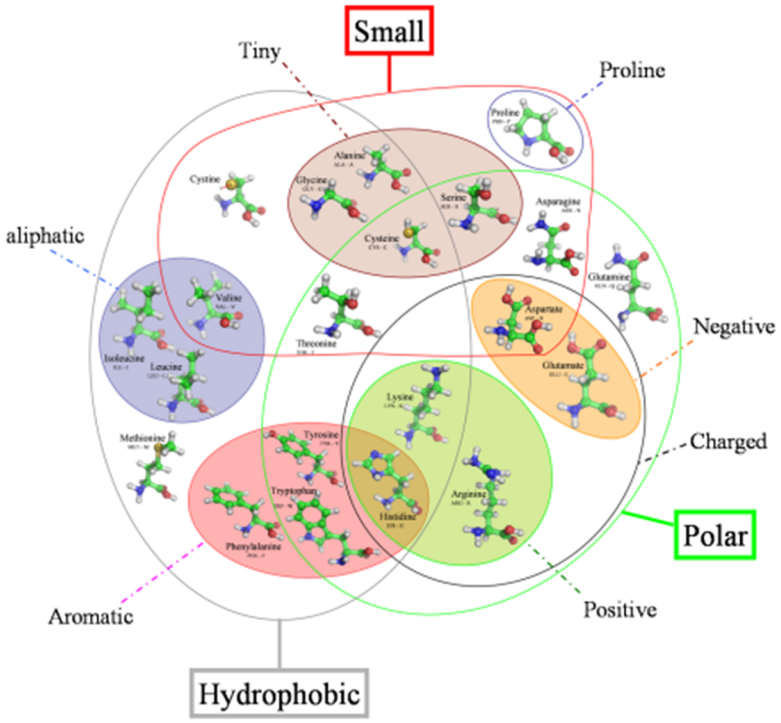
20. Protein Folding

- Composed of 50–500 amino acids linked in 1D sequence by the polypeptide backbone
- The amino acid physical and chemical properties of the 20 amino acids dictate an intricate and functional 3D structure.
- Folded structure is energetic ground state (Anfinsen)



Reprinted from
http://swift.cmbi.ru.nl/teach/Wageningen/IMAGE/aa_venn_diagram.png

Many proteins spontaneously refold into native form in vitro with high fidelity and high speed.

Different approaches to studying this phenomenon:

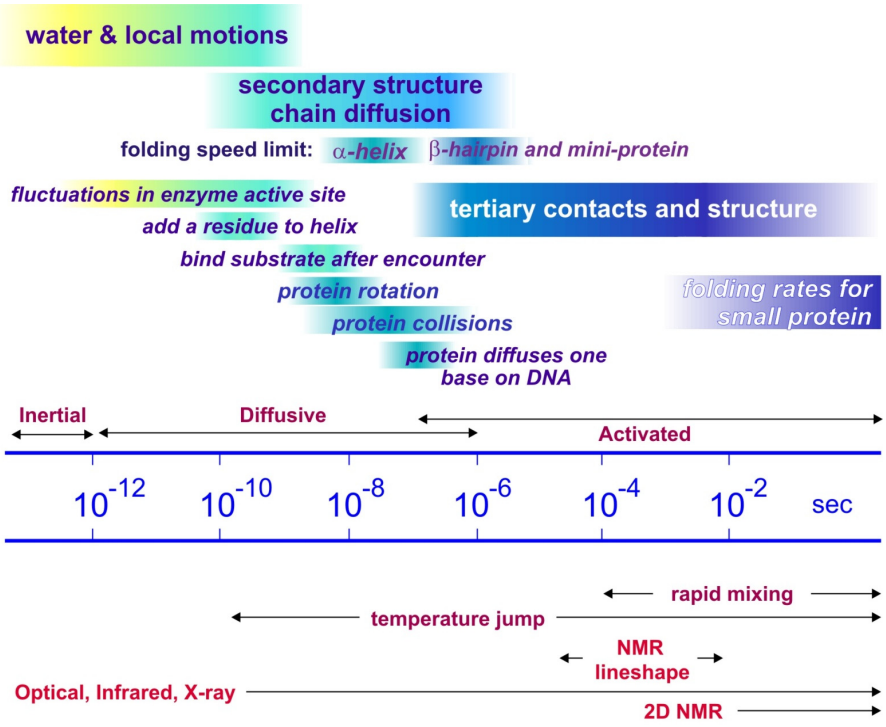
- How does the primary sequence encode the 3D structure?
- Can you predict the 3D fold from a primary sequence?
- Design a polypeptide chain that folds into a known structure.
- What is the mechanism by which a disordered chain rapidly adopts its native structure?

Our emphasis here is mechanistic. What drives this process? The physical properties of the connected pendant chains interacting cooperatively give rise to the structure.

It is said that the primary sequence dictates the three-dimensional structure, but this is not the whole story, and it emphasizes a certain perspective. Certainly we need water, and defined thermodynamic conditions in temperature, pH, and ionic strength. In a sense the protein is the framework and the solvent is the glue. Folded proteins may not be as structured from crystal structures, as one is led to believe.

Kinetics and Dynamics

Observed protein folding time scales span decades. Observations for protein folding typically measured in ms, seconds, and minutes. This is the time scale for activated folding across a free-energy barrier. The intrinsic time scale for the underlying diffusive processes that allow conformations to evolve and local contacts to be formed through free diffusion is ps to μs. The folding of small secondary structure happens on 0.1–1 μs for helices and ~1–10 μs for hairpins. The fastest folding mini-proteins (20–30 residues) is ~1 μs.



Cooperativity

What drives this? Some hints:

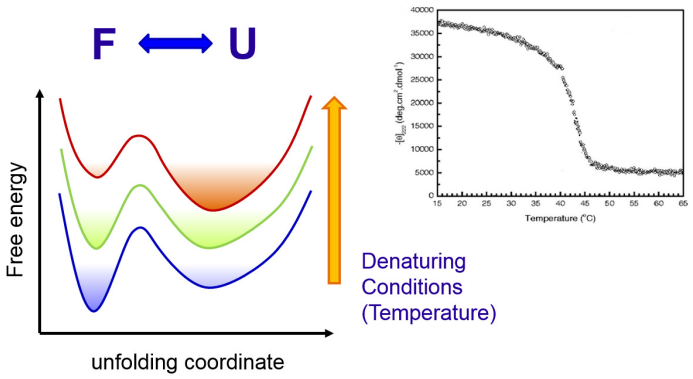
Levinthal's paradox[1]

The folded configuration cannot be found through a purely random search process.

- Assume:
  - 3 states/amino acid linkage
  - 100 linkages
- $3^{100} = 5 \times 10^{47}$ states
  - Sample 10^{-13} sec/state
- 10^{27} years to sample

Two-state thermodynamics

To all appearances, the system (often) behaves as if there are only two thermodynamic states.



Entropy/Enthalpy

ΔG is a delicate balance of two large opposing energy contributions ΔH and $T\Delta S$.

1. C. Levinthal, Are there pathways for protein folding?, J. Chim. Phys. Phys.-Chim. Biol. **65**, 44-45 (1968).

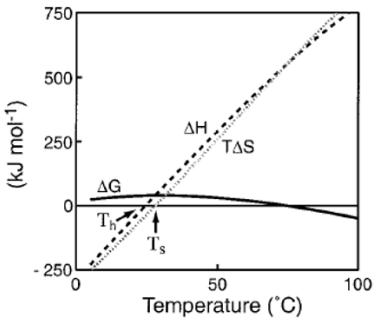
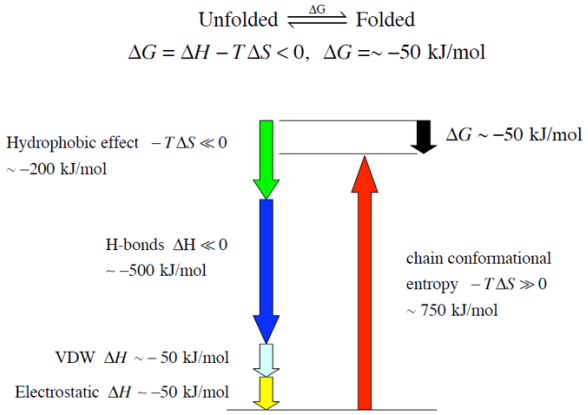


Figure 3. Protein unfolding free energy, $\Delta G = G_u - G_f$, entropy, ΔS, and enthalpy, ΔH, versus temperature. For proteins, $T_s \approx T_h$. Data on myoglobin from Makhatadze, G. I. and Privalov, P. L. *Biophys. Chem.* **1994**, *51*, 291.

Reprinted with permission from N. T. Southall, K. A. Dill and A. D. J. Haymet, J. Phys. Chem. B **106**, 521-533 (2002). Copyright 2002 American Chemical Society.

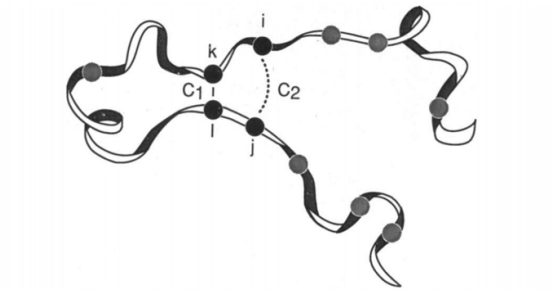


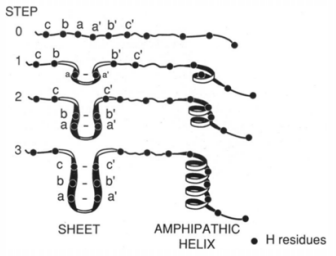
Reprinted from James Chou (2008). http://cmcd.hms.harvard.edu/activities/_media/bcmp201/lecture7.pdf.

Cooperativity underlies these observations

Probability of forming one contact is higher if another contact is formed.

- Zipping
- Hydrophobic collapse





Reprinted from K. A. Dill, K. M. Fiebig and H. S. Chan, Proc. Natl. Acad. Sci. U. S. A. **90,** 1942-1946 (1993). Copyright 1993 PNAS.

Protein Folding Conceptual Pictures

Traditional pictures rooted in classical thermodynamics and reaction kinetics.

- Postulate particular sequence of events.
- Focus on importance of a certain physical effect.
    1) **Framework or kinetic zipper**
    2) **Hydrophobic collapse**
    3) **Nucleation–condensation**

Framework/Kinetic Zipper Model

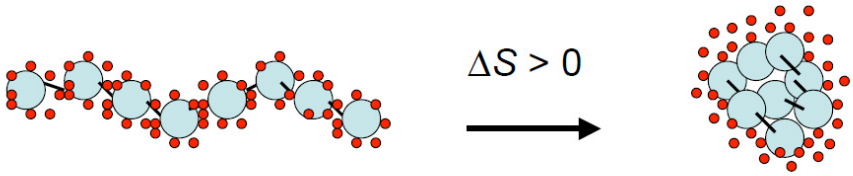
- Observation from peptides: secondary structures fold rapidly following nucleation.

- Secondary structure formation precedes tertiary organization.
- Emphasis:
    - Hierarchy and pathway
    - Focus on backbone, secondary structure

Hydrophobic Collapse

- Observation: protein structure has hydrophobic residues buried in center and hydrophilic groups near surface.
- An extended chain rapidly collapses to bury hydrophobic groups and thereby speeds search for native contacts.
- Collapsed state: molten globule
- Secondary and tertiary structure form together following collapse.

U ⟷ Molten Globule ⟷ F



Unfolded — Folded

More Hydrocarbon-Water Interfacial Area, More Water Ordered

Less Hydrocarbon-Water Interfacial Area, Less Water Ordered

Nucleation–Condensation

Nucleation of tertiary native contacts is important first step, and structure condenses around that.

Some observations so far:

- *Importance of collective coordinates*
- *Big challenge: We don't know much about the unfolded state.*

Models for Simulating Folding

Our study of folding mechanism and the statistical mechanical relationship between structure and stability have been guided by models. Of these, simple reductionist models guided the conceptual development from the statistical mechanics side, since full atom simulations were initially intractable. We will focus on the simple models.

- Reductionist Models
  - Lattice Models
  - Gō Models
  - Coarse Grained
- Atomistic
  - Force fields

Increasing level of molecular detail

HP Model[2]

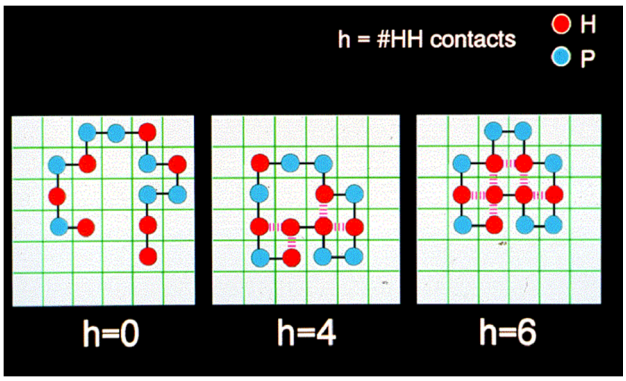
- Chain of beads. Self-avoiding walk on square lattice.
- 2 types of beads: Hydrophobic (H) and polar (P).
- H-H contacts are energetically favorable to H-P contacts.

more H → collapse to compact state, but many collapsed structures

more P → well-solvated, doesn't fold

~1:1 H:P optimal

Can be used for folding mechanism using Monte Carlo.



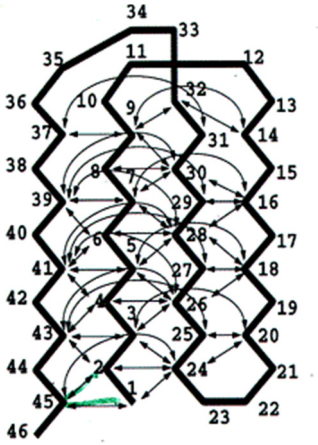
Coarse-Grained Models[3]

Hierarchy of various models that reduce protein structure to a set of interacting beads.

Gō Models[4]

2. K. F. Lau and K. A. Dill, A lattice statistical mechanics model of the conformational and sequence spaces of proteins, Macromolecules **22**, 3986-3997 (1989).
3. V. Tozzini, Coarse-grained models for proteins, Curr. Opin. Struct. Biol. **15**, 144-150 (2005).
4. Y. Ueda, H. Taketomi and N. Gō, Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme, Biopolymers **17**, 1531-1548 (1978).

Gō models and Gō-like models refer to a class of coarse-grained models in which formation of structure is driven by a minimalist interaction potential that drives the system to its native structure. The folded state must be known.



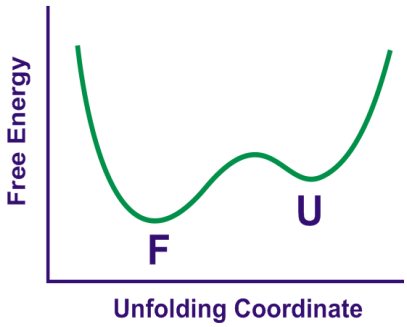
- Coarse grained
    - Original: one bead per AA
    - “Off-lattice model”
- Native-state biasing potential
    - Multiple forces in single interaction potential
    - Need to know folded structure
    - Increased simulation speed
    - Doesn’t do well metastable intermediates or non-native contacts

Perspectives on Protein Folding Dynamics

These models have helped drive theoretical developments that provide alternate perspectives on how proteins fold:

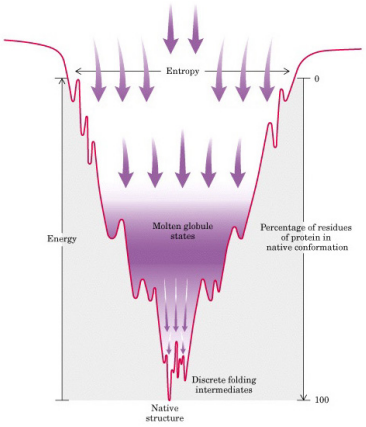
State Perspective

- Interchange between states with defined configurations
- What are the states, barriers and reaction coordinates?



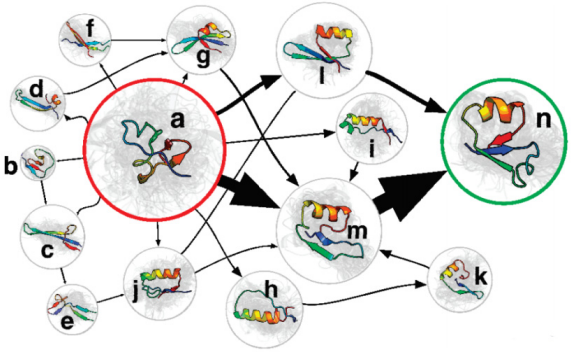
Statistical Perspective

- Change in global variables
- Configurational entropy

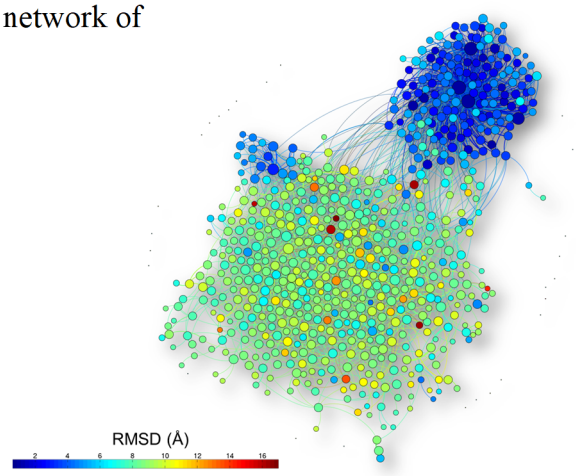


Networks

- Characterize conformational variation and network of connectivity between them.



Reprinted with permission from V. A. Voelz, G. R. Bowman, K. Beauchamp and V. S. Pande, J. Am. Chem. Soc. **132**, 1526-1528 (2010). Copyright 2010 American Chemical Society.

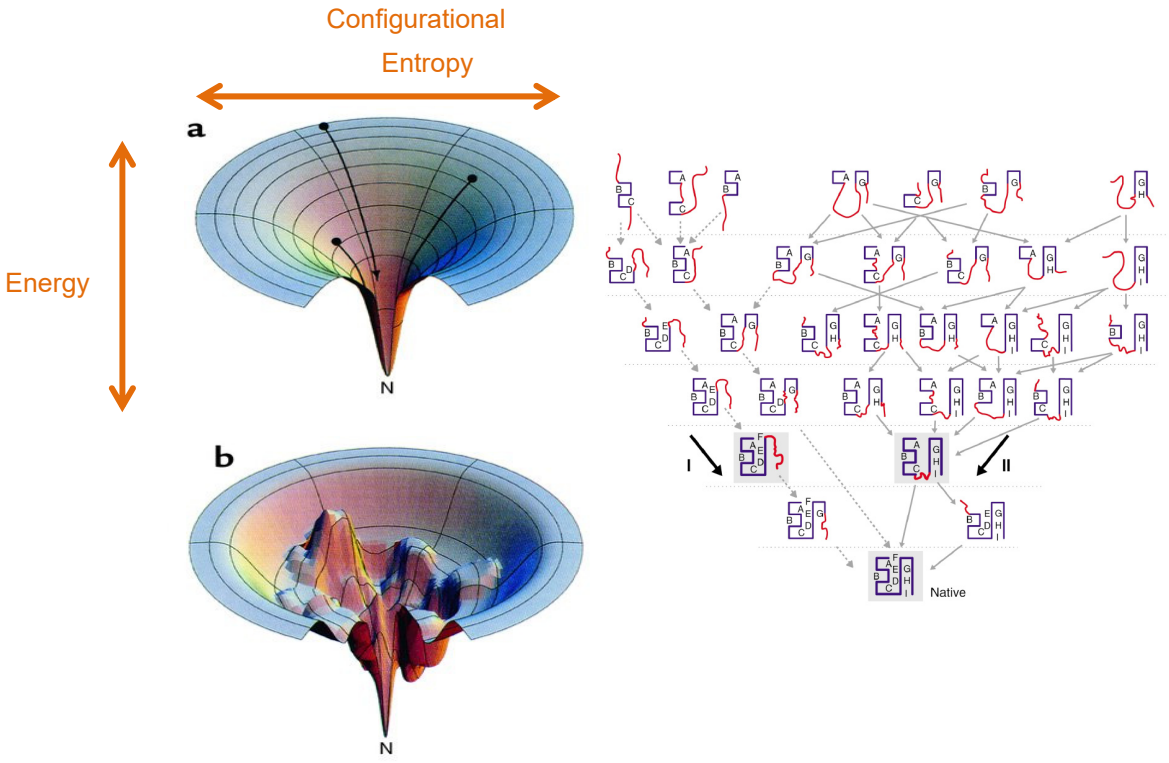


Reprinted with permission from C. R. Baiz, Y.-S. Lin, C. S. Peng, K. A. Beauchamp, V. A. Voelz, V. S. Pande and A. Tokmakoff, Biophys. J. **106**, 1359-1370 (2014). Copyright Elsevier 2014.

The statistical perspective is important. The standard ways of talking about folding is in terms of activated processes, in which we describe states that have defined structures, and which exchange across barriers along a reaction coordinate. And the emphasis is on molecularly interpreting these states. There is nothing formally wrong with that except that it is an unsatisfying way of treating problems where one has entropic barriers.

Folding Funnels and Configurational Entropy

Helps with entropic barriers[5]



5. K. A. Dill, Polymer principles and protein folding, Protein Sci. **8**, 1166-1180 (1999).

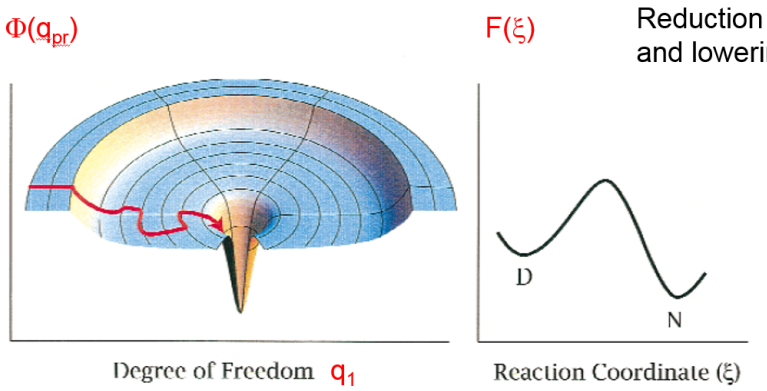


Fig. 5. (A) Energy landscape vs. (B) reaction diagram. A landscape is a free energy F_{micro} of each individual chain conformation vs. the many microscopic degrees of freedom. A reaction diagram is a free energy F_{macro} of an ensemble of molecules, and includes the chain conformational entropy. Here F_{macro} is a function of a single variable, ξ, such as a reaction coordinate. The reaction coordinate is usually not known for protein folding. The red arrow on the landscape indicates a possible micropath, an individual folding trajectory. In this case, the micropath never involves an uphill step, and yet the reaction diagram has a free energy barrier. The barrier is due to the slow entropic search of many different chains seeking the entry to the central steep funnel.

Reprinted with permission from K. A. Dill, Protein Sci. **8**, 1166-1180 (1999). John Wiley and Sons 1999.

Transition State vs Ensemble Kinetics

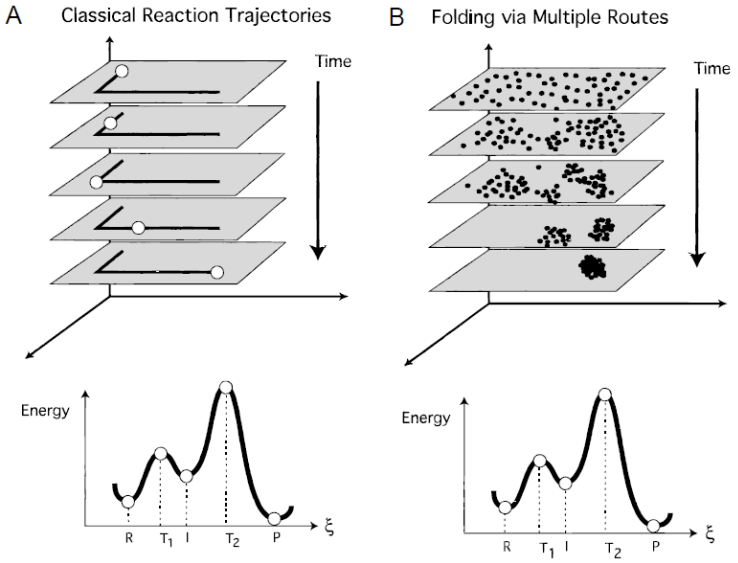


Fig. 8. **A:** For chemical reactions (energies $\gg kT$), the macrostates on reaction coordinate diagrams correspond to the time series of microstates on the energy landscape. **B:** For folding processes (energies per interaction $\approx kT$), the observed macrostates may not uniquely specify the time series of microstates on the energy landscape.

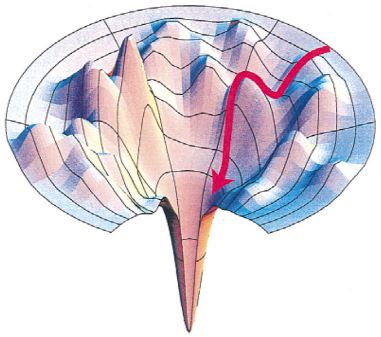


Fig. 9. An uphill micropath (red line) is surrounded by more favorable routes that do not involve uphill steps to reach the native state.

Reprinted with permission from K. A. Dill, Protein Sci. **8**, 1166-1180 (1999). John Wiley and Sons 1999.