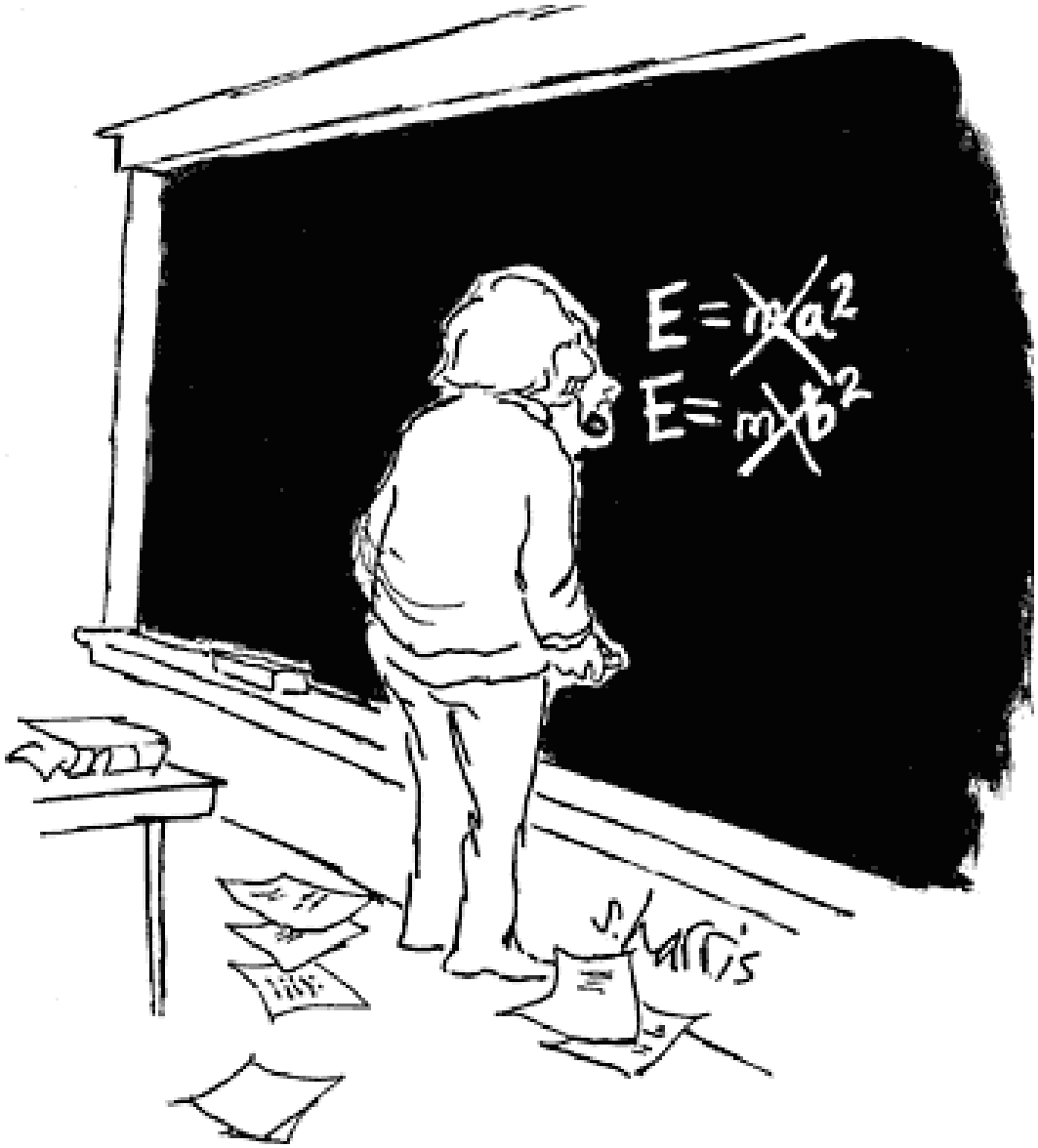
# Computer Simulations of the folding/unfolding of proteins and nucleic acids

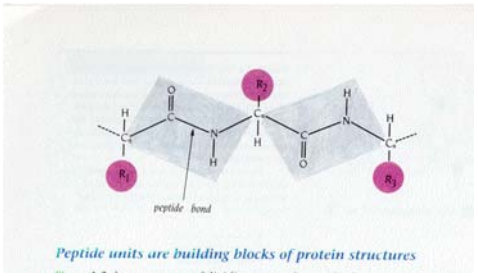
Angel E. Garcia
Department of Physics
and
Center for Biotechnology and Interdisciplinary Studies
Rensselaer Polytechnic Institute

- We use theoretical and computational methods to study biological systems.
- We are interested in understanding fundamental biological processes at the molecular level using chemical and physical principles.
- Our research has applications in health and Bioscience and Bioengineering

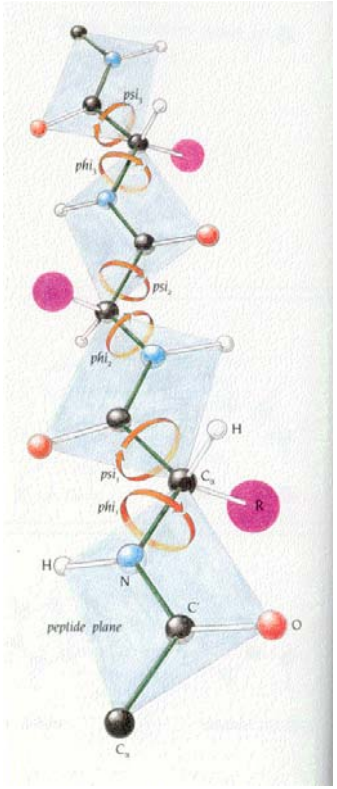


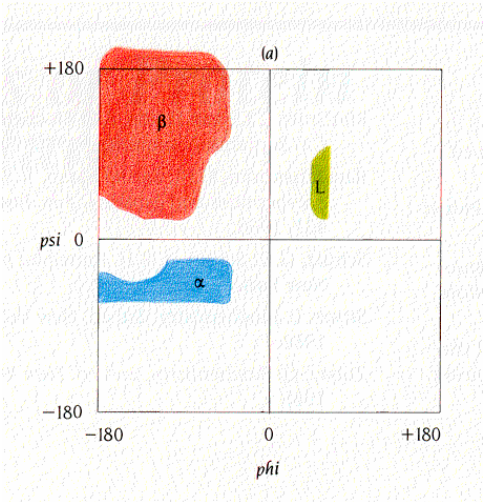
# What is a protein?

*Folded protein*


*Peptide bond*




*Ramachandran Plot*
*(Secondary structure)*

# The old view

## Levinthal Paradox

**Statistical improbability of proteins finding a unique stable folded conformation by random searches**

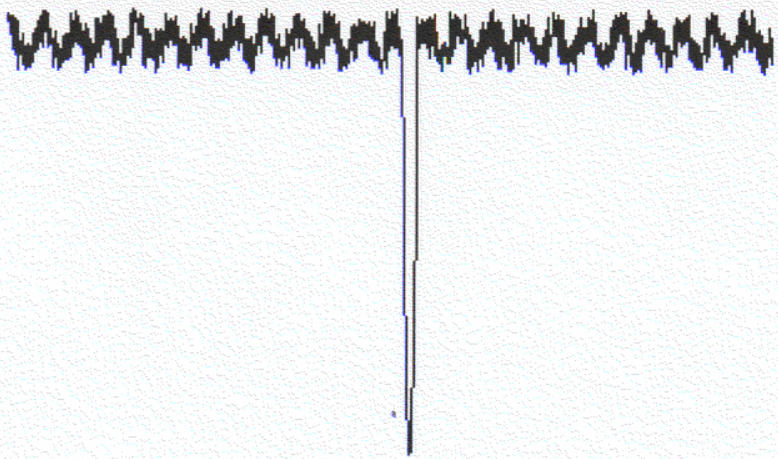
**dihedral angle space:**

$\phi, \psi, \chi_1, \chi_2 \cdots$

**Simplest assumption:**
**3 states $g^+, g^-, t$ :**

$3^N$

**N= 154 myoglobin**
**104 cyt c**
**327 R. dehalogenase**



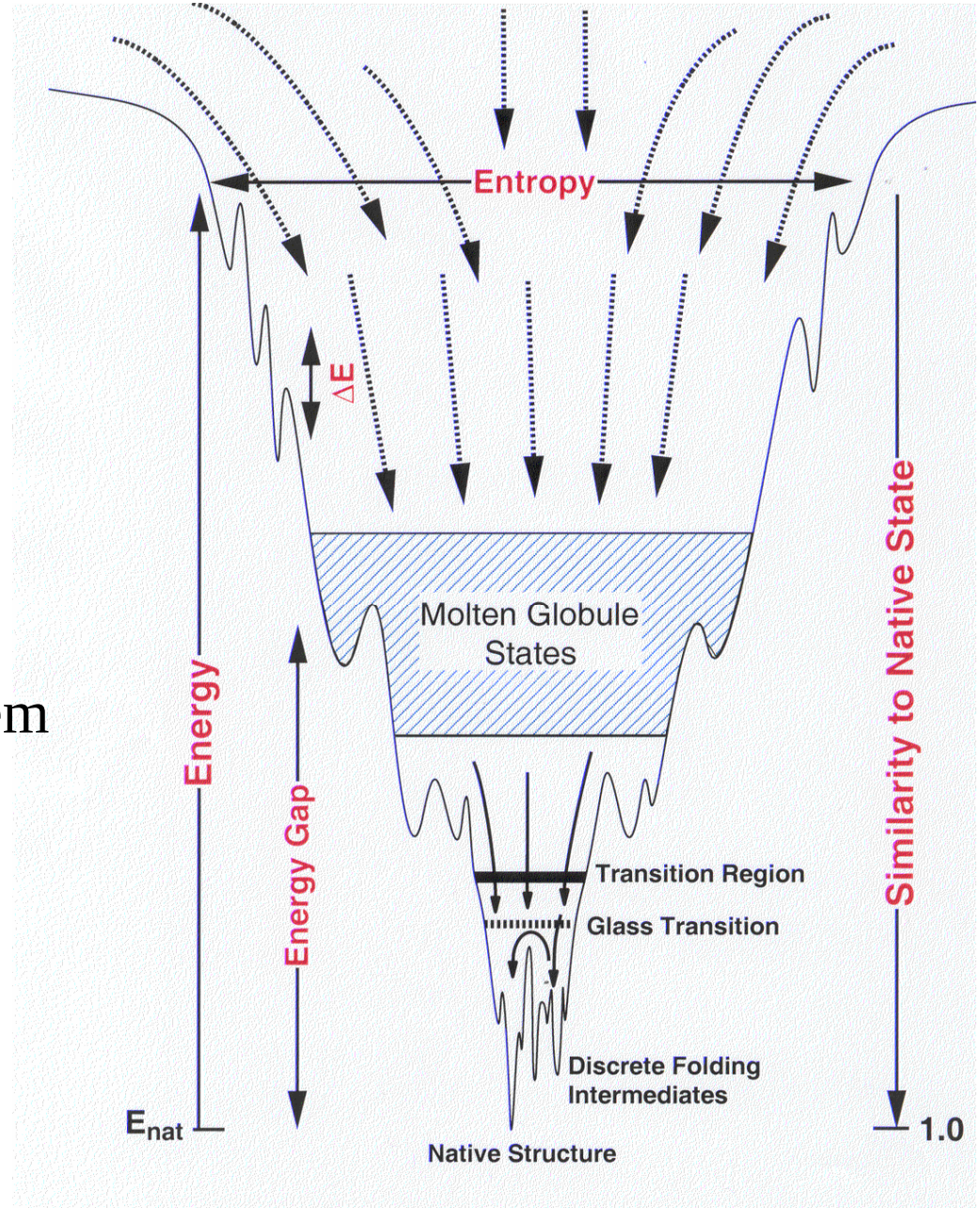
: A cartoon golf-course like landscape. A one-dimensional landscape is shown for simplicity. In the case of extreme frustration there is no correlation between structural similarity to the ground state and energy. In other words, by just measuring the energy there is no way to tell how close you are to the ground state.

# The new view

***Funnel theory***

Entropy and enthalpy are correlated

Energy decreases as the system becomes *'native like'*




Bryngelson, J., Onuchic, J., Socci, N. & Wolynes, P.
Funnels, Pathways, and the Energy Landscape of Protein-folding - A Synthesis.
*Proteins-structure Function And Genetics* **21**, 167-195 (1995).

Can we fold a protein in the computer starting from first principles?

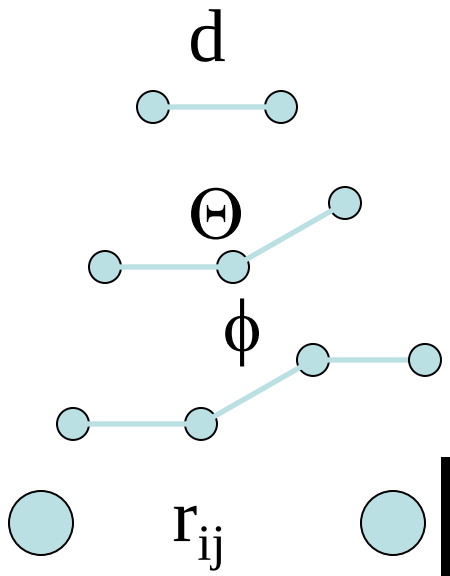
# Model (Energy Function)

- $U = K_d (d - d_0)^2 +$

$K_\theta(\Theta - \Theta_0)^2 +$

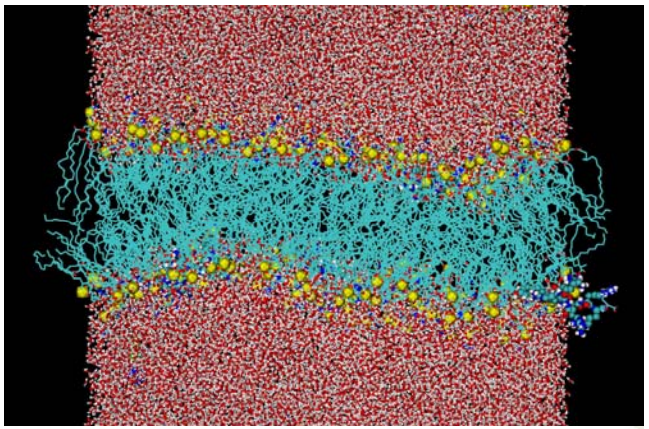
$\Sigma_{n,i} \; V_n[\, 1 + \cos(\, n\phi_\iota + \delta_n)] +$

$V_{LJ}(r_{ij}) \quad + \quad V_{Coul}(r_{ij})$

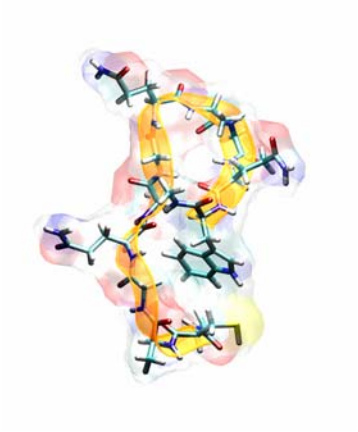



Periodic boundary conditions, T and P couplings

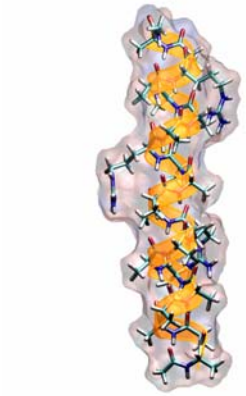
*(Protein + solvent)*



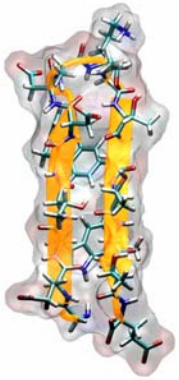
# Folding Time scales



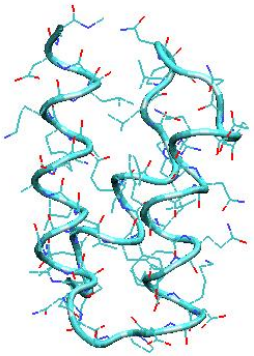
Loop closing in a random coil
~ 10 ns



Alpha helix formation
~ 200 ns



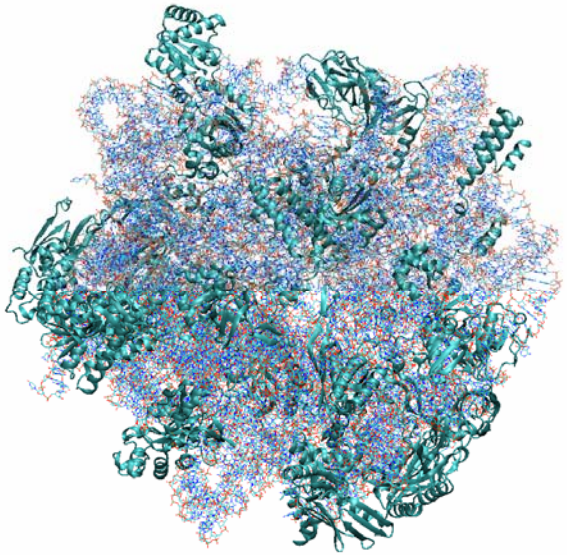
beta hairpin
~ 1-10 μs



Folding of proteins:
1 μs – sec

System size:

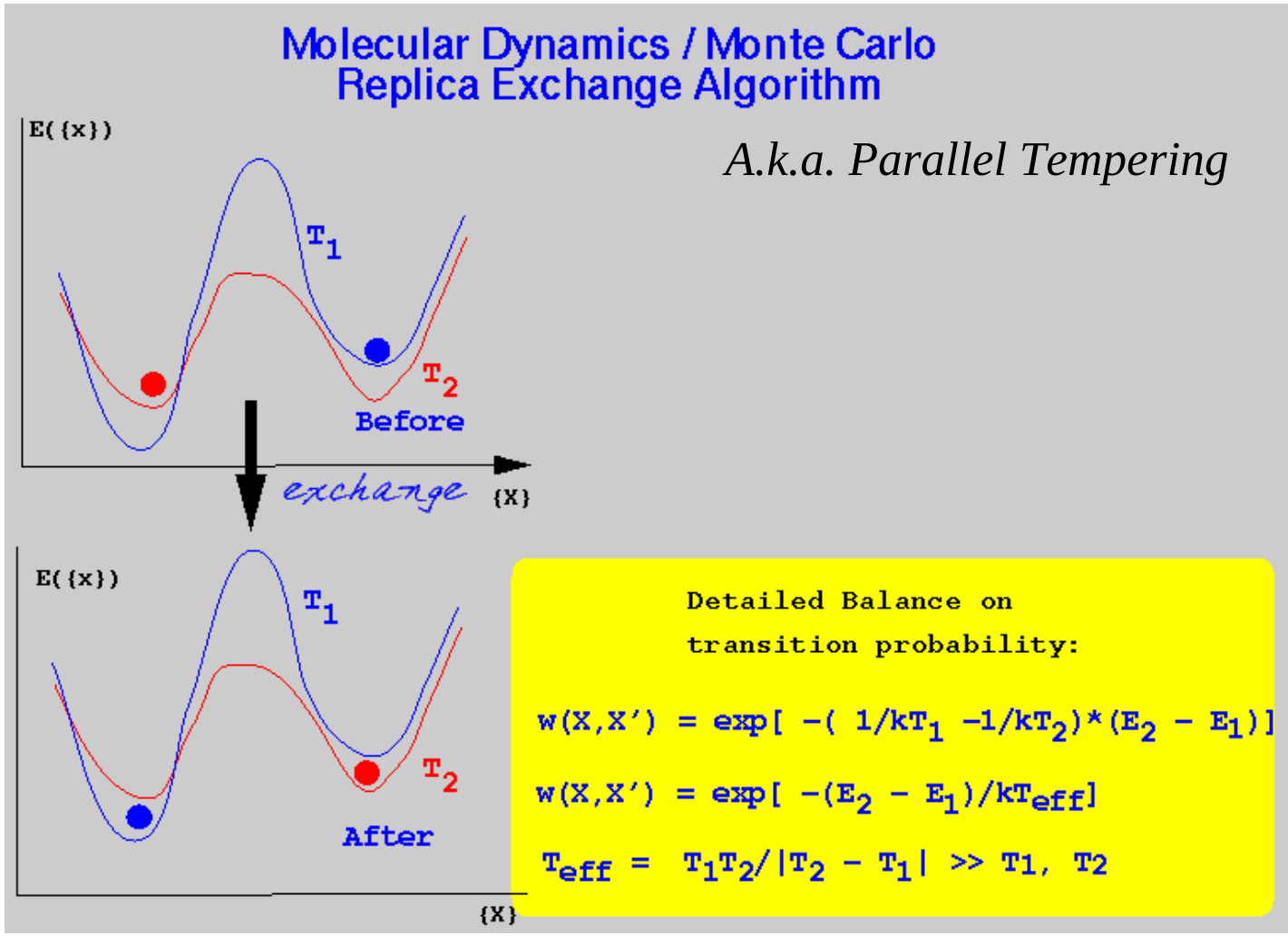
Protein A
56 amino acids protein



Ribosome; ~2000 nuclei acid bases
21 proteins

# Methods

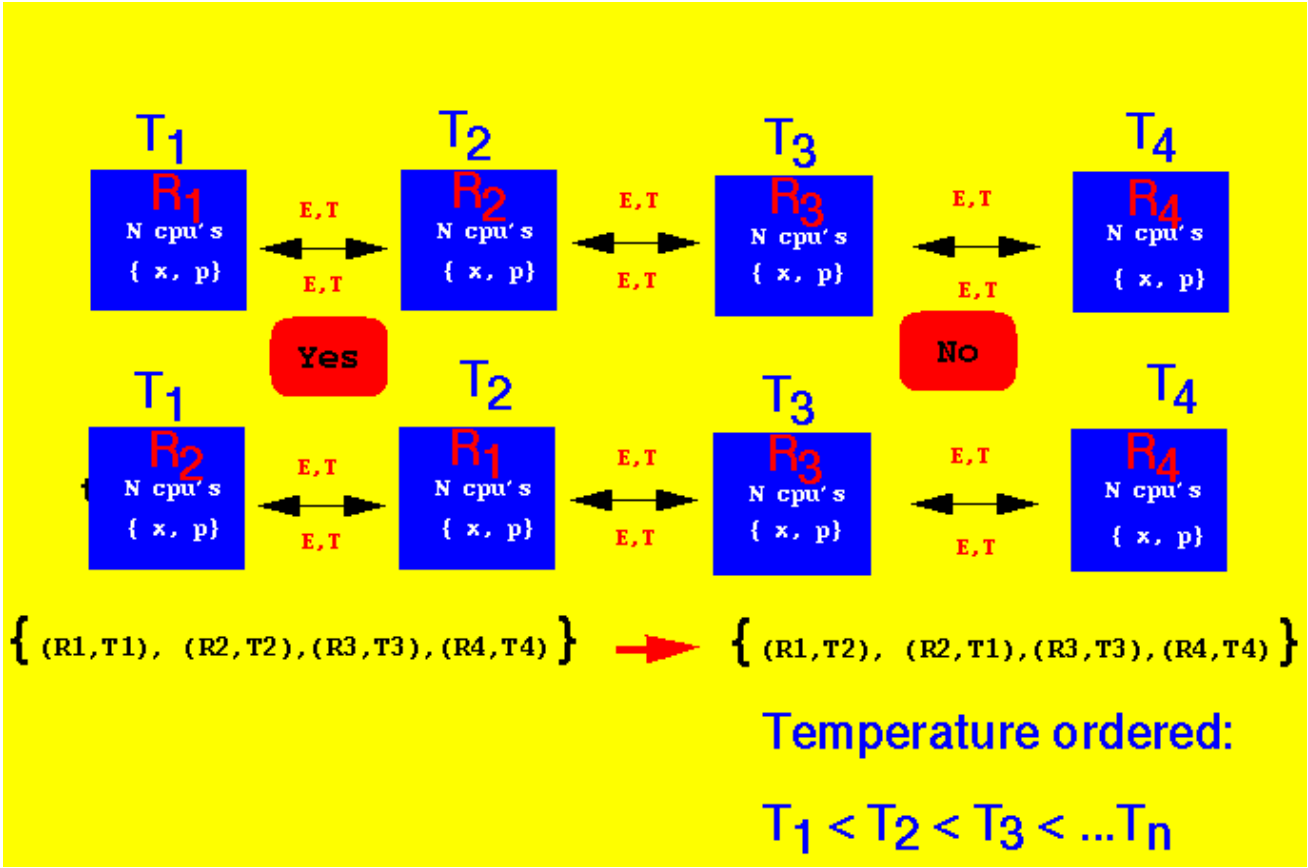
- Thermodynamics
  - Replica exchange (parallel tempering) molecular dynamics simulations
- Kinetics
  - Replica dynamics (A. Voter)
  - folding@home
- Large systems
  - Parallel molecular dynamics simulations




*Hukushima K and Nemoto K, "Exchange Monte Carlo Method And application to spin glass simulations"*
*J. Phys. Soc. Japan 65: 1604-1608 (1996)*

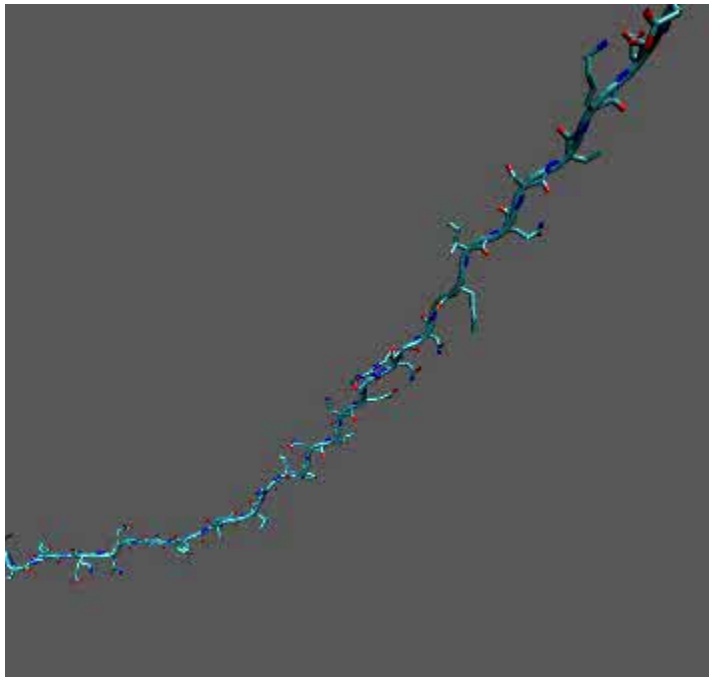
*Sugita and Okamoto, Chem. Phys. Lett. 314: 141 (1999)*

# Exchange algorithm

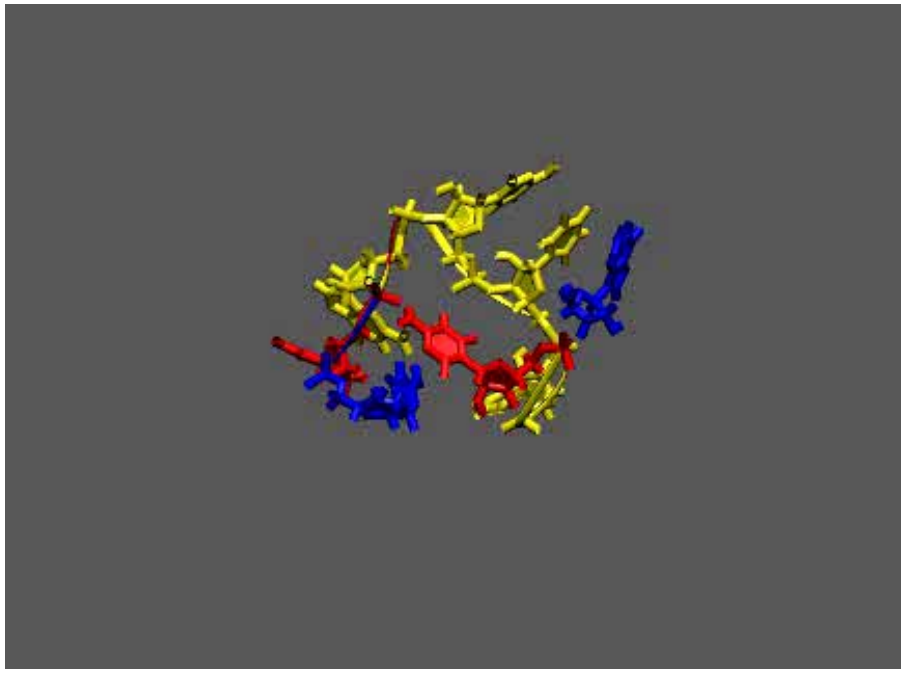


# Protein & RNA folding thermodynamics

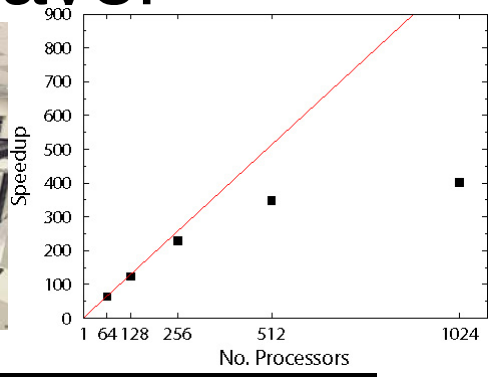
Molecular dynamics folding of protein A

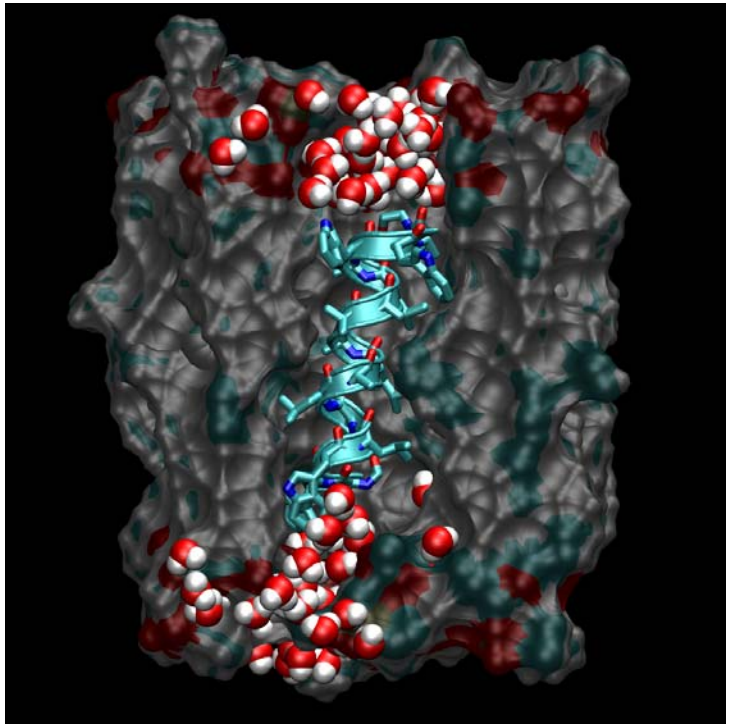


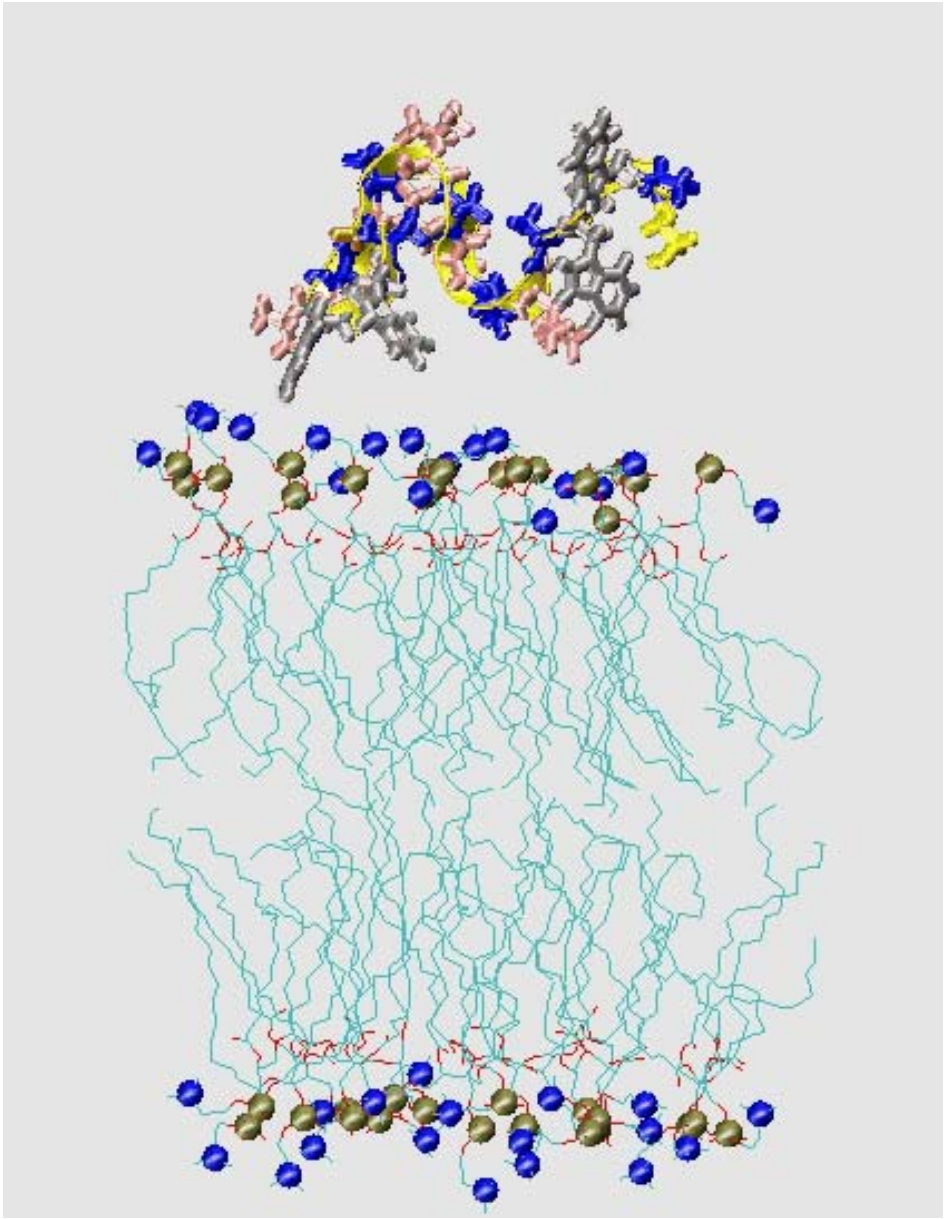
Molecular dynamics folding of an RNA tetraloop



# Insertion and folding of peptide into a lipid bilayer

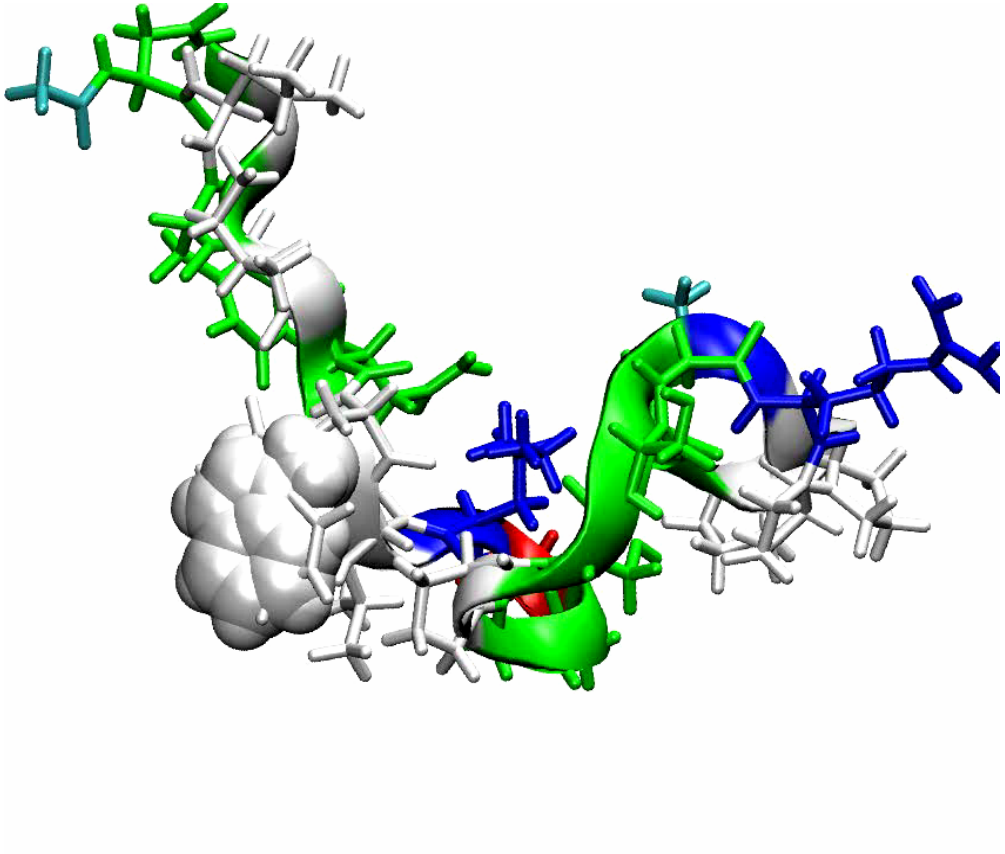




Nymeyer, Woolf and Garcia, Proteins 2005

# Folding of the Trp-cage protein



Replica ID # 02: 0…50 ns

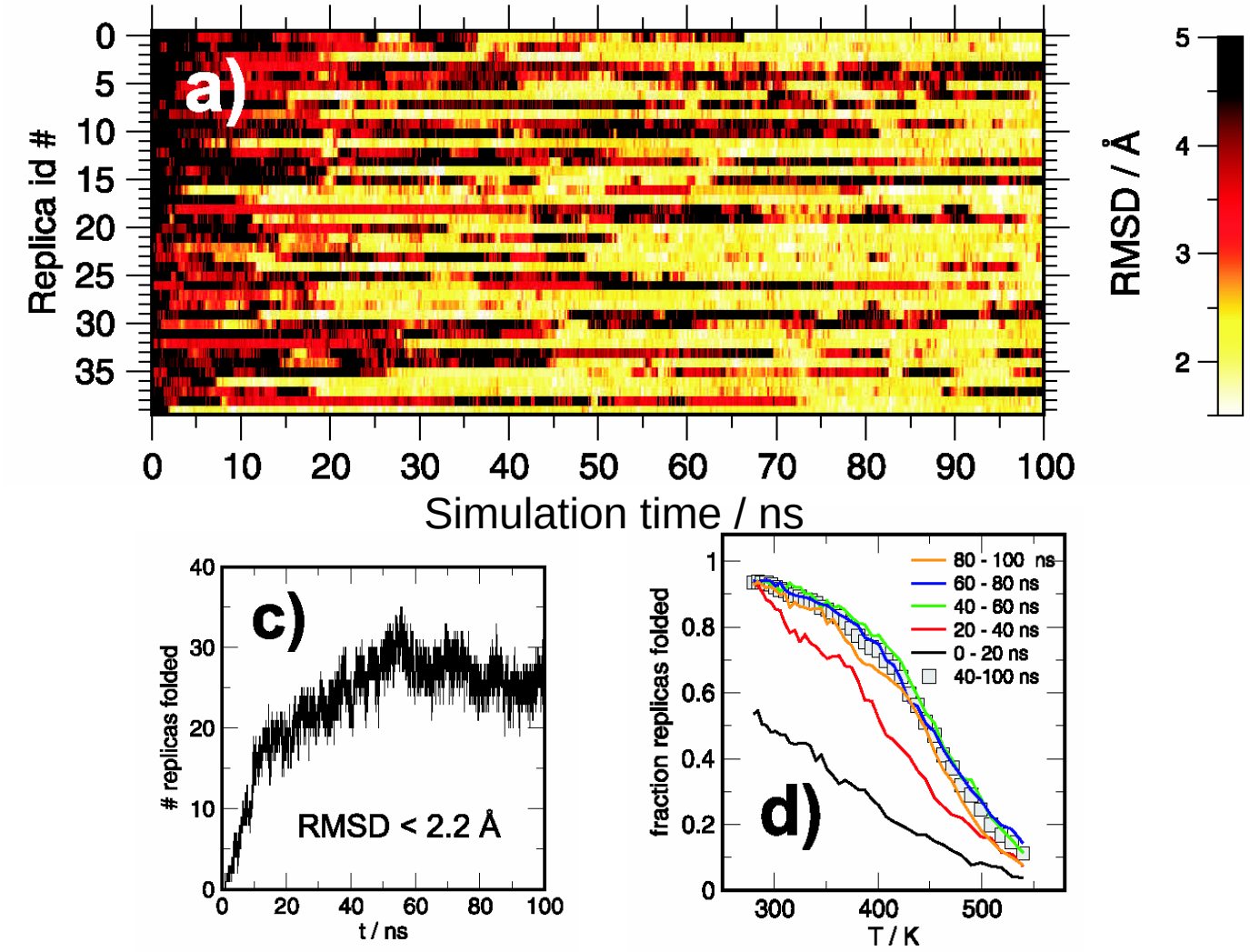
**REMD-Simulation:**

40 Replicas
280.0 K- 539.8K
AMBER94
2637 TIP3P
Ewald (PME)
100 ns/replica (4 μs accumulated)

Starting from an extended configuration

Paschek, Nymeyer, Garcia, J Str Bio 2007

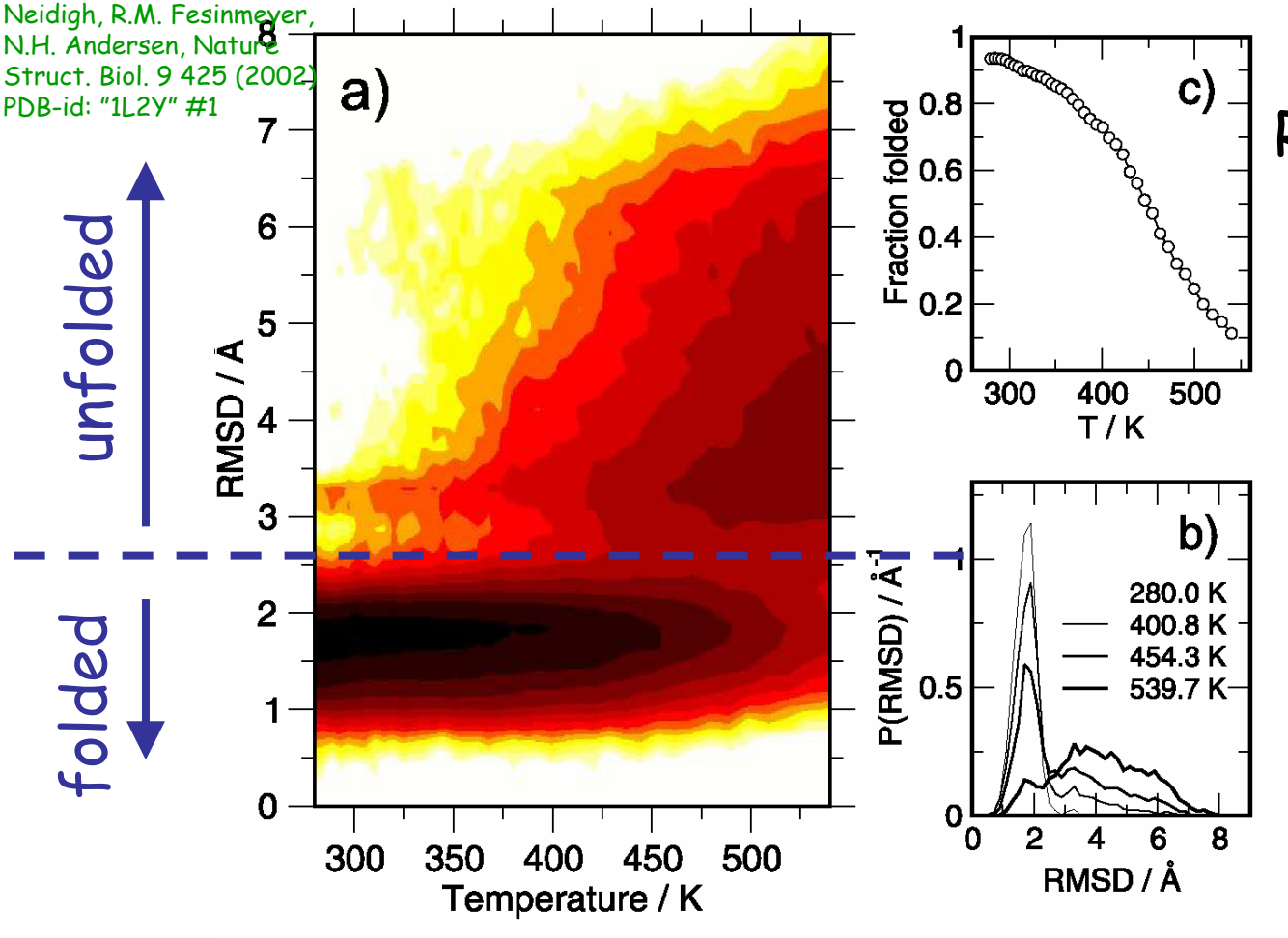
# Can we access the folding/unfolding equilibrium?




Average Lifetime of folded states in REMD-Ensemble: ~30 ns

# Folding: Free Energy Map of the Trp-cage

RMSD: NMR: J.W. Neidigh, R.M. Fesinmeyer, N.H. Andersen, Nature Struct. Biol. 9 425 (2002) PDB-id: "1L2Y" #1

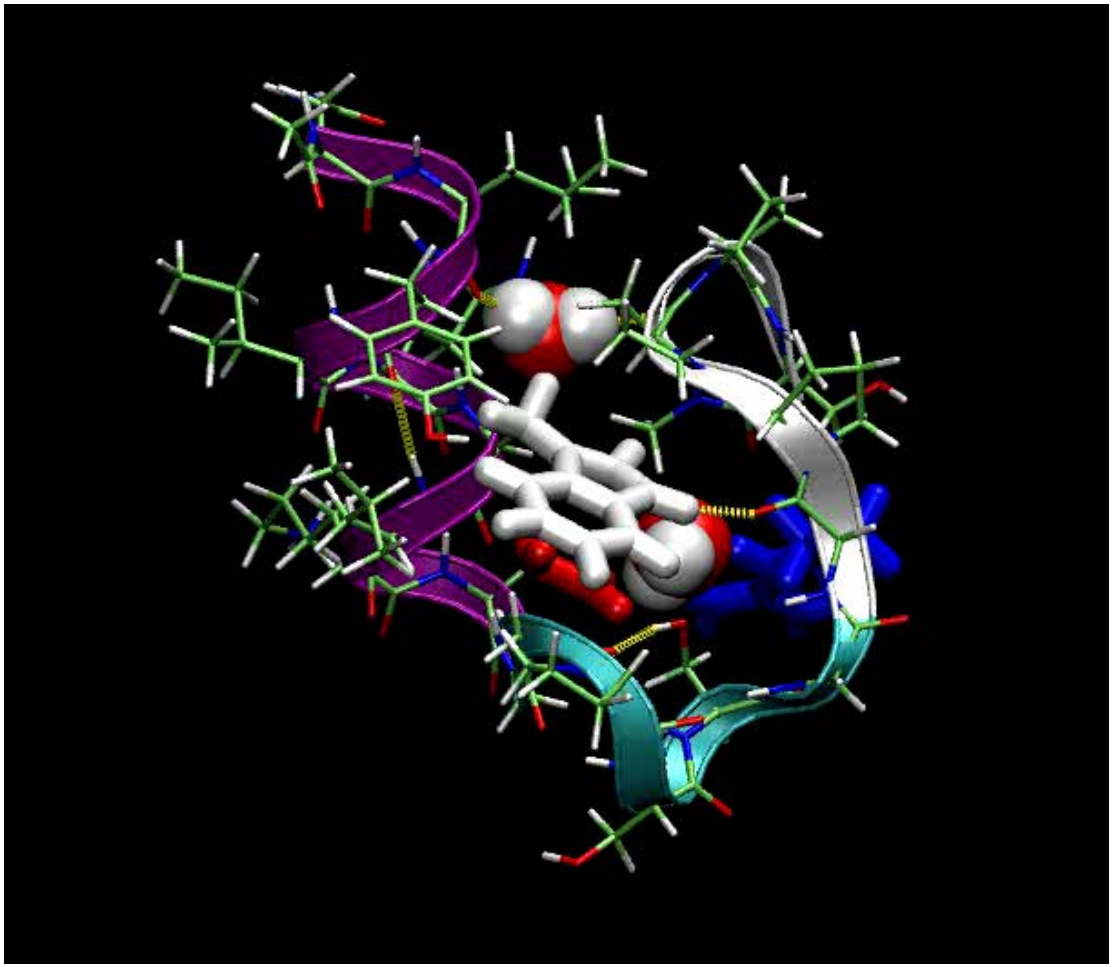



RMSD < 2.2 A

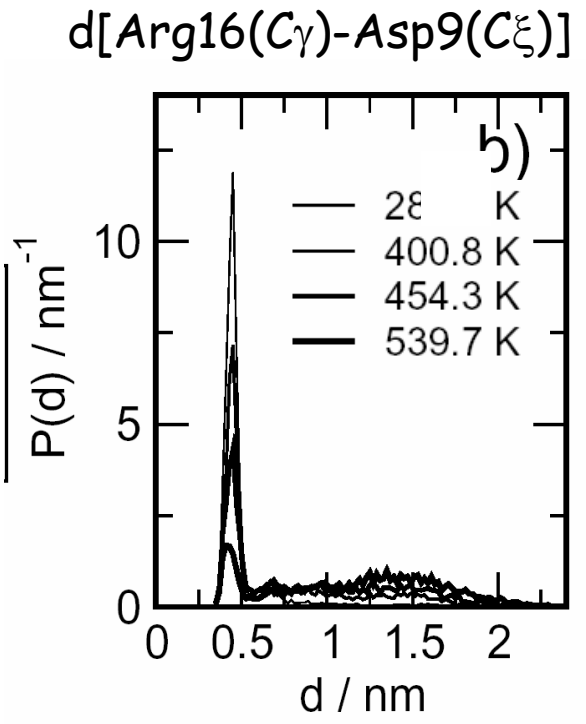
Two state folder: Qiu et al., *JACS* **124**, 12952 (2002)
Intermediate states: Neuweiler et al., PNAS **102**, 16650 (2005)

Paschek, Nymeyer, Garcia, JSB, 2007

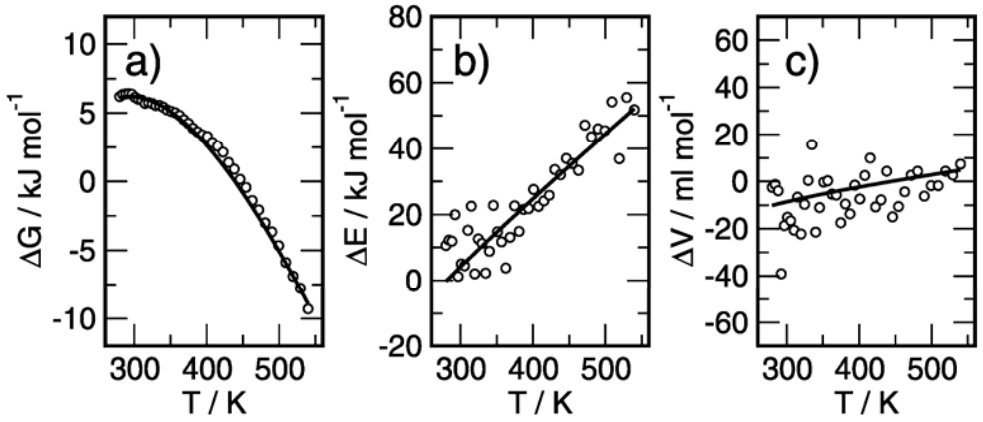
# Sources of Stability:

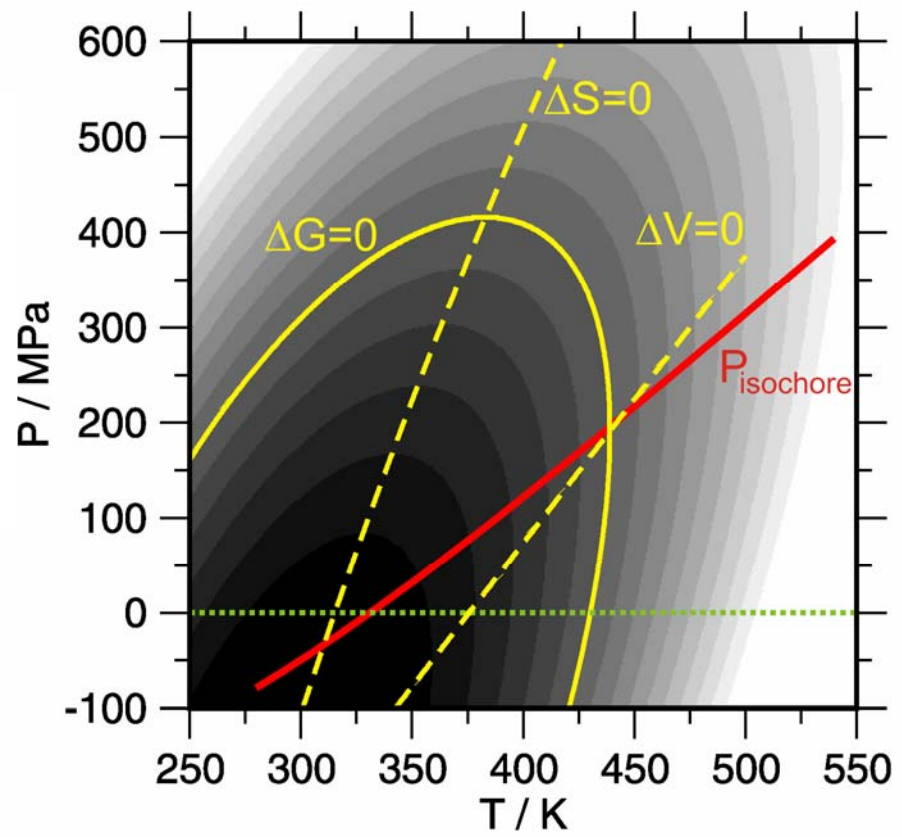


1. Salt-bridge: ***Asp9-Arg16***
2. Internal Water



# Model Trp-cage: Plausible Stability Diagram




$$\begin{aligned}\Delta G &= \Delta\beta/2(P-P_0)^2 - \Delta C_P\left[T\left(\ln\frac{T}{T_0}-1\right)+T_0\right] \\ &\quad +\Delta\alpha(P-P_0)(T-T_0) \\ &\quad +\Delta V_0(P-P_0) - \Delta S_0(T-T_0) + \Delta G_0\end{aligned}$$

$T_0 = 331\,\mathrm{K}$ $\quad P_0 = 0\,\mathrm{MPa}$

$\Delta G_0 = 5.7\,\mathrm{kJ\,mol^{-1}}$ $\quad \Delta S_0 = 15\,\mathrm{J\,K^{-1}mol^{-1}}$ $\quad \Delta V_0 = -6.5\,\mathrm{ml\,mol^{-1}}$

$\Delta Cp = 0.31\,\mathrm{kJ\,K^{-1}\,mol^{-1}}$ $\quad \Delta\alpha = 14.5\,10^{-2}\,\mathrm{ml\,K^{-1}\,mol^{-1}}$ $\quad \Delta\beta = -4.8\,10^{-5}\,\mathrm{kJ\,mol^{-1}\,MPa^{-2}}$

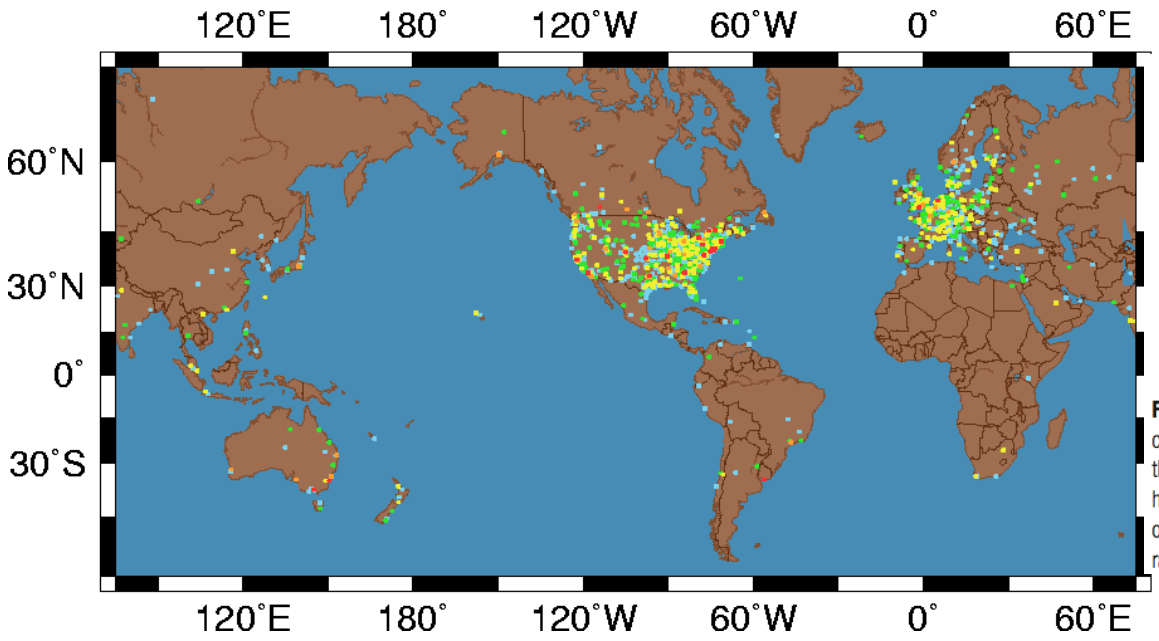
REMD are constant V calculations

# Kinetics (time domain)

# Folding@home distributed computing

V Pande, Stanford

- Similar to SETI@home, volunteers run screensaver software to donate unused CPU cycles.
- Since October 1, 2000, over 1,000,000 CPUs throughout the world have participated in Folding@Home.



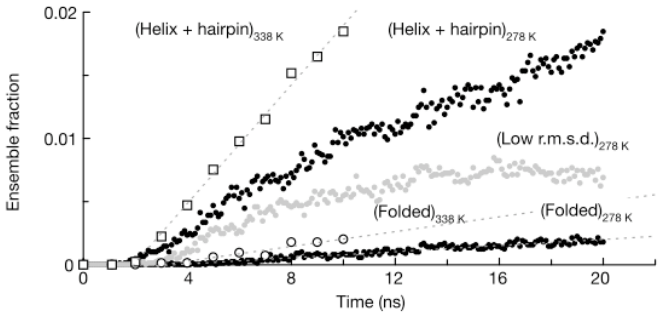
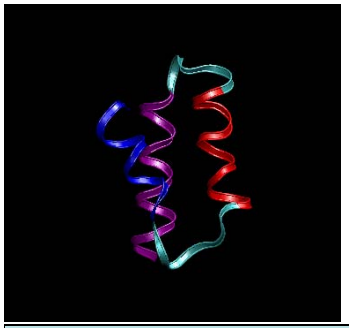



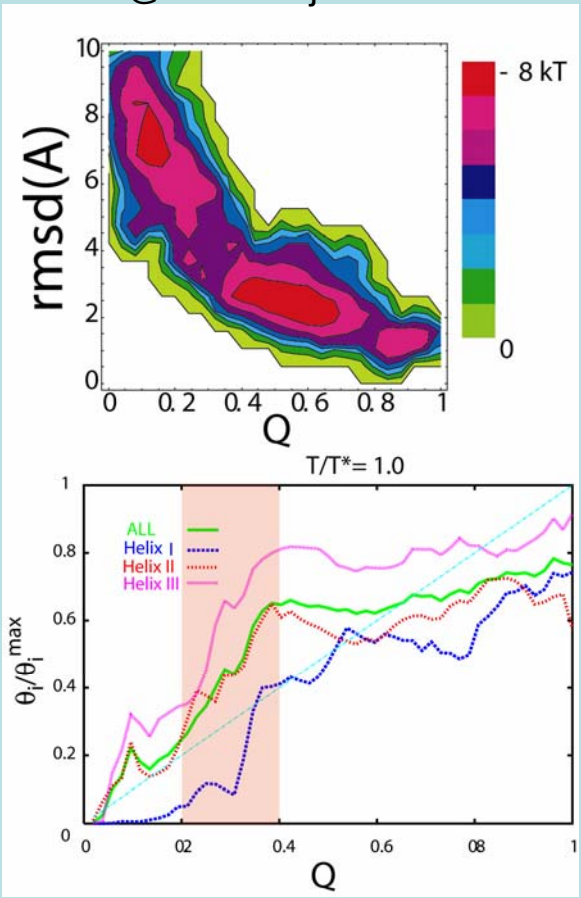

**Figure 3** Folded population growth. The growing population of folded double mutant conformations, at 278 K (filled circles, dotted line) and at 338 K (open circles, dotted line), the population with low r.m.s.d.$_{C\alpha}$ (grey circles), and the population with both helix and hairpin at 278 K (filled circles) and 338 K (open squares, dotted line), are shown. The scale of the *y* axis illustrates the need to simulate thousands of trajectories. Note the increased rate of folding at high temperature.



# Folding/unfolding kinetics and thermodynamics

## Replica exchange

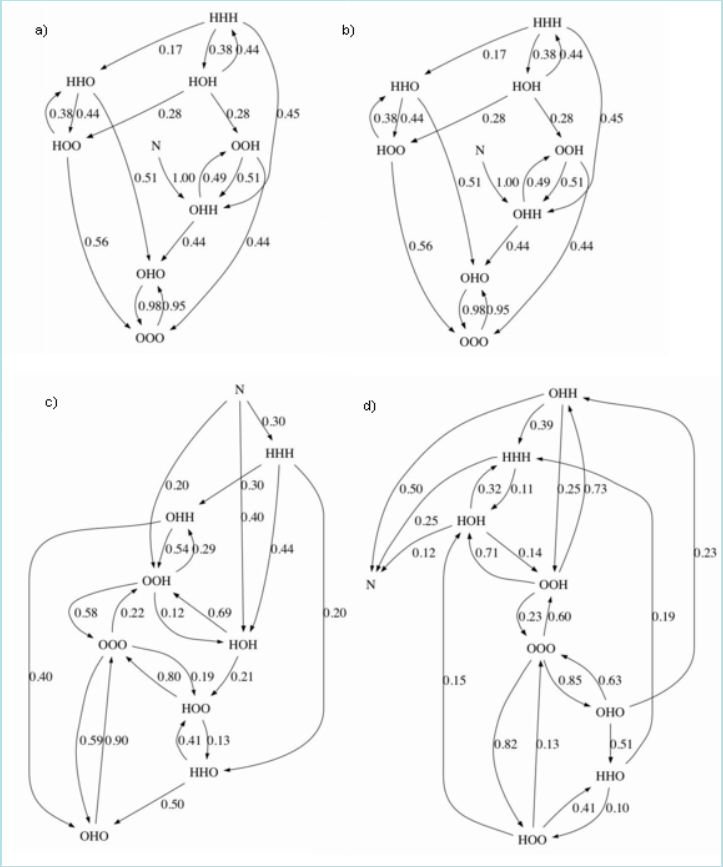
64@15 ns trajectories




A.E. Garcia and J.N. Onuchic,PNAS. 100:13898-13903 (2003).

## Folding@home

2129@35 ns trajectories




Jayachandran, Vishal, Garcia and Pande, J Str. Bio. 2007

# Multi-time scale modeling of folding

Start from reaction diffusion equation:

$$\frac{\partial P(R,\,t)}{\partial t} = -\frac{\partial}{\partial R}\left[v(R)P - \frac{\partial}{\partial R}D(R)P\right] \qquad (1)$$

Construct distributions for small time steps from REMD

$$P(R,\,t) = \frac{1}{\sqrt{4\pi D(R_c)t}}\exp\left(-\frac{(R - R_c - v(R_c)t)^2}{4D(R_c)t}\right) \qquad (2)$$

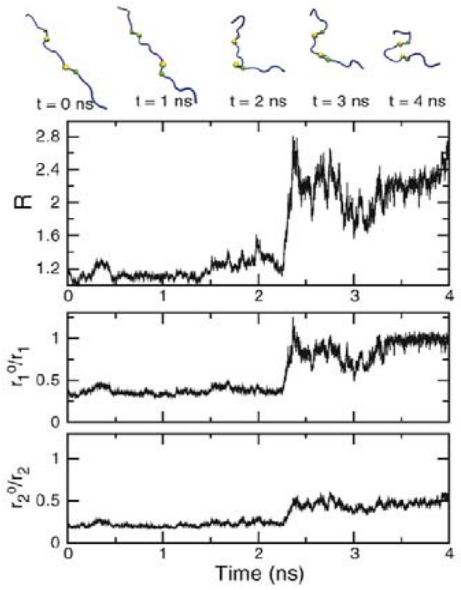
Obtain **drift velocity** and **diffusion coefficients** as a function of reaction coordinate R

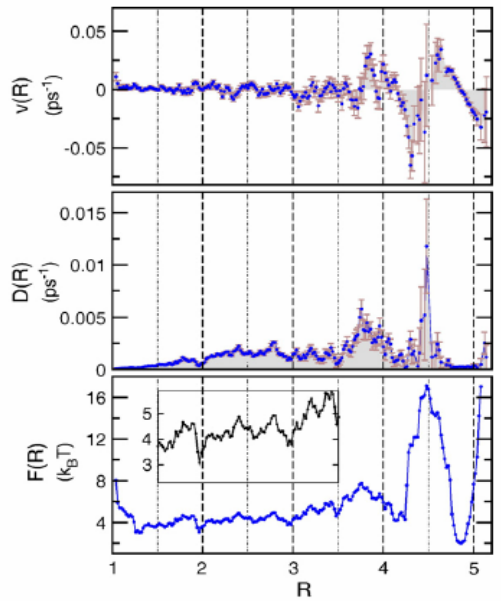
$$v(R) = \frac{R_c(t_2) - R_c(t_1)}{\Delta t} \qquad (3)$$

$$D(R) = \frac{\sigma^2(t_2) - \sigma^2(t_1)}{2\Delta t} \qquad (4)$$

Free energy (PMF) is defined by

$$F(R)/k_BT = -\int_{R_{ref}}^{R} \frac{v(R)'}{D(R')}dR' + \log D(R) + \text{const}$$



Given v( R) and D( R) solve Langevin dynamics

$$\frac{dR}{dt} = v(R) + \eta(R, t) \qquad (7)$$

With noise defined by

$$P[\eta] \propto \exp\left(-\int^{\delta t} \frac{\eta^2(t)}{4D} dt\right) \overset{\delta t \to 0}{\approx} \exp\left(-\frac{\eta^2 \delta t}{4D}\right) \qquad (8)$$

Such that

$$R(t + \delta t) = R(t) + v(R)\delta t + \tilde{\eta}(R)\sqrt{\delta t} \qquad (9)$$

Yang, Onuchic, Garcia & Levine, JMB
In press 2007

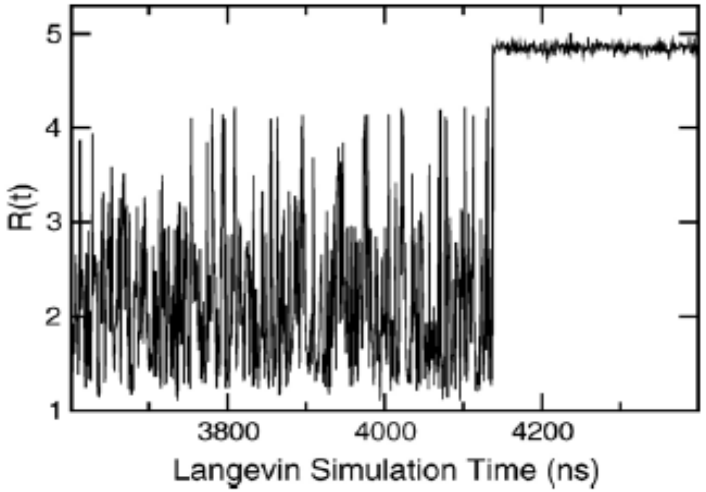



**Figure 5.** A typical Langevin simulation trajectory

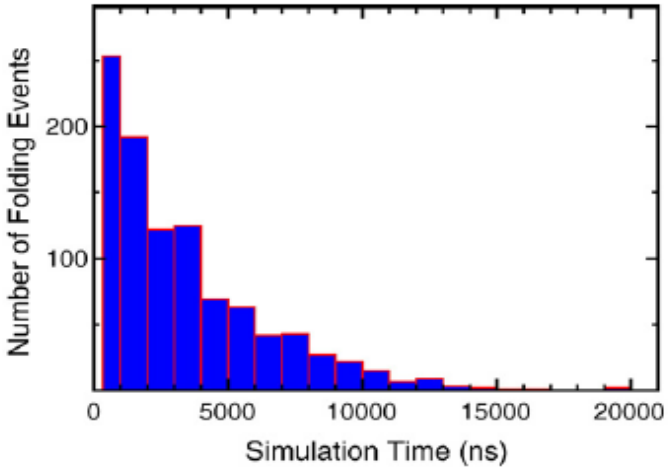
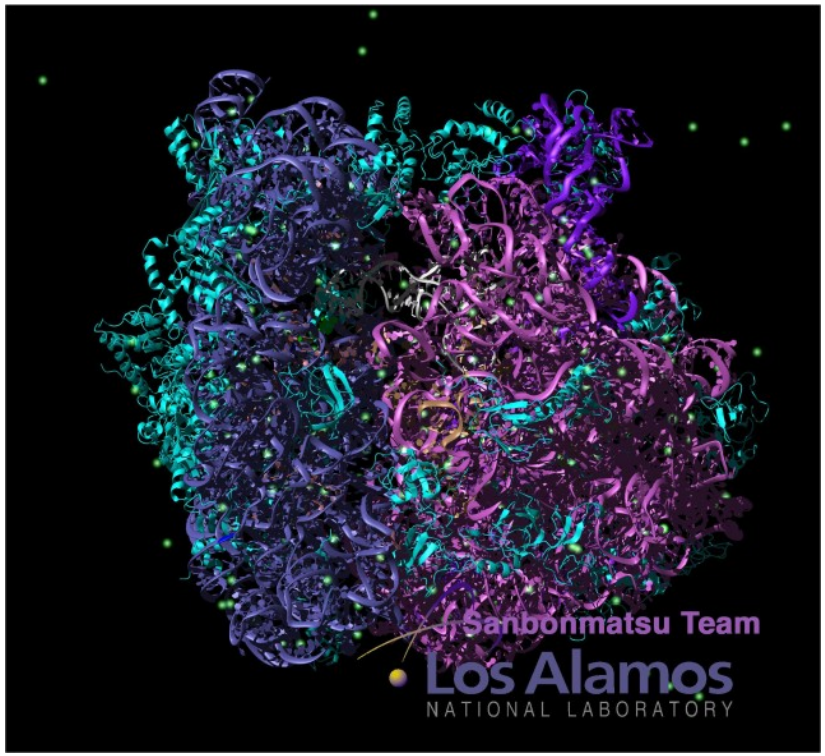
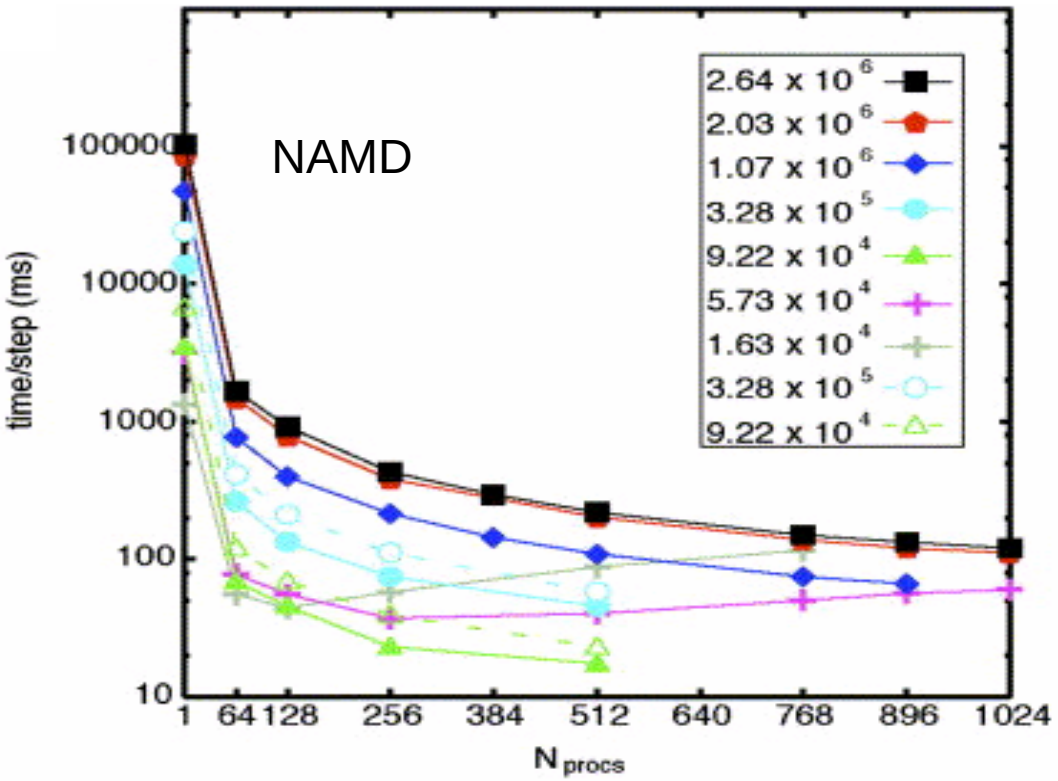



**Figure 6.** The mean simulated folding time is (3.4± 0.1) μs from 1000 folding events by using Langevin simulations, in which the initial starting point is $R=1.2$ and the system is considered to be folded when it reaches $R=4.8$. The experimental folding time is 6 μs.

# Modeling Complex Biomolecular systems

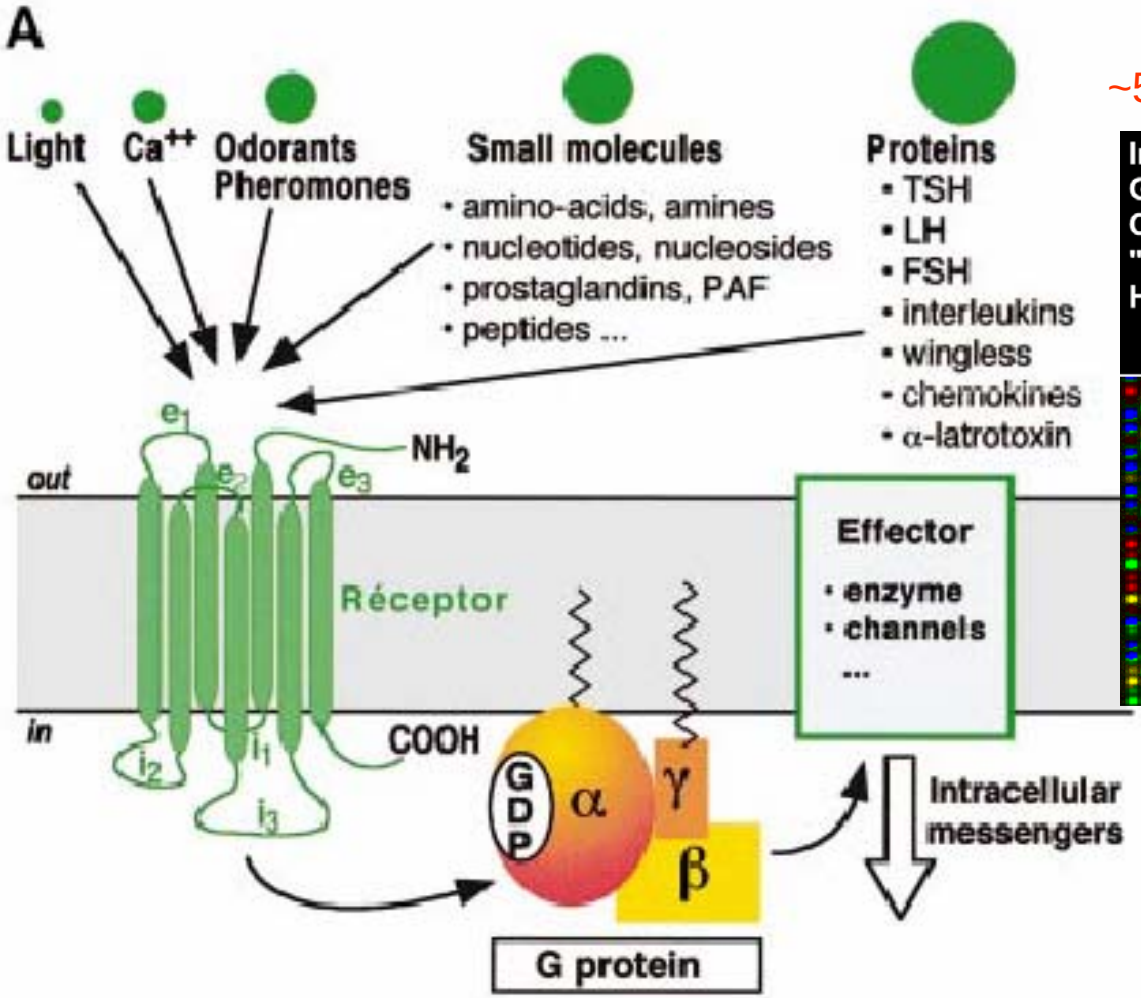
# Scalability



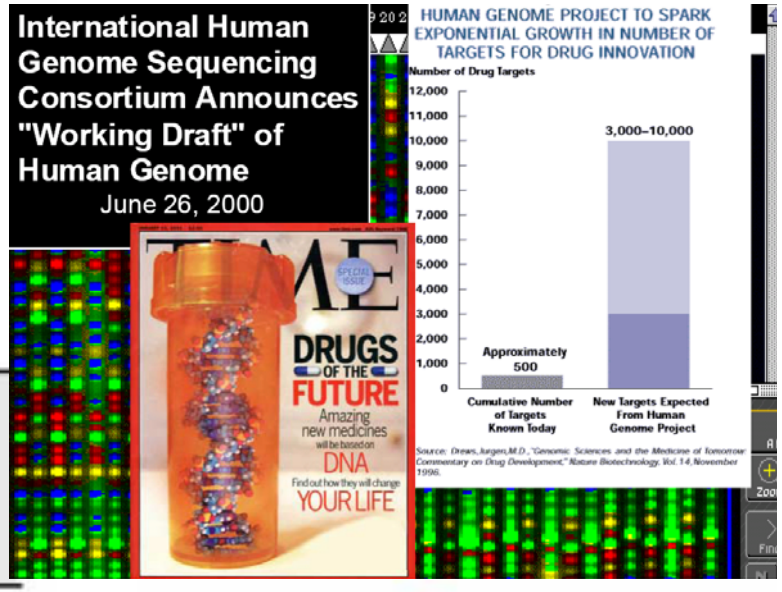



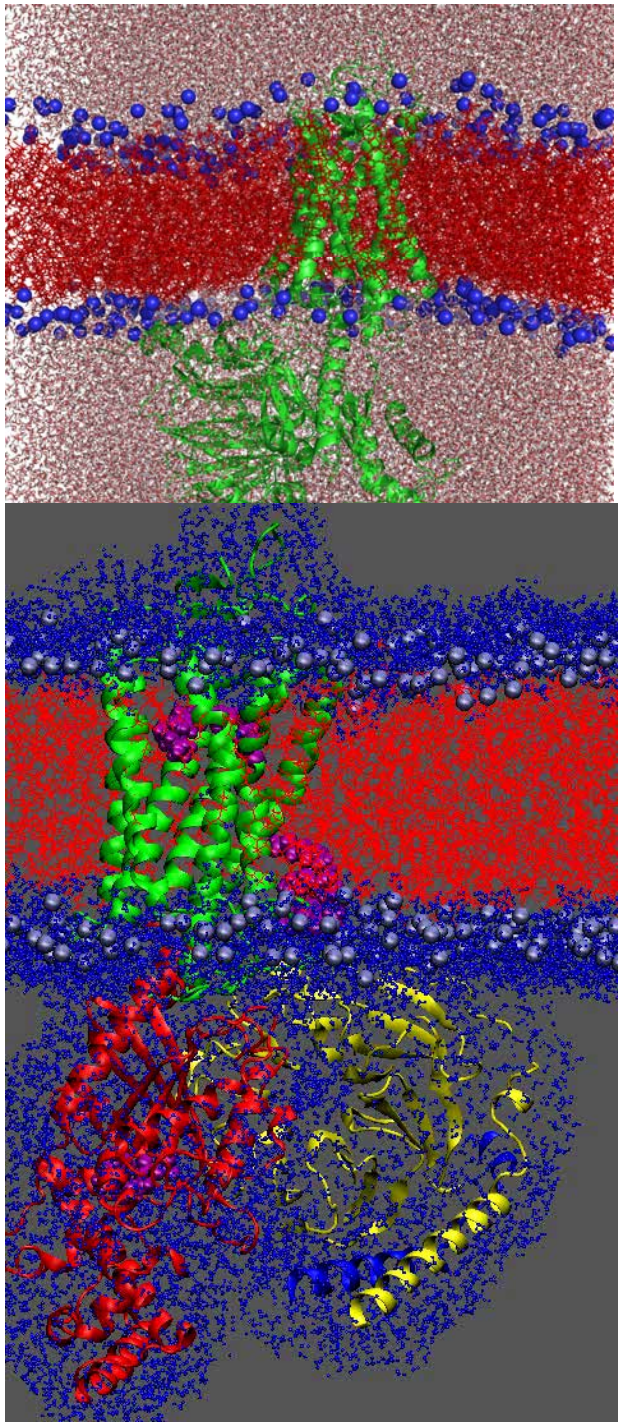

*Sanbonmatsu and Tung, 2006*

# G-protein coupled receptors



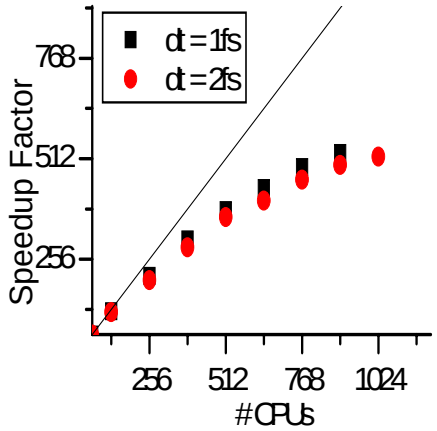

~50% of known drugs act on GPCRs!

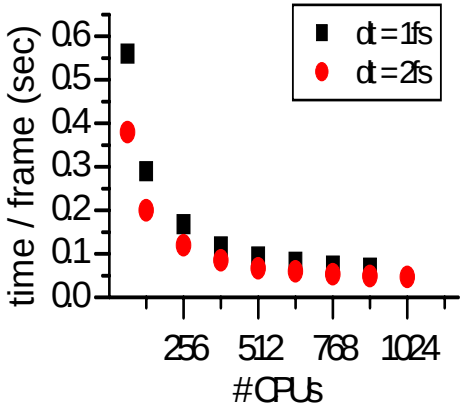


# Modeling a GCRP complex

MD code: NAMD (UIUC)




400k atoms

Docked conformations provided by Dr. Francesca Fanelli and coworkers, Dulbecco Telethon Institute, Italy

# Conclusions

- computer simulations allow the study, in full atomic detail,  of simple steps in biological processes
- Developed models can be used in the design of drugs, vaccines, and nanomaterials.

# Acknowledgements


- Nikolaos Sgourakis and Henry Herce, RPI
- J.N Onuchic, H Levine, S Yang, UCSD
- V Pande, Stanford
- D Paschek, U of Dortmund

NSF (MCB 0543769)
NIH Molecular Libraries Roadmap (1P20HG003899-01)
RPI
IBM SUR grant (BG/L rack)


Rensselaer

IBM