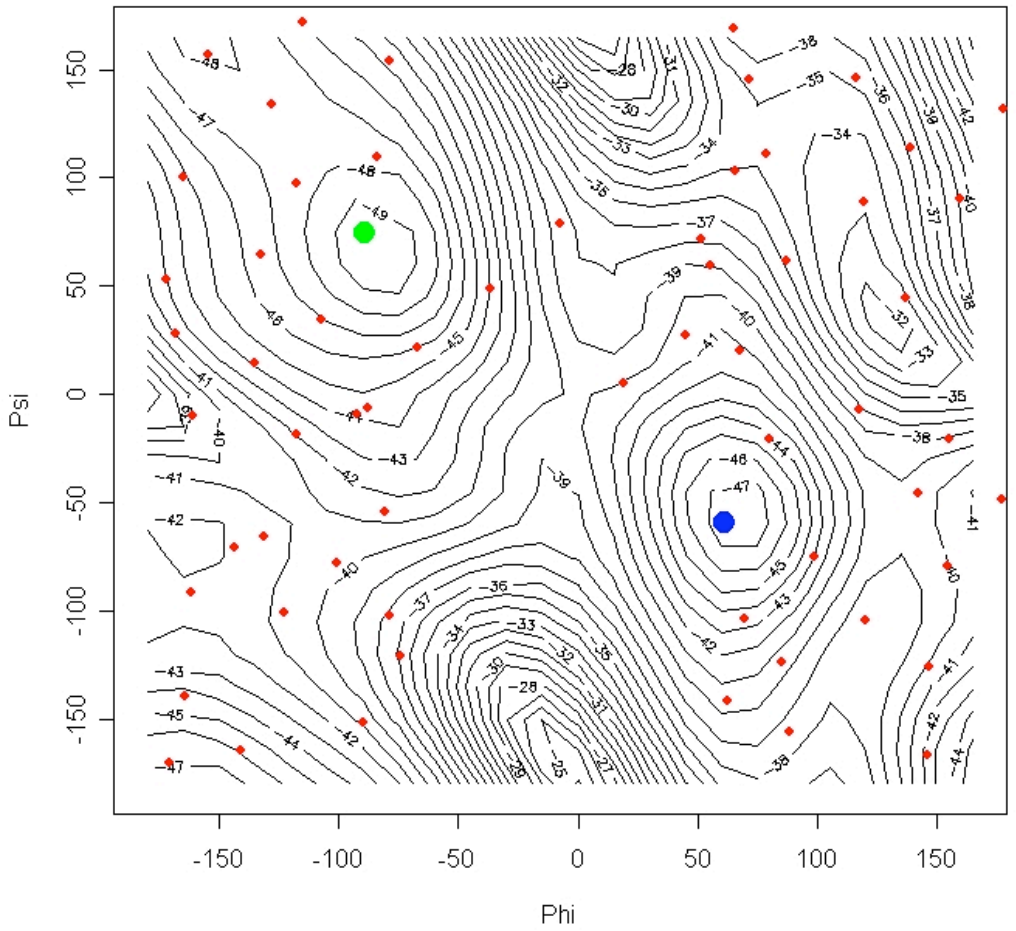
Alanine dipeptide: Milestoning by “mindless” images/Voronoi cells (Peter Majek)

method	mean first passage time from blue -> green	total simulated time*
brute force MD calculation	320 ns +- 20 ns	80 micro secs
2000 - 6000 trajectories / milestone	311 ns +- 49 ns	6 micro secs

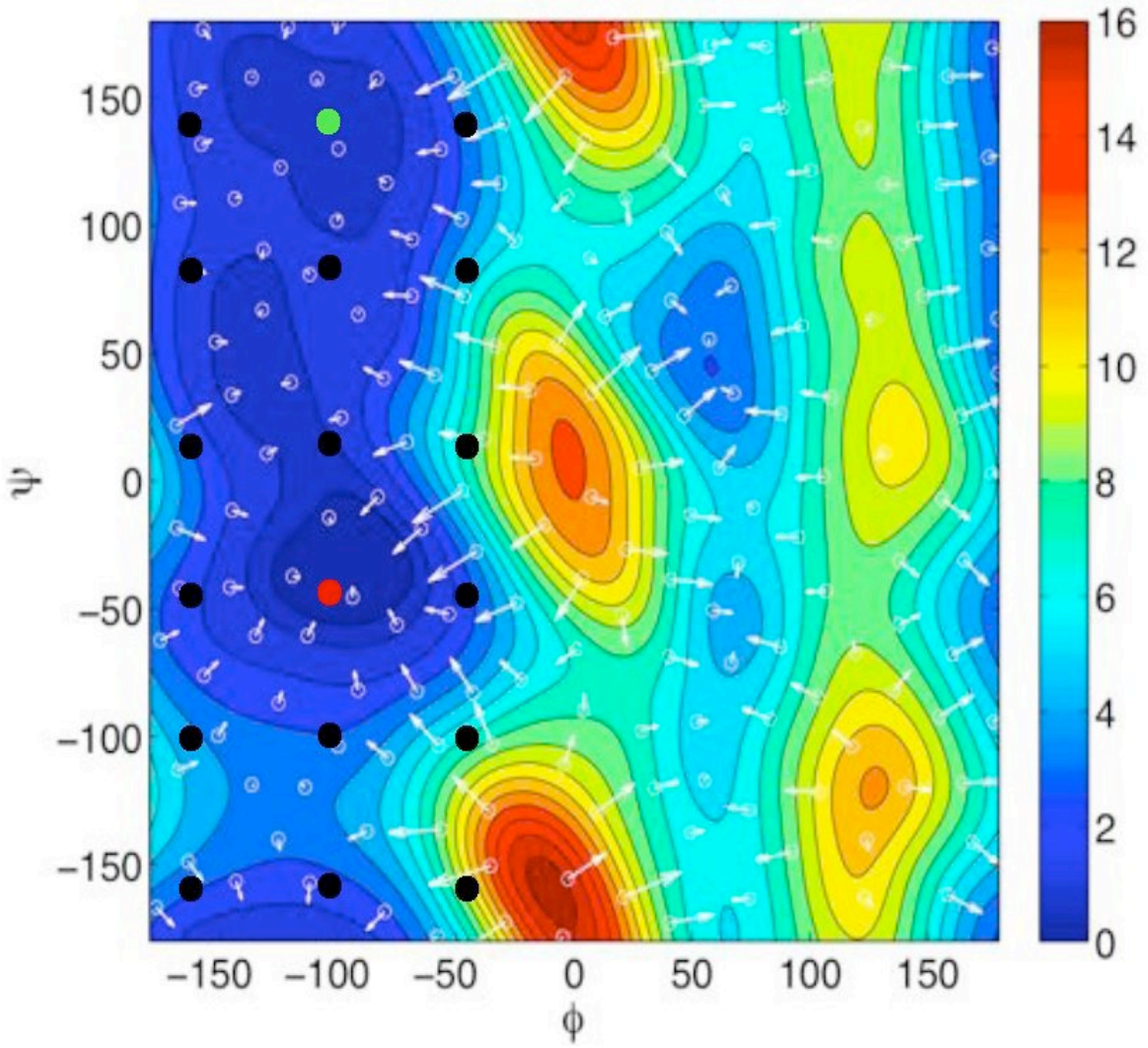
A point belongs to the nearest image

Compute first Passage time distributions Between cells - $K_{ij}(\tau)$

Brownian dynamics



Alanine dipeptide in water: Voronoi cells



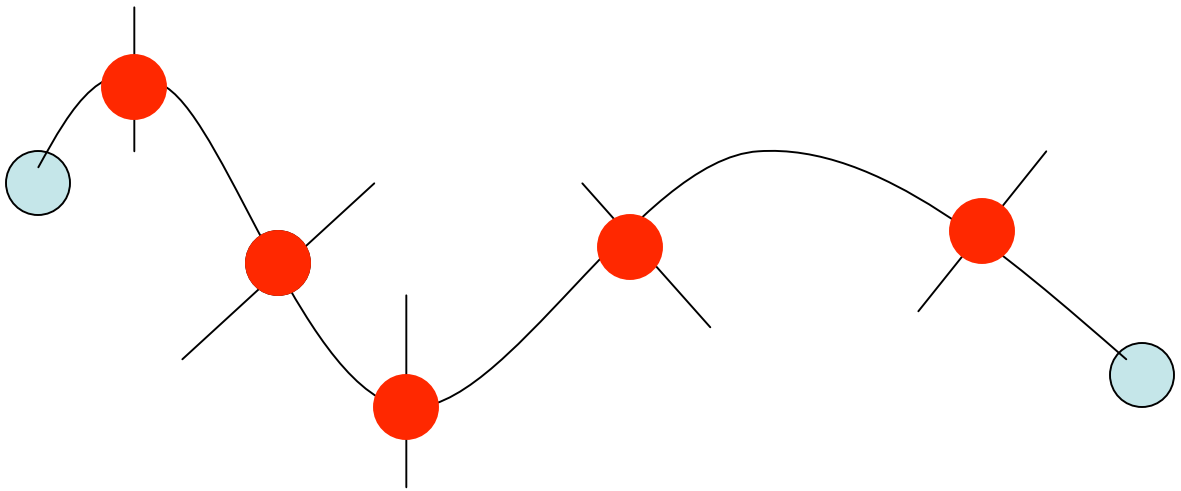
Alanine dipeptide in water

method	**MFPT 1/MFPT2 [ps], (sd)**	**total cost [ns]**
brute force MD	60 / 51	30
milestoning, 100 trajectories/milestone	42 (6) / 42 (5)	1.5
milestoning, 200 trajectories/milestone	47 (5) / 44 (4)	3.0
milestoning, 400 trajectories/milestone	49 (5) / 42 (3)	6.0

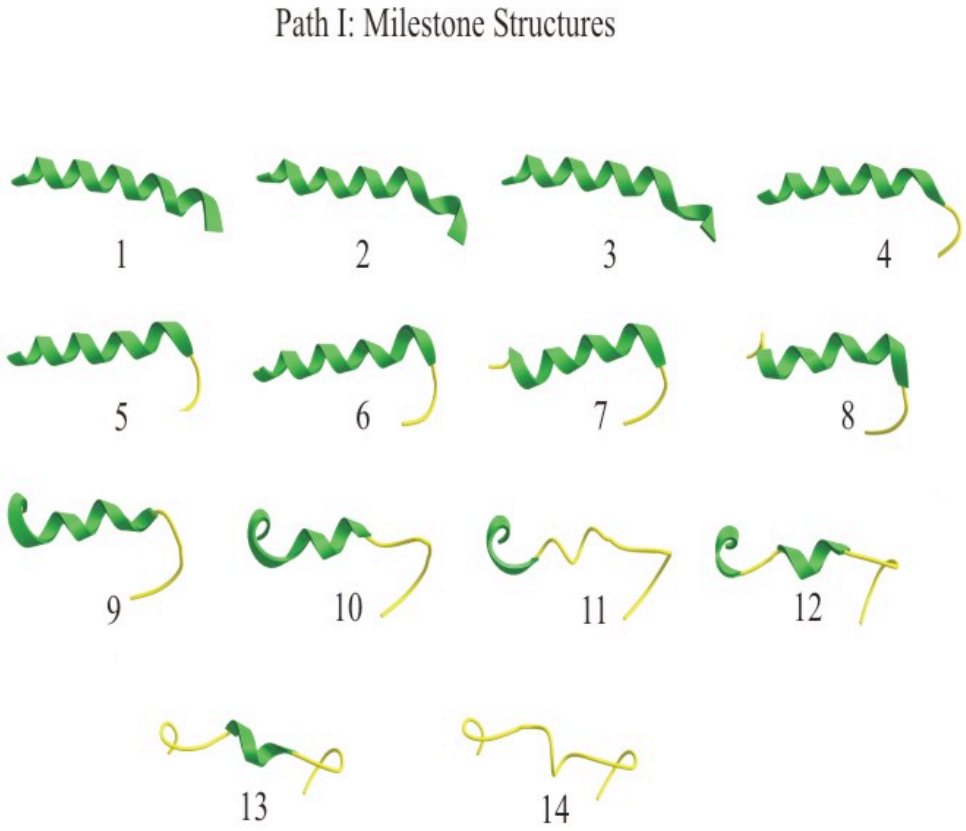
Numerically determine reaction coordinate to define milestones in high dimension

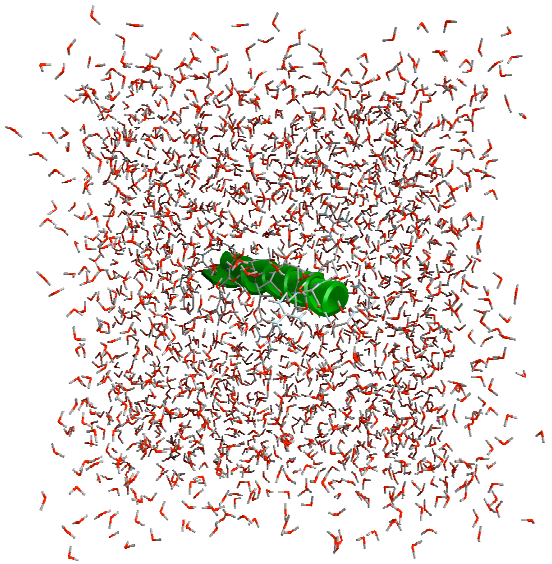
- Hyperplanes perpendicular to optimized paths (Olender & Elber 96)

$$S = \int_{X_R}^{X_P} \sqrt{\nabla U^t \cdot \nabla U}\, dl + C\left(X(l)\right)$$



Folding of a helical peptide WH21 (with Profs. G. Jas and K. Kuczera)





REACTION COORDINATE CRITICAL:

Paths determined from Replica Exchange + functional optimization

Olender and Elber 96

$$S = \int \sqrt{\nabla U^t \nabla U}\, dl$$

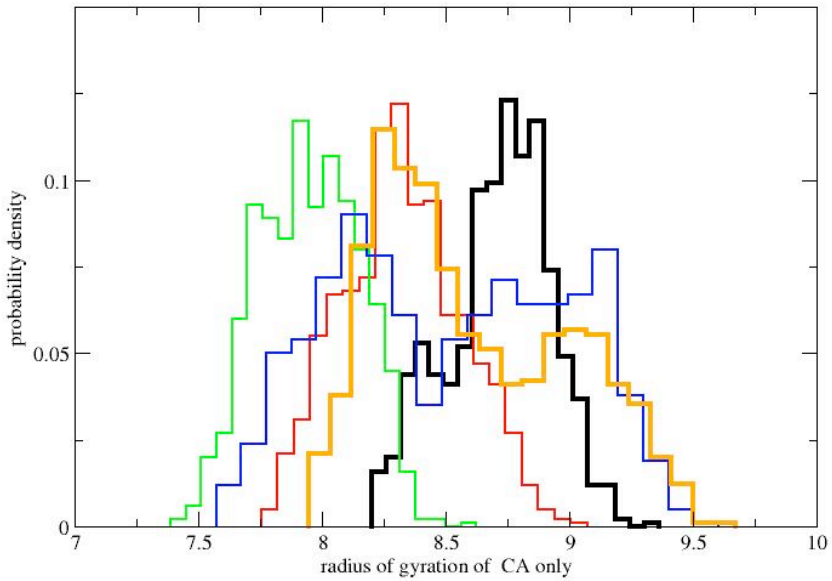
WH21

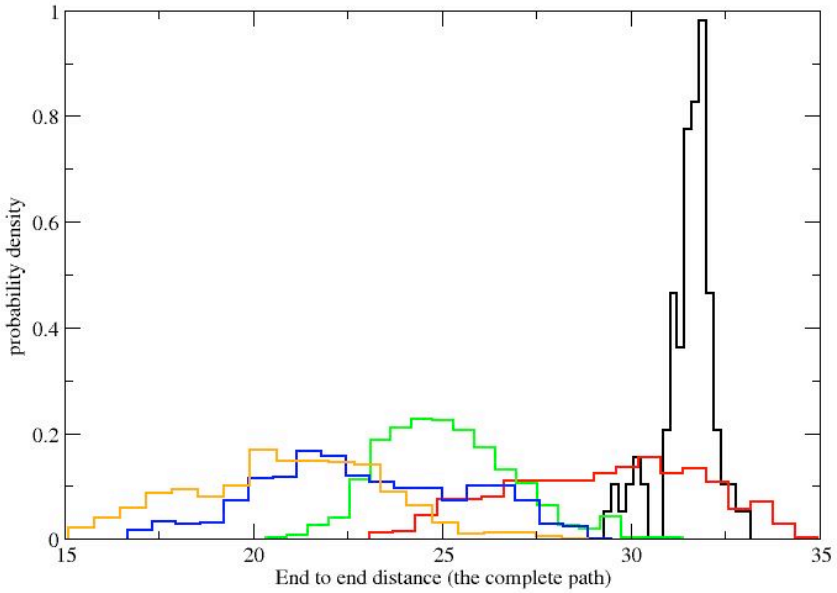
Mean first passage time (exp. 300 ns)

		Unfolding time	Elementary step
Path 1	(N terminal)	280ns	455ps
Path 2	(C terminal)	7μs	1.58ns
Path 3	(middle)	86μs	8.9ns

Elementary steps comparable to Sorin and Pande, BJ 2005

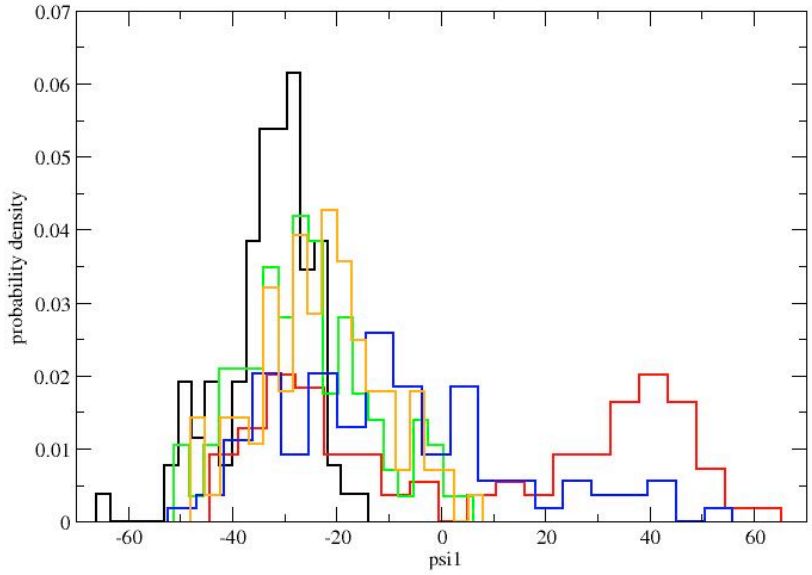
WH21: choice of RC, radius of gyration? End to end distance

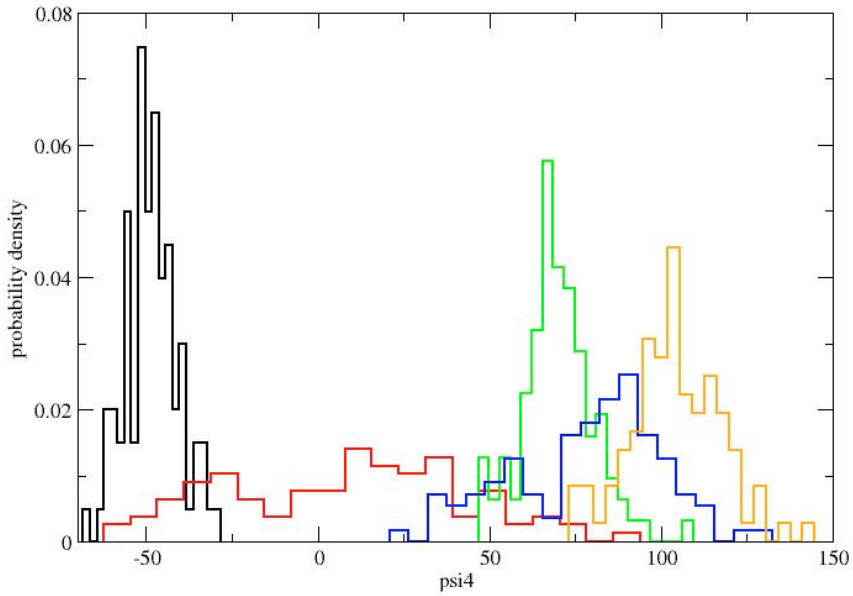




Time sequence in color: black, red, green, blue, orange

Elementary event: WH21 unfolding from the N terminal
Not a single (or two) torsion process





Time sequence in color: black, red, green, blue, orange

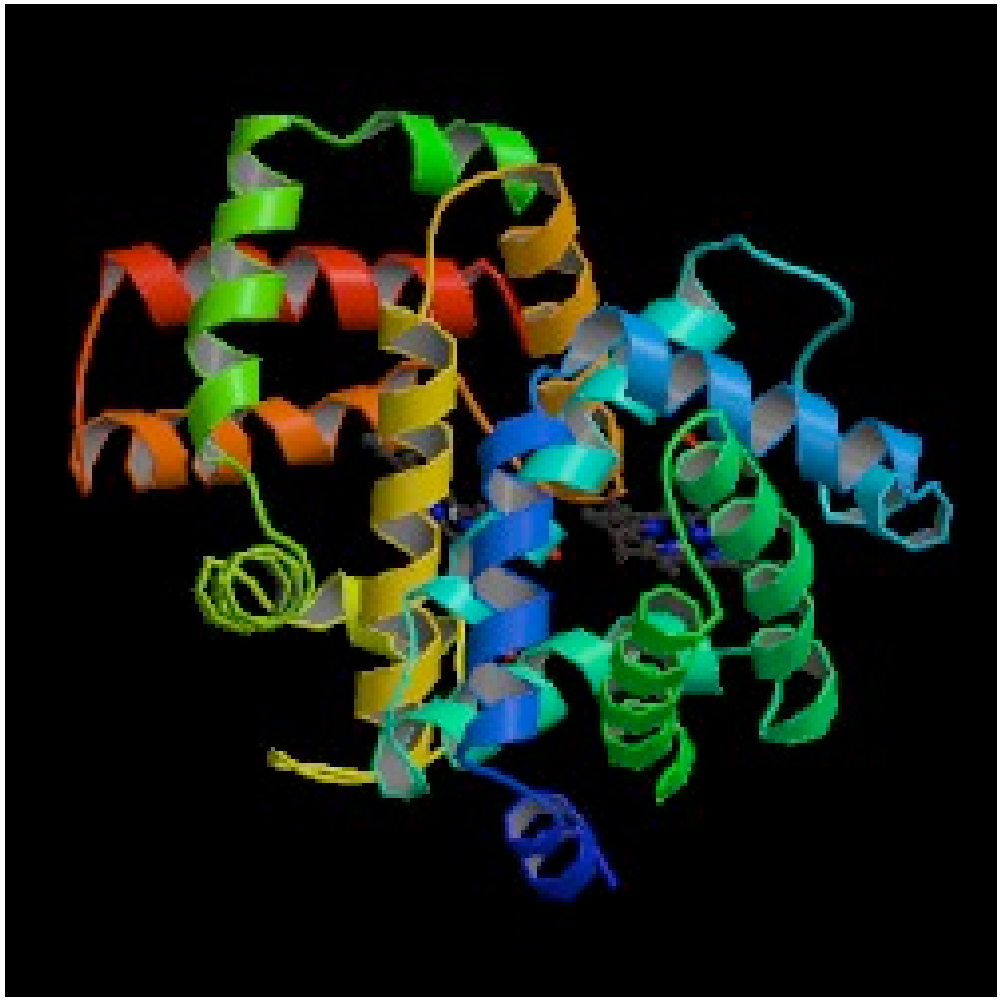
Scapharca Hemoglobin (4SDH)

Bill Royer solved oxy and deoxy structures

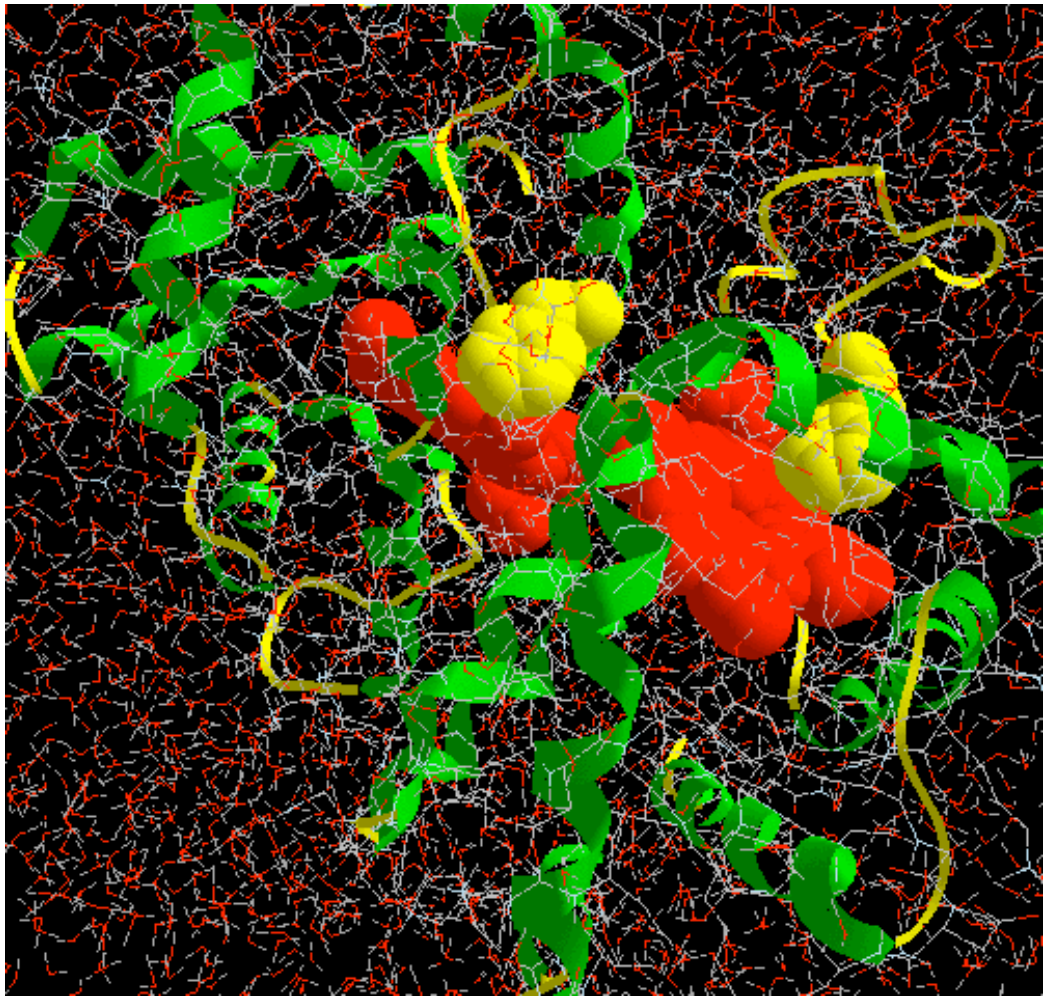
Proposed allosteric mechanism:
1. Phe side chain flips
2. Water organization at interface

Theory:
1. Ligand diffusion: Gibson et al
2. Zhou & Karplus: MD traj of transition

No kinetics



The reaction coordinate from minimization of SPW action



Phenylalanine transitions along the reaction coordinate

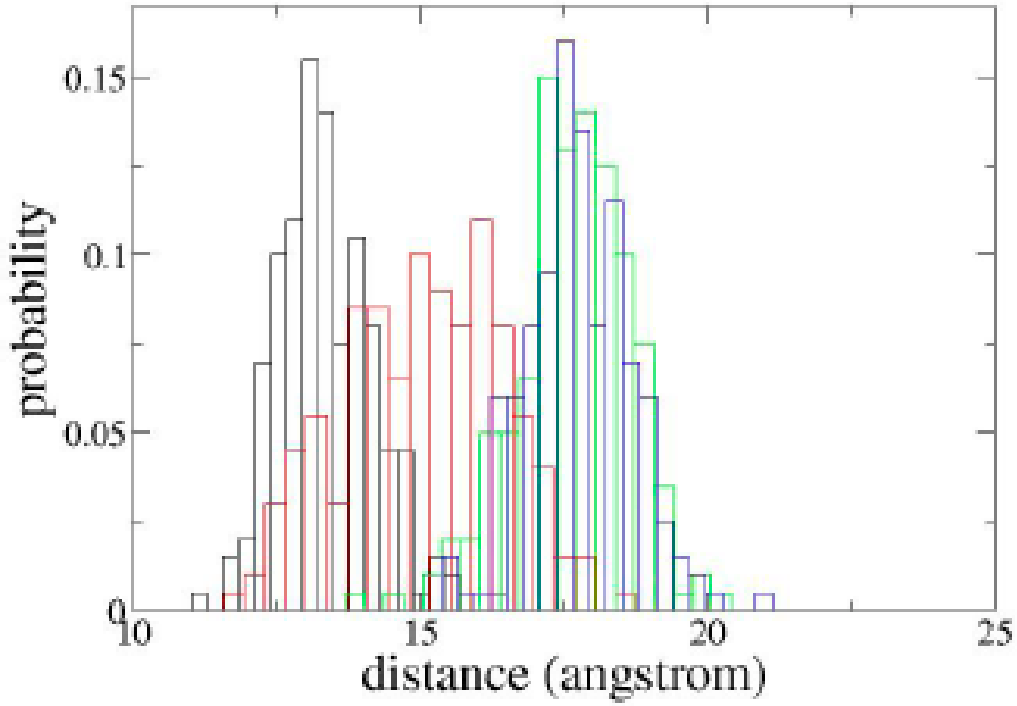


Figure 3: The distributions of the Cz–Cz distances of the two Phenylalanines Phe-97 for milestones 3 (black), 5 (red), 7 (green), and 9 (blue). This distance is a useful measure for the R (or T) transition. Most of the side chain transitions are performed in the neighborhood of milestone 5. The distribution of milestone 7 is similar to milestone 9 suggesting that the side chain transitions are complete at milestone 7.

Scapharca Hemoglobin: Milestone 7 LFPTD

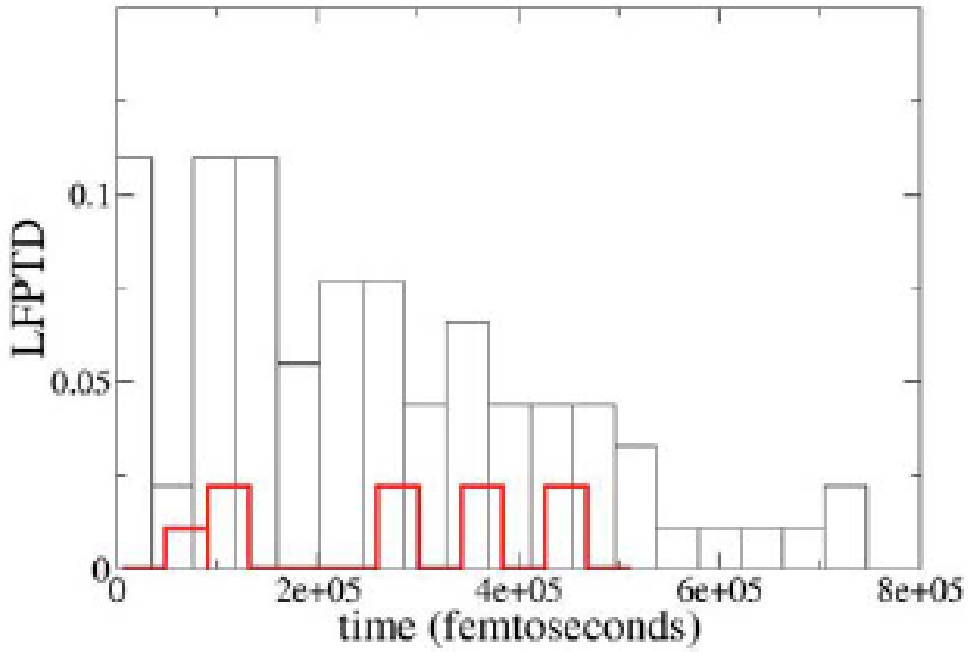


Figure 2: The Local First Passage Time Distribution (LFPTD) of milestone 7. The times of trajectories which were initiated at milestone 7 and terminated at the previous or the next milestone are binned. The times from milestone 7 to 6 are black and from 7 to 8 are red. Moving forward (from 7 to 8) is significantly less likely suggesting a free energy barrier (only 9 trajectories made it to milestone 8).

Water interface adjustments in the final step of the transition in accord with time resolved crystalography

rate 10 microseconds in accord with exp

Ron Elber, "A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin", Biophysical J. ,2007 92: L85-L87

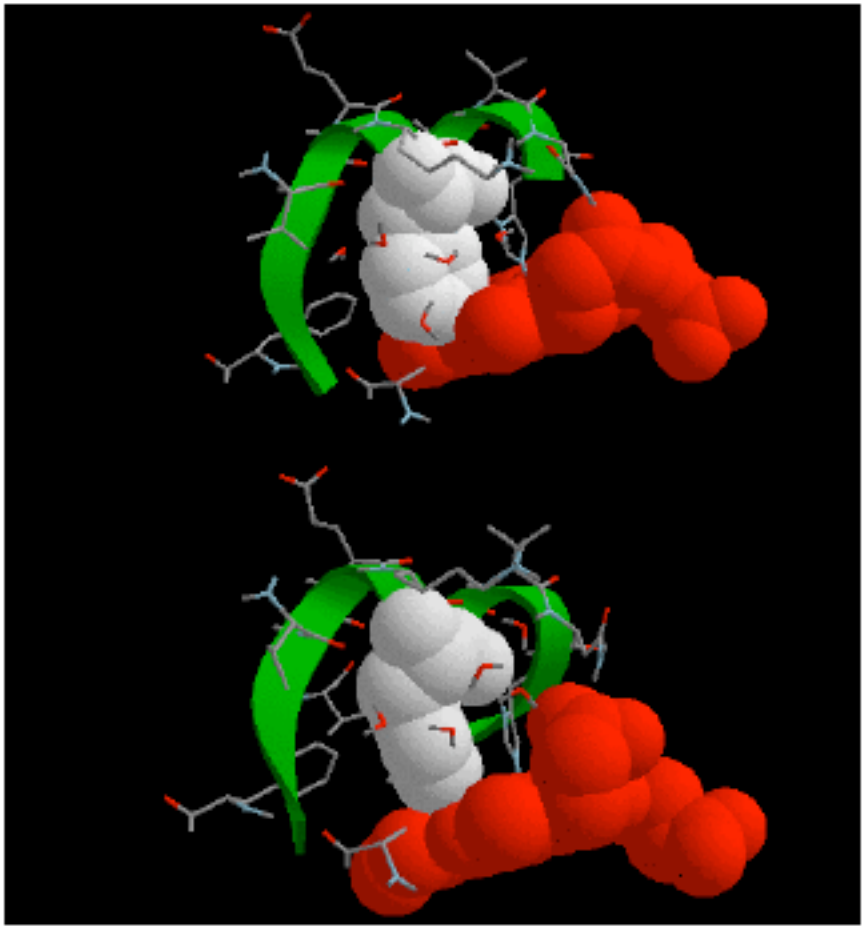
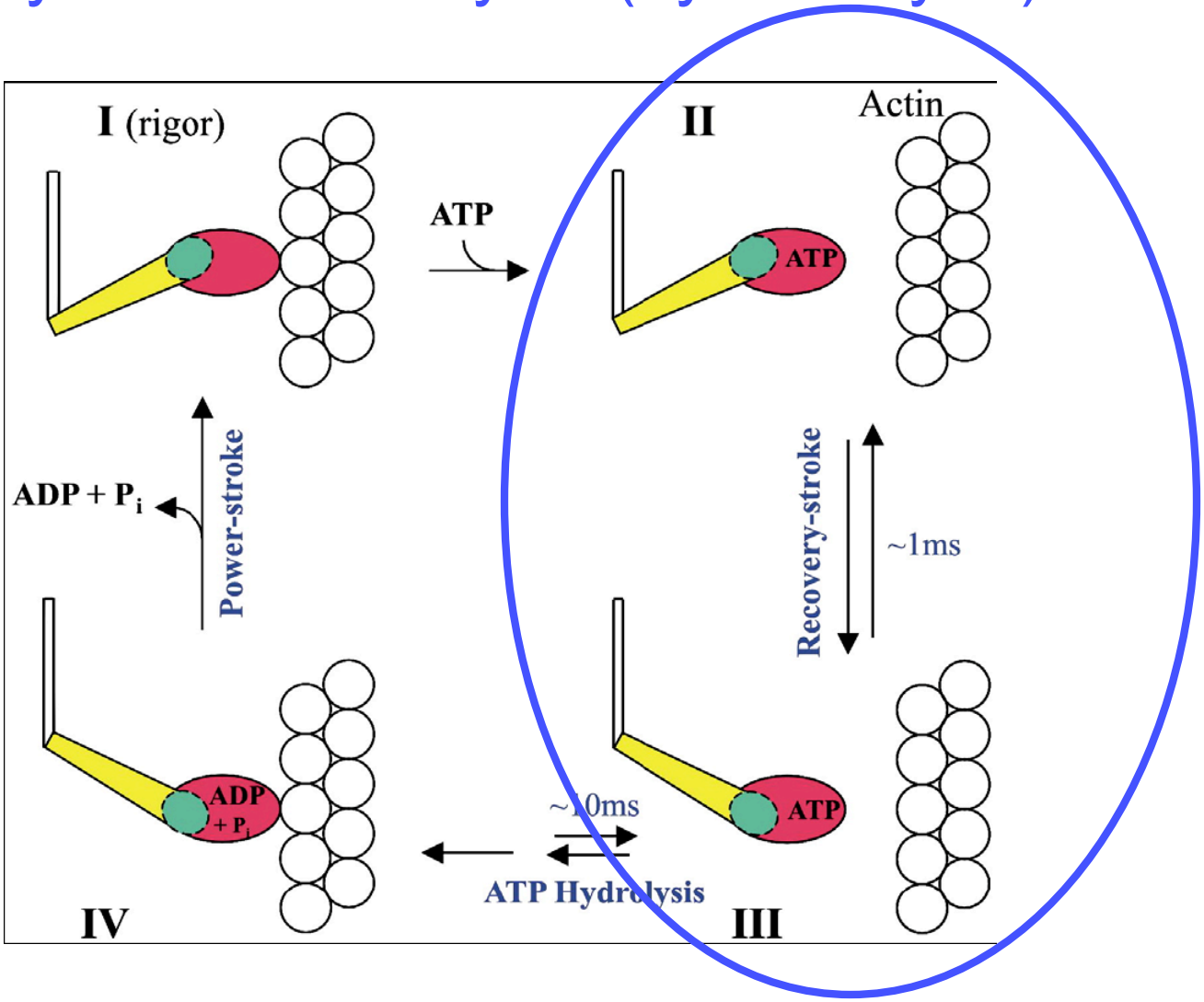


Figure 4: The difference in the packing of Phe-97 (white) against the heme (red) in milestone 7 (top figure) and the deoxy state (lower figure). While the phenylalanine packing is essentially complete, there remain numerous adjustments. The lysine and the water molecules should come higher and Phe-97 requires a slight rotation. The carboxyl groups of the heme also change orientation.

Myosin II

- Structural & metabolic unit of muscle tissue (with actin)
- Implicated in cytokinesis, chromosome separation, and DNA transcription
- Converts energy from ATP hydrolysis into mechanical work: *but how?*
- ATPase structure very similar to those of kinesin & dynein (flagellar motility & intracellular transport) and to Ras, G proteins

Myosin Stroke Cycle (Lymn-Taylor)



S. Mesentean et al., *JMB*, **367**, 591–602 (2007)

Recovery stroke mechanism still a mystery

- Proposals of Fischer et al. based on minimum energy path (1200-point MEP), and MD at endstates.
- Q. Cui et al. (*PLoS CompBio*, **3**, e23, Feb 2007) use 3 TMD trajectories, 1.2 ns: rotation occurs *without* H-bonding between ATP and Gly457; relay helix "unwinding" happens much later than in MEP.
- Milestoning simulates the *entire* stroke with unbiased MD: OEQ ensembles and FP trajs can suggest a mechanism with good statistics
- By comparing to experiment, resulting rate and free energy profiles are tests of the RC & mechanism

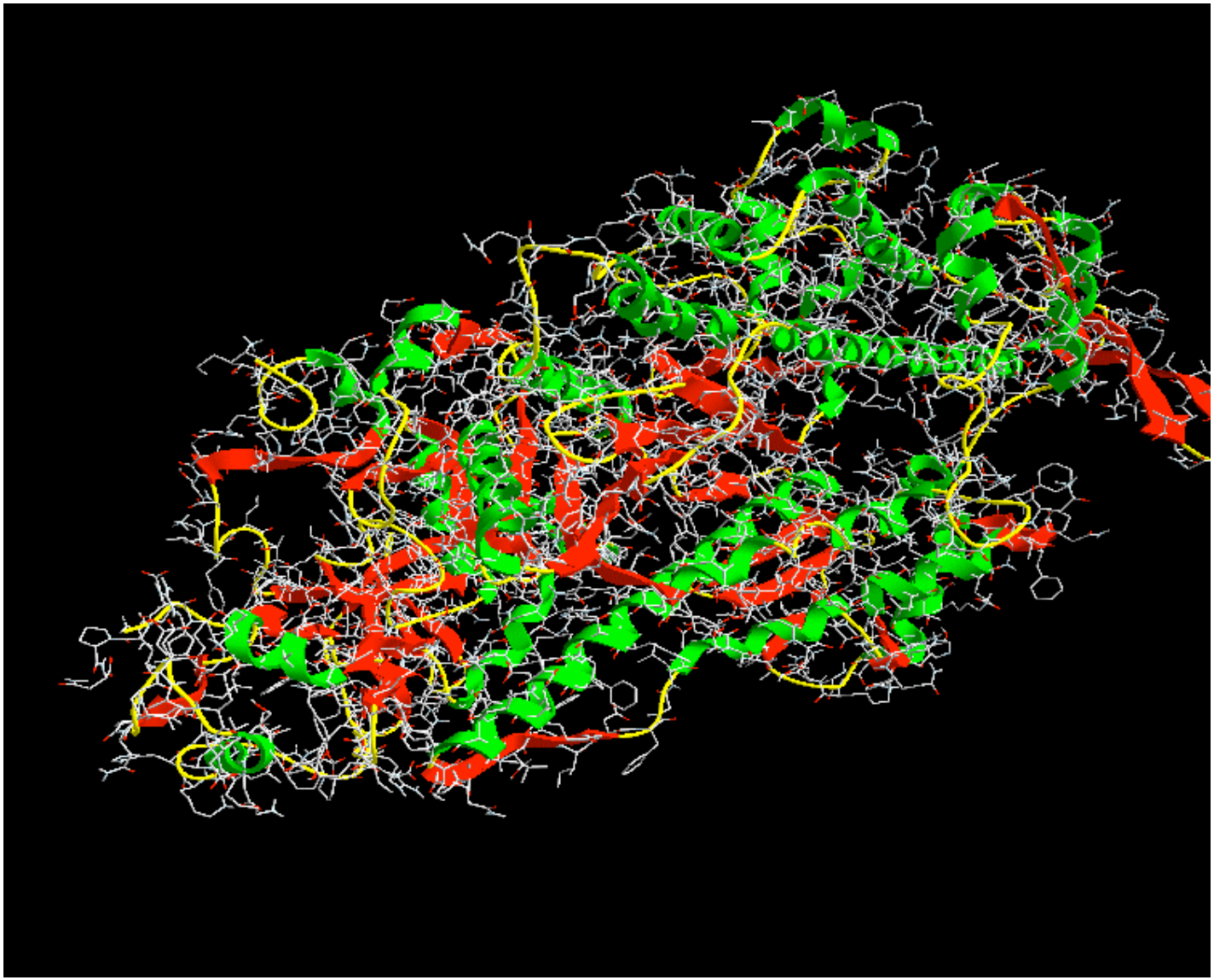
Reaction coordinate for the recovery stroke

- Use action optimization (provide the SDP for C_α coordinates as collective variables, Olender & Elber, 96)

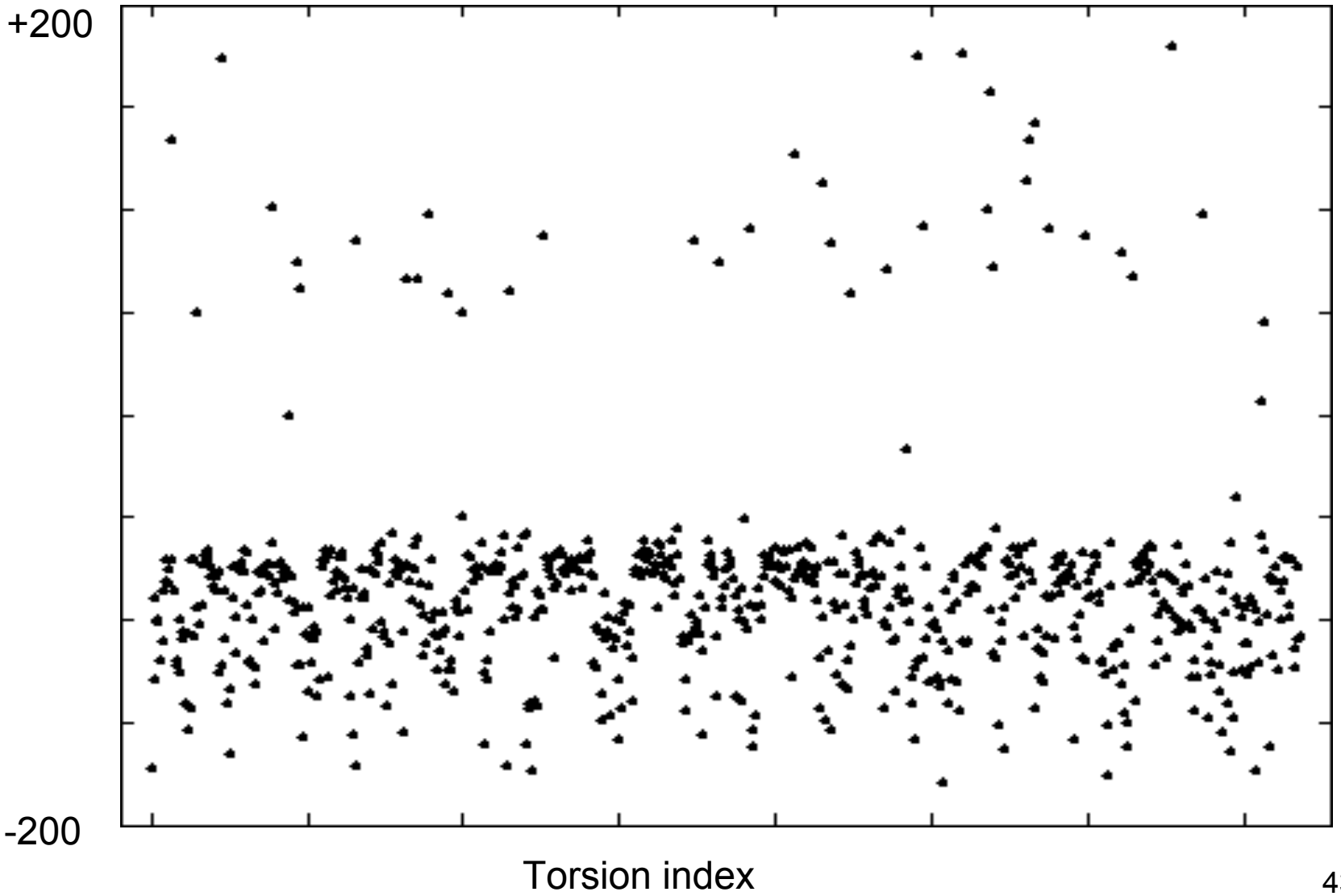
$$S = \int_{X_R}^{X_P} \sqrt{\nabla U^t \cdot \nabla U}\, dl + C\left(X(l)\right)$$

- optimize a path with 241 grid points (milestones)
  - Use GBSA for solvation
  - Note Milestoning is done with explicit solvation (TIP3P water molecules)

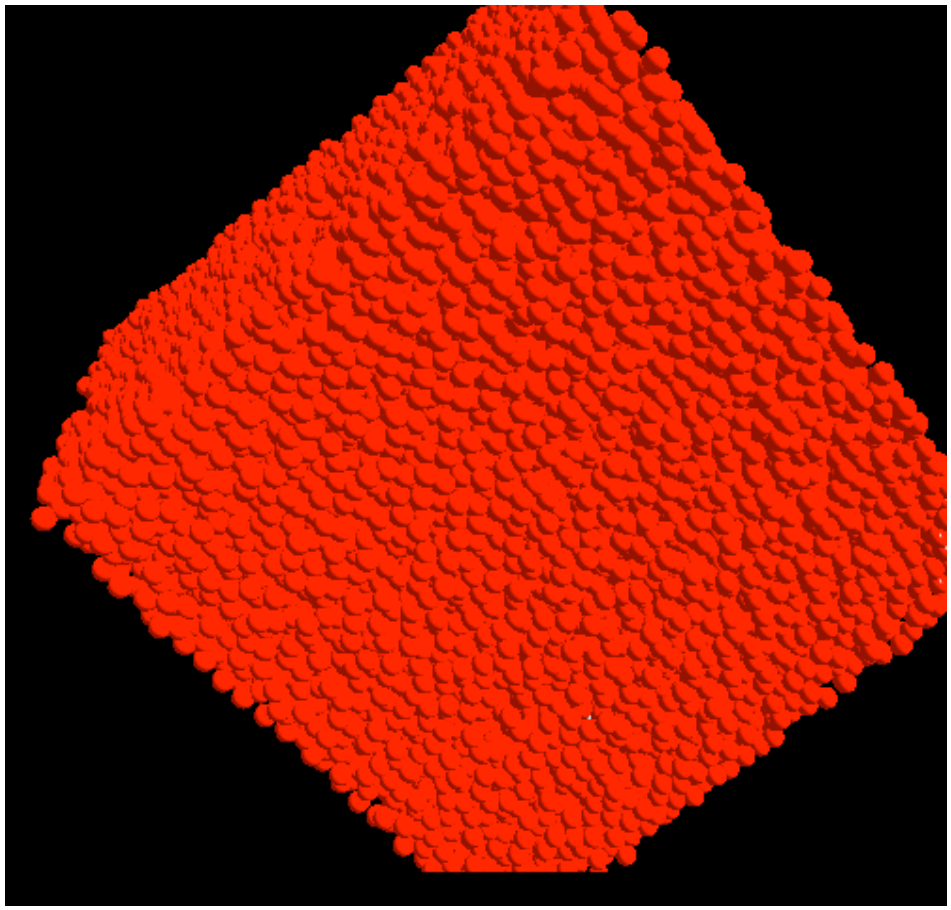
Movie of the reaction path for the recovery stroke in myosin



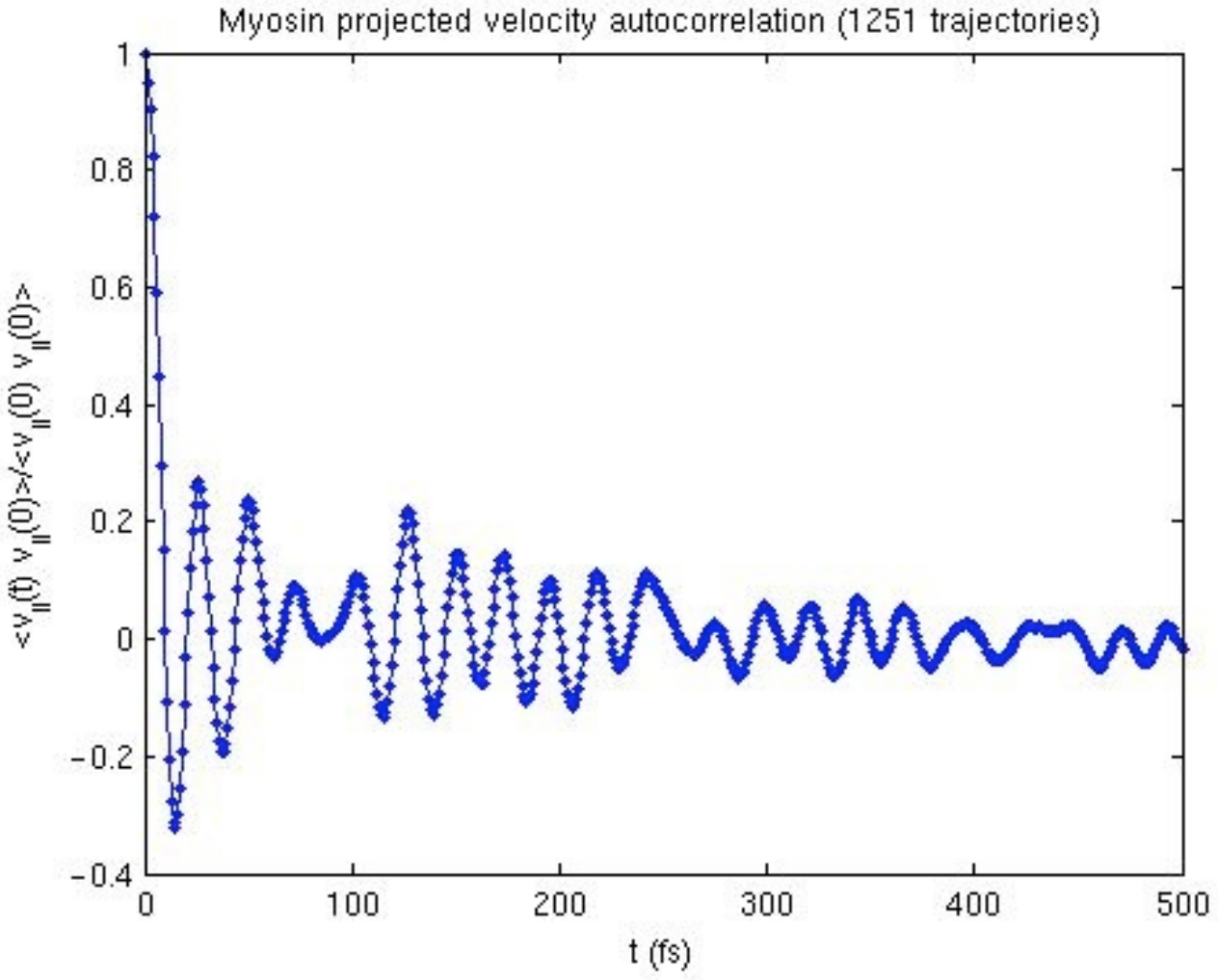
Contributions of different ψ dihedrals to the reaction coordinate



For Milestoning that path was placed in a box of water



Velocity correlation function along the reaction coordinate



Some results for myosin recovery stroke

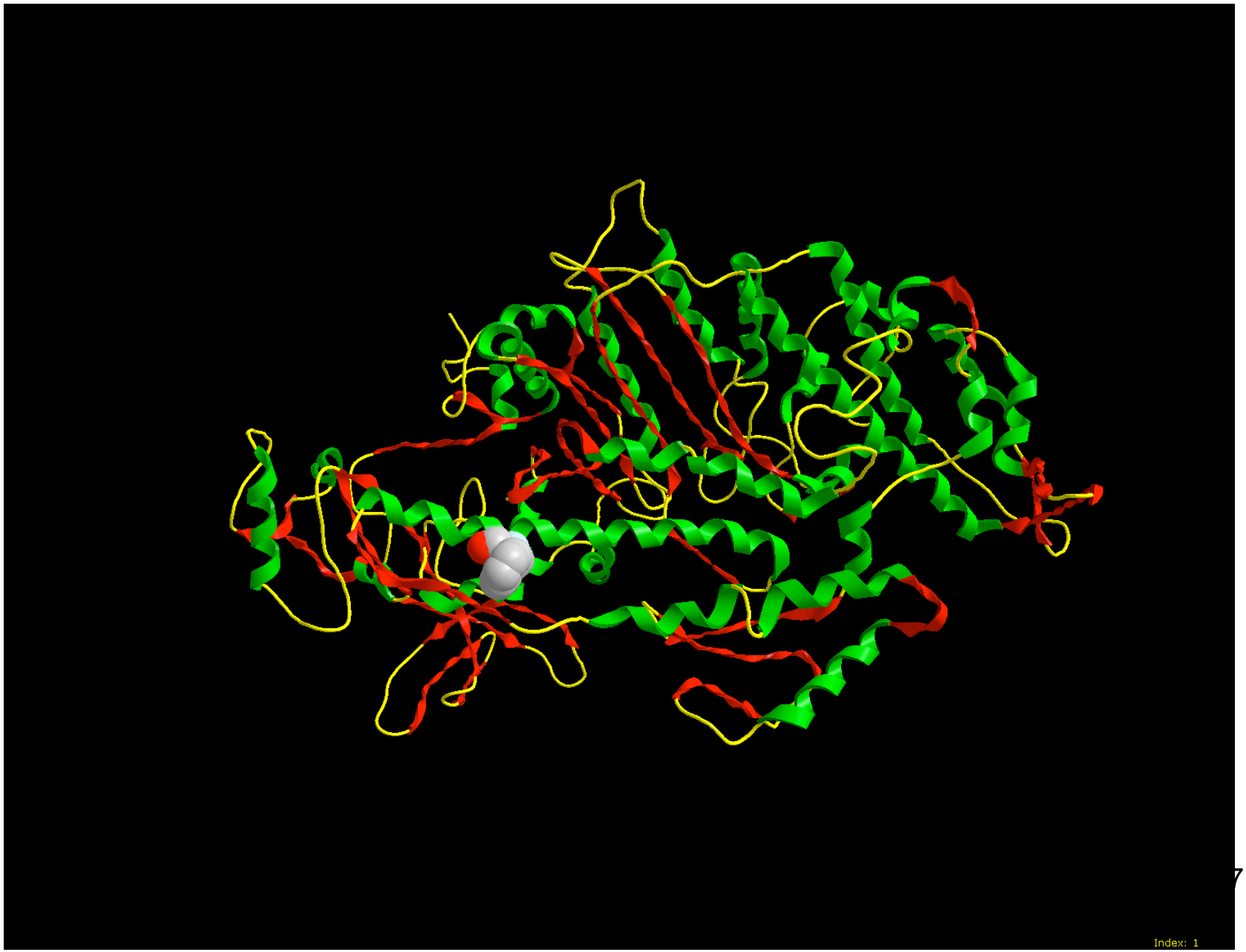
- Overall first passage time estimated as 0.05-0.5 ms for different conditions (experiment about 1-10ms). Statistical errors -- 10%

To gauge magnitude of force field errors two water models were used in two different Milestoning calculations --

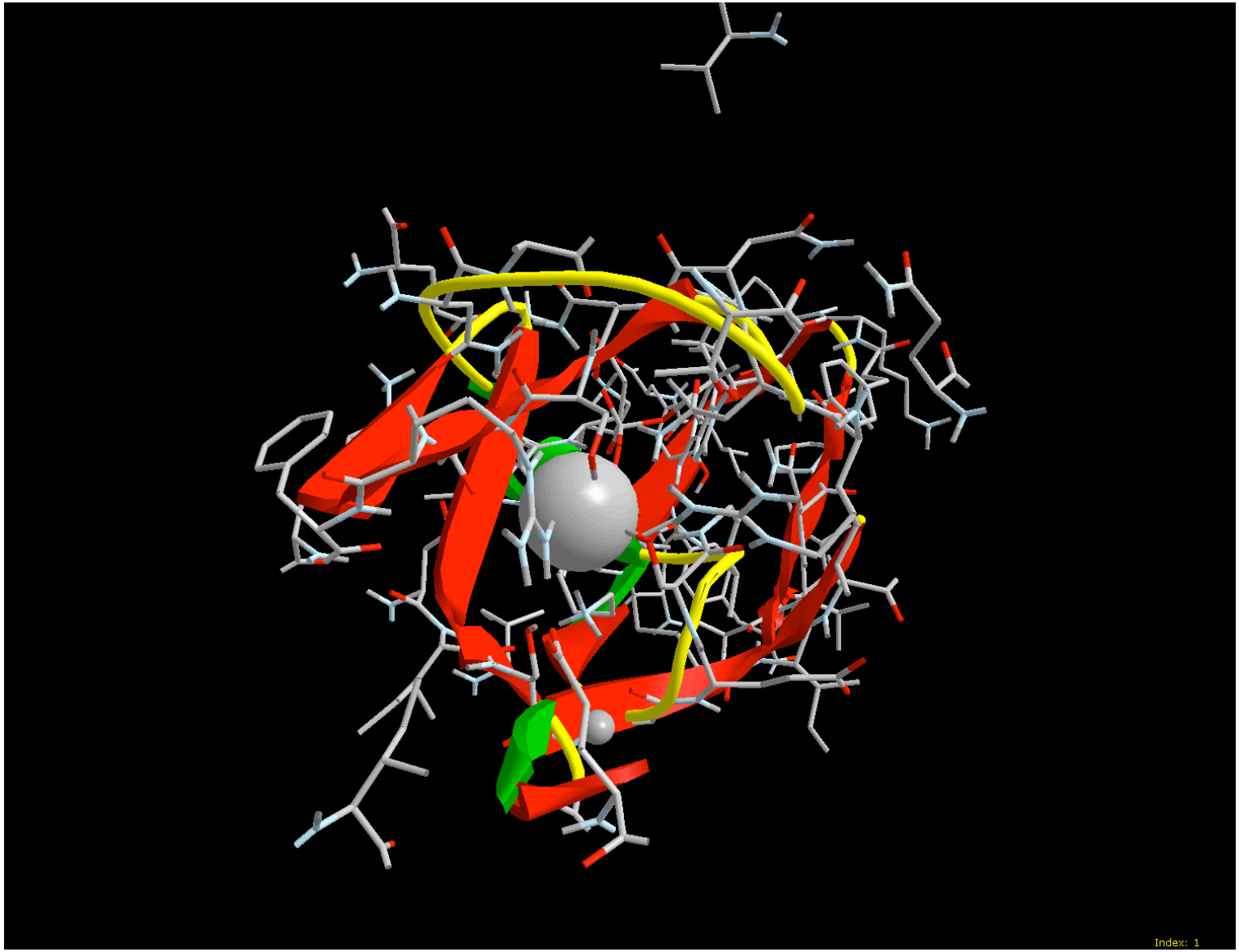
TIP3P 0.05ms

SPC/E 0.5ms

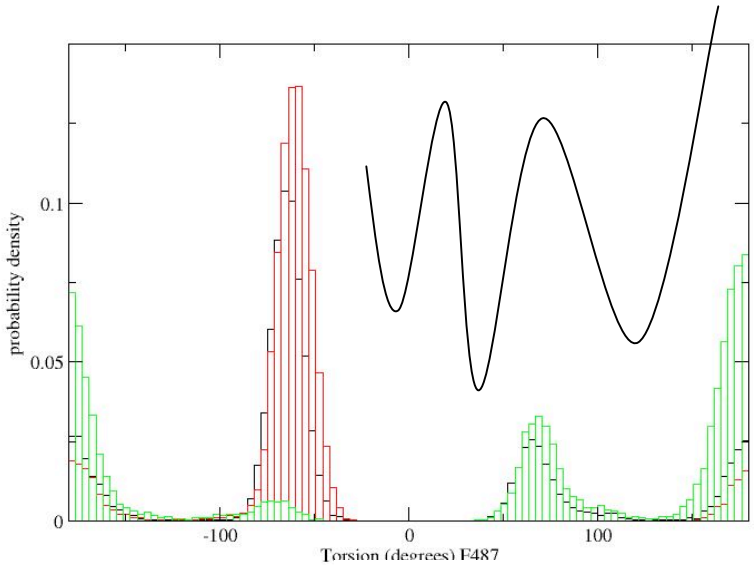
Coupling of local and global motions: Myosin and Phe 487

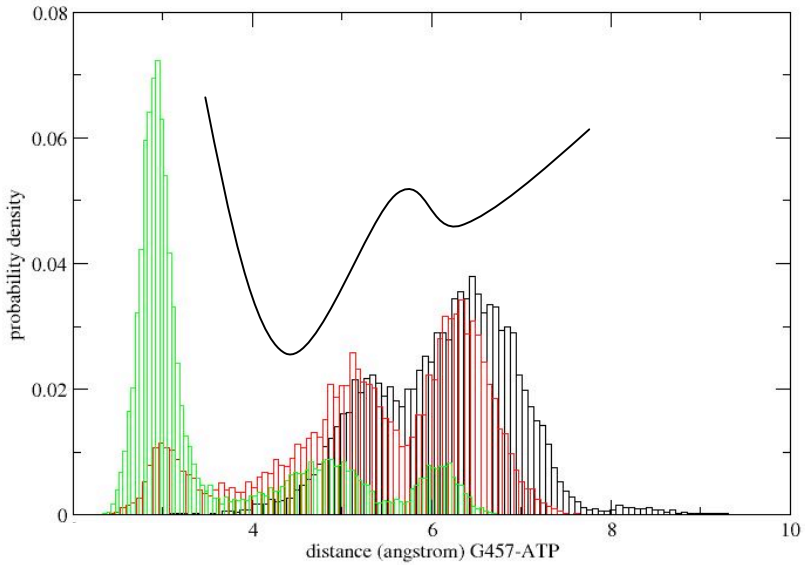


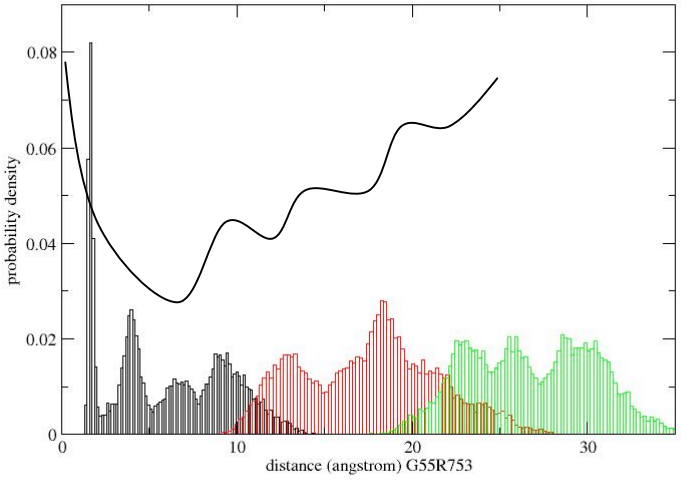
The kiss of death of glycine 457 to ATP

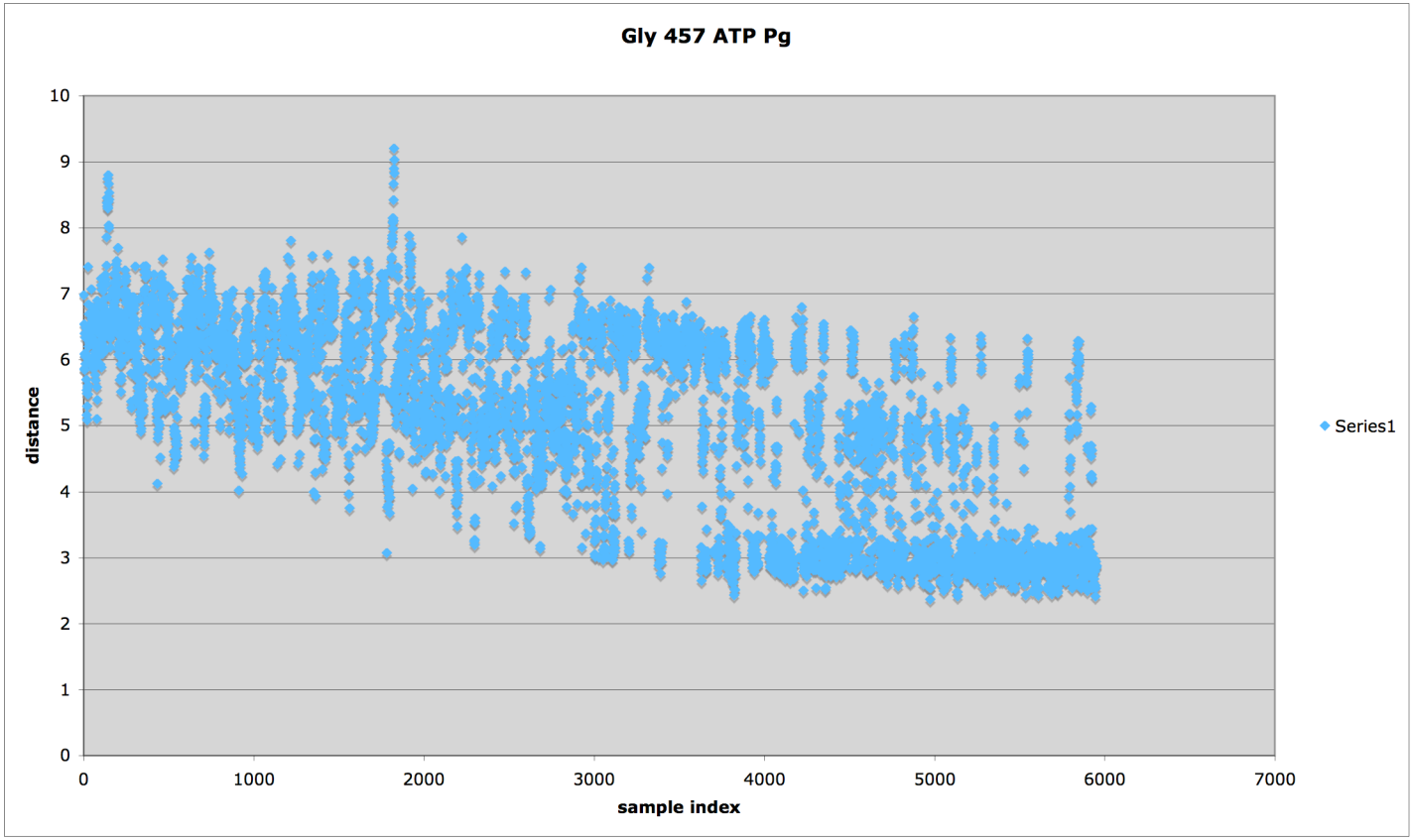


Simultaneous progression of local (activated) and global (diffusive) molecular events








Gly 457 ATP Pg
distance
sample index
0
1
2
3
4
5
6
7
8
9
10
0
1000
2000
3000
4000
5000
6000
7000
Series1

Summary

- Milestoning divides RC into fragments whose kinetics can be computed independently then "glued" together
- Provides factor of M improvement in computational efficiency on serial machines, plus exp bootstrapping
- Uses LFPTDs from microscopic dynamics: $K_{s,s'}(\tau)$
- System distribution $P_s(t)$ given by simple integral equations that can be easily solved numerically
- Correct kinetics for solvated alanine dipeptide (x 9 speedup)
- Predicts microsecond *Scapharca* rate with ~10 ns total serial time
- Sub-millisecond rate for myosin recovery stroke with total run time (serial) - speedup : more than 10,000

$$200\,sample \times 241 mlst \times 0.01 ns + 0.1 ns \times 241 = 501.6 ns$$

Final thanks

- **Dr. Tony West**
- Peter Majek

- Prof. Gouri Jas
- Prof. Krzysztof Kuczera

- Prof. Eric Vanden Eijnden
- Dr. Maddelena Venturoli

Milestoning papers

- Krzysztof Kuczera, Gouri Jas and Ron Elber, “The kinetics of helix unfolding: Molecular dynamics simulations with Milestoning”, J. Phys. Chem A, in press.
- Vanden Eijnden Eric, Maddalen Venturoli, Giovanni Ciccotti, and Ron Elber, “On the assumptions underlying Milestoning”, J Chem. Phys. 129,174102(2008)
- Ron Elber, "A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin", Biophysical J. ,2007 92: L85-L87
- Anthony M.A. West, Ron Elber, and David Shalloway, "Extending molecular dynamics timescales with milestoning: Example of complex kinetics in a solvated peptide", J. Chem. Phys. 126,145104(2007)
- Anton K. Faradjian and Ron Elber, "Computing time scales from reaction coordinates by milestoning", J. Chem. Phys. 120:10880-10889(2004)
- Code (moil + zmoil) available from https://wiki.ices.utexas.edu/clsb/wiki

Comparison to TPS

- TPS provide a rate constant for a high energy barrier
  - Rate constant exists
  - Based on trajectory enrichment
  - Trajectory computed sequentially (must be short)
  - No need for a reaction coordinate

• Milestoning computes arbitrary kinetics on rough energy landscapes

- •Requires a reaction coordinate
- •Non-exponential kinetics is fine.
- •Trajectory computed in parallel (can be long)

Comparison to PPTIS

PPTIS

- Use reaction coordinate
- Markovian process in time (exponential kinetics)
- Trajectory enrichment (sequential calculation)

Milestoning

- Use reaction coordinate
- Non-Markovian model in time
- Lost of spatial memory between milestones (allows for efficient parallelism)

Markov Approximation

$$j_s^{\pm} \equiv \frac{\int_0^{\infty} K_s^{\pm}(\tau)\,d\tau}{\int_0^{\infty} \tau\left(K_s^{+}(\tau)+K_s^{-}(\tau)\right)d\tau}$$

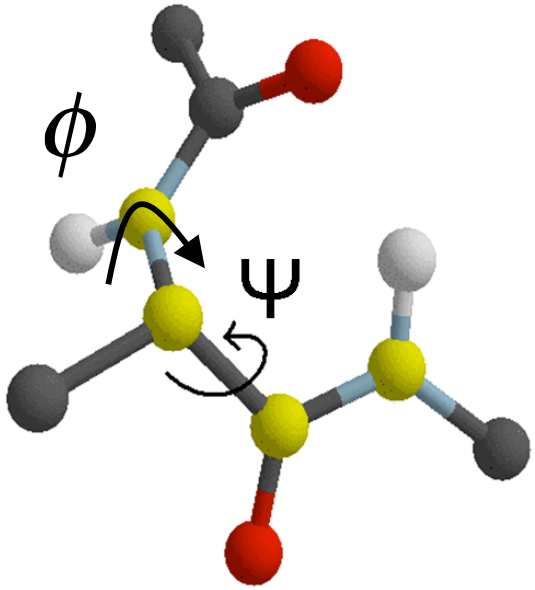
$$W = \begin{bmatrix} -j_1 & j_2^- & 0 & & & & \\ j_1^+ & -j_2 & j_3^- & & & & \\ 0 & j_2^- & -j_3 & \cdots & & & \\ 0 & 0 & j_3^+ & \ddots & & & \\ & \vdots & & & j_{M-1}^- & 0 \\ & & & & -j_{M-1} & j_M^- \\ & & & & j_{M-1}^+ & -j_M \end{bmatrix}$$

$$\dot{P}_s(t) = \sum_{s'} W_{ss'} P_{s'}(t)$$

MFPT is preserved, but 2nd FPTD moment is not

Haenggi & Talkner, *PRL*, **51**, 2242 (1983)

Alanine dipeptide

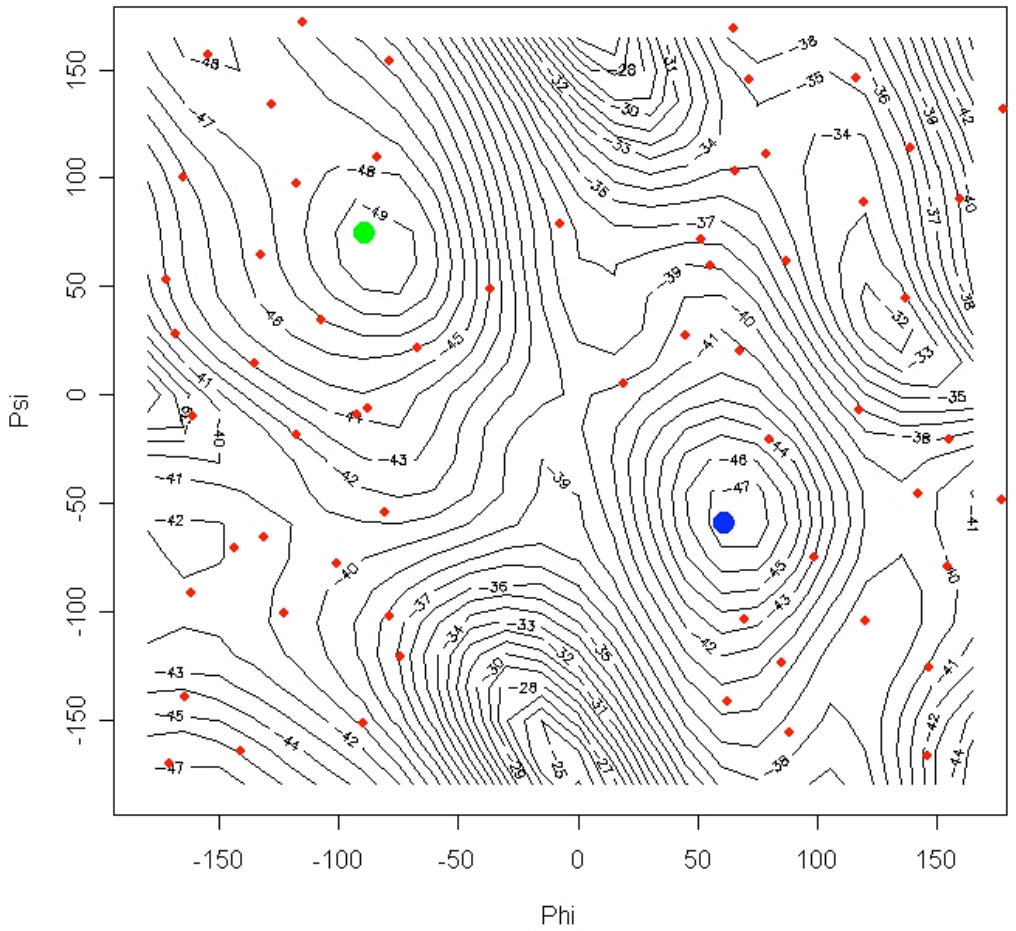


Alanine dipeptide: Milestoning by images/Voronoi cells (Peter Majek)

method	mean first passage time from blue -> green	total simulated time*
brute force MD calculation	320 ns +- 20 ns	80 micro secs
2000 - 6000 trajectories / milestone	311 ns +- 49 ns	6 micro secs

A point belongs to the nearest image

Compute first Passage time distributions Between cells - $K_{ij}(\tau)$



Milestoning: From atomically detailed simulations to stochastic equations with memory

- Why add memory?
  - Rigorous theory
  - More detailed information
  - No need to identify metastable states
  - Compute overall time dependence (not only rate constants)
  - Computing memory effects is actually straightforward.
- Building on the connection between CTRW (continuous time random walk) and GME (Generalized Master Equation). (Montroll)

Milestoning:

- Determine reaction coordinate (q)/ cell structure between A and B
- Define hypersurfaces / cells along the guiding path.
- Compute short trajectories to estimate first passage time distribution between hypersurfaces / voronoi cells
- Use the short time trajectories to compute overall rate.