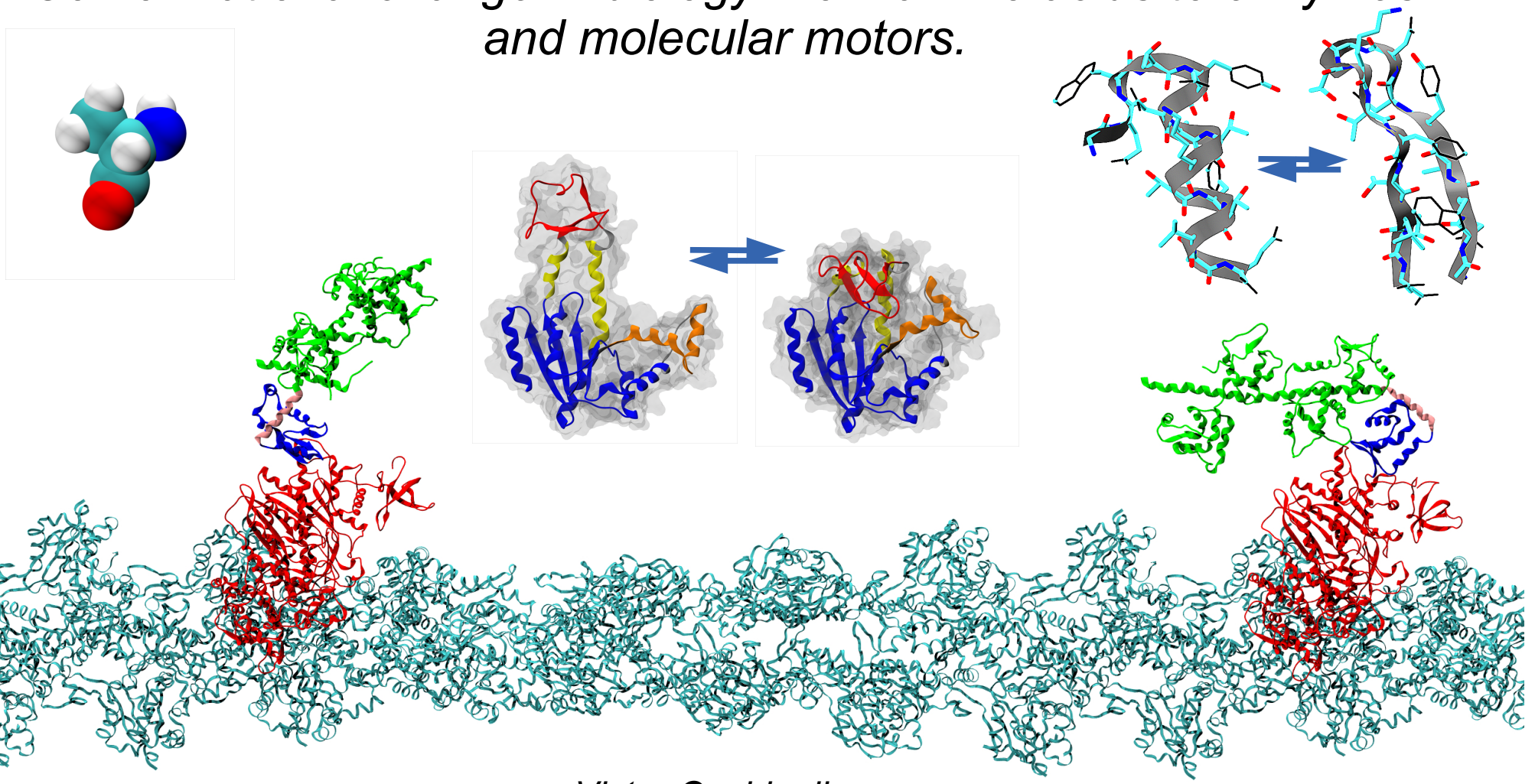
Conformational change in biology: from amino acids to enzymes and molecular motors.



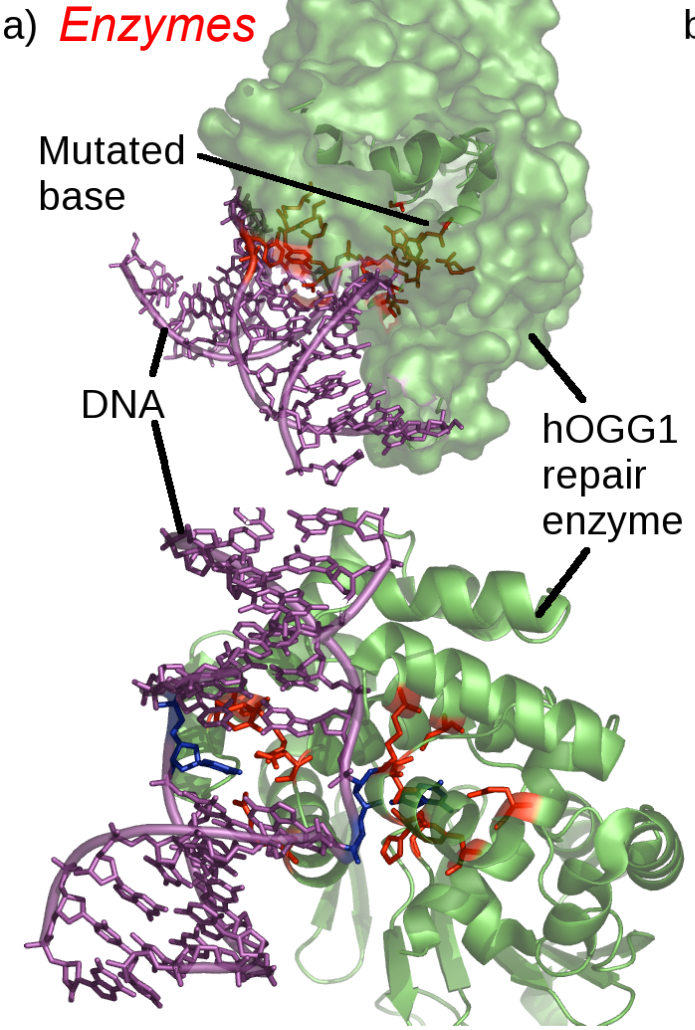
Victor Ovchinnikov

Computational Facilities: NERSC
Collaborators: Martin Karplus, Eric Vanden-Eijnden, Kwangho Nam, Anne Houdusse, Robert Sauer
Financial support: NIH

Introduction

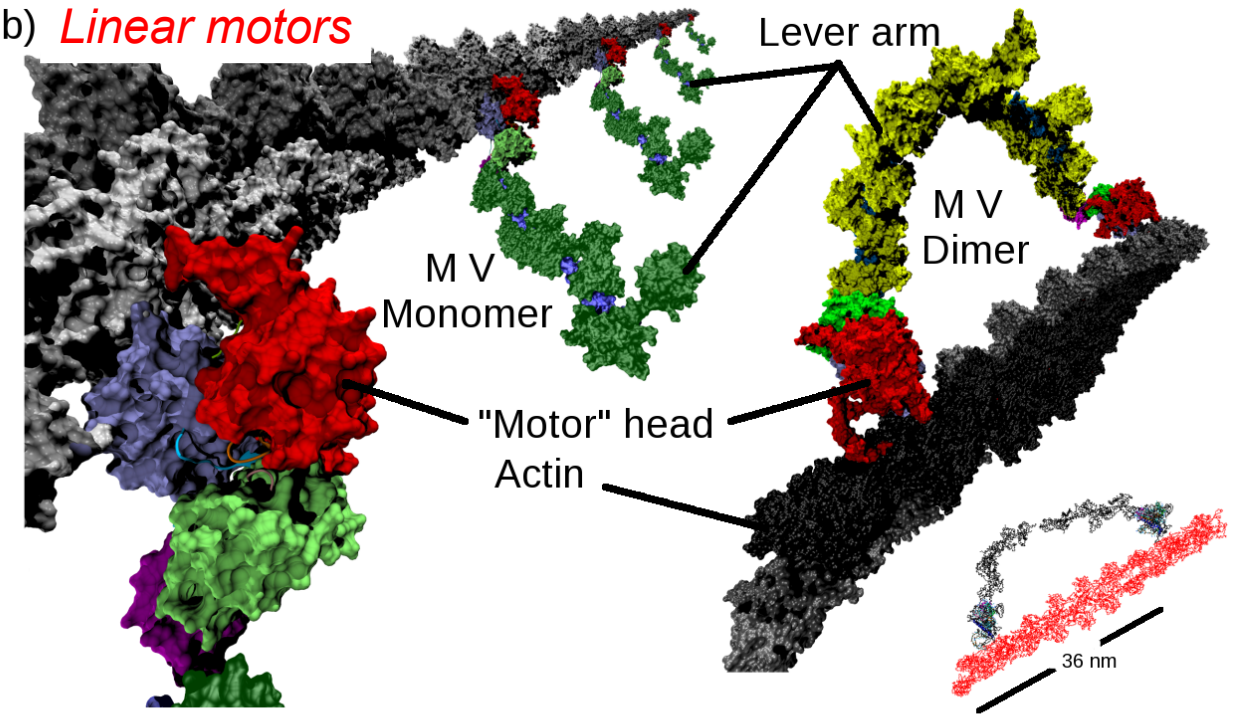
- Conformational motions in biomolecules define all living things
  - Transport across membranes
  - Enzyme reactions (from proton transfer to DNA replication and repair)
  - Linear (Myosin, Kinesin) and Rotary (F_1ATPase, ClpX Dynein) motors
- We would like to understand how chemical energy is used to generate force and motion at the molecular level
  - Biological processes are
    - Inherently “renewable”
    - Efficient (light harvesting by photosynthetic bacteria)
    - Robust *w.r.t.* environmental perturbations (*e.g.* from temperature changes to antibiotics)

a) *Enzymes*



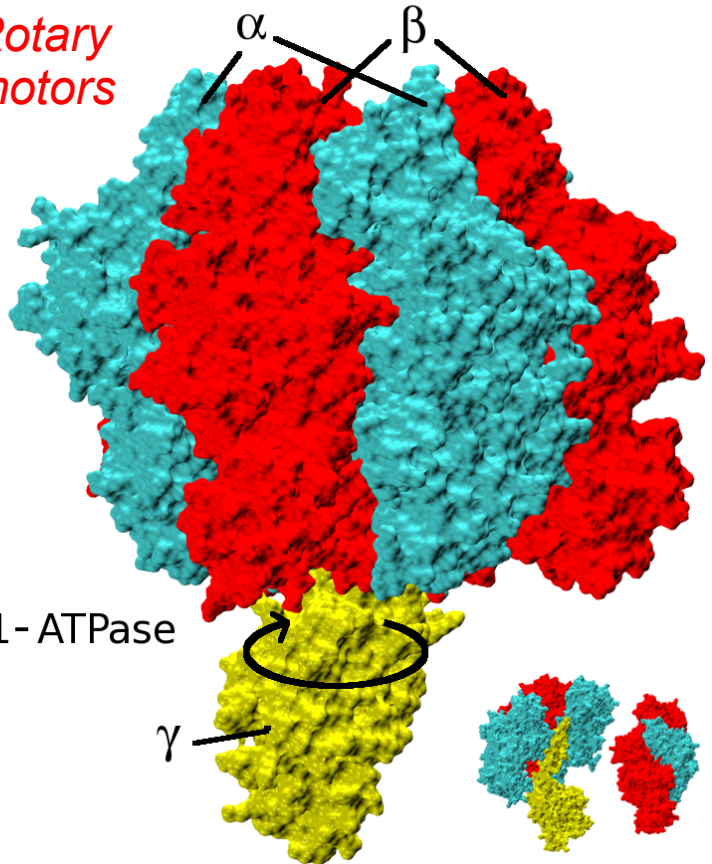
Human DNA repair enzyme hOGG1 bound to site of DNA mutation

b) *Linear motors*



Myosin V dimer taking variable steps

c) *Rotary motors*



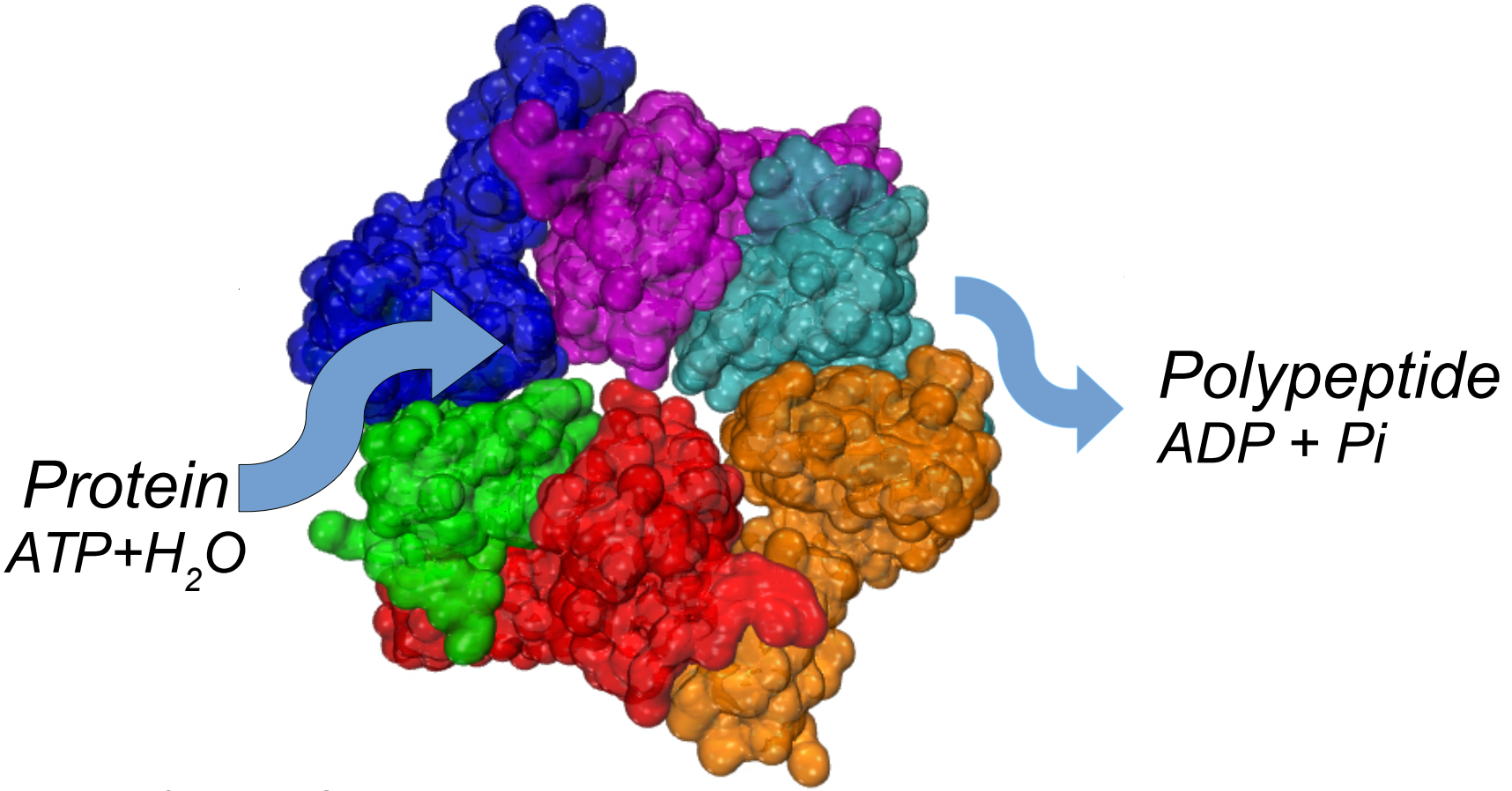
(With Profs. Kwangho Nam, Xray structures : Greg Verdine)

$$ATP + H_2O \rightleftharpoons ADP + Pi + work$$

Example: ClpX protein unfolding machine

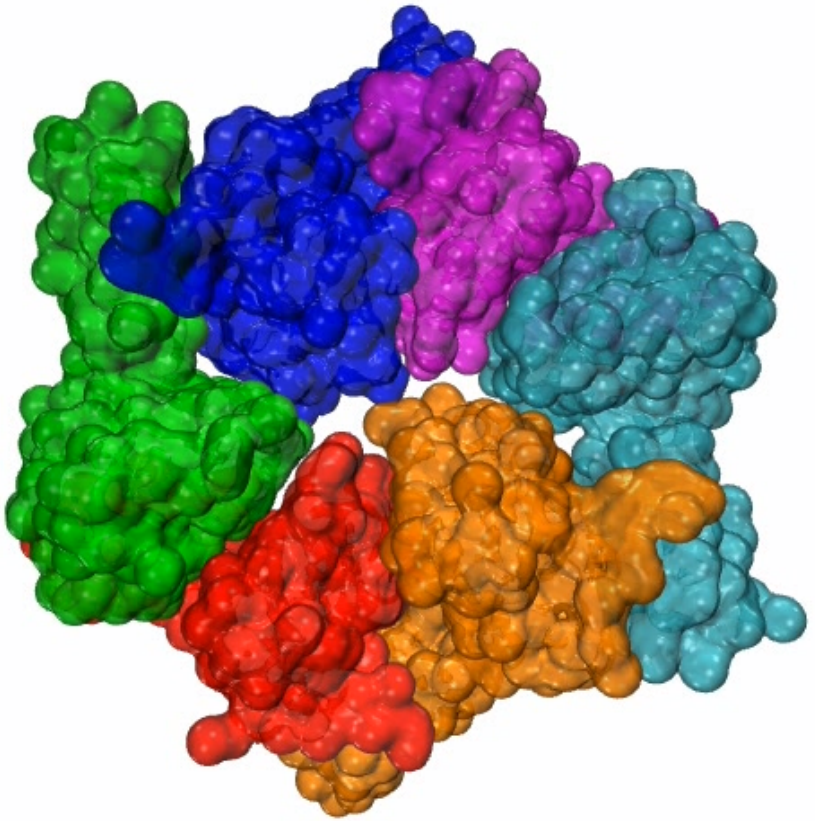
- Part of protein “recycling” complex
- Six subunits (identical in sequence) adopt different nucleotide-dependent (ATP *vs.* ATP conformations
- Energy of ATP hydrolysis coupled to “threading” motion

$ATP + H_2O \rightarrow ADP + Pi + work$



Collaboration with Prof. Robert Sauer, MIT

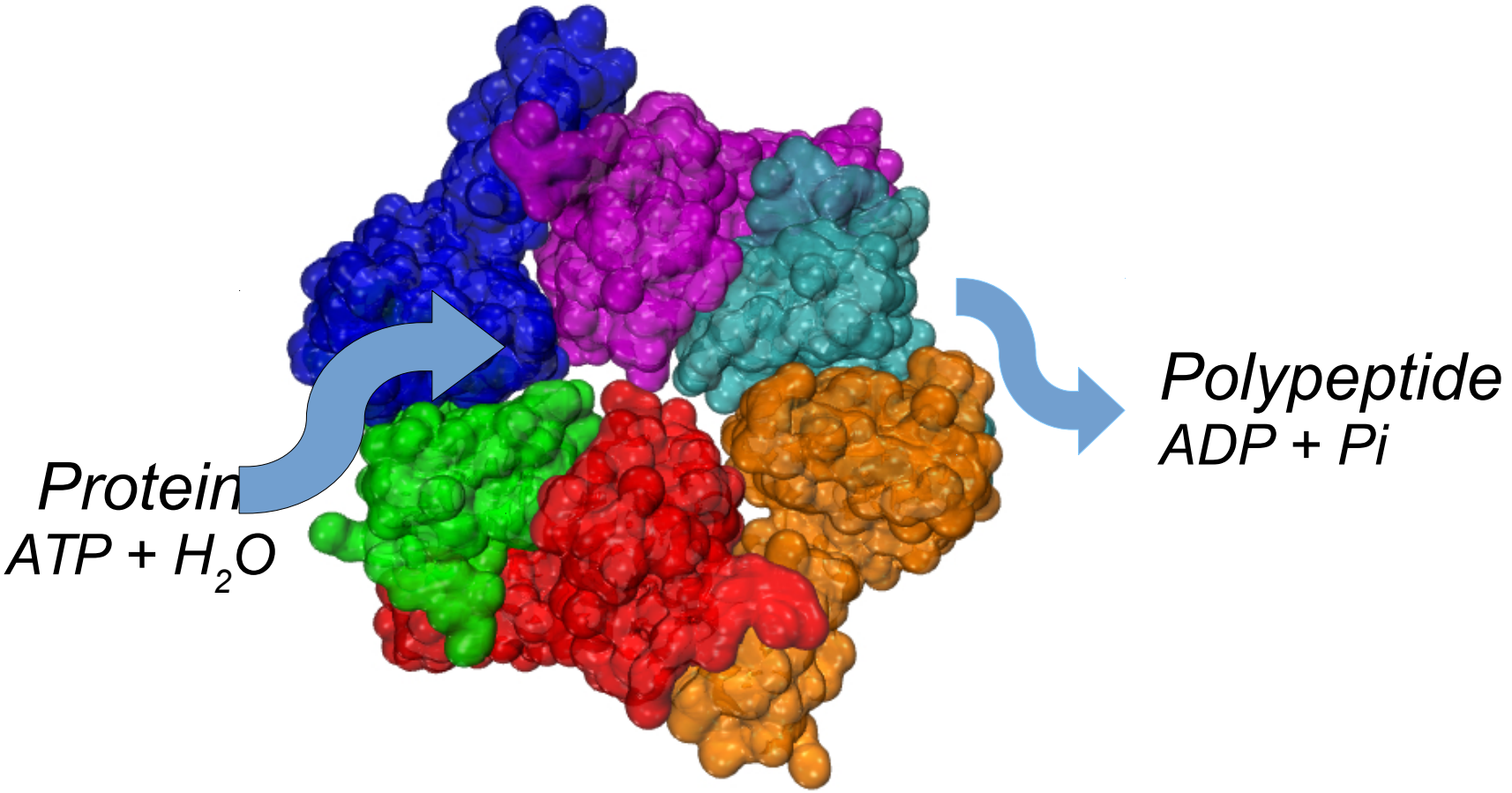
Example: ClpX protein unfolding machine



Collaboration with Prof. Robert Sauer, MIT

Example: ClpX protein unfolding machine

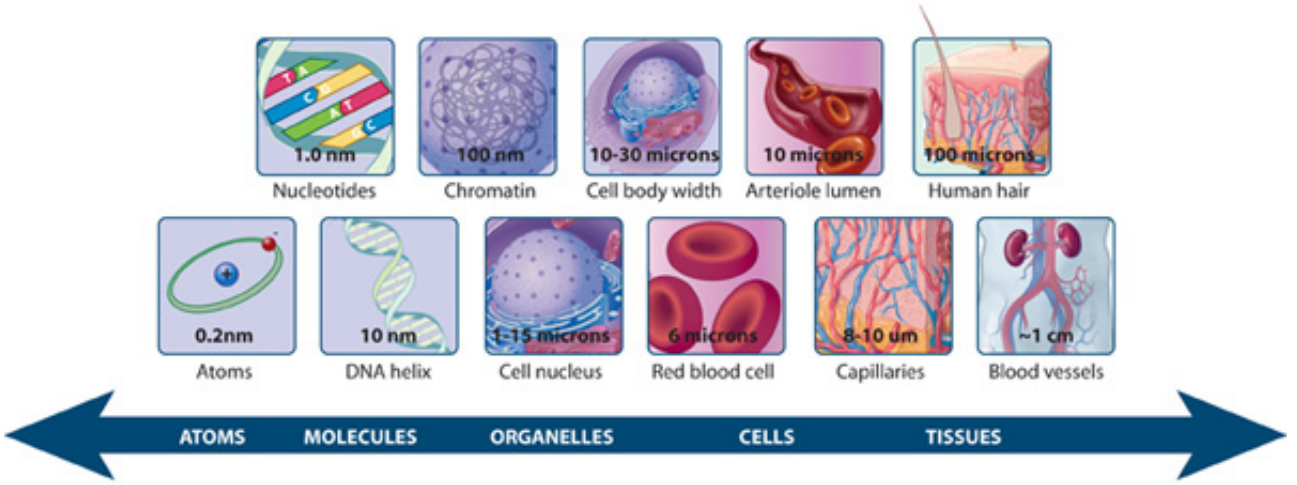
- How are force generation and motion coupled to hydrolysis ?



Collaboration with Prof. Robert Sauer, MIT

Challenges

- Experiments can be difficult
- Small spatial scales : ~10^{-9} m, variable temporal scales: 10^{-12} – 10^{-3} s



- Thermal fluctuations can complicate measurements
- Carefully parametrized computer simulations can help elucidate basic thermodynamics and kinetics, *e.g.* :

1 Stability of intermediates ($\Delta F = \Delta H - T\Delta S$)

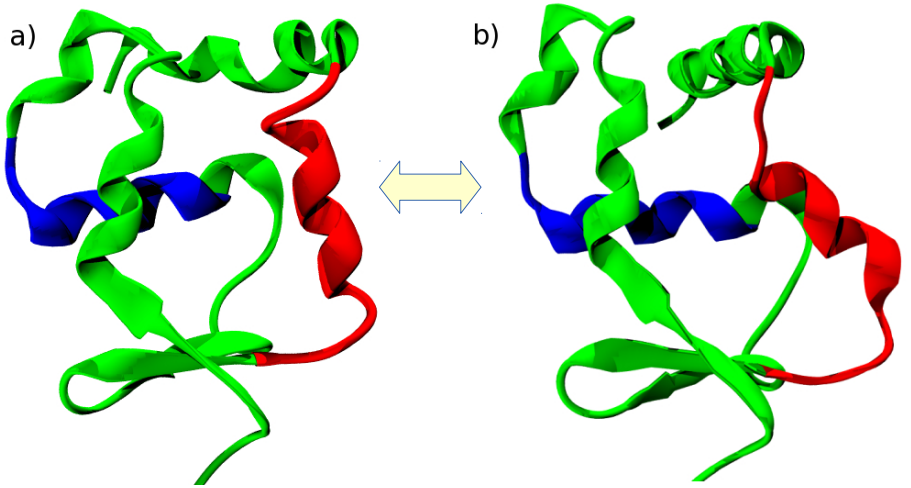
2 Pathways of transition

3 Rate of transition, *e.g.* Markov state models or TST: $k_{\mathrm{TST}} = \kappa e^{-\frac{\Delta F_{\mathrm{TST}}}{k_B T}}$

Image http://www.nature.com/scitable/content/the-relative-scale-of-biological-molecules-and-14704956

Theoretical/Computational Challenges (I)

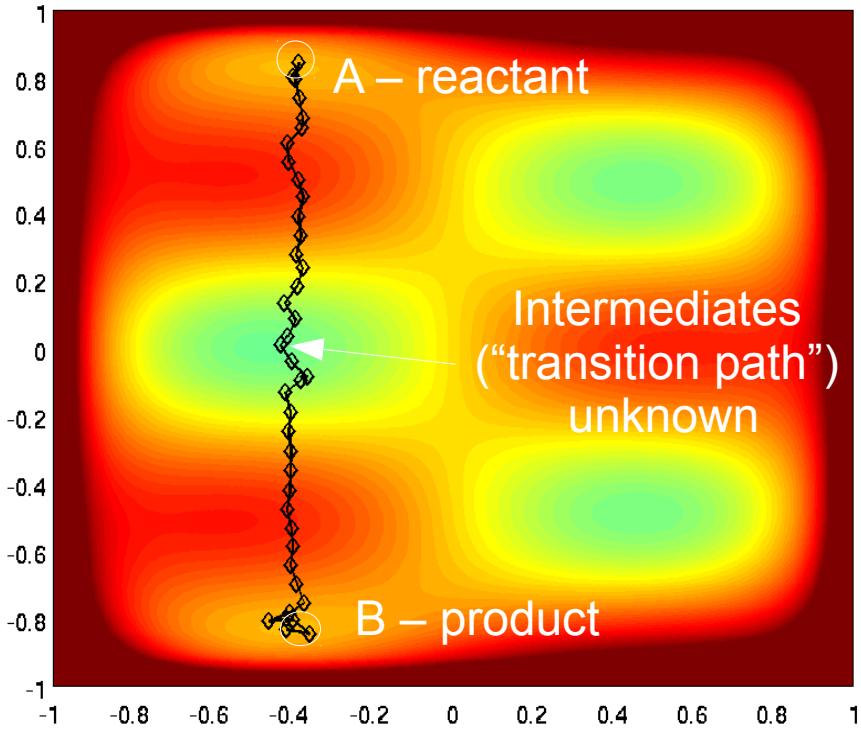
- Endpoints are provided from crystallography (or NMR)



- Intermediates must be found (computed)

String methods (2D landscape example)

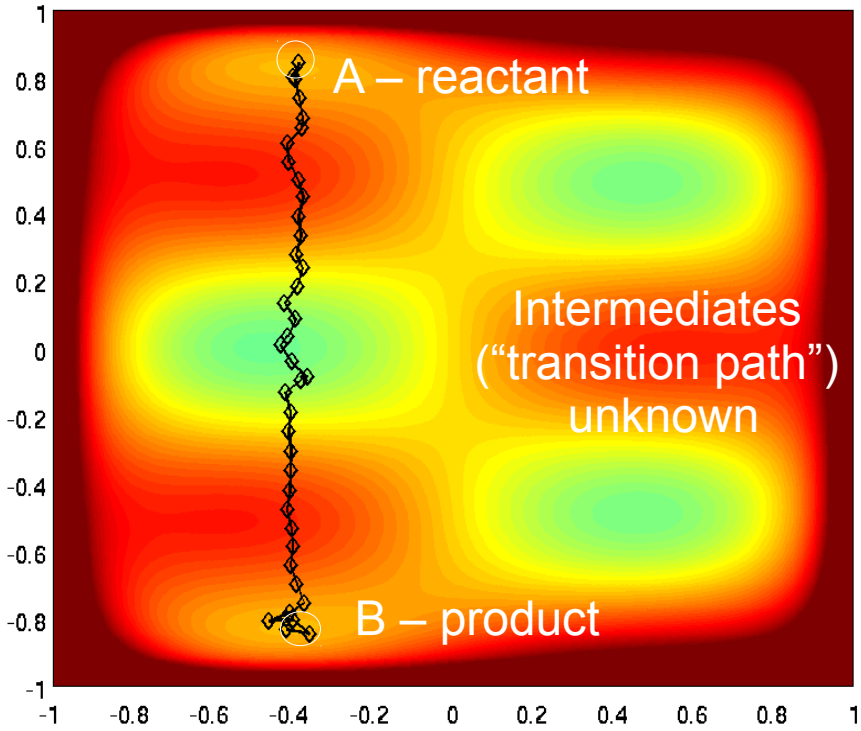
- Transition pathways between endpoint coordinates can be obtained using equilibration in “path space”

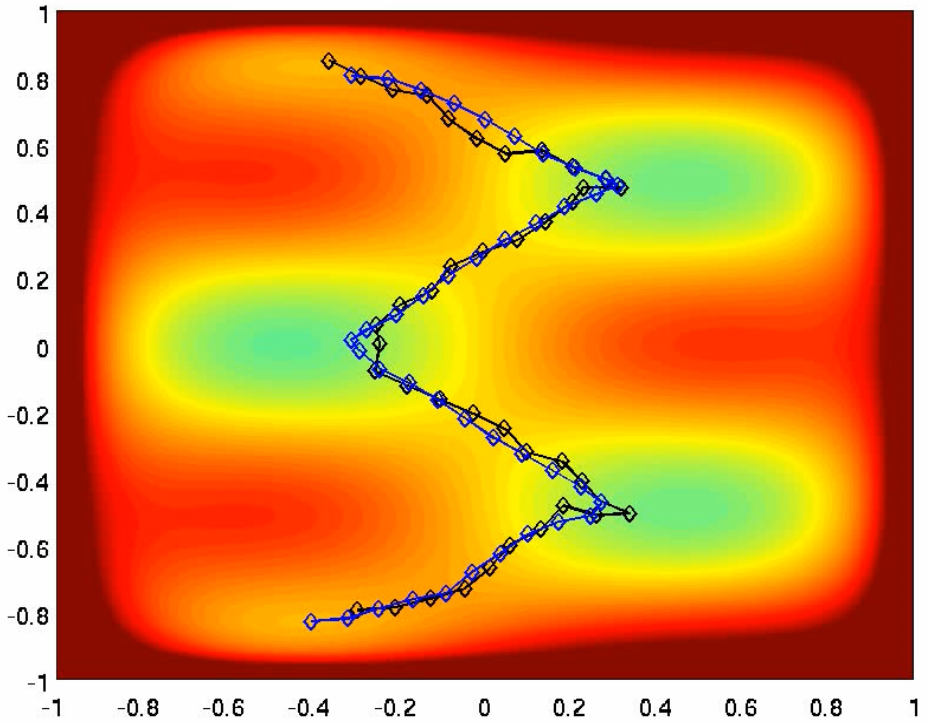


- Note: a medium-sized protein, e.g., calcium-binding calmodulin has 3N-6 ~ 7000 dof

String methods (2D landscape example)

- Each string point is an (almost) independent MD simulation





- Note: a medium-sized protein, e.g., calcium-binding calmodulin has 3N-6 ~ 7000 dof

Theoretical/Computational Challenges (II)

- "States" { B_i } must be defined precisely:
- *e.g.* reactant, product, transition state

$$F_{Bi} = -k_B T \log \int_{B_i \in \mathbb{R}^{3N-6}} e^{-\beta E(\vec{r})} d\vec{r}$$

(free energy of state X_i)

$$\cup B_i \supseteq \mathbb{R}^{3N-6}$$

- Many configurations r comprise a state (but which ones?)

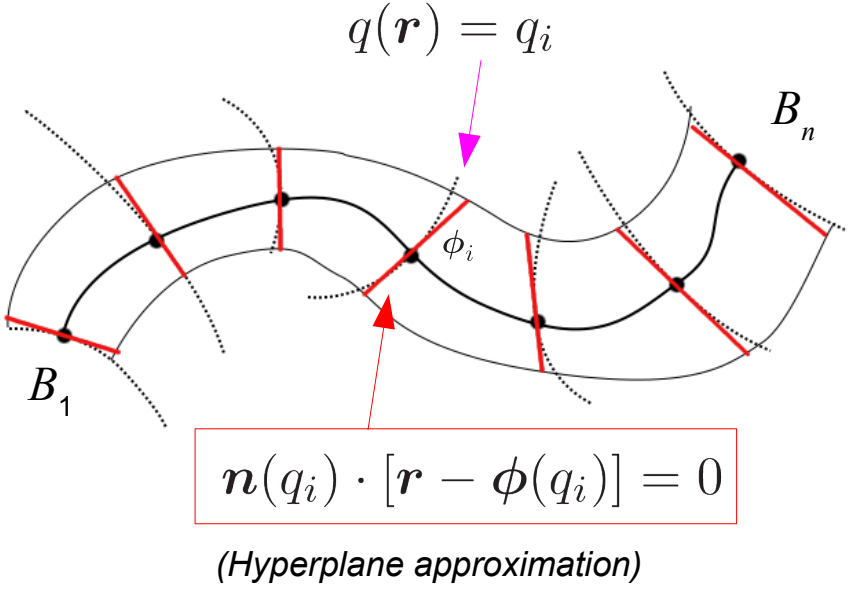
- Also known as the "reaction coordinate problem": which r correspond which stage of reaction (reactants, products, intermediates) ?

Transition path theory and string methods

$\phi(q) \in \mathbb{R}^{3N-6}$ average path (string)
$q \in [0,1]$ string parameter
$\boldsymbol{n}(q)$ normal to hyperplane

- For overdamped LD:
  states are locally separated by hyperplanes perpendicular to "average" path

$$\boldsymbol{n}(q_i) \parallel \boldsymbol{\phi}'(q_i)$$



(Hyperplane approximation)

- Can also change variables

$$\boldsymbol{r} \to \boldsymbol{\xi}$$

$$\boldsymbol{n}_{\boldsymbol{\xi}}(q_i) \parallel \underbrace{\langle \nabla_{\boldsymbol{r}}(\boldsymbol{\xi})^T \nabla_{\boldsymbol{r}}(\boldsymbol{\xi}) \rangle^{-1}_{\xi=\phi}}_{M^{-1}} \boldsymbol{\phi}'_{\boldsymbol{\xi}}(q_i)$$

(*M*: Metric tensor)

- This is the basis for the string method in collective variables

* W. E, W. Ren & E. Vanden-Eijnden, 2005. Finite-temperature String Method for the Study of Rare Events. *J. Phys. Chem. B*, **109**, 6688-6693)

Free energy from tessellations

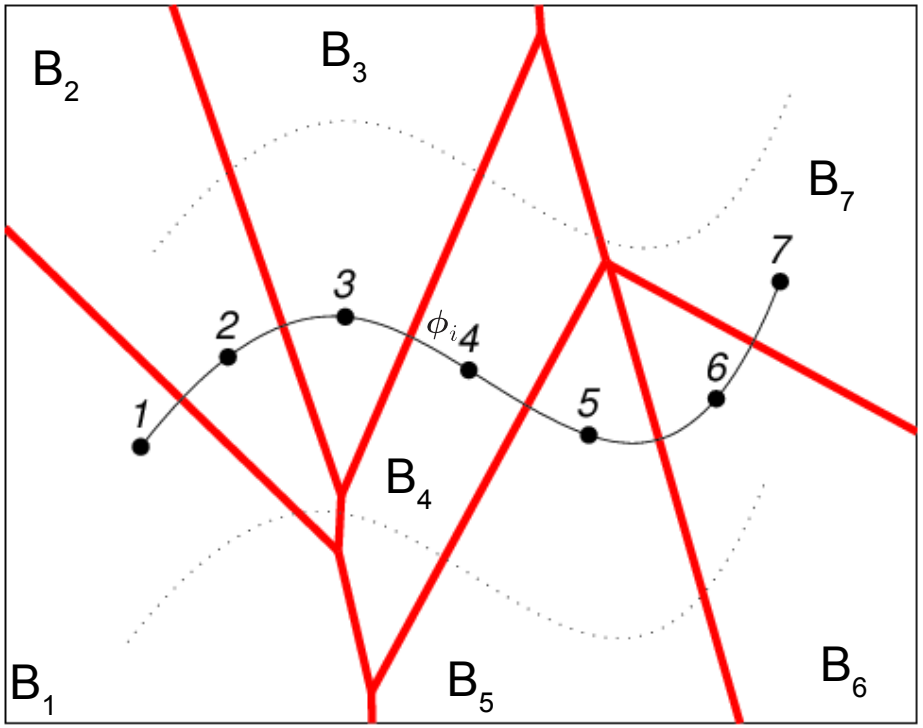
- The free energy can be computed from a *tessellation*, e.g.:

$$B_i = \{\boldsymbol{r} : \|\boldsymbol{r} - A\boldsymbol{\phi}_i\| < \|\boldsymbol{r} - A\boldsymbol{\phi}_j\|,\ \forall j \neq i\}$$

by constrained MD in each B_i and flux matching

$$0 = -\sum_{k \neq i} P_i R_{i \to k} + \sum_{k \neq i} P_k R_{k \to i}$$

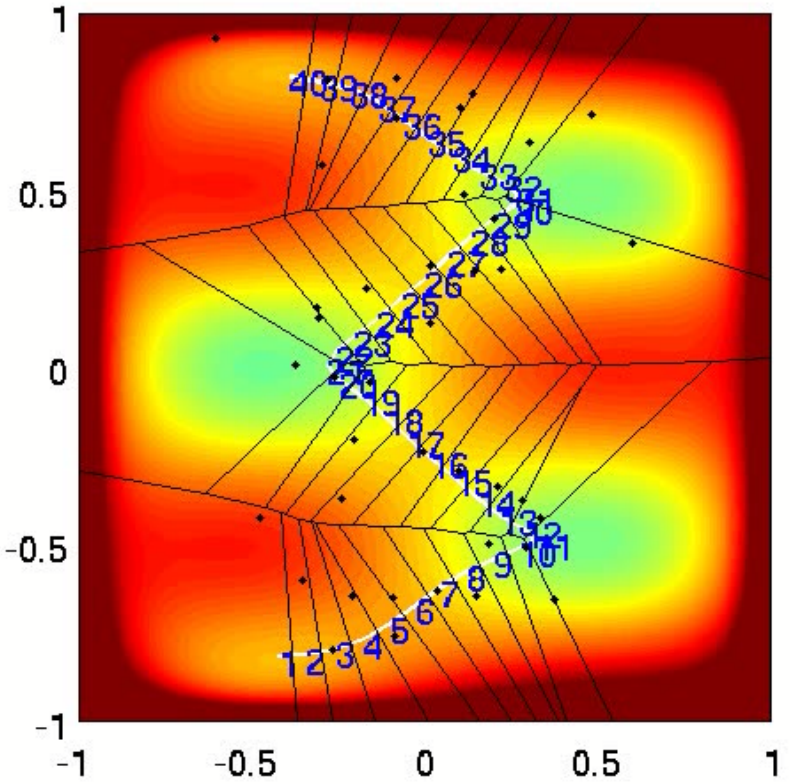
$$F_i = -kT \ln P_i$$

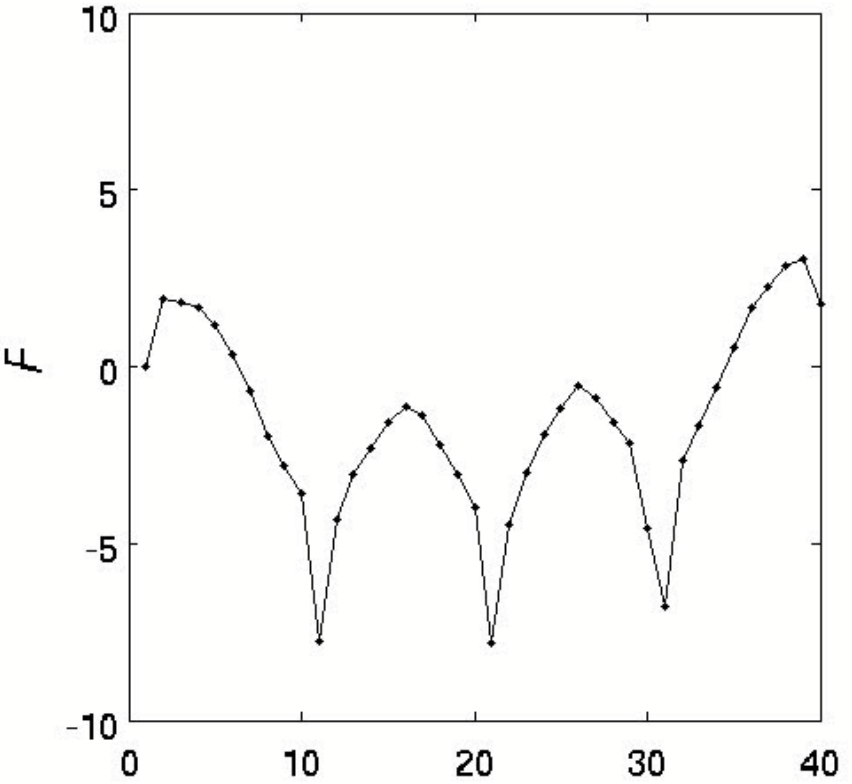


- (Note: P is the invariant distribution of a Markov state model with transition matrix R)
- Does not allow sampled surfaces to cross (by construction)

String methods (2D landscape example)

- Free energy profile can be computed from many short quasi-equilibrium simulations



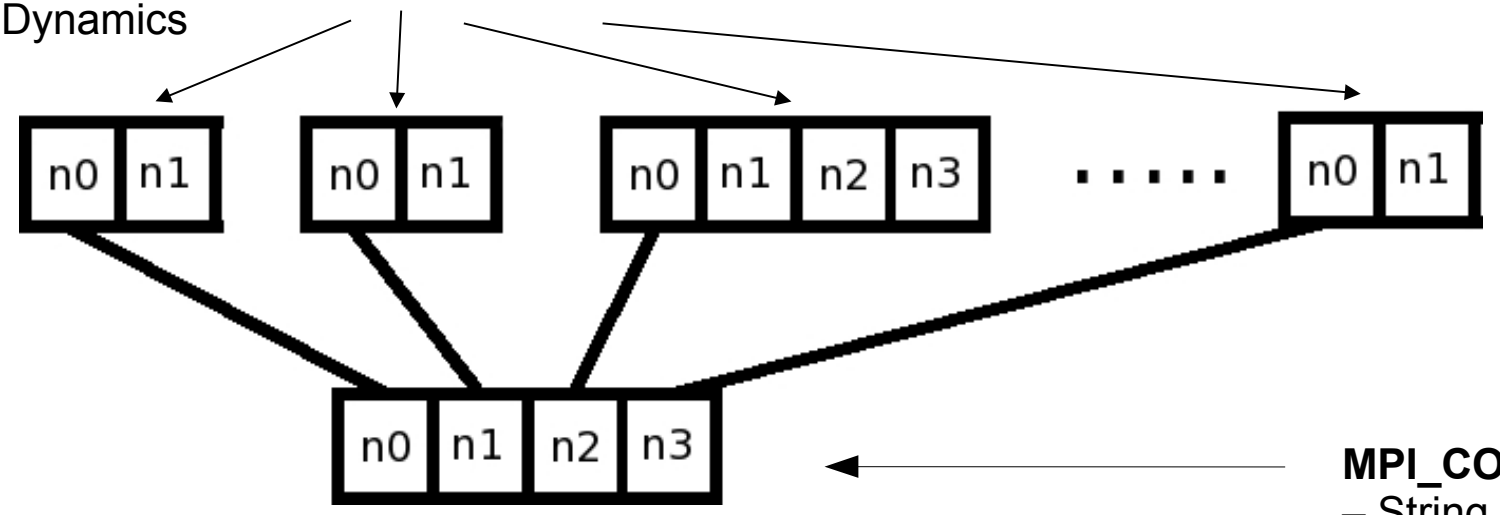


Parallel string method (CHARMM)

- MULTICOM:
  - Interactive module to add/modify/assign MPI communicators at runtime

MPI_COMM_WORLD
[=COMM_CHARMM]

MPI_COMM_LOCAL [=COMM_CHARMM] (assigned to different communicators on different nodes)
– Dynamics

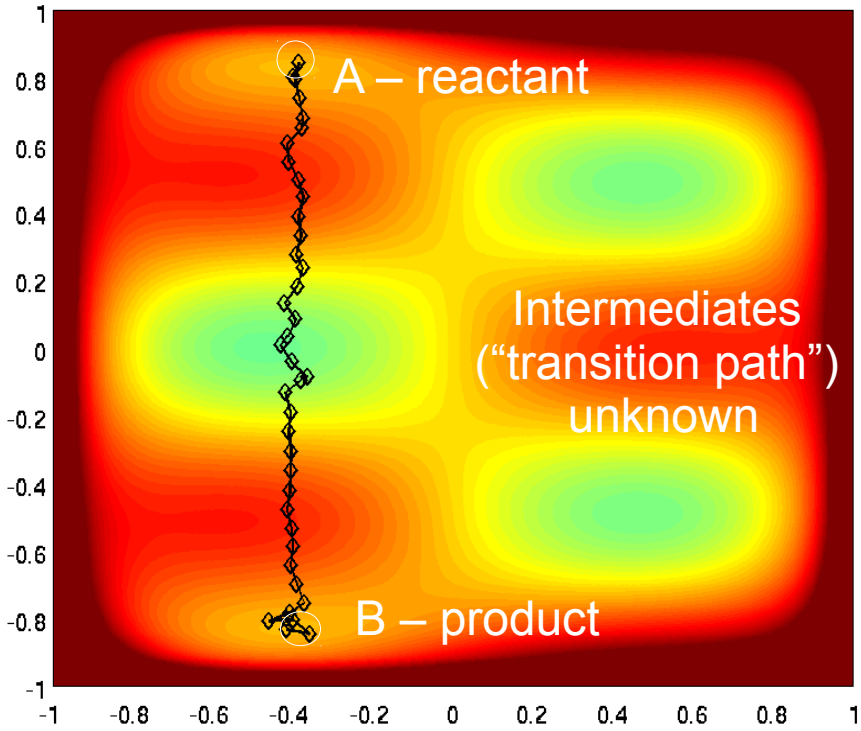


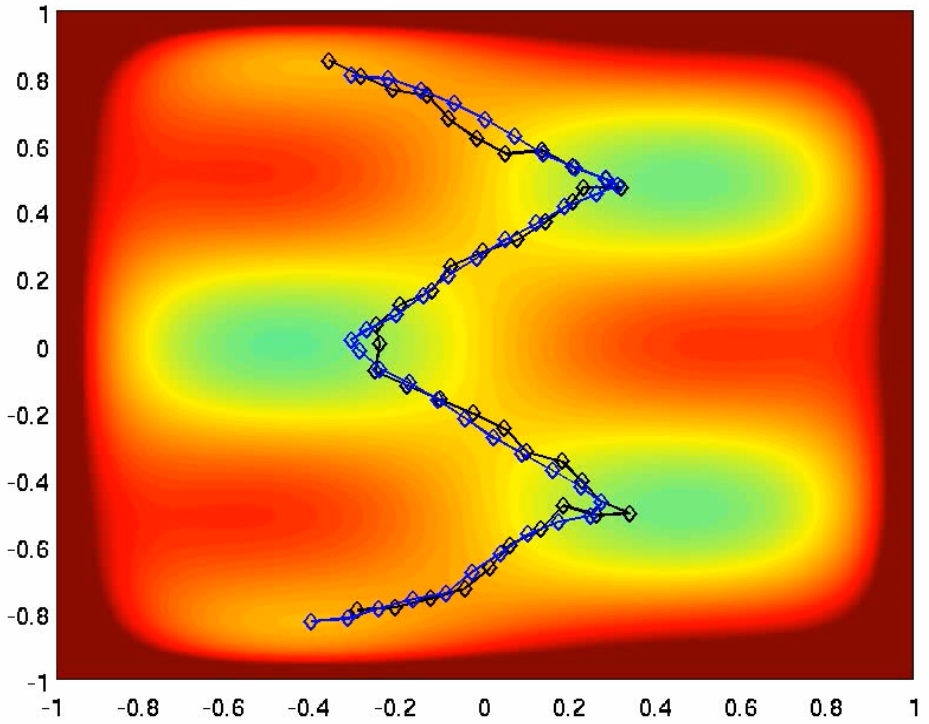
MPI_COMM_STRINGM
– String distance constraint
– Tangent Vector
– FE computation

- Each _LOCAL group only needs $\phi_i,\ \phi_{i\pm1}$

String methods (2D landscape example)

- Transition pathways between endpoint coordinates can be obtained using equilibration in “path space”





- Note: a medium-sized protein, e.g., calcium-binding calmodulin has 3N-6 ~ 7000 dof

Parallel performance

- TEST SYSTEM: Calmodulin (2272 atoms)
  FACTS implicit solvent model
- No slowdown with increasing resolution (weak scaling)

test case	total cores	replicas	cores/rep	total atoms	steps	time (m)	steps/m	speed-up
FTSM simulations								
1.	128	4	32	9088	20000	10	2000	x 2.9
2.	4096	128	32	290816	138100	72	1980	x 2.8

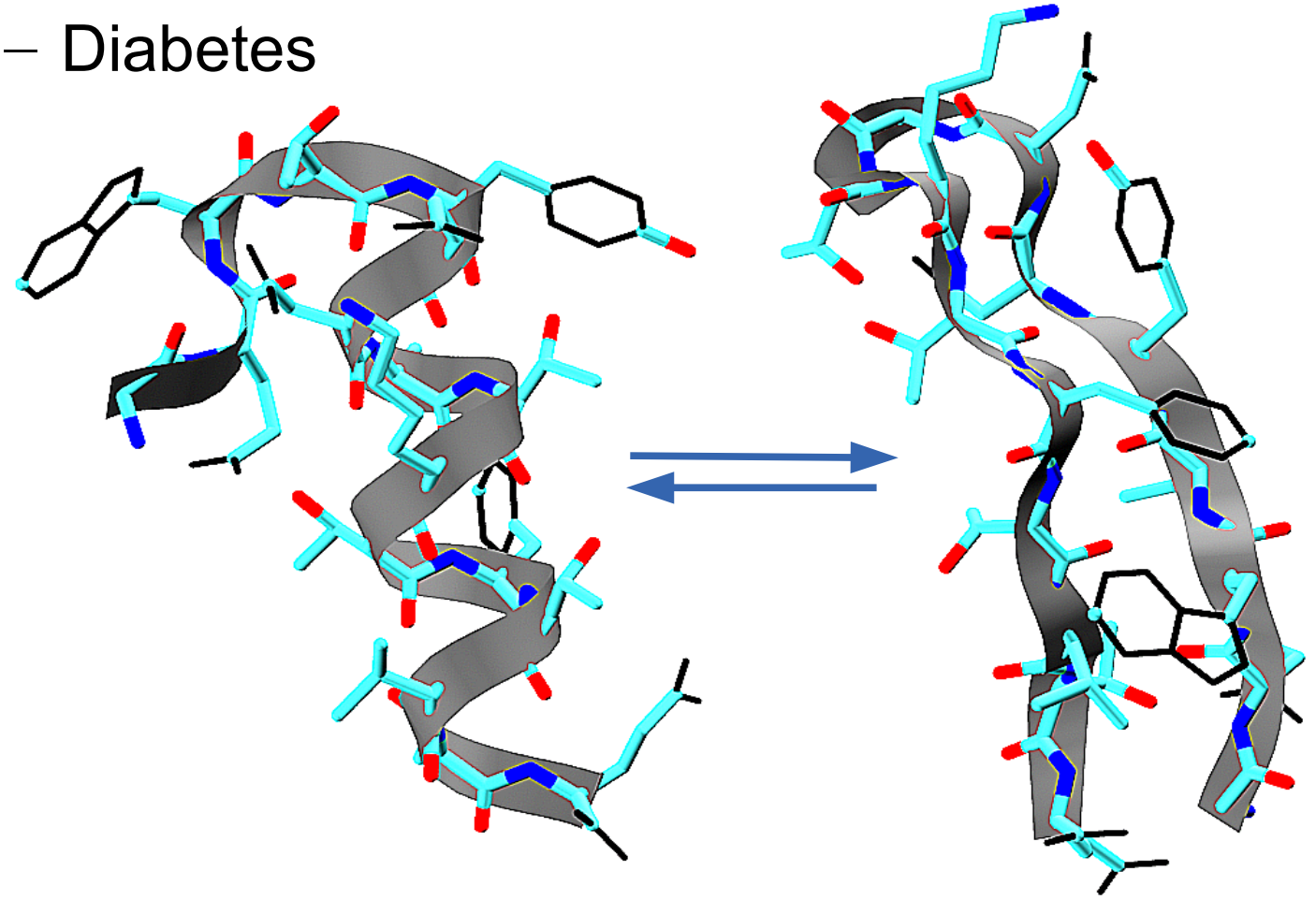
- String calculations are 3-4 times slower relative to simple MD

test case	total cores	replicas	cores/rep	total atoms	steps	time (m)	steps/m	speed-up
FTSM simulations								
3.	1024	128	8	290816	7000	10	700	x 1
MD simulation (non-interacting replicas)								
4.	160	20	8	45440	150000	69	2173	x 3.1

- Force calculations are expensive
- Differentiation of coordinate transformations can be made faster
- Smaller subsets of coarse-grained variables will increase speed
- Use of explicit solvent will mask additional string overhead

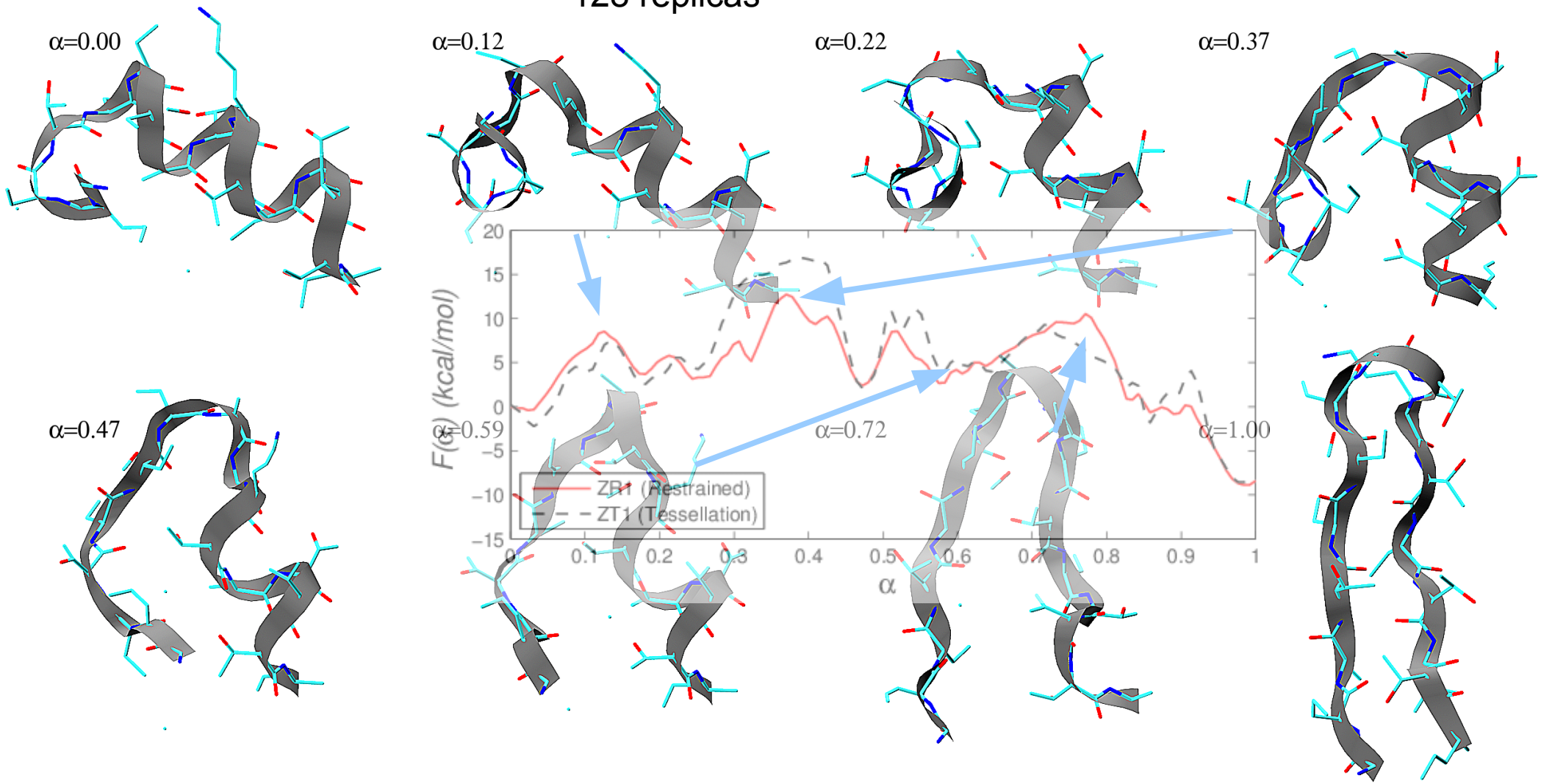
Application: α-helix $\rightarrow$ β-sheet transition

- Plays a role in
  - Alzheimer's (β-amyloid aggregation)
  - Diabetes



Transition path and free energy profile

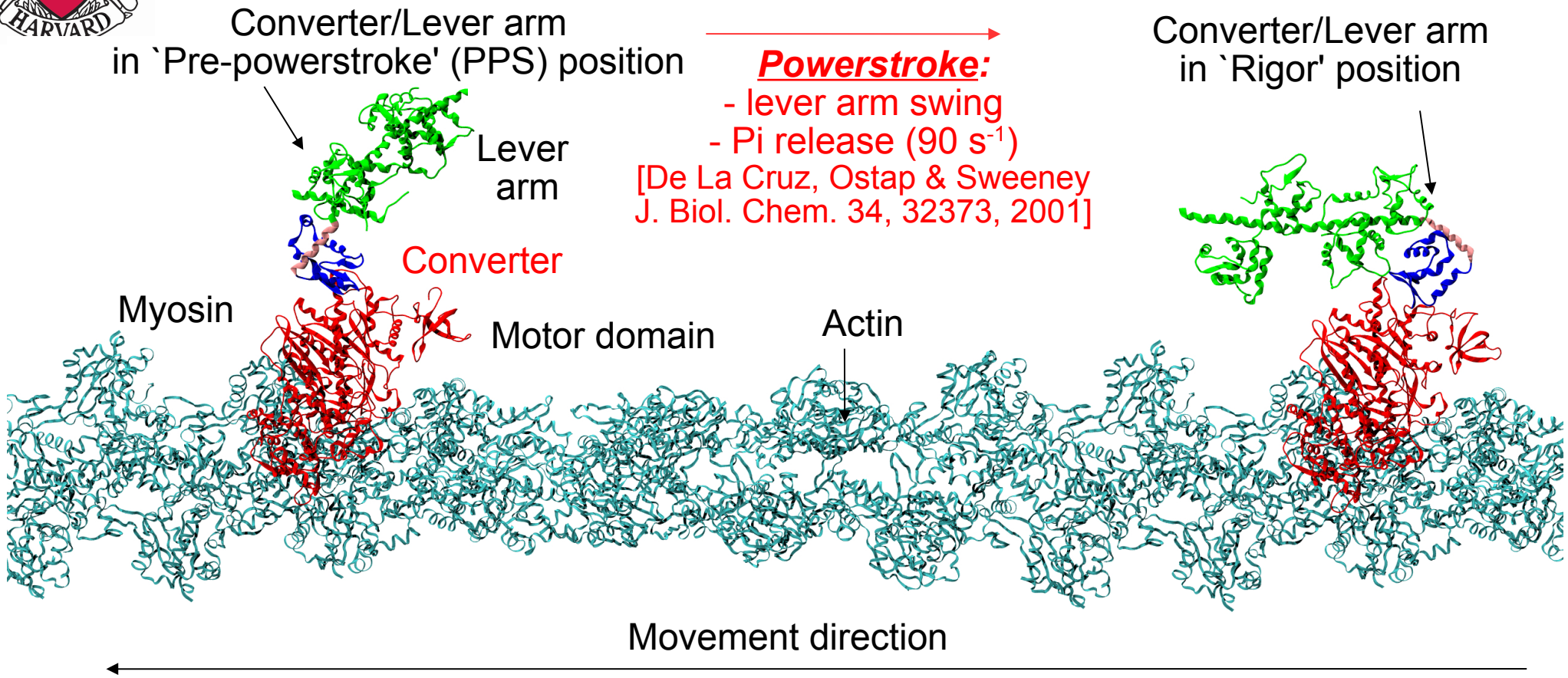
128 replicas



- Barriers are due to the sequential unwinding of α-helical turns
- Highest barrier corresponds to hairpin formation
  - Occurs in the “middle” of the path

	α-helix	β-sheet	$\Delta_{\beta\rightarrow\alpha}$
G	-75.9±0.3†	-82.6±0.3†	6.7±0.4
$\overline{E}$	-332.6±0.5	-346.2±0.5	13.6±0.7
TS	-37.6±0.6†	-44.5±0.6†	6.9±0.8

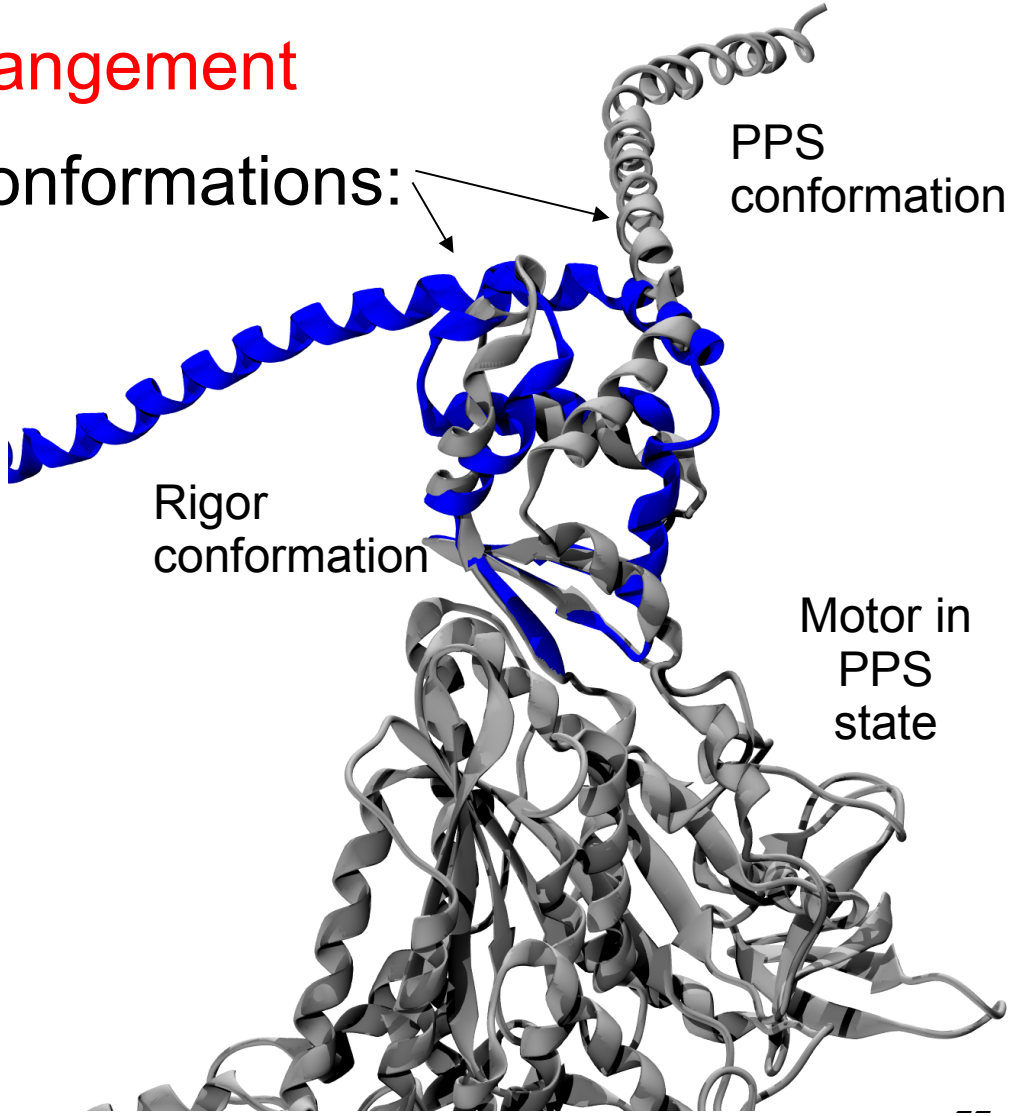
Application: Myosin VI motor



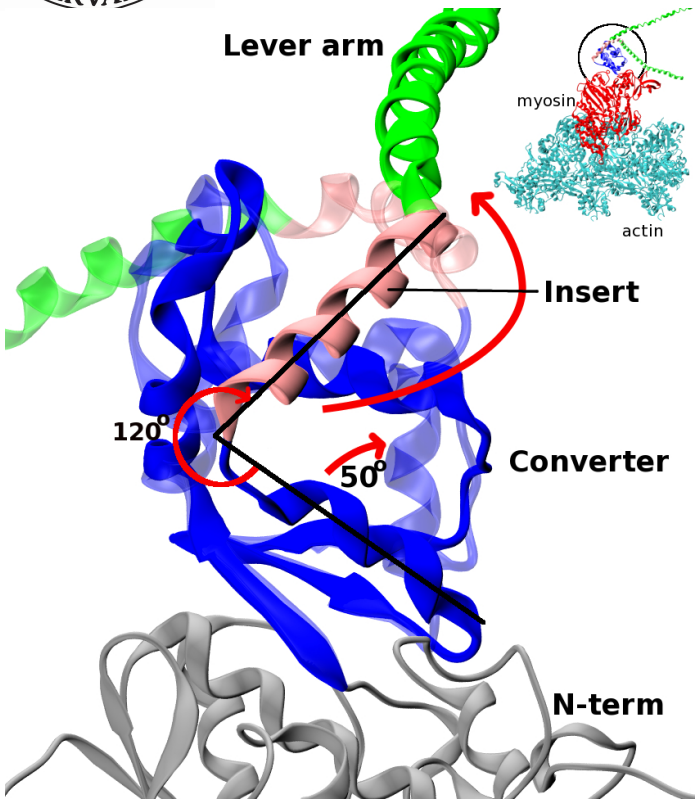
- ATP-driven “motor”
- Present in all Eukaryotic cells
- MVI dimers “walk” on actin filaments in 36 nm steps
- Transports various cellular cargo (organelles, membranes)
- Walks “backwards”: i.e. In the direction opposite to that of other myosins

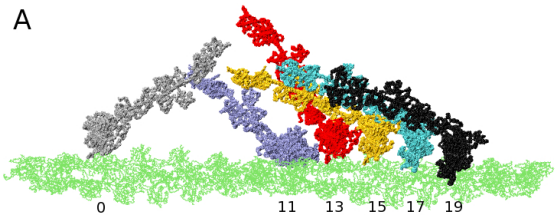
How does MVI walk backwards?

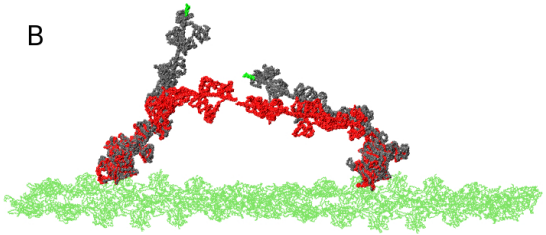
- We would like to understand the powerstroke
- No detailed computational studies have been done
- (1) Small-scale converter rearrangement
  - MVI converter adopts two conformations:
  - Which one is more stable and why?
  - What is the mechanism and rate of isomerization?
    - Is this step rate-limiting in the powerstroke?
    - What mutations could change the rate?
- (2) Large-scale rearrangement
  - Future studies

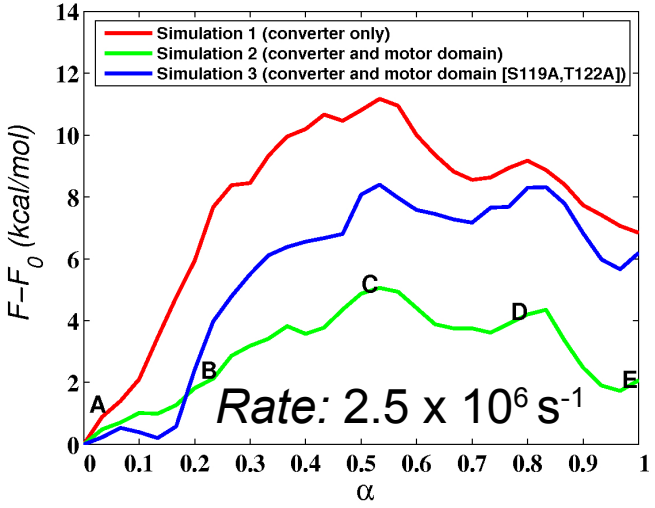


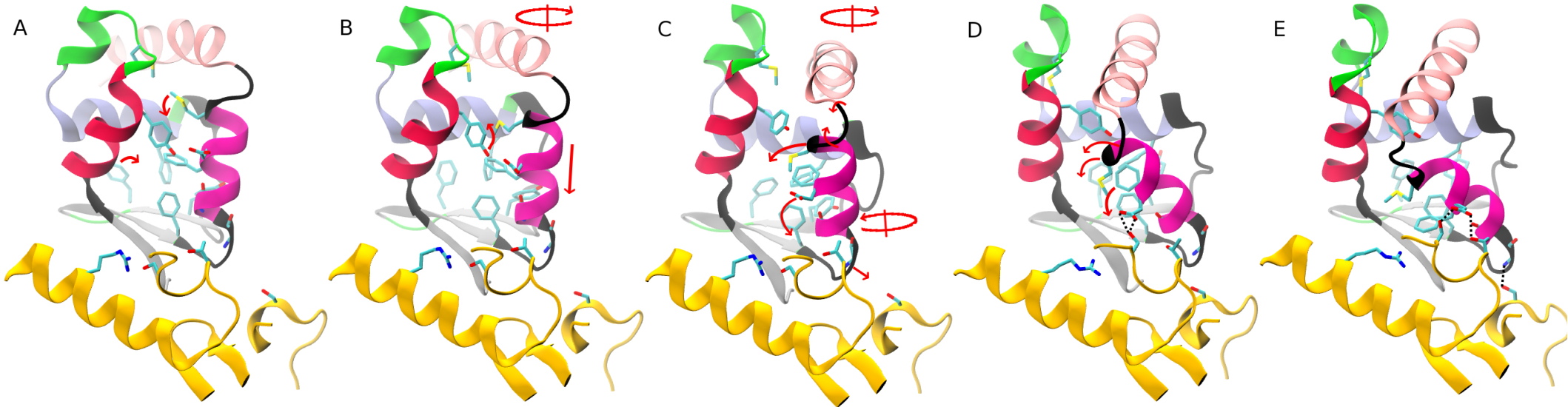
Myosin VI converter rearrangement











Summary and Ongoing work

- String method is a powerful tool to study biomolecular systems:
- Simulates entire “transition” paths
- (almost) trivially parallel
- Independent of force-field (MM, QM/MM)
- Can be applied Cartesian space or in coarse variable space (e.g. distances between amino acid COMs)

- Ongoing application to other biological/chemical problems:
- Proton transfer reactions using QM/MM (with Prof. Qiang Cui)
- DNA remodeling by topoisomerases (with Prof. Ioan Andricioaei
- DNA repair enzymes (with Prof. Kwangho Nam)
- Triose phosphate isomerase (TIM) enzyme (with Dr. Guishan Zheng)