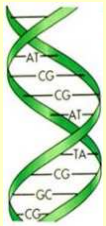
Theoretical Biophysics
Florida International University



Protein Folding: Energy, Entropy, and Prion Diseases

Bernard S. Gerstman
Department of Physics
Florida International University
University Park
Miami, FL 33199
305-348-3115
gerstman@fiu.edu

Abstract

Living systems are the epitome of self-organized complexity. The self-organization occurs on all scales, from the molecular up to the organismal level. The machines responsible for maintaining organization are protein molecules that receive energy and convert it to work. However, protein molecules themselves must self-organize into highly specific shapes. The folding of proteins is a self-organizing process in which a long chain heteropolymer in a disorganized configuration spontaneously changes its shape to a highly organized structure in milliseconds. I explain how the energy and entropy landscape of protein chains is shaped to allow self-organization. I also show how these principles can be used in molecular level investigations of protein-protein interactions that lead to both beneficial dimerization or disastrous, disease producing and potentially fatal protein aggregation.

Contributors

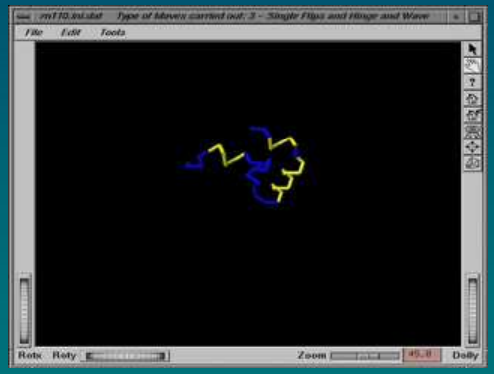
Dr. Prem Chapagain
Jose Parra
Yanxin Liu
Yonni Garbourg
Cary Baker

First: energy and entropy of self-organization of individual proteins

Later: protein-protein interactions and disease causing protein aggregation

Folding simulation: self-organization of 4-helix bundle protein

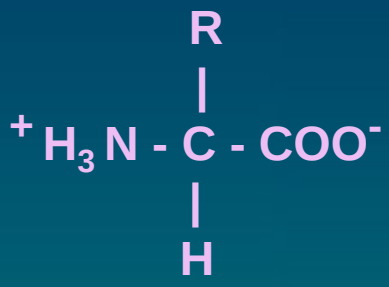
yellow =
helical turns

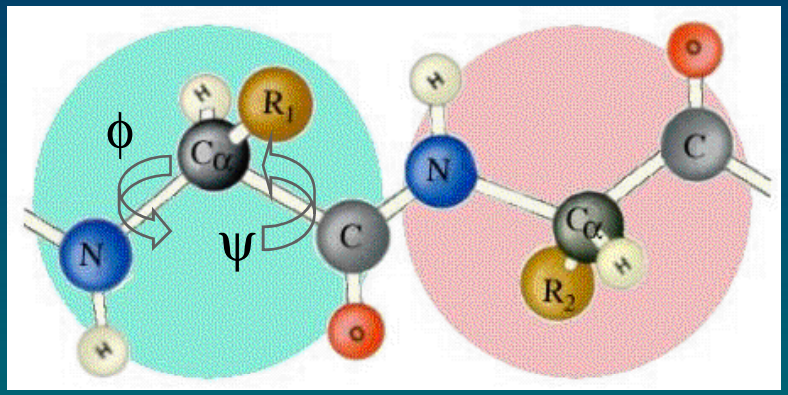


Proteins: structure and dynamics

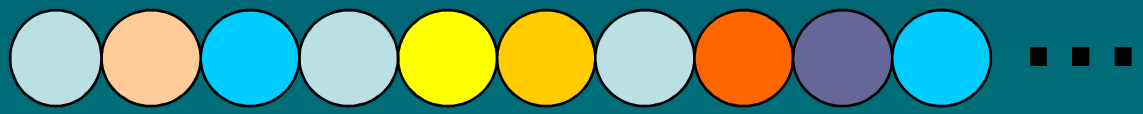
Protein structure

Protein: heteropolymer chain made of amino acid residues





Each amino acids has three flexible degrees of freedom: Φ, Ψ, sidechain



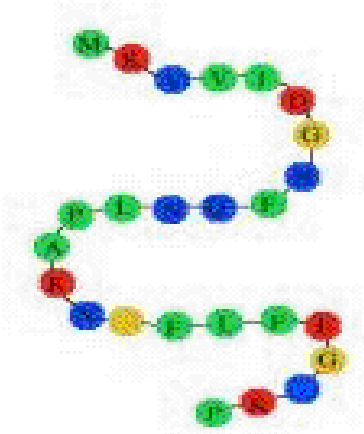
Chain of amino acid residues

20 different amino acids

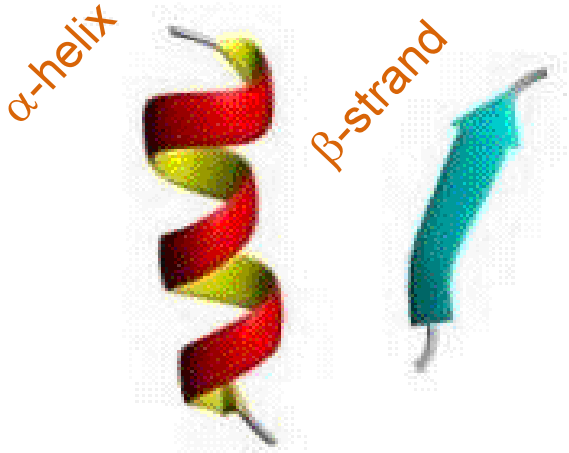
More than 50,000 different proteins in human body alone

Protein structure

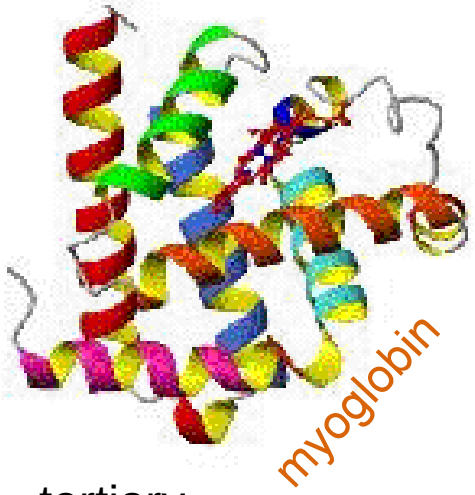
Hierarchical levels of structure:



primary



secondary



tertiary

Linear chain of amino acids (info in DNA: clever biology)

Local regular structures

3-D compact structure with long-range contacts

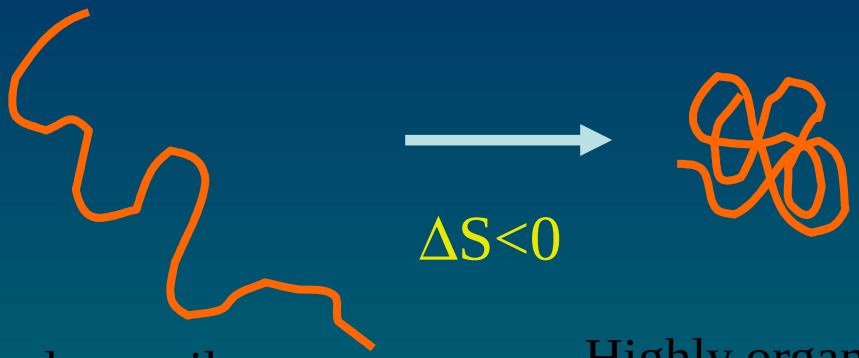
clever physics

The biological function is determined by shape.

The shape is determined by primary sequence of amino acids. How? Efficient folding route through configuration space also determined by primary sequence → Protein Folding Problem

Protein Folding

"Protein folding": Primary sequence → Native state



Random coil

Highly organized
Compact 3-d structure

Huge variation in the possible primary sequence:
20^N (20 different amino acids, *N* is # of amino acids in a chain)
Most sequences do not fold; primary sequence must be carefully chosen

Methods for finding primary sequences that fold to specific shapes:

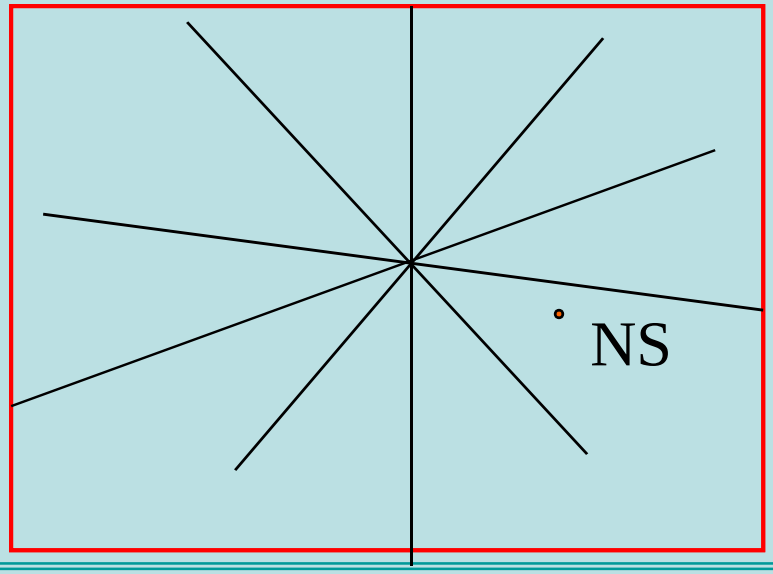
- Evolution: trial and error, requires lots of time
- Engineering: Understand underlying principles of Self-organization

Protein Folding Problem

Proteins are long (>50) chains of amino acid residues

- Biological functioning requires protein chain to fold to very specific compact shape: "native state"
- Chain is very flexible: each amino acid has internal degrees of freedom (Φ, Ψ, sidechain, e.g. 4 states each) $\Rightarrow$ > 64 configurations

Ex: Myoglobin (153 amino acids) $64^{153} = 10^{276}$ configs !!



Native State

Number of native configurations is a tiny fraction of total configuration space.

Protein Folding Problem

- Paradox: Even with super-fast sampling rate 10^{-12}sec/config , 10^{276} configs $\Rightarrow$ 10^{264} seconds (10^{256} yrs) to randomly find native state.
  (degeneracy of native state reduces this to merely10^{118} years)

Actual real protein folding times: milli-seconds !! How ??

- Folding must be a guided deterministic process, not random.

Configuration space is frustrated, ultra-metric.

FACTS

- Different initial configurations converge to native state.
- Interactions are non-linear

→Anti-chaotic dynamics !?

(B. Gerstman and Y. Garbourg, Journal of Polymer Science B: Polymer Physics, **36**, 2761-2769, 1998.)

Ultimate Physics Aim: Determine which aspects of 1-D sequence of amino acids in peptide chain determine efficient folding pathway and the final shape (native state).

Immediate Aim: For simple, small proteins, investigate dynamics of folding to known native state configurations.

Approach: Use computer lattice simulations to determine relative importance of various biophysical forces.

Why use computer model?

The system is complex

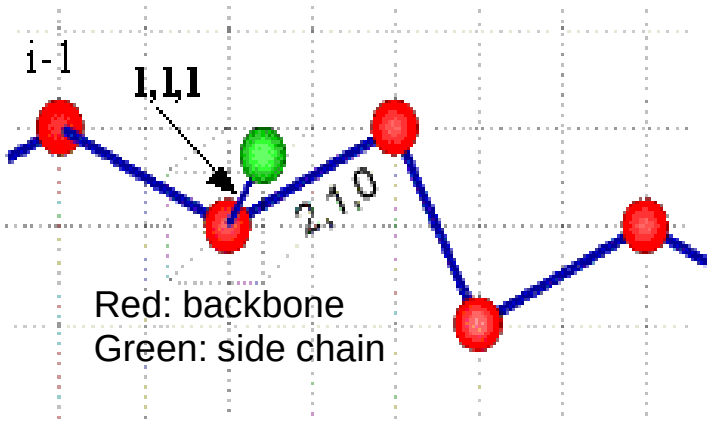
- Huge number of degrees of structural freedom
- Many terms in the Hamiltonian
- System is not solvable analytically

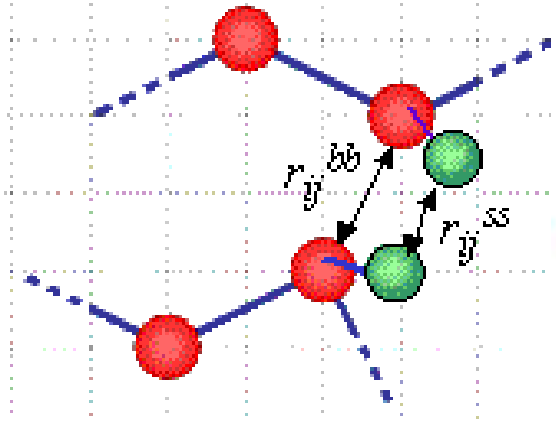
Monte Carlo simulations are very useful for these kinds of systems

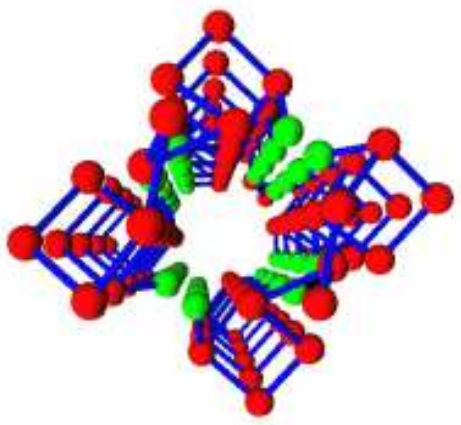
- Interested in relaxation times (non-equilibrium dynamics), as well as final configurations (equilibrium).

Computer simulation and lattice model

Lattice model and interaction Hamiltonian







Interaction Hamiltonian:

$$H = \sum_i \left(\sum_{j>i} \left(a_{ij}^{ss} \sum_p E_{ij}^{ssp} + a_{ij}^{bb} E^{bb} + a_{ij}^{rep} E_{rep} \right) + \sum_l a_i^l E_l + \sum_m a_i^m E_m \right)$$

Close enough contact (or preferred state)? Yes: $a = 1$; No: $a = 0$.

ss : sidechain-sidechain
bb: backbone-backbone
l : local
m : cooperative
p : hydrophobic or polar or hydrophilic

Protein Configuration Energy Determined by Interaction Hamiltonian

$$H = \sum_i \left(\sum_{j>i} \left(a_{ij}^{ss} \sum_p E_{ij}^{ssp} + a_{ij}^{bb} E^{bb} \right) + \sum_l a_i^l E_l + \sum_m a_i^m E_m \right)$$

i,j : amino acid residue number in the primary sequence.

a_{ij}^{ss}: are sidechains of i and j close enough to interact; yes = 1, no = 0.

E_{ij}^{ssp}: sidechain-sidechain energy (p = 1 hydrophobic-hydrophobic, p = 2 hydrophilic-hydrophilic, p = 3 hydrophobic-hydrophilic).

a_{ij}^{bb}: are backbones i and j close enough to interact; y=1, n=0.

E^{bb}: backbone-backbone interaction energy (hydrogen bond, dipole, soft core repulsion combined together)

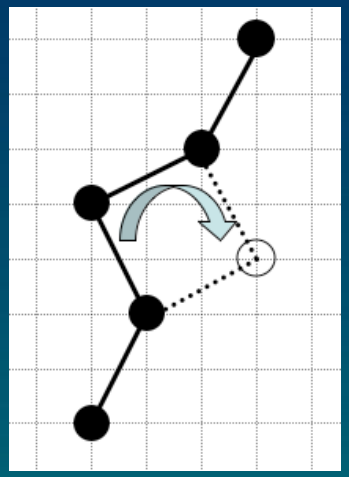
a_i^l: are residues i-1, i, i+1, arranged so that 'i' is in its preferred user-defined local configuration (i.e. α-helix, β-sheet, turn); y=1, n=0.

E_l: local propensity energy.

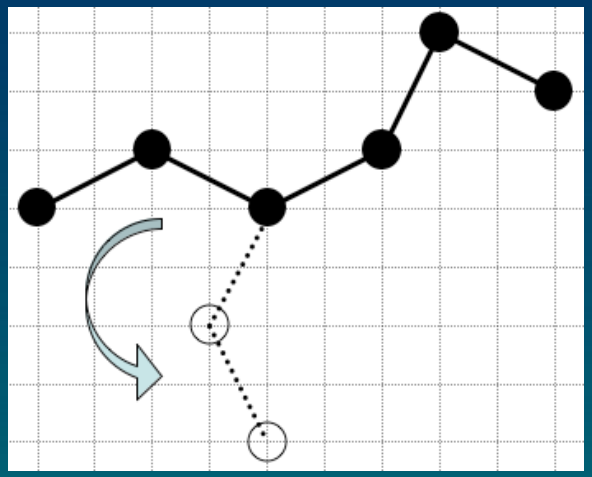
a_i^m: are residues i-1, i, i+1, i+2 arranged so that i and i+1 are both in the same preferred local configuration; y=1, n=0

E_m: medium range (cooperative) propensity energy

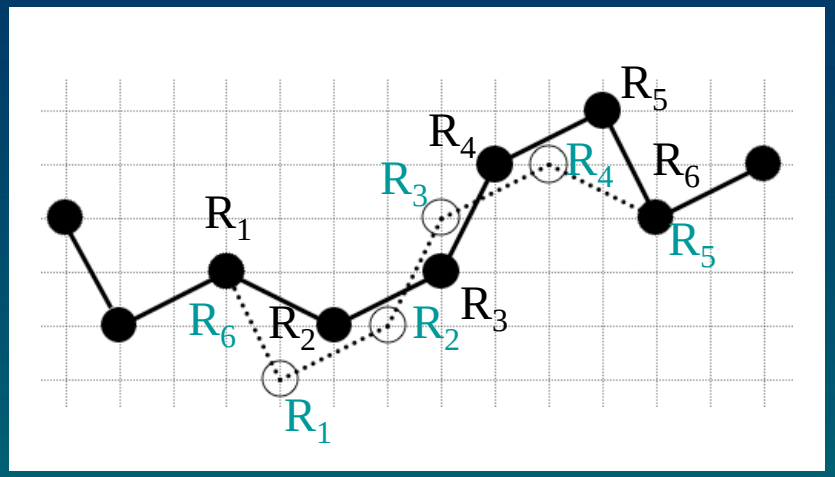
Types of attempted moves on the lattice



Singleton moves



Hinge-like move



Wave-like move

Multi residue moves

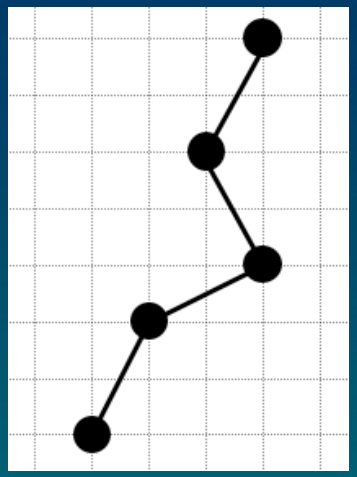
Each Monte Carlo step (~10^{-9}sec) consists of:

N attempts of single residue moves (*N* is # of residues in chain)

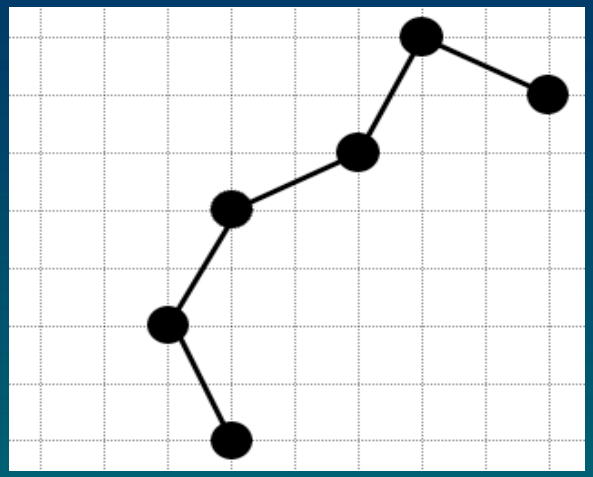
One attempt of multi-residue hinge move

One attempt of multi-residue wave move

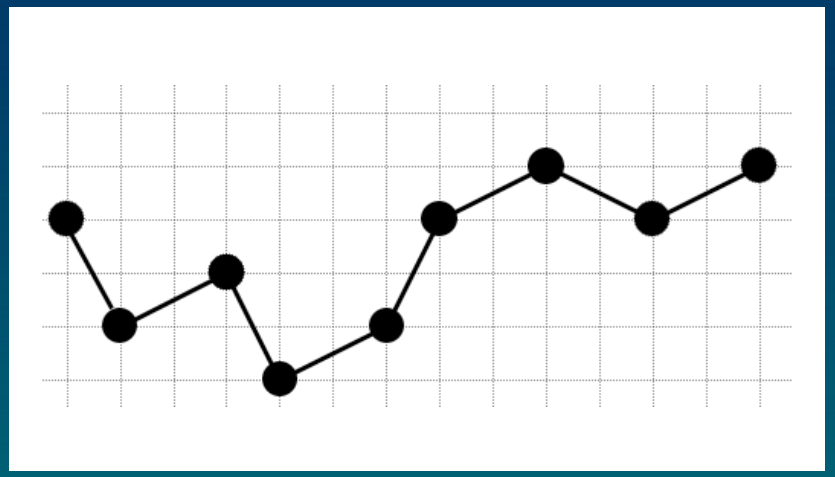
Types of attempted moves on the lattice



Singleton moves



Hinge-like move



Wave-like move

Multi residue moves

Each Monte Carlo step (~10^{-9}sec) consists of:

N attempts of single residue moves (*N* is # of residues in chain)

One attempt of multi-residue hinge move

One attempt of multi-residue wave move

Statistical Mechanics Acceptance criteria: Metropolis Algorithm

1. Calculate the Energy of the chain, E_{old}
2. Attempt to move the chain
3. Calculate the Energy of the new chain configuration, E_{new}
4. If $E_{new} <= E_{old}$, accept the move and update the chain config.
5. If $E_{new} > E_{old}$, calculate Boltzmann factor $e^{-\Delta E/kT}$; $\Delta E = E_{new} - E_{old}$
   Compare with a random number, r.
   If $e^{-\Delta E/kT} > r$, accept the move and update config.
   If $e^{-\Delta E/kT} < r$, reject the move and go to step 2.

$$P_{accept} = \begin{cases} e^{-\Delta E / kT} & \text{if } \Delta E \text{ is positive} \\ 1 & \text{otherwise} \end{cases}$$

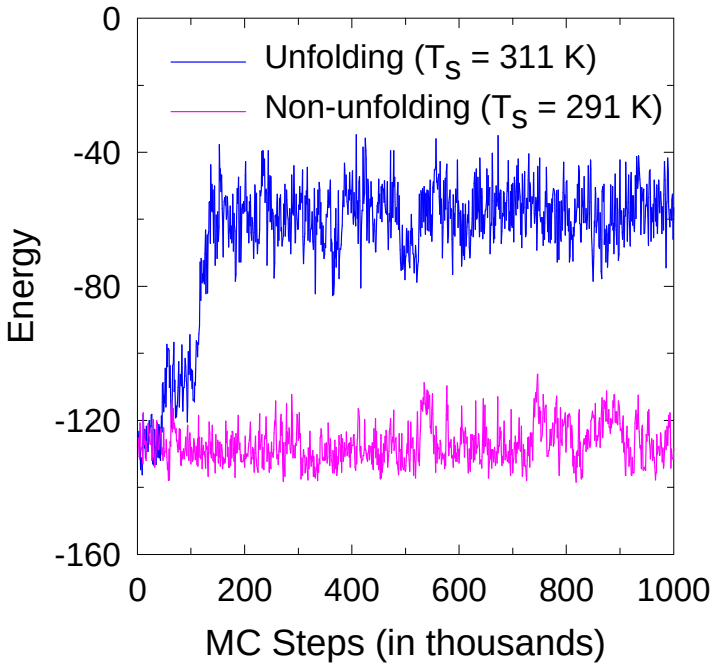
The scheme is ergodic (any configuration can be reached from any other configuration)

Generates canonical ensemble: conformations weighted by Boltzmann probability

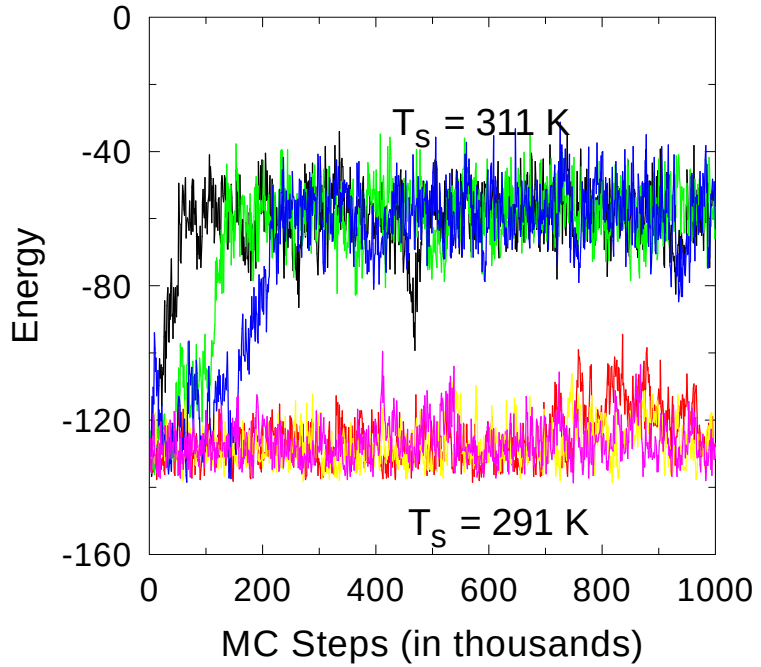
Simulated Data from Model: Time Series of Structure, Energy, etc.

Folded four-helix bundle

Unfolded random coil



Multiple runs with different random numbers

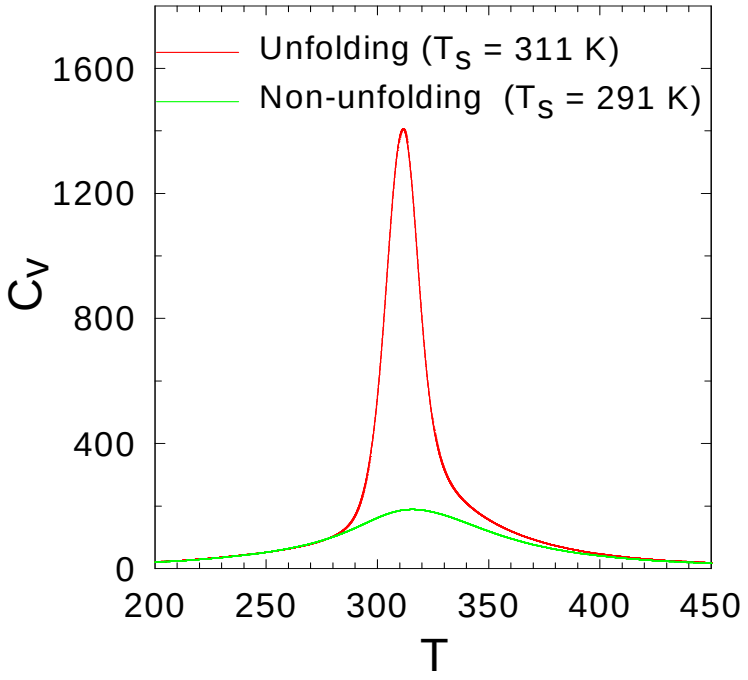


Check that Simulations of Model are Physically Realistic: Temperature dependence of Heat Capacity calculated using Monte Carlo histogram technique.

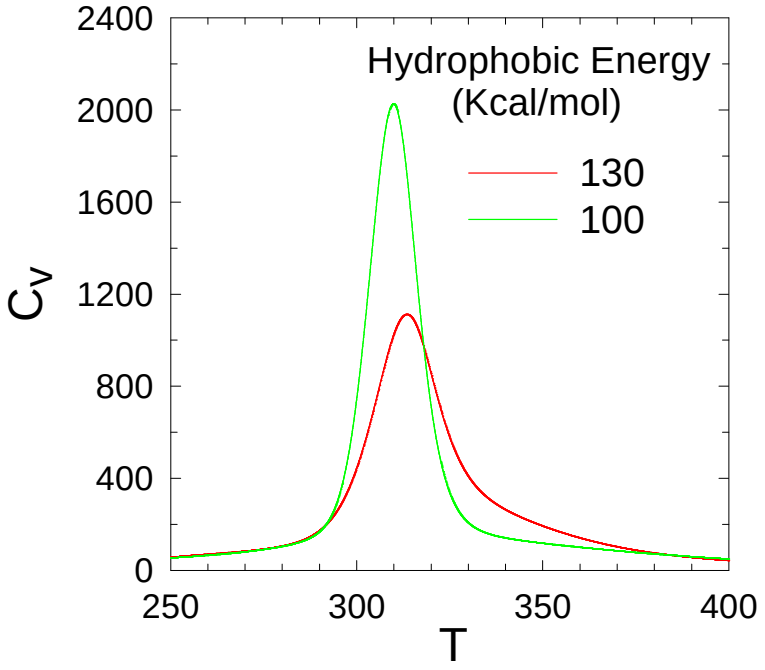
$$C_v = \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2}$$

where the average quantity is calculated using:

$$\langle Q \rangle = \frac{\sum_E h(E, T_s) Q(E) e^{-\frac{E}{T} + \frac{E}{T_s}}}{\sum_E h(E, T_s) e^{-\frac{E}{T} + \frac{E}{T_s}}}$$



As Expected: Unfolding runs show a sharp peak in the heat capacity curve indicating a first order like phase change where-as non-unfolding runs give a much flatter curve.



As Expected: As the strength of hydrophobic interaction is increased, the peak shifts towards the higher temperature implying increased stability.

Free energy landscape and kinetics: Protein engineering to enhance folding

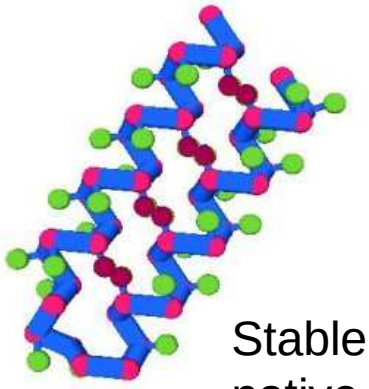
P. Chapagain and B. Gerstman, Biopolymers, **81(3)**, 167-178, 2006.

Protein engineering (designer drugs): strategic placement of specific amino acids to enhance folding to desired shapes.

Investigation of physics of underlying dynamics

Sequence design: substitute just two H for P → big difference

Seq A
Trap-containing (bad)



-H:H-
-H:H-
-H:H-
-H:H-

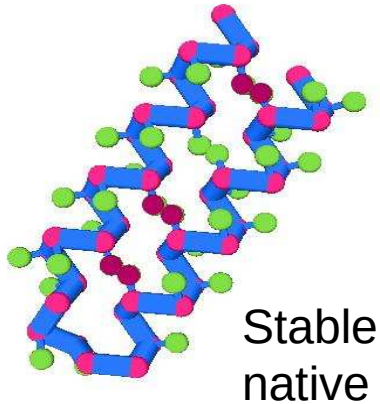
Native State Bonds

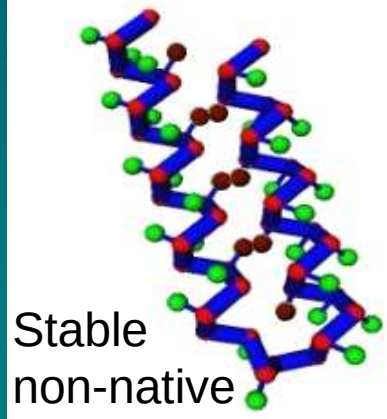
4 strong H:H bonds: stable

3 strong H:H Bond: unstable

-H:H-
-P P-
-H:H-
-H:H-

Seq B
Trap free (good)

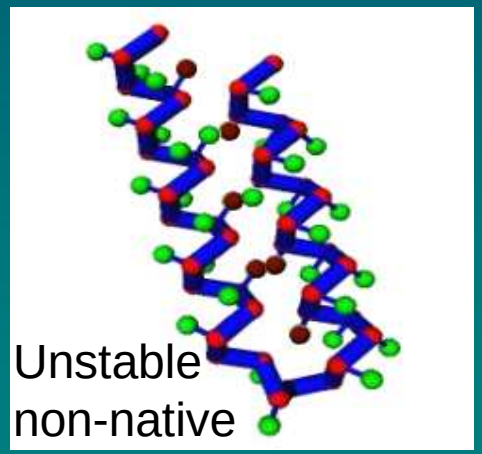




-H
-H:H-
-H:H-
-H:H-
H-

Non-Native State

-H
-P↔H-
-H↔P-
-H:H-
H-



Non-obligatory traps:
Folding is slow and unreliable

Trap free: folding is fast and reliable

Folding to stable native state: Seq A with traps takes 2x as long as trapless Seq. B

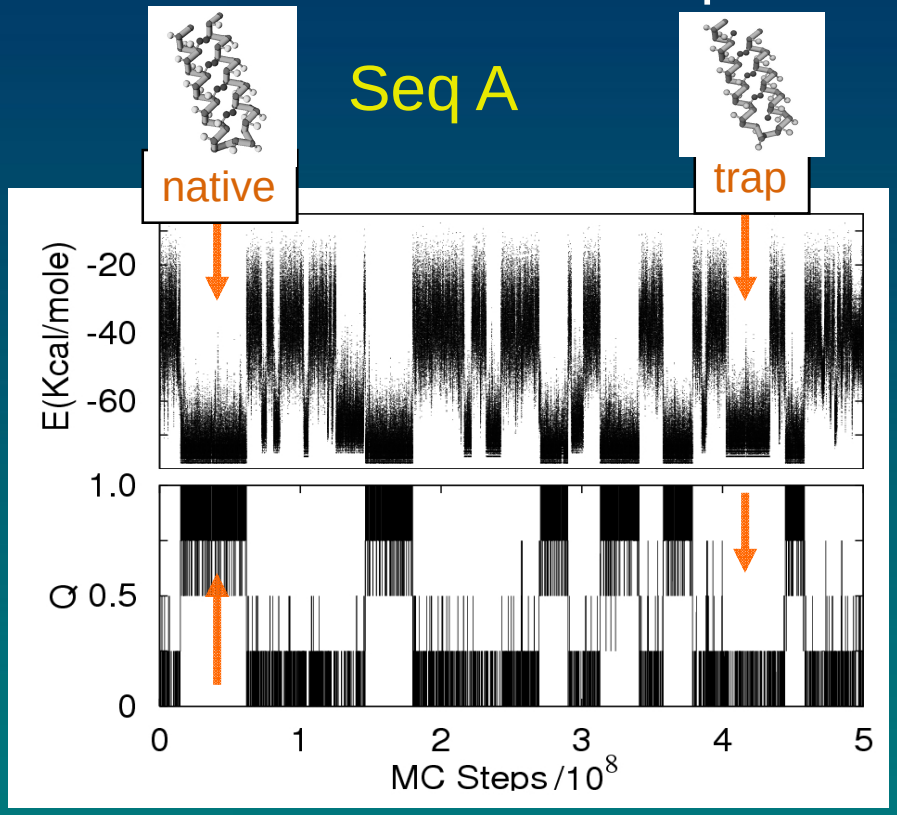
Unfortunately: Seq A also Mis-Folds to long-lived, non-native trap configuration which looks similar but is biologically useless

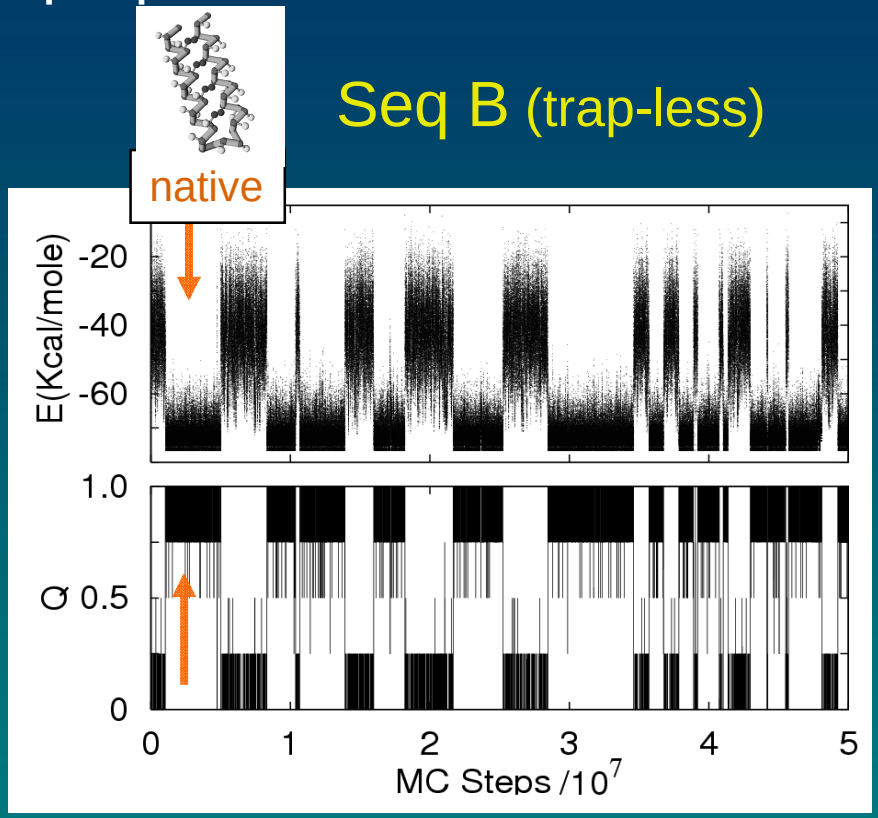
QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

Seq B wastes little time in the non-native trap configuration and folds quickly and reliably: Why?

Connecting Thermodynamics, Statistical Mechanics, Kinetics of Sequence design: Explain trap-containing vs. trap free folding

Equilibrium properties

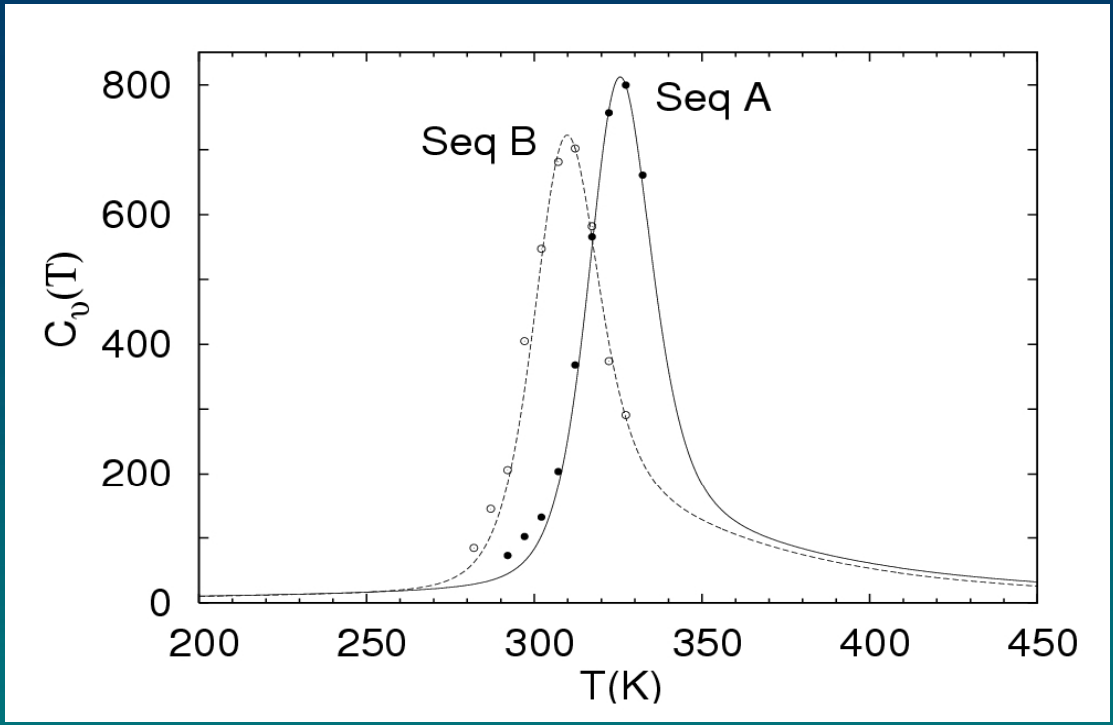




- Native state: Q~1 lowest energy conformation E=-77.8
- Long-lived low energy conformations Q~0 (non-native traps)
- Thermodynamically three states

- Native state: Q~1 lowest energy conformation E=-76.55 (~same)
- No long-lived low energy non-native (Q~0) conformations
- Thermodynamically two-state

Thermodynamics: heat capacity

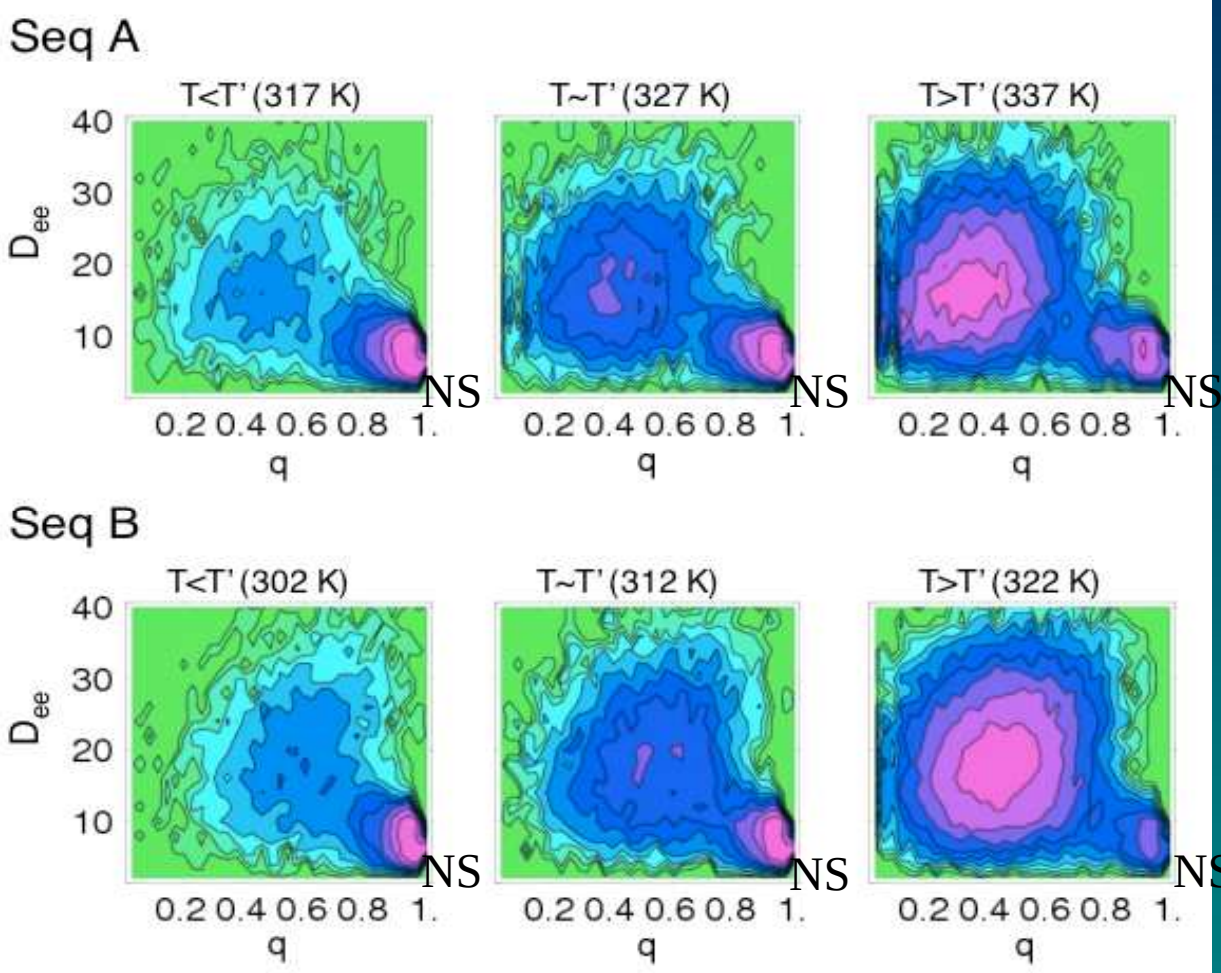


Small difference in stability of native state because Seq A has 4 H:H bonds whereas Seq B has only 3

Heat Capacity Peaks: First order like transitions

T' ~310K (Seq B) T' ~325K (Seq A)

Deeper level of understanding the differences in folding: Free energy landscapes $F(q,D_{ee}) = -kT \ln P(q,D_{ee})$



Pink: low *F*

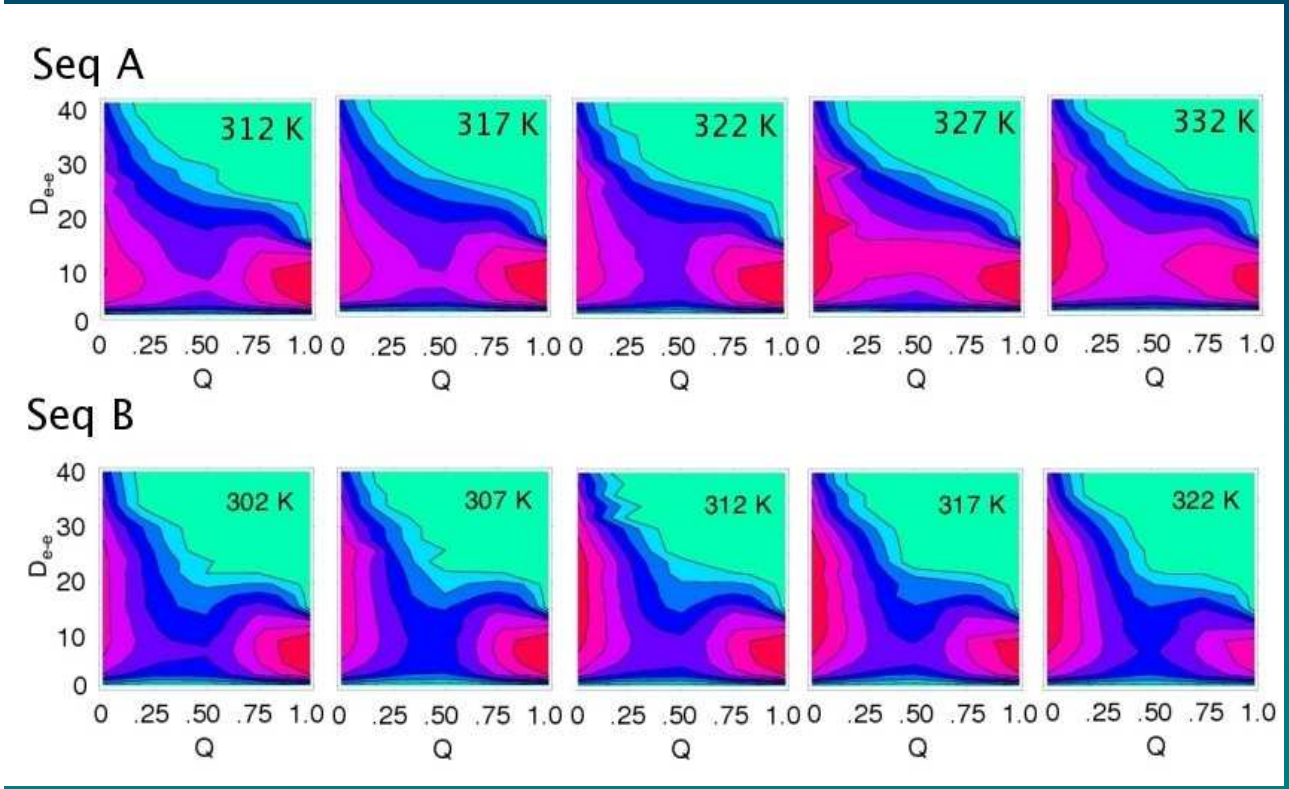
Similar free energy landscapes from end-to-end distance, D_{ee}, and helicity, q (secondary structure) →

H to P substitutions have little effect on secondary structure formation

Helical secondary structure formation is NOT the reason for differences in folding

Free energy landscapes $F(Q,D_{ee})= -kT \ln P(Q,D_{ee})$

Amino acid substitutions cause subtle but important changes in free energy landscapes in terms of tertiary contacts, Q.



Seq A:

Lower T: deep non-native (trap) minimum localized at low D_{e-e} and Q~0

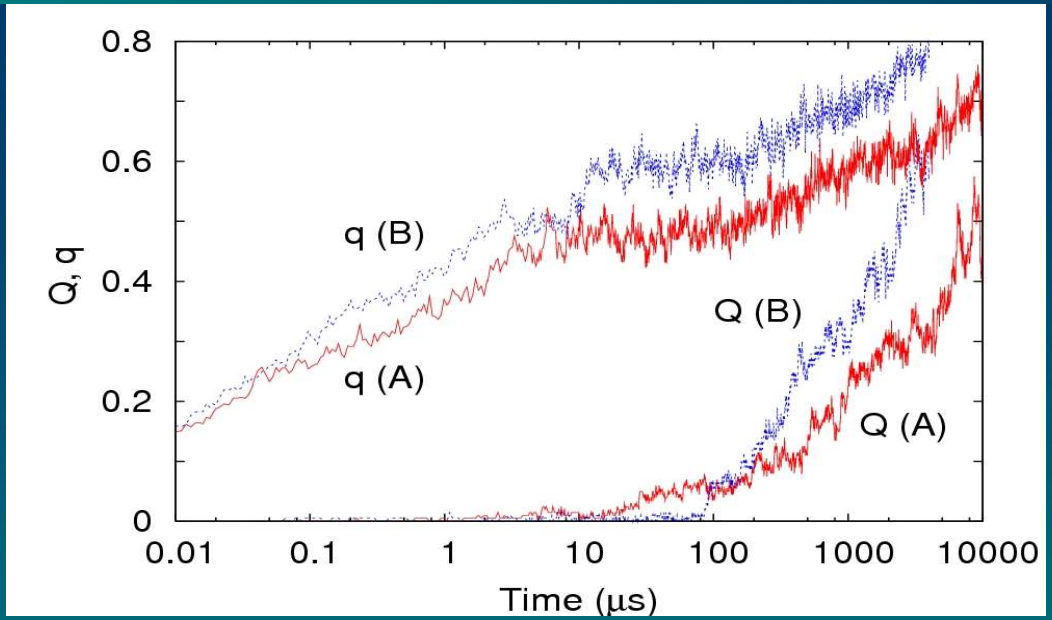
Folding is slow when native state is stable (T<T')

Seq B:

Low T: shallow non-native minimum widespread over a range of D_{e-e} (easy to get out of: not a trap)

Folding is fast when native state is stable (T<T')

Kinetics: time evolution



Time evolution of the parameters q (secondary helicity) and Q (fraction of native tertiary contacts)

Secondary structure formation is similar:

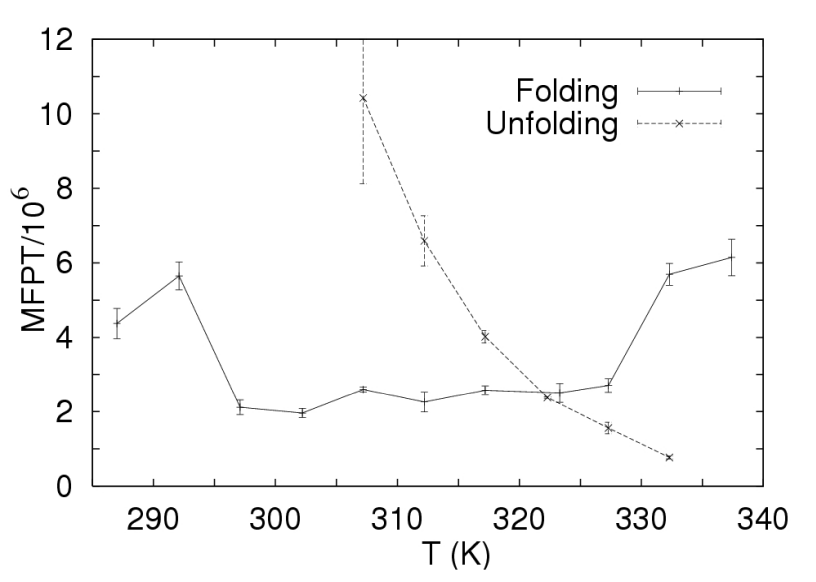
~ 50% helicity (q) exist for both A and B by 10 μs

Tertiary structure formation is different:

interhelical contacts (Q) form at different rates;
Q(B) much faster than Q(A)

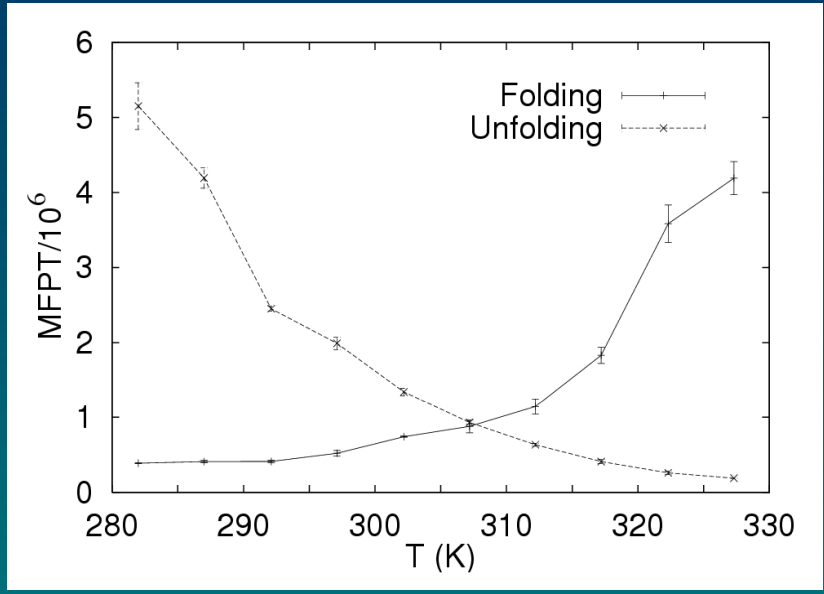
Kinetics: Median First Passage Time (MFPT)

Seq A



MFPT increases at low temperatures
⇒Presence of traps

Seq B



Folds faster than Seq A

MFPT follows a monotonic decrease with decrease in temperature
No traps ==> Fast folding

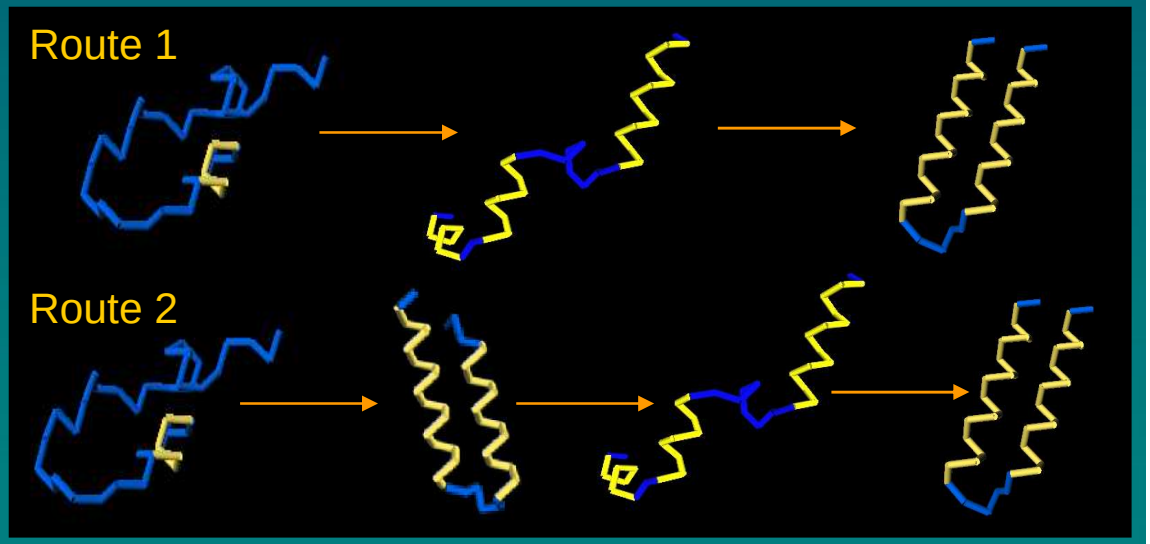
Folding Kinetics

Seq B: Follows single exponential kinetics
All folding routes are trap free: kinetics are described by a single exponential

Seq A: Follows double exponential kinetics
Two folding routes

Route 1: Trap free.
Directly reach native state.
Seq B and fast Seq A

Route 2: First fall into trap then fold to native state.
Slow Seq A



Conclusions from folding of 2-helix bundle

- α-t-α hairpin peptide: important model system for studying protein folding kinetics and thermodynamics
- Small changes in the primary sequence may lead to a significantly different free-energy landscape
- Engineering sequence to remove traps smoothes free-energy landscape and makes protein fold significantly faster
- non-two state to two state
- Important from protein engineering point of view

Next: Investigate at deepest level; statistical mechanics of micro-states

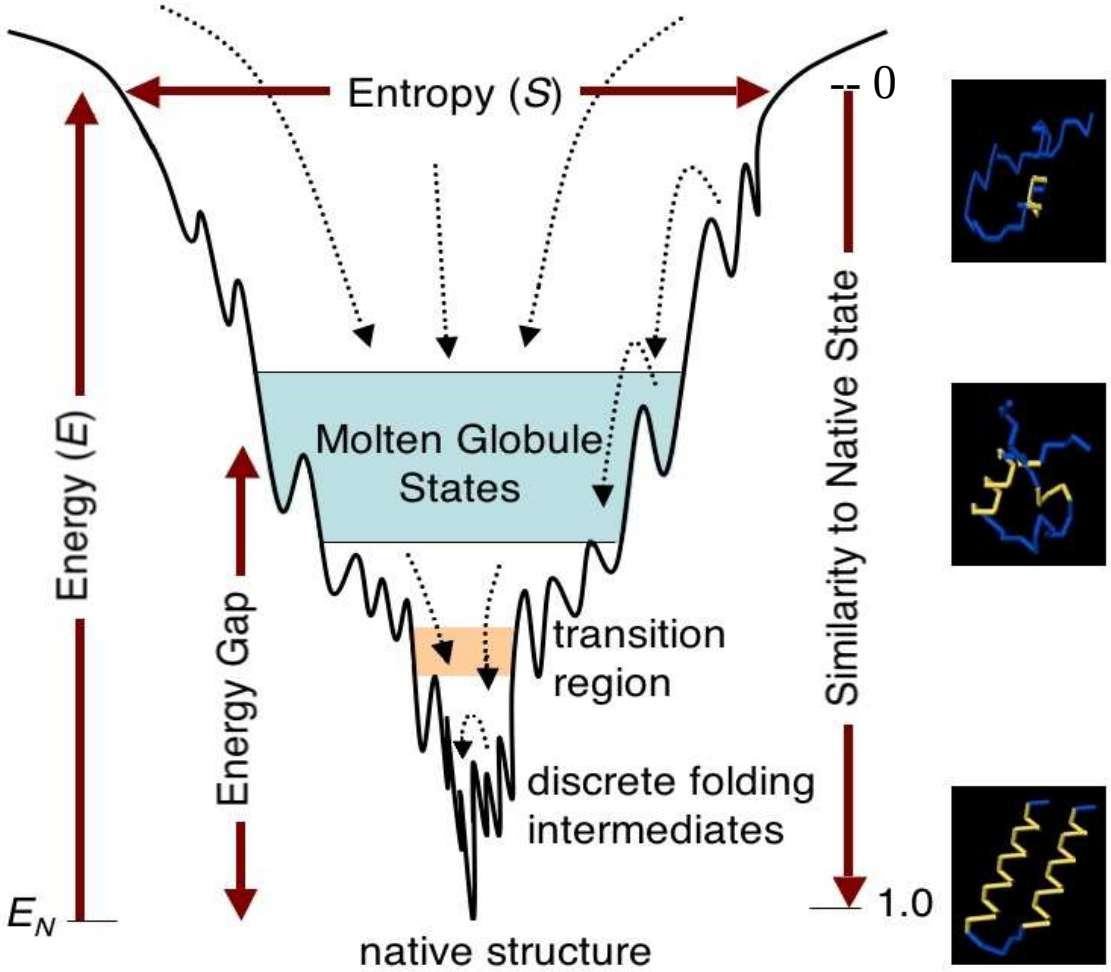
Deepest Understanding of Deterministic Guidance of Folding

Statistical Mechanical Landscapes: microstates, folding funnels

(P. P. Chapagain, J. L. Parra, B. S. Gerstman and Y. Liu, JCP, **127**, 075103, 1-7, 2007.)

Many other axes are necessary to represent all the structural degrees of freedom.

Which are most important?



Counting States to get $S(E) \equiv k \ln \Omega(E)$

Desire: look at each configuration and determine E.

Create 2-column table: [Configuration | E] → Sort table to get $\Omega(E)$

Problem: Too many configurations (Levinthal Paradox for Computer Modeling)

Each residue: 18 possible configurations

Small chain: 33 residues, 31 residues have configurations (R-states)

Total number of possible configurations: $31^{18} \sim 10^{27}$.

Too large for a computer to look at each state and record E.

Note: Most of these configurations are self-intersecting and therefore not allowed.

However, if only 1% are non-self-intersecting and allowed, still 10^{25} possible.

If computer can examine 10^9 configurations/second, still requires 10^{16} secs

$\sim 10^9$ yrs

to determine the energy of each possible configuration to get $\Omega(E)$.

Need a better way to sample states to get $\Omega(E)$.

Method to Get $\Omega(E)$

Run simulations to canonically (Boltzmann) sample configuration.

Each simulation $T \rightarrow \langle E \rangle$

For each configuration in simulation, get: $P_r = N_r / N_{total}$

(e.g. $N_{total} = 200{,}000{,}000$)

First get S using $S = -k \Sigma P_r \ln P_r$

Then get $\Omega(E) \equiv e^{S/k}$,

Change to different T to change $\langle E \rangle$

Can use these relationship to create table

$\langle E \rangle$	$S(\langle E \rangle)$	$\Omega(\langle E \rangle)$
….	….	….
….	….	….
….	….	….

Results Make Sense

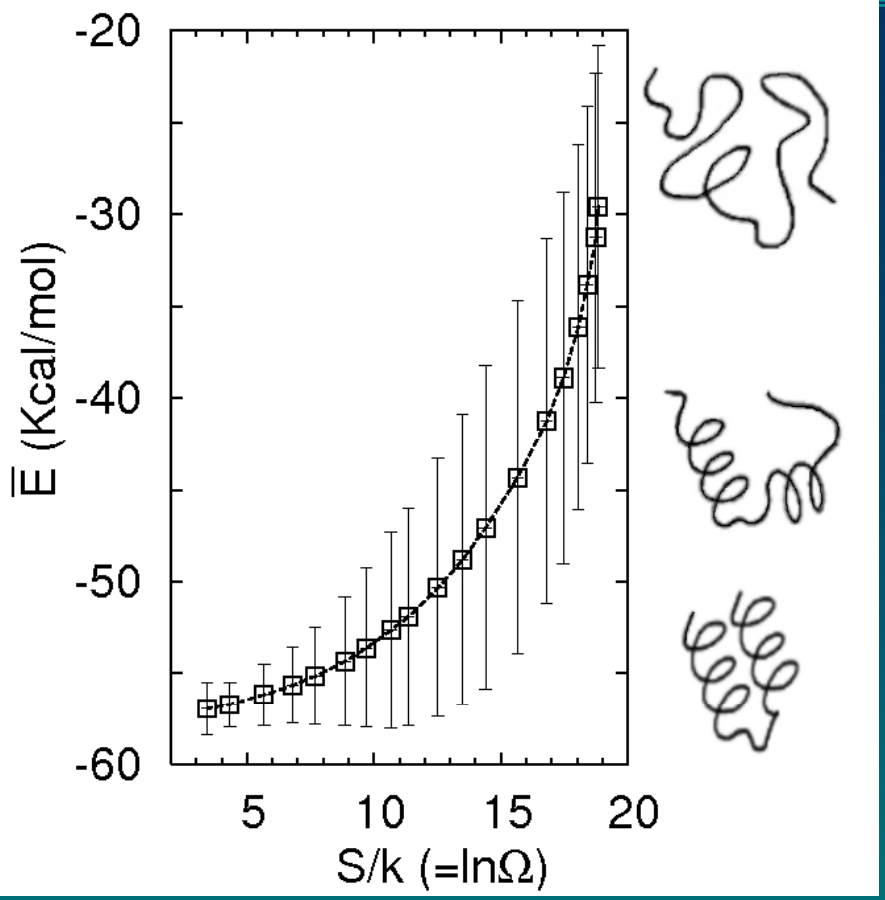
Seq B (Trap-less): Two helices, each with three turns; interface sidechains H-P-H

Total number of MC steps used: 200,000,000.

Temp (kcal/mol)	Average <E>	No. of distinct configurations	Sum [P Ln(P)]	Entropy S=-k[Sum]	Ω(<E>) exp(S/k)
0.60	-33.865	164074759	-18.4010	2.5406e-22	98046600
0.58	-38.911	137989785	-17.4677	2.4118e-22	38559900
0.57	-41.278	122502977	-16.8048	2.3202e-22	19871100
0.56	-44.336	100149056	-15.6600	2.1622e-22	6324580
0.55	-47.070	78045712	-14.3954	1.9876e-22	1785780
0.54	-48.833	63249352	-13.4846	1.8618e-22	718292
0.53	-50.343	49369314	-12.5160	1.7281e-22	272654
0.52	-51.918	35005745	-11.3520	1.5674e-22	85134
0.50	-53.613	18474192	-9.7029	1.3397e-22	16364
0.48	-54.341	11996221	-8.8594	1.2232e-22	7040
0.46	-55.146	5917739	-7.6813	1.0606e-22	2167
0.44	-55.650	2878088	-6.7973	0.9385e-22	895
0.40	-56.171	1143981	-5.6634	0.7819e-22	288
0.35	-56.718	281235	-4.3207	0.5966e-22	75

Result: detailed micro-state folding funnel landscape for trap-less Seq. B.

As expected, folding funnel is smooth with no traps.



Additional Considerations

Ω(E) reliable only if each configuration is visited multiple times to get reliable P_r.

At high E, huge Ω (10^{20}) → impossible to sample fully → Unreliable at high E.

Not so bad: shape at top probably not important: every route leads down towards native state.

Good sampling at low E because small Ω.

Future Work

- Compare folding funnel of trap-less Seq. B to Seq. A to elucidate traps
- Mid E is important: transition region where protein chooses route:

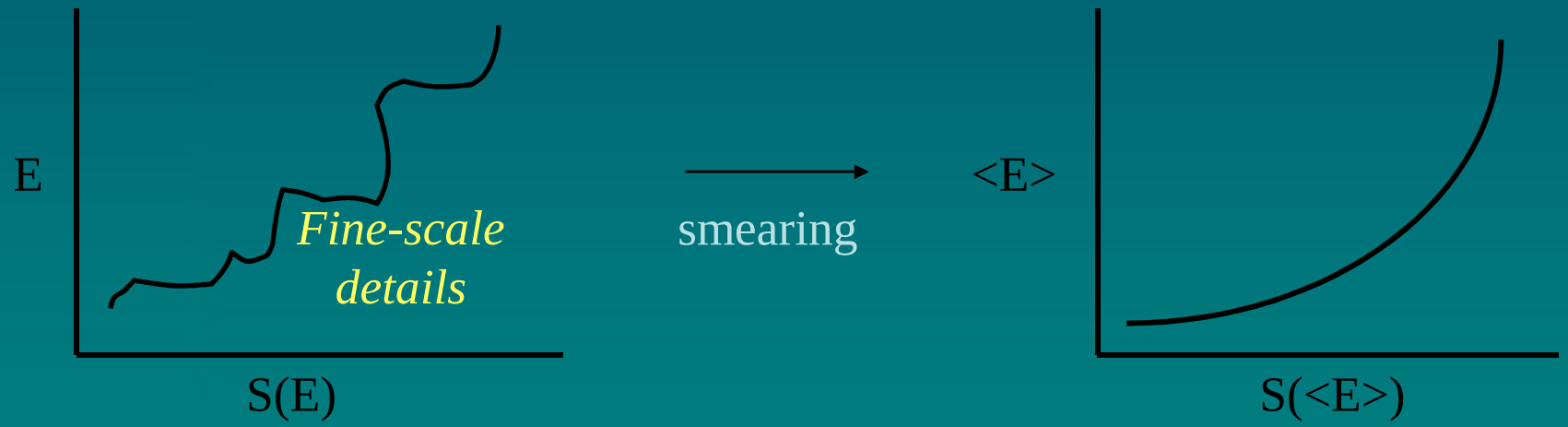
folding towards native state vs

molten globule traps vs mis-folding to non-native states

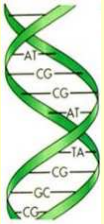
→ Want Good sampling at Mid E

- Fixed T gives <E> and therefore Ω(<E>)

Problem: Thermal Boltzmann fluctuations help by allowing us to get S via P_r, but smears out details at different E → cannot get fine-scale details of Ω(E)



Also, for ultimate details,, want Ω(E) at specific E, not Ω(<E>)



How are diseases related to protein folding?

Protein function depends on its specific fold and stability

Altering the sequence may change its binding property (chemical,electronic,etc)

May completely change its fold → Big consequences when folding goes wrong!!

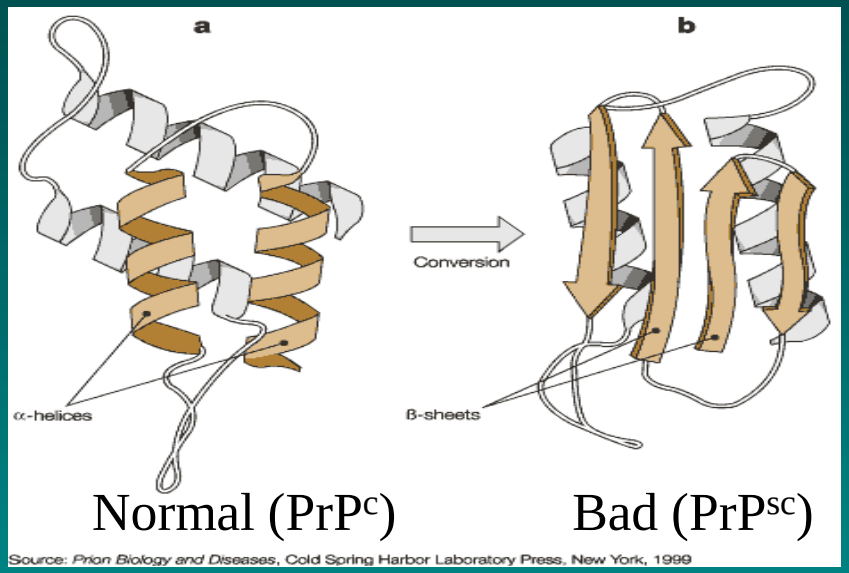
Protein Dimerization (often necessary for proper functioning)

Protein Aggregation: Causes serious diseases

Prions -- Protein misfolding, and Structure conversions. Amyloid fibrils formation, Alzheimer's Disease, Parkinson's Disease, Mad Cow Disease

Prion Proteins (bad guys spoiling normal ones)

1997 Nobel Prize -Stanley Prusiner, along with Carleton Gajdusek)



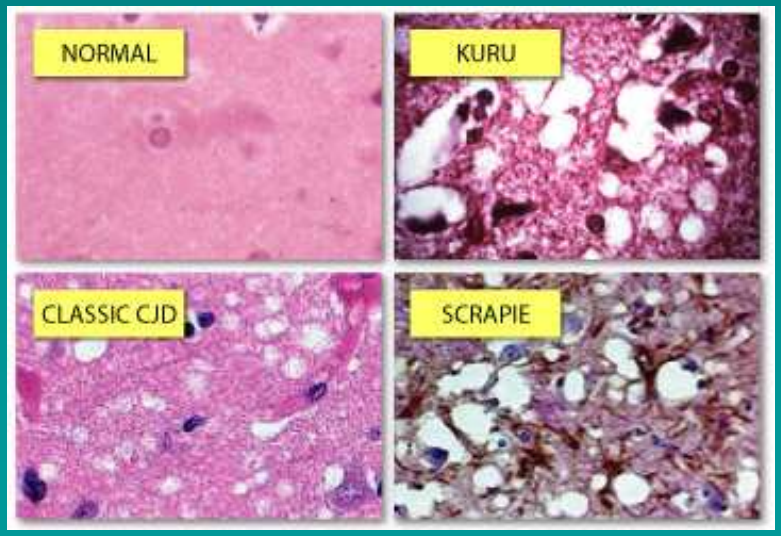
normal prion protein PrP^C is characterized by 4 α-helices

PrP^{Sc}

(disease) prion protein PrP^{Sc} is in loss of 2 α-helices, replaced by beta-sheets

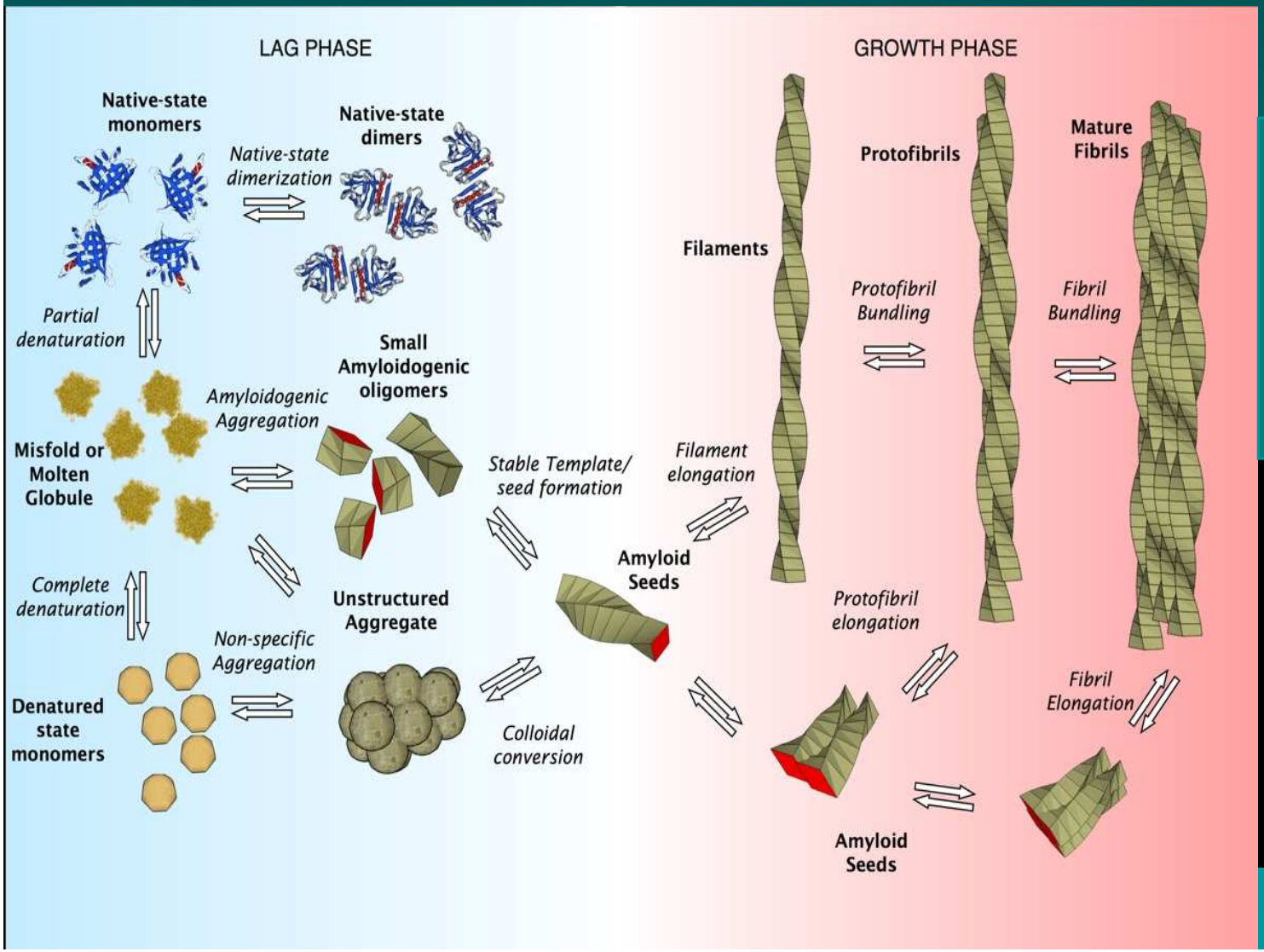
Even worse !

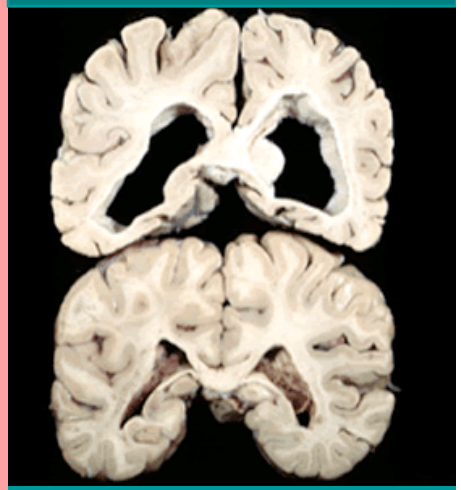
Misfolded proteins form amyloid fibrils (Alzheimer's, Scrapie, BSE, TSE- CJD)



More than 100 diseases associated with formation of amyloid fibers.

Alzheimer's, Huntington's, cystic fibrosis, BSE (Mad Cow disease), CJD, many cancers





Protein-Protein Interactions and Aggregation

- Protein Quaternary Structure formation
- Protein Dimerization for Proper Functioning
- Protein Misfolding and Aggregation - Prions
- Degenerative Diseases

  Example: Alzheimer's Disease, Parkinson's Disease, Mad Cow Disease

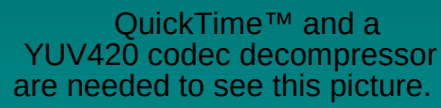
Modify our computer simulation to investigate multi-chain aggregation.

We can change the identity (properties) of any amino acid in the chain and re-run the simulation to determine the affect on the aggregation dynamics and structure.

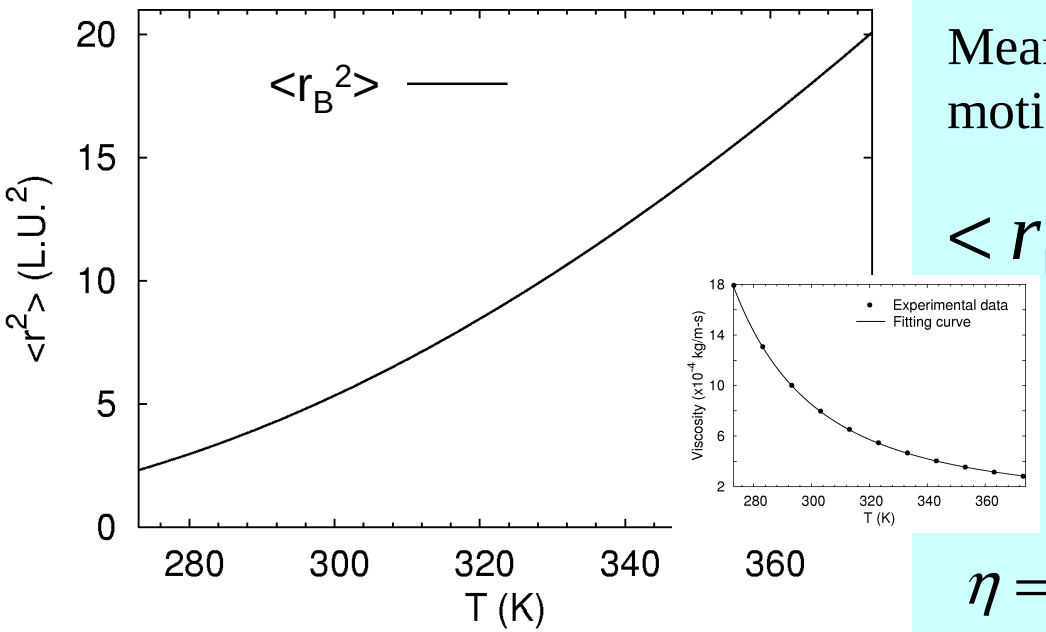
Protein engineering: Encourage beneficial dimerization

Prevent dangerous aggregation.

GCN4-p1 Leucine Zipper Folding simulation

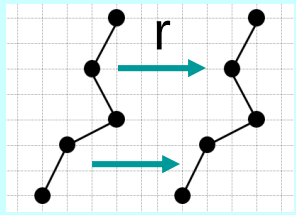


Translation Move: One Chain Relative to Other Chain



Mean square displacement from Brownian motion theory:

$$< r_B^2 >= \frac{2kT}{6\pi\eta(T)a} t$$



where

η(T): viscosity of water

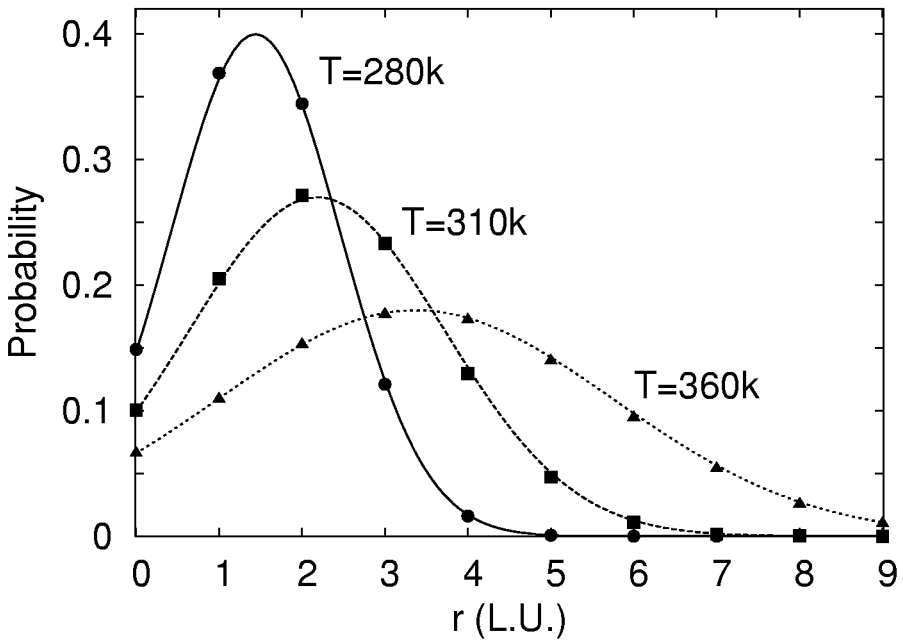
$$\eta = 3.15209\times10^{-5}\times e^{\frac{484.177}{T-153.067}} (kg/m\cdot s)$$

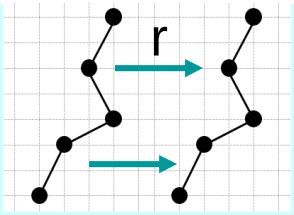
To make quantitative calculations for translation: assume protein chain is a *sphere*. (?)
a: radius of protein (20 lattice units for chain that is 39 residues long, 1 lattice unit=0.169nm)

Computer Simulations: Use Gaussian (±x,y,z) distribution that depends on T through r_B

$$P(r) = \exp\left(\frac{-\left(r - \sqrt{\frac{2}{3} < r_B^2 >} \right)^2}{2\left(\sqrt{\frac{1}{3} < r_B^2 >} \right)^2} \right)$$

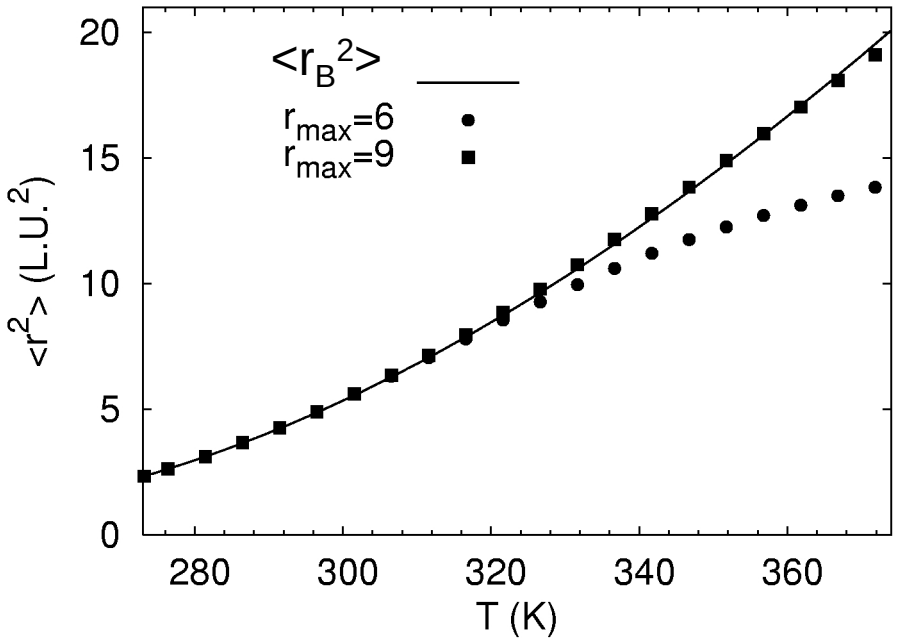
Must confirm that this "guessed" P(r) will, in computer simulations, give proper $<r^2>=<r_B^2>$ expected from Brownian Motion Theory (above), including T dependence.





Use random numbers in simulations to pick attempted integer translation distances.

← P(r) for different T



← Comparison of computer $\langle r^2 \rangle$ to $\langle r_B^2 \rangle$

When using P(r), what is smallest r_{max} that can be used?

r_{max}=6 is OK at low T, but not large enough at high T.

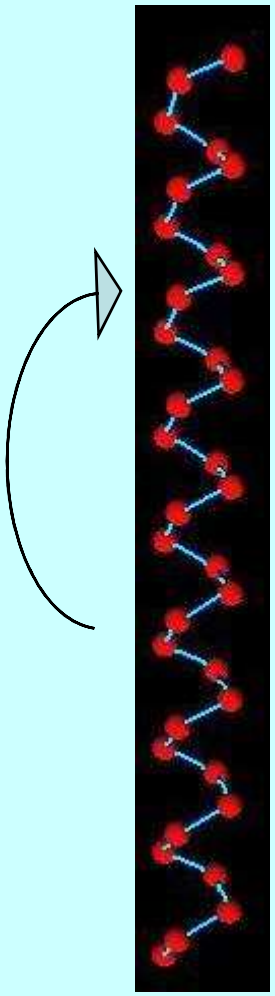
r_{max}=9 is necessary to match Brownian

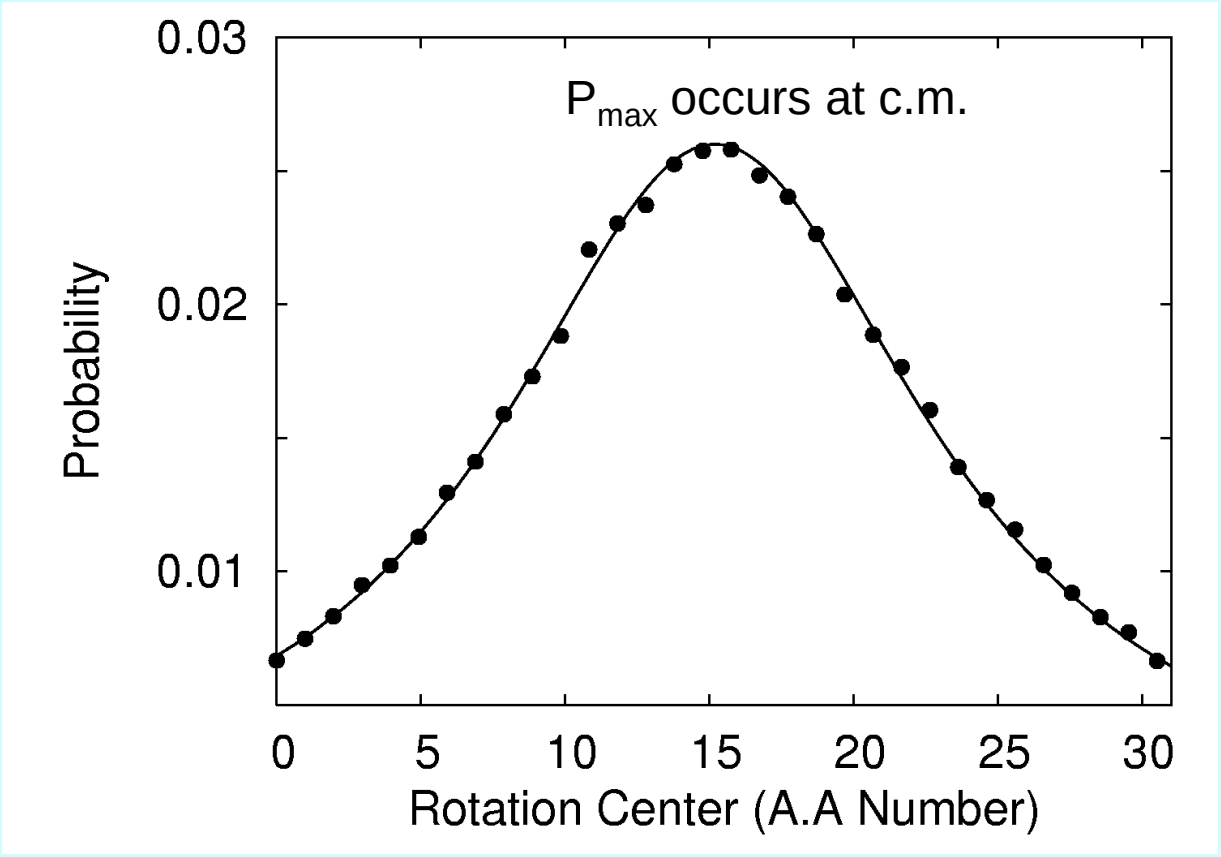
Rotation Move

Probability to rotate chain around amino acid ‘i’ depends on Moment of Inertia (I) relative to i

$$P(i) = \frac{I_{cm}}{I_i} = \frac{I_{cm}}{I_{cm} + Mh_i^2}$$

$\vec{h}_i = \vec{r}_{cm} - \vec{r}_i$





Interaction Hamiltonian

intra- and inter-chain bonds

$$H = \sum_i \left(\sum_{j>i} \left(a_{ij}^{sc} \sum_p E_{ij}^{scp} + a_{ij}^{es} \sum_e E_{ij}^{ese} + a_{ij}^{hb} E^{hb} + a_{ij}^{dip} E^{dip} + a_{ij}^{rep} E^{rep} \right) + \sum_l a_i^l E^l + \sum_m a_i^m E^m \right)$$

each chain separately (secondary structure)

i,j : amino acid residue number in the primary sequence.

a_{ij}^{sc}: are sidechains of i and j close enough to interact; yes = 1, no = 0.

E_{ij}^{scp}: sidechain-sidechain energy

a_{ij}^{es}: are sidechains of i and j close enough to have electrostatic interaction; yes = 1, no = 0.

E_{ij}^{ese}: sidechain-sidechain electrostatic interaction energy (saltbridge)

a_{ij}^{hb}: are backbones or side chain i and j close enough to interact; y=1, n=0.

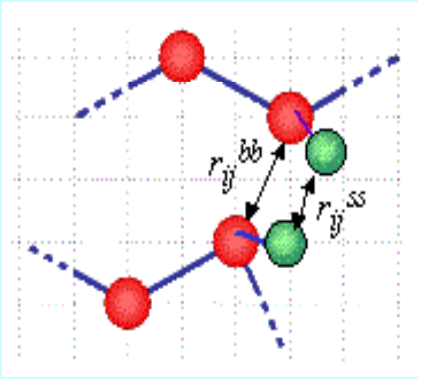
E^{hb}: Hydrogen bond interaction energy

a_{ij}^{dip}: are backbones i and j close enough to interact; y=1, n=0.

E^{dip}: dipole interaction energy

a_{ij}^{rep}: are backbones i and j close enough to have repulsion; y=1, n=0.

E^{rep}: backbones repulsion;



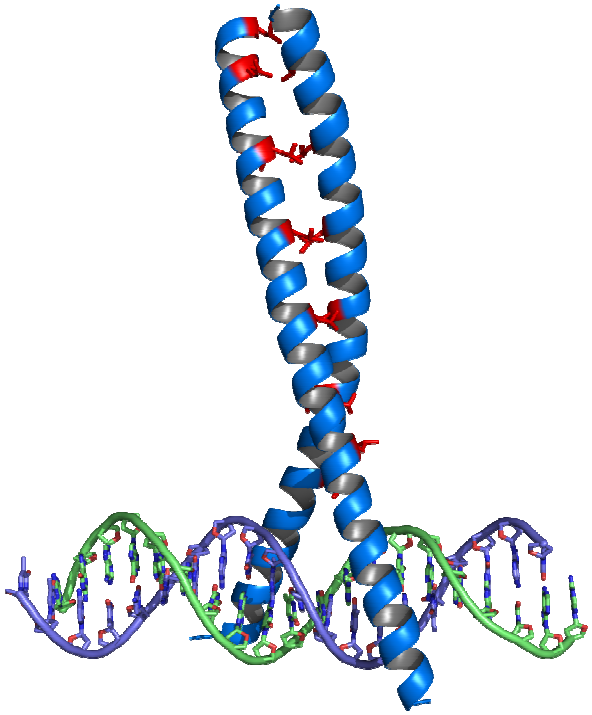
a_i^l: are residues i-1, i, i+1, arranged so that 'i' is in its preferred user-defined local configuration (i.e. α-helix, β-sheet, turn); y=1, n=0.

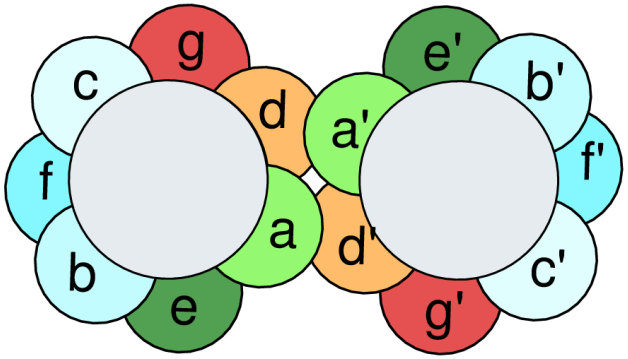
E^l: local propensity energy.

a_i^m: are residues i-1, i, i+1, i+2 arranged so that i and i+1 are both in the same preferred local configuration; y=1, n=0

E^m: medium range (cooperative) propensity energy

Computer Simulations allow us to investigate the effects of individual amino acids on protein dynamics and aggregation





Results from Simulations of GCN4-leucine zipper

Leucine zipper: a common structural motif

GCN4 is a yeast transcriptional activator protein which contains coiled coil GCN4 leucine zipper that can bind with DNA as a parallel homodimer.

Main features of GCN4 Leucine Zipper:

Leucine residues repeat in every seventh (d) position (heptad repeat)

Residues at positions a and d are hydrophobic and form a hydrophobic core

Residues at e and g are charged (may form inter-helical salt-bridge)

First Results: Importance of "Trigger Sequence" for Dimerization

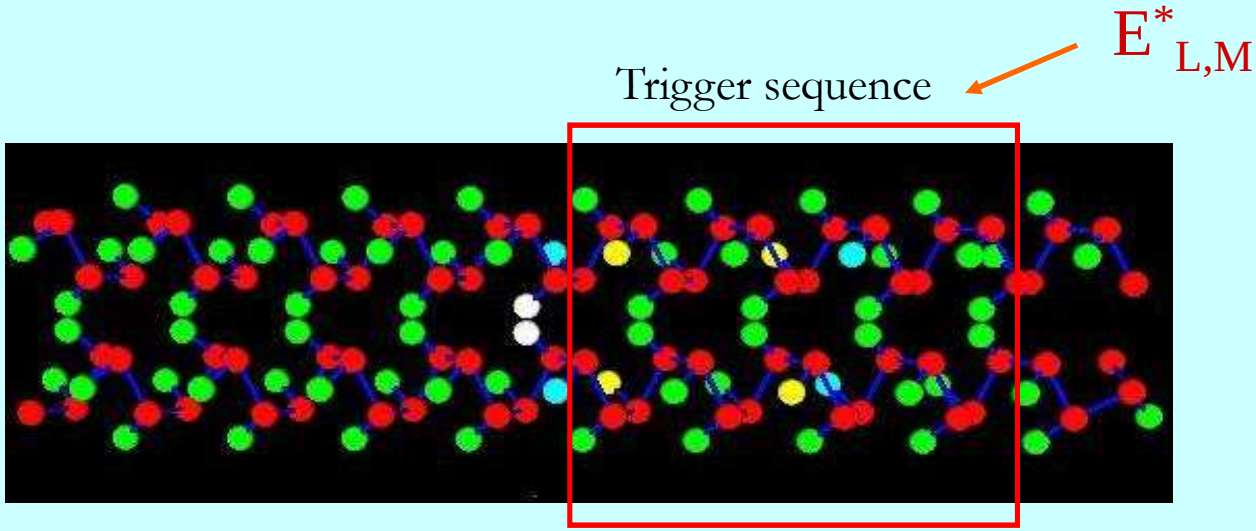
Trigger Sequence: a helical segment with higher $E^*_{L,M}$ stability that helps folding

Experiments* have conflicting results on the trigger sequence and the folding mechanisms

Questions About Dimerization:

1) Is a "trigger sequence" with enhanced α-helix $E^*_{L,M}$ necessary?

2) Is it a two-state, or multi-step, process?



*Lee, D.L etc. *J. Mol. Biol.* **306**, 539-553(2001)
*Kammerer, R.A etc. *Proc. Natl. Acad. Sci. USA*, **95**, 13419-13424(1998)

First Results: Importance of "Trigger Sequence" for Dimerization

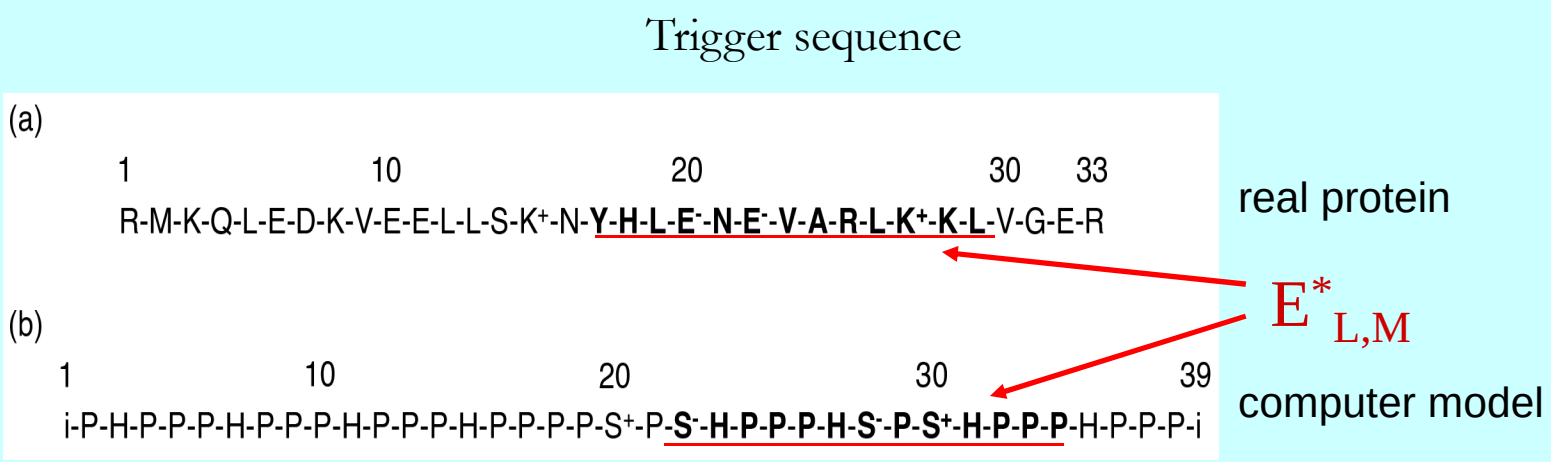
Trigger Sequence: a helical segment with higher $E^*_{L,M}$ stability that helps folding

Experiments* have conflicting results on the trigger sequence and the folding mechanisms

Questions About Dimerization:

1) Is a "trigger sequence" with enhanced α-helix $E^*_{L,M}$ necessary?

2) Is it a two-state, or multi-step, process?



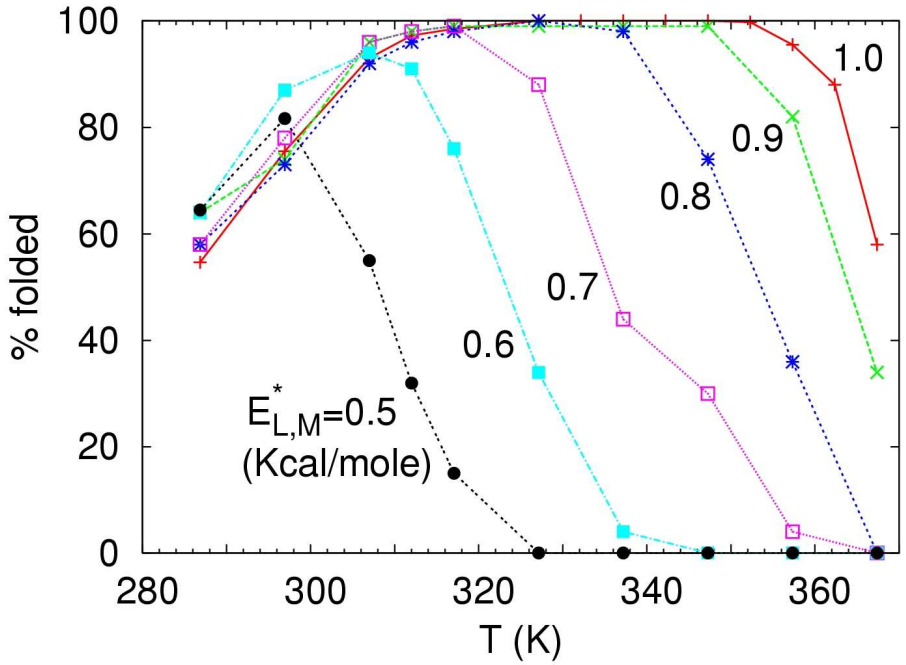
*Lee, D.L etc. *J. Mol. Biol.* **306**, 539-553(2001)

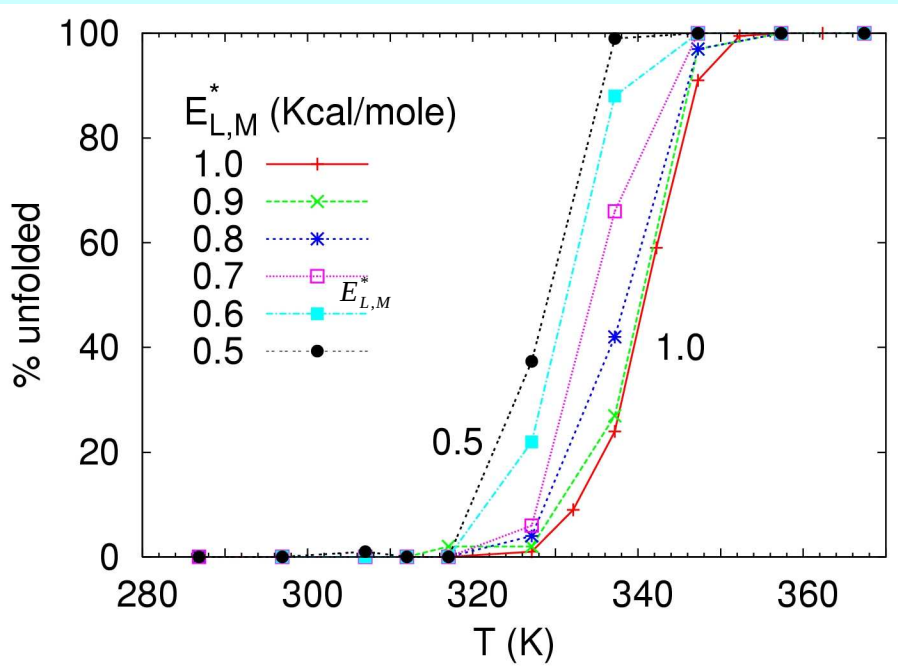
*Kammerer, R.A etc. *Proc. Natl. Acad. Sci. USA*, **95**, 13419-13424(1998)

Simulation Results: Trigger sequence is important

Dimerization Folding/Unfolding Probability depends on helical propensity of trigger sequence

Computations: Increase $E^*_{L,M}$ trigger sequence from 0.5 to 1.0 Kcal





Start with both chains unfolded. Determine probability to dimerize within 20ms.

Weak $E^*_{L,M}$ → Low probability

Strong $E^*_{L,M}$ → high probability

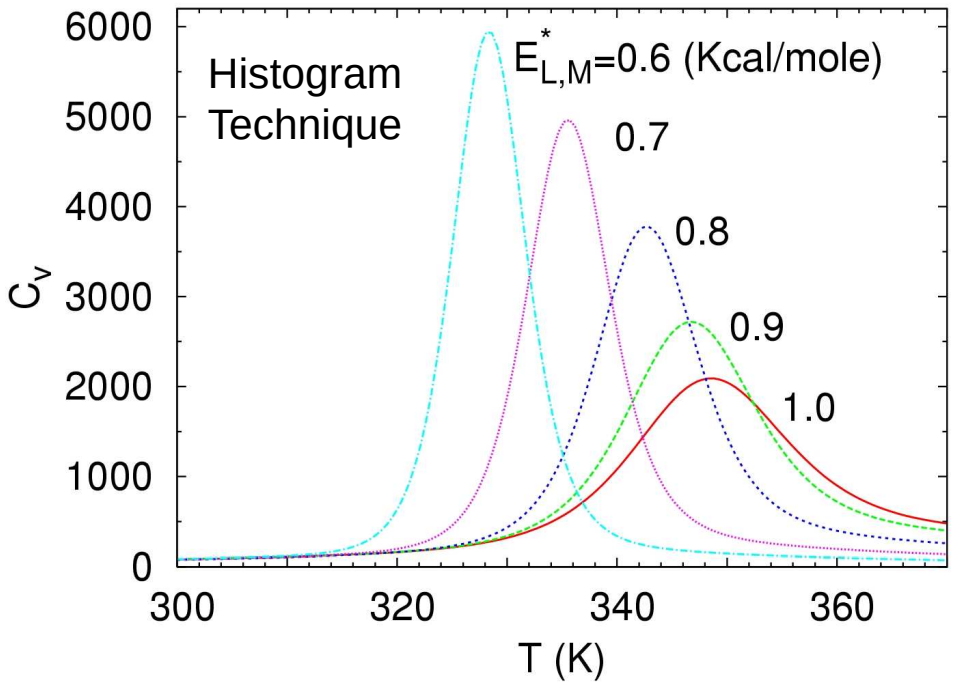
Stability: Start with dimer. Determine probability to undimerize within 20ms.

Weak $E^*_{L,M}$ → High probability (unstable)

Strong $E^*_{L,M}$ → Low probability (stable)

→Trigger sequence is important for GCN4 leucine zipper dimerization and stability!

Heat Capacity C_v → Info about transition



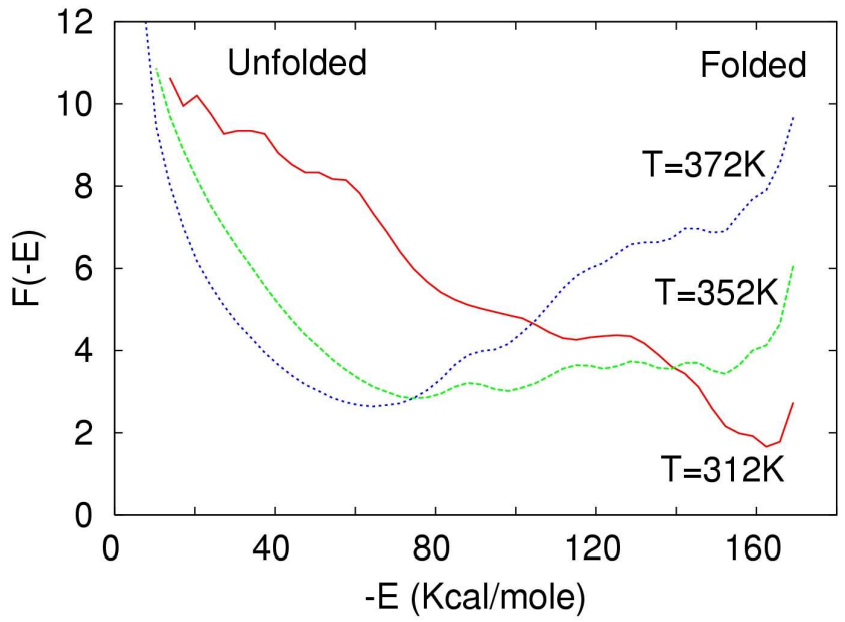
C_v curves for different $E^*_{L,M}$ allow a comparison of stability of native state as well as the nature of structural transition

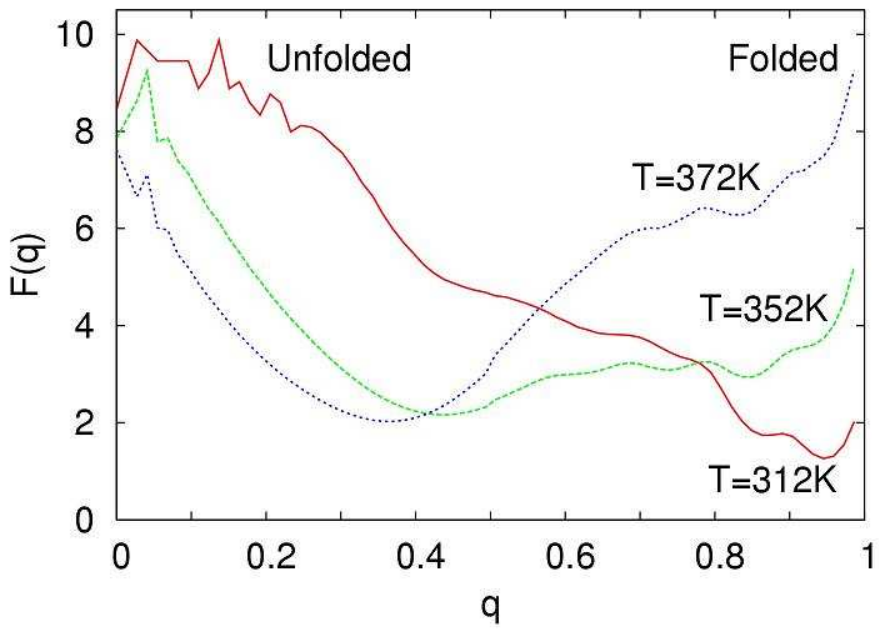
Thermodynamic results consistent with kinetics (dimer ←→ undimer times):

Stronger α–helical $E^*_{L,M}$ raises transition temperature T'
→At T=310K, dimer is more stable for higher $E^*_{L,M}$

But: for stronger $E^*_{L,M}$, dimer←→undimer transition not as sharp
→ Dimerization is less like sharp, first-order two state transition

Deeper Thermodynamics: Free Energy F=E-TS





Free Energy as a function of a reaction coordinate x. (We use x=E or q; same results)

$$P(x) \propto e^{-F(x)/kT}$$

$$\downarrow$$

$$F(x;T) = -kT \ln P(x;T)$$

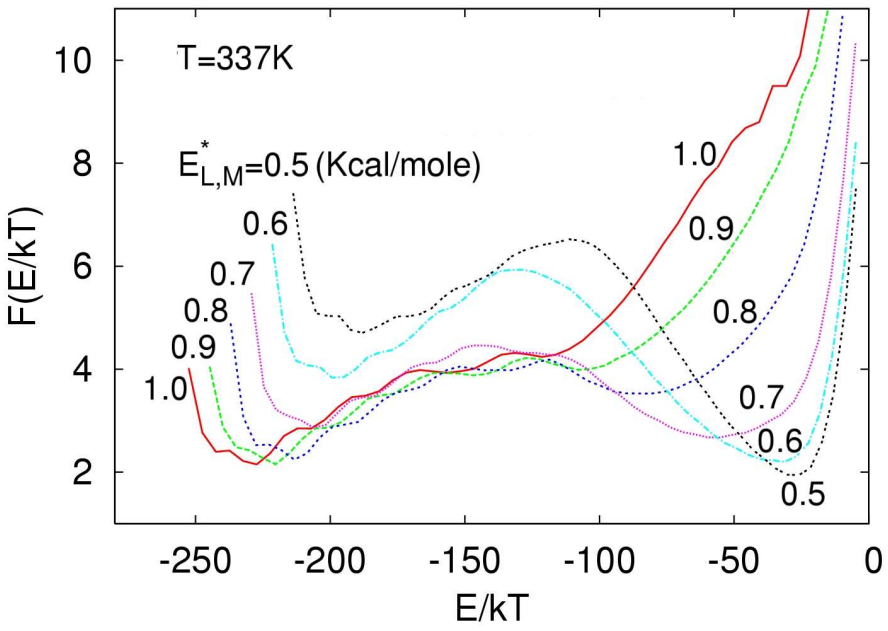
-- Low T (312 K, ~Body Temp): Folded state is more stable (free energy minimum)

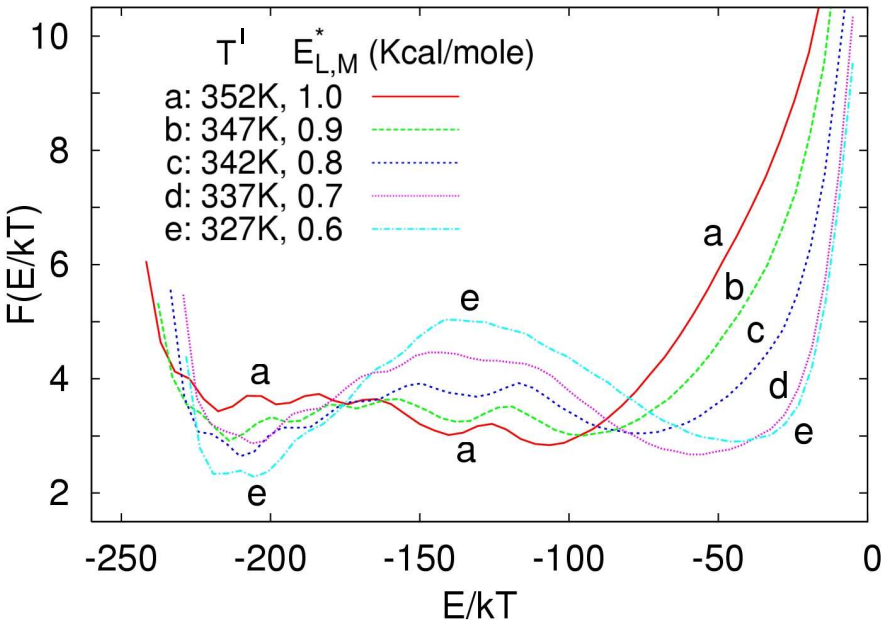
-- High T (372K): Unfolded state is more stable (free energy minimum)

Curve at Transition Temp. T' (352K): Folded and unfolded states are equally stable and several intermediate states present → multiple states during transition

(not first-order two state transition)

Summary: Effect of $E^*_{L,M}$ on Free energy landscape





Alpha helical propensity $E^*_{L,M}$ of trigger sequence can dramatically change the landscape at constant T
-- Trigger sequence is important

Reliable dimerization requires strong $E^*_{L,M}$ which also changes nature of dimerization transition

Free energy for different helical propensities $E^*_{L,M}$ at their respective transition temperatures, T'.

-- Multiple states for high $E^*_{L,M}$

-- More two-state like for low $E^*_{L,M}$

→ Alpha helical propensity of trigger sequence determines the folding mechanism.

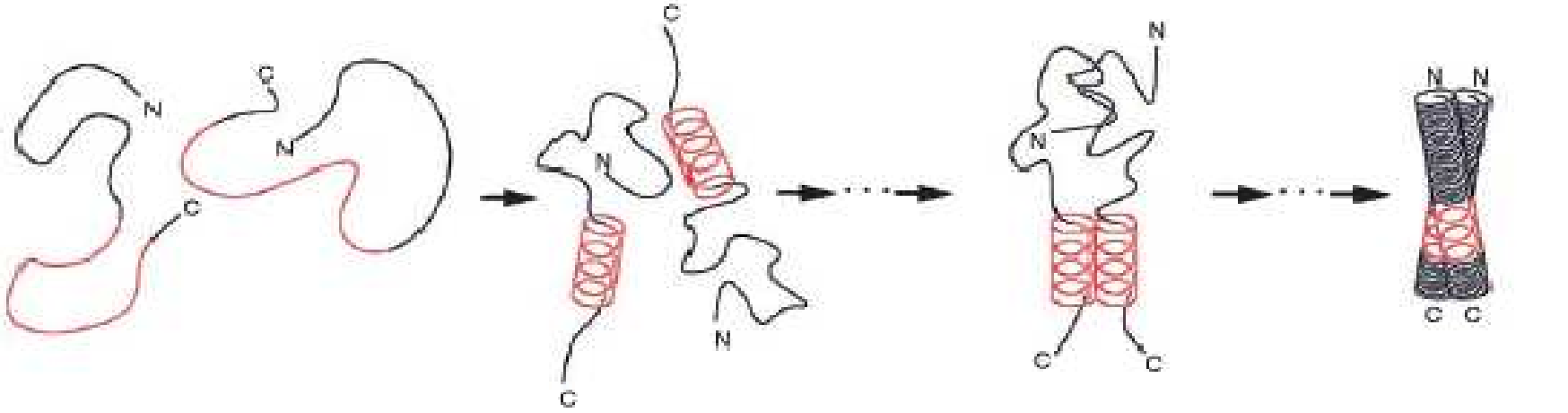
Investigation of Folding Mechanism

Is dimerization 2-state or multi-step process?

Can be determined through thermodynamics and statistical physics:

Folding and Unfolding Kinetics

Results: Folding and dimerization is a multi-step, four state process



Conclusions

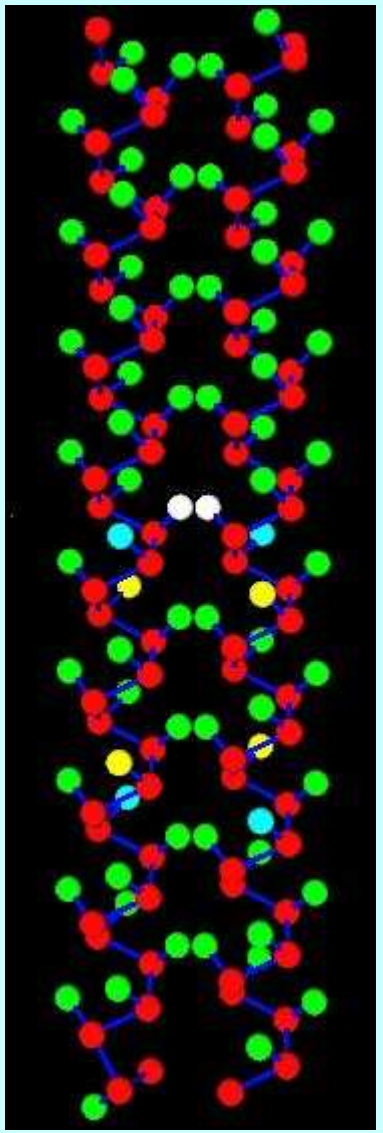
- The modified MC model allows realistic simulation of two peptide chain system (protein-protein interactions)
- Trigger sequence with stronger alpha helical propensity is important for the folding and stability of GCN4-leucine zipper
- Alpha helical propensity in the trigger sequence also determines the folding mechanism.

  → Folding of Leucine zipper with trigger sequence is not a simple two state transition. The pathways to the native state involve multiple states

Future Directions

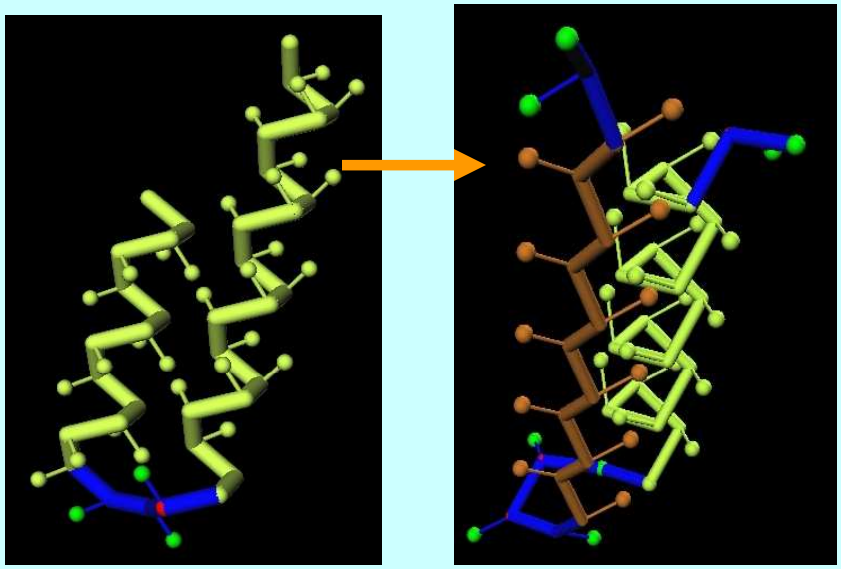
Protein Engineering:
Novel proteins that fold faster and are more stable → selective Amino Acid substitution to increase stability and foldability (speed and reliability)

- Effect of Salt bridge on dimerization
- Concentration effect: change lattice size
- Chain length effect
- Priors
- Multi Protein aggregations (e.g. Amyloid Fibrils)



Theoretical Biophysics
Florida International University

Aggregation / Amyloid formation



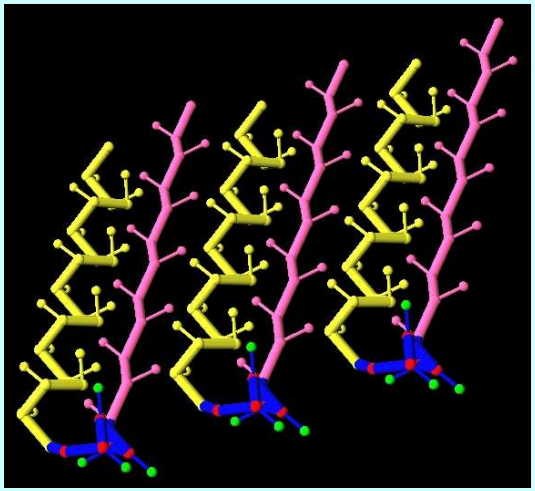
In Simulation

Part of the helical residues can be made to have preference of forming either all helical or all beta structure by changing cooperative energy

aaaaaaaaaaaa → bbbbbbbbbbbb

Q: is normal one-- a kinetic trap and bad one -- the native?

Interested in the competition between folding and protein aggregation



GCN4-p1 Leucine Zipper Folding simulation

QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

Physics will provide the molecular level understanding of protein dynamics to usher-in a brave new world of biomedicine.

Thank you !