Lectures 3, 4, 5

- Roderick Clayton: Purification of photosynthetic reaction center (1968)
- George Feher: Reaction center structure (1974)

"*For his many contributions to the understanding of the physics of photosynthesis; specifically, for his role in the pioneering of the concept of reaction enters in photosynthetic bacteria, their isolation, their spectroscopy and their structural characterization.*" (1982)

- Peter Wolynes: Protein folding, energy landscape (1987)

"*For his conceptual breakthroughs in protein dynamics and protein folding, and his critical insights toward the understanding of how proteins work at the most fundamental level.*" (2004)

- Jose Onuchic: Protein folding routes (1995)
- Ken Dill:

"*For independent contributions to a new view of protein folding, from the introduction and exploration of simple models, to detailed confrontations between theory and experiment.*"

Lecture 4: Protein folding

Peter Wolynes

• **energy landscape theory** of protein folding which brought the perspective of modern statistical mechanics to this central problem of molecular biology leading to new approaches for predicting protein structure from sequence.

• to fold into that special native state, a protein molecule must be able to bypass a continuum of glassy states. (1974)

"*For his conceptual breakthroughs in protein dynamics and protein folding, and his critical insights toward the understanding of how proteins work at the most fundamental level.*" (2004 Delbruck Prize)

Lecture 4: Protein folding

Jose Nelson Onuchic and Ken Dill

• Jose Onuchic: His research considers theoretical methods for molecular biophysics and gene networks, and he has introduced the concept of protein folding funnels. **Energy landscape theory and the funnel concept** provide the framework needed to pose and to address the questions of protein folding and function mechanisms. He developed the tunneling pathways concept for electron transfer in proteins, and is also interested in stochastic effects in genetic networks with applications to bacteria decision-making and cancer. Currently he is modeling the 3D-structure of the genome.

• Ken Dill: His research is at the intersection of statistical physics with structural and cell biophysics. He has worked on the **physics of protein folding, including the role of the hydrophobic effect in the folding code, and the understanding of their funnel-shaped energy landscapes**. Dill has also elucidated the structures and properties of water, and has contributed to the foundations of nonequilibrium statistical physics, through an entropy-like principle called Maximum Caliber.

" *For independent contributions to a new view of protein folding, from the introduction and exploration of simple models, to detailed confrontations between theory and experiment.*" (2019 Delbruck Prize)

Lecture 1-2

What information do we need to decide if something is a breakthrough?

- context
- details about the work (facts)
- impact
  - citations on a topic before and after
  - scientific landscape before and after

What makes a breakthrough?

- nonlinear change in understanding / paradigm shift
- fills a knowledge gap
  - which was unfilled for a long time
  - better than other attempts by some-fold
- opens a new direction or create a new field. creative.

Protein folding problem

(i) What is the physical code by which an amino acid sequence dictates a protein's native structure?

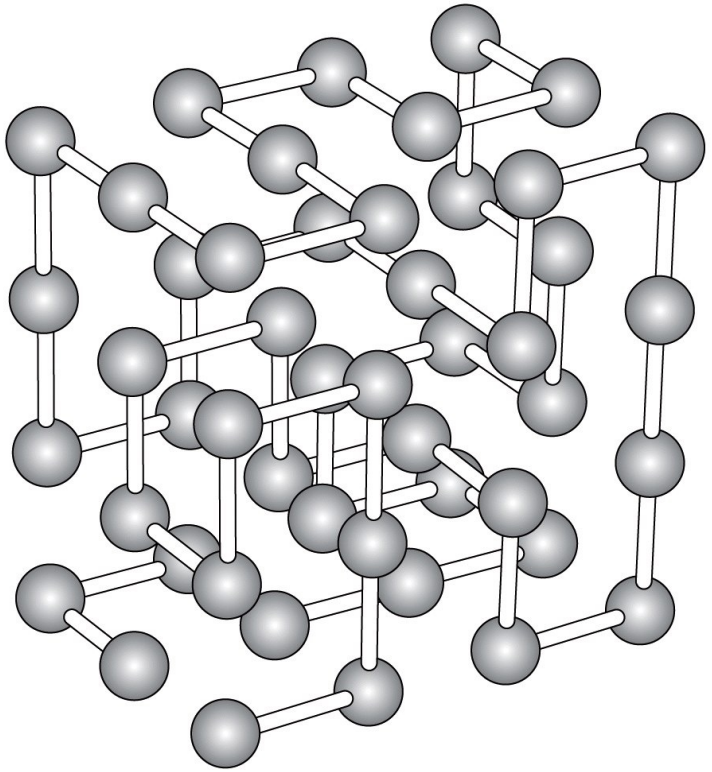
(ii) How can proteins fold so fast?

(iii) Can we devise a computer algorithm to predict protein structures from their sequences?

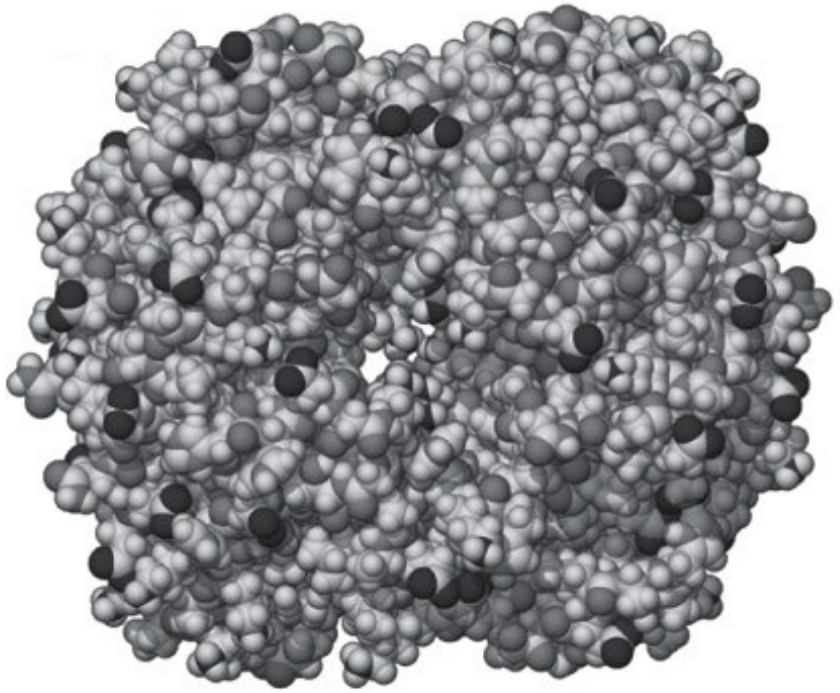
(iv) others?

Theories of protein folding

polymers vs proteins



random walks, disorder



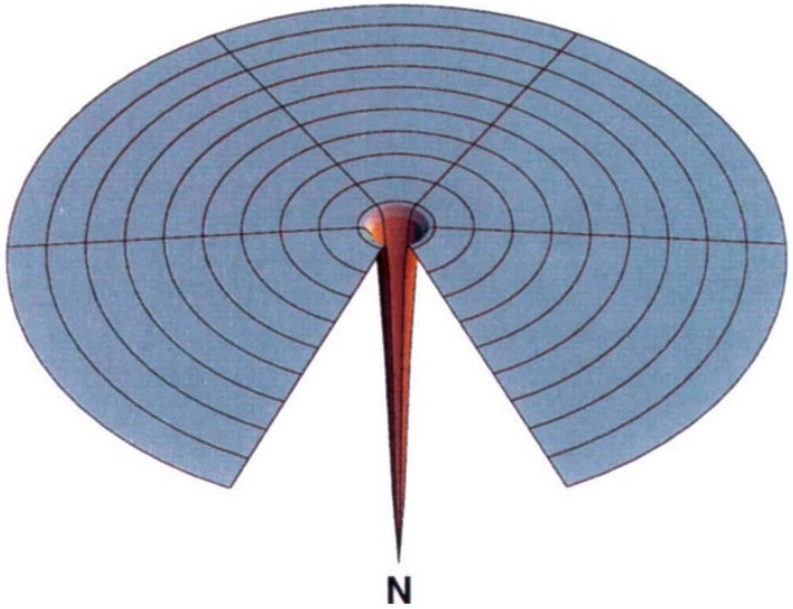
structural specifics, atomistic details

Theories of protein folding

How can proteins fold so fast?

Afinsen and the thermodynamic hypothesis (early 1960s):

- Proteins can fold reversibly: amino acid code contains all information needed
- Native structures are thermodynamically stable states
- Conformations at the global minima of their accessible free energies in the cell
- 1972 Nobel Prize "for his work on ribonuclease, especially concerning the connection between the amino acid sequence and the biologically active conformation"
- proteins are two-state structures and can only occupy, at equilibrium, either their fully native or fully unfolded condition



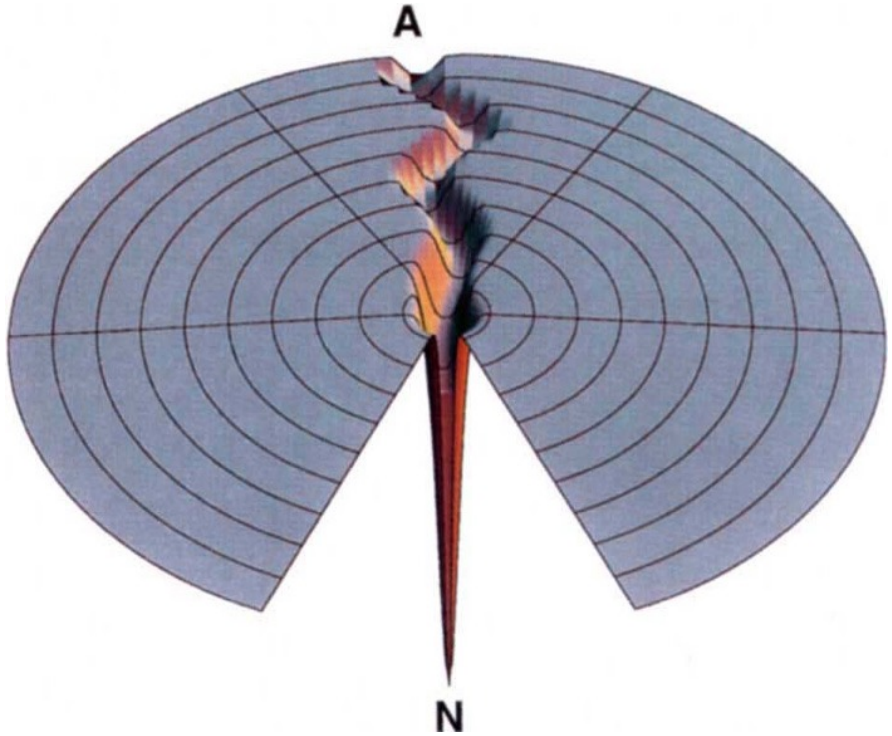
energy landscape

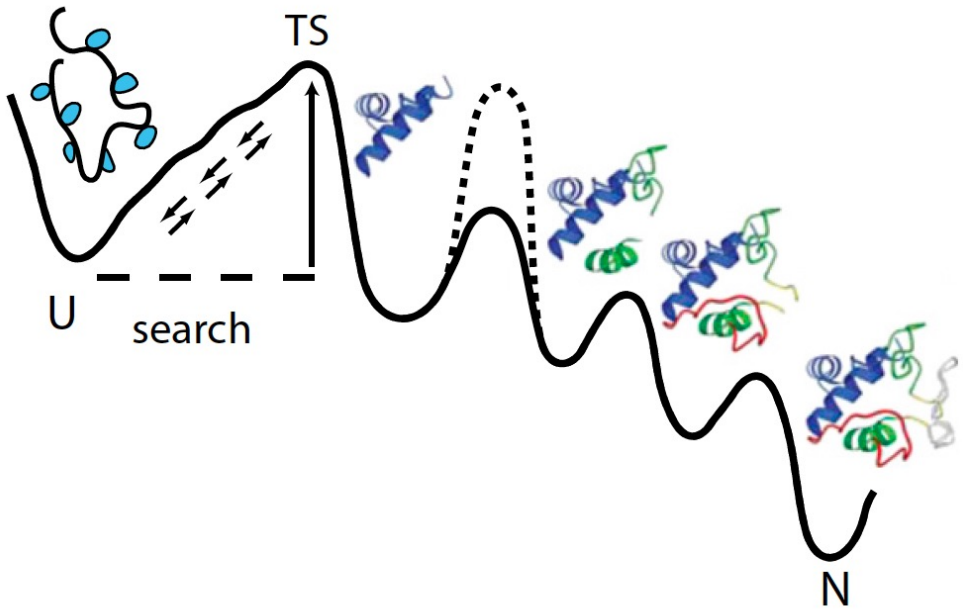
Theories of protein folding

What is the physical code by which an amino acid sequence dictates a protein's native structure?
How can proteins fold so fast?

Levinthal's paradox (late 1960s)

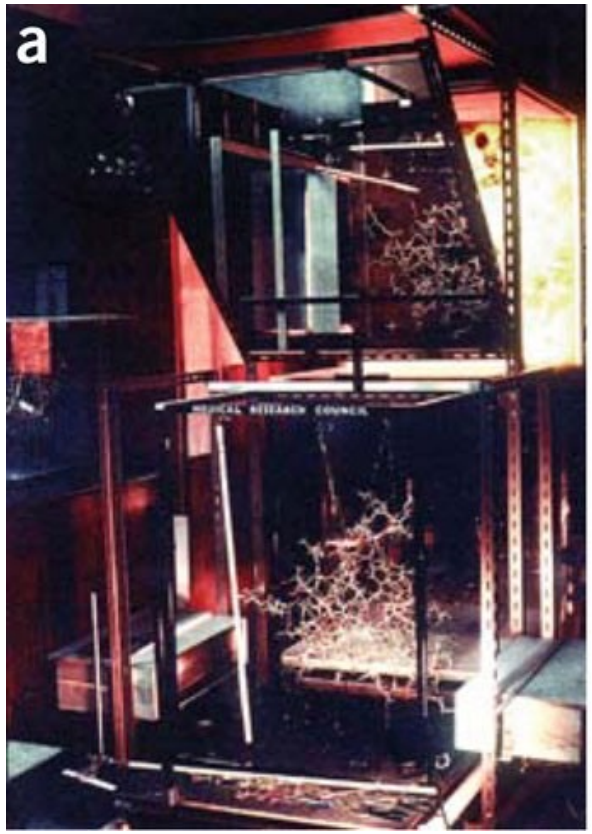
- Too many possible configurations to fold through random search
- Native structure in the conformational space is like the needle in a haystack
- Proposed proteins fold by a unique **folding pathway** with defined and sequential intermediates

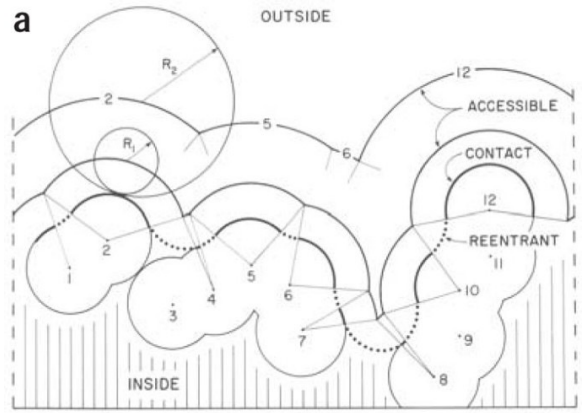


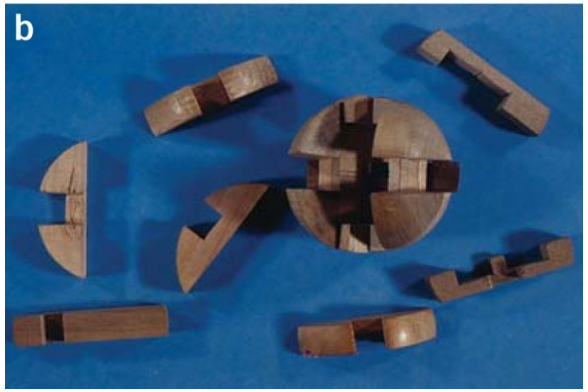


Theories of protein folding

"protein folders can be divided into **'minimizers'** and **'packers'**. The former seek to minimize the interaction energy among atoms or groups of atoms, whereas the latter concentrate on probable geometry, guided by both excluded volume limitations and structural motifs seen in proteins of known structure."





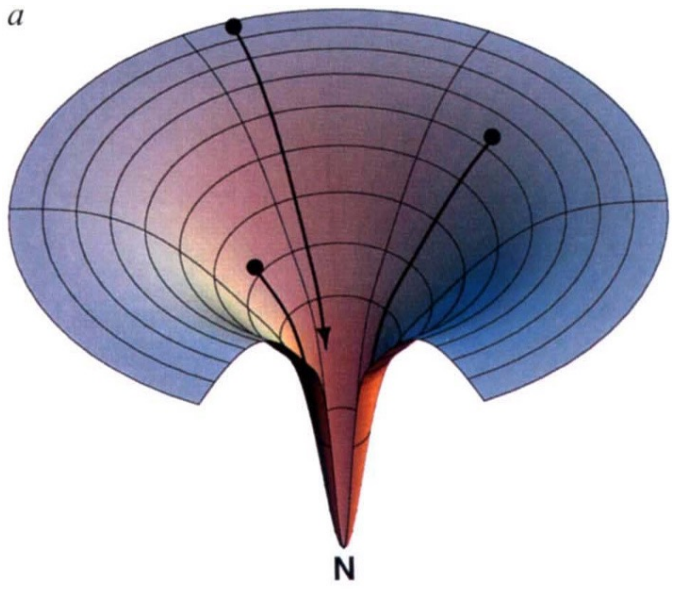


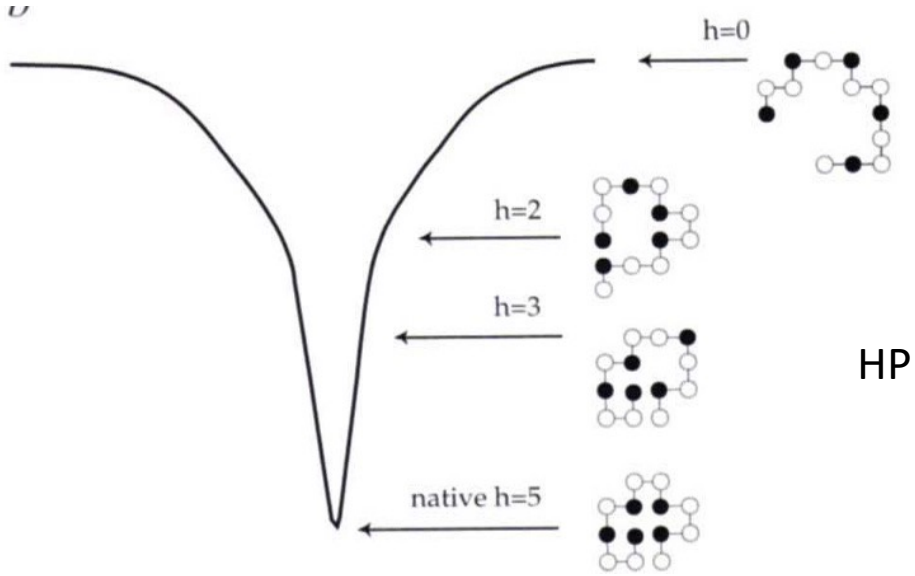
Theories of protein folding

What is the physical code by which an amino acid sequence dictates a protein's native structure?
How can proteins fold so fast?

Energy landscape, folding funnel (mid 1990s)

- Neither the folding pathway nor the set of folding intermediates is unique (**many pathways, disordered**)
- Reduction in number of accessible states as proceed down the funnel





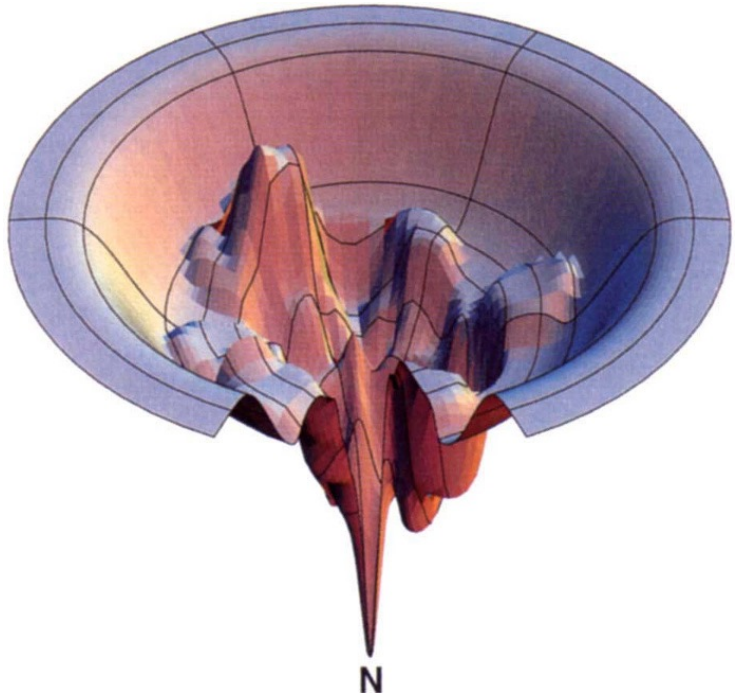
HP model

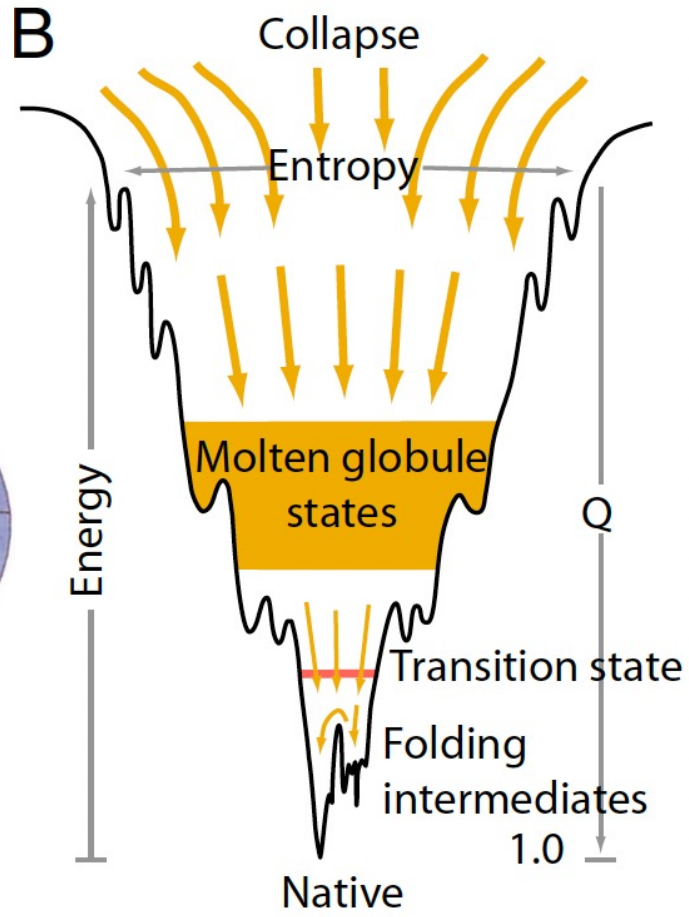
Theories of protein folding

What is the physical code by which an amino acid sequence dictates a protein's native structure?
How can proteins fold so fast?

Energy landscape, folding funnel (mid 1990s)

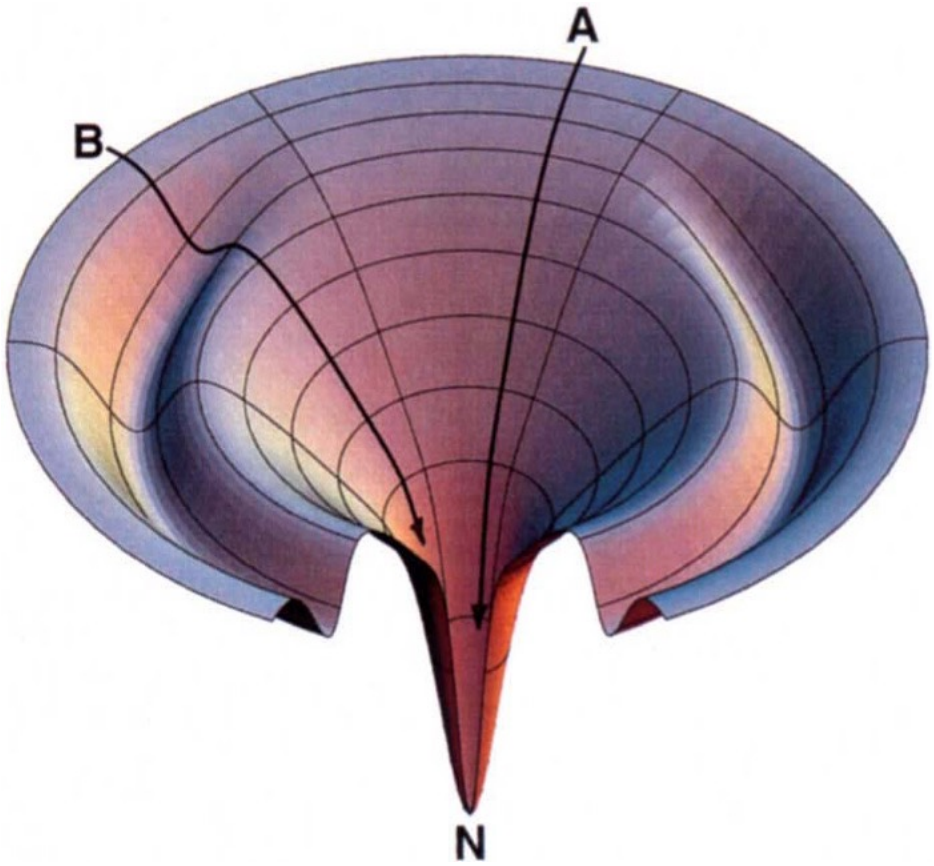
- A rugged energy landscape with kinetic traps, energy barriers, and some narrow throughway paths to native.
- Folding can be multi-state.
- Folding intermediates accumulate because of kinetic traps caused by partial misfolding



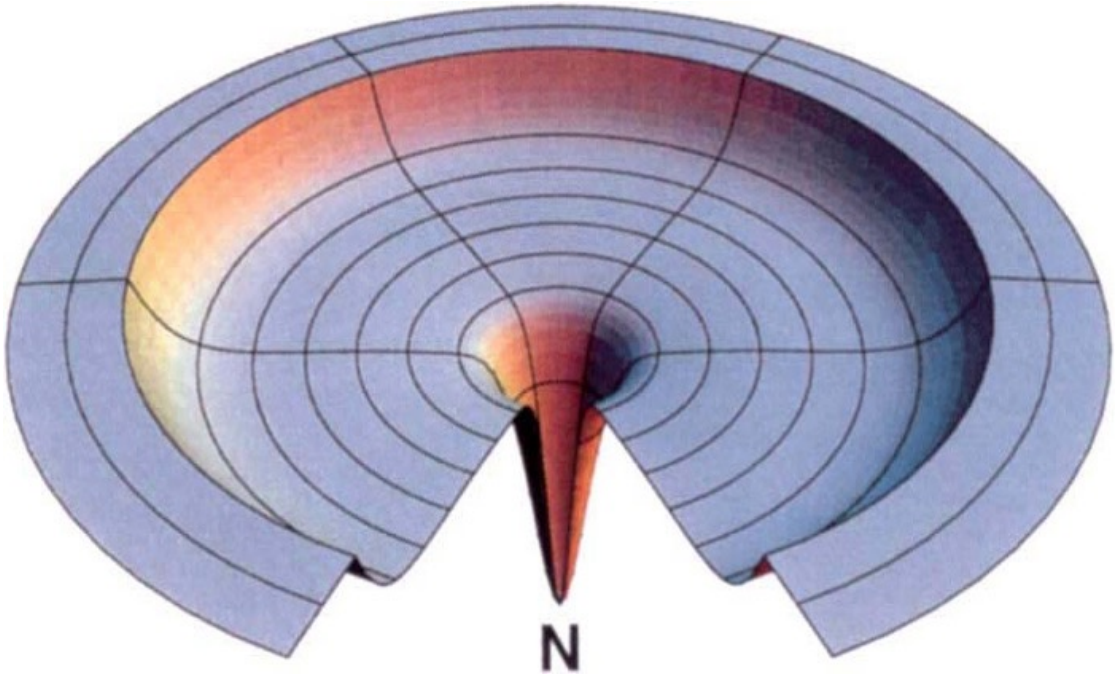


Theories of protein folding

Barriers to folding



energy barrier (kinetic trap)



entropy barrier (flat landscape)

Protein folding problem

(i) What is the physical code by which an amino acid sequence dictates a protein's native structure?

(ii) How can proteins fold so fast?

(iii) Can we devise a computer algorithm to predict protein structures from their sequences?

(iv) others?

Protein folding problem

(i) What is the physical code by which an amino acid sequence dictates a protein's native structure?
(ii) How can proteins fold so fast?
(iii) Can we devise a computer algorithm to predict protein structures from their sequences?

How could theory and computation address these issues? In the early days, modeling aimed to test whether the inter-atomic potentials of the day could explain native structures. Over time, three roles emerged:
(1) Database models. Using the growing database of protein structures (Protein DataBank, PDB), could we determine unknown structures using sequence homologies to known structures?
(2) Computational molecular physics. As Molecular Dynamics (MD) with semiempirical force fields improved, they joined with experiments to tell protein microstories, i.e. the details of motions, binding and other actions, roughly one protein at a time, one condition at a time.
(3) Theory & coarse-grained simulations. Statistical mechanics and polymer theories were developed to tell macrostories, i.e. to address more global questions of principle, of commonalities and differences among proteins.

Theories of protein folding

Challenges:

The question could be resolved by determining experimentally the structure of the intermediate forms that bridge between unfolded and native states in real proteins, but this effort has turned out to be exceptionally difficult. The usual methods, crystallography and NMR, cannot define partial structures that form and decay in less than 1 s. Experimentalists have been forced to depend on spectroscopic methods (fluorescence, CD, IR) that can follow kinetic folding in real time but are blind to the specifics of structure and so allow the possibility of alternative folding mechanisms.

Theorists have attempted to avoid these difficulties by simulating the folding process in computers. Theory-based computer simulations can be remarkably powerful. For example, one can compute the path of a multiton rocket through 150 million miles of free space to a pinpoint landing on Mars. The equations that govern space flight are known precisely (22), computer power is ample, and the track to be controlled is clear. Computing the structural journey of minuscule protein molecules through submicrons of space has proved to be more difficult. The computer power required to track the folding process at the level of thermally driven residue-level dynamics is immense. The forces that direct protein folding are delicately balanced, interlocking, and not describable in exact terms. The reaction path(s) to be mined from the mass of computer data are unknown.