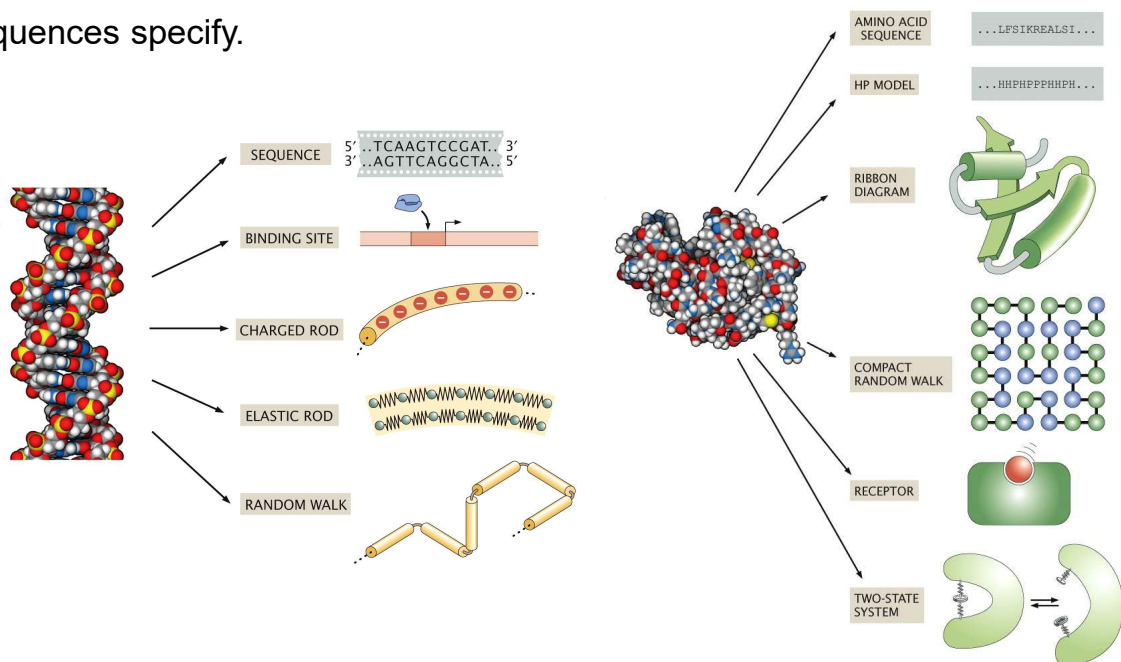


Quantitative principles in biological systems

8. Sequences and spin glass models

Spring 2025

Sequences specify.



Phillips et al

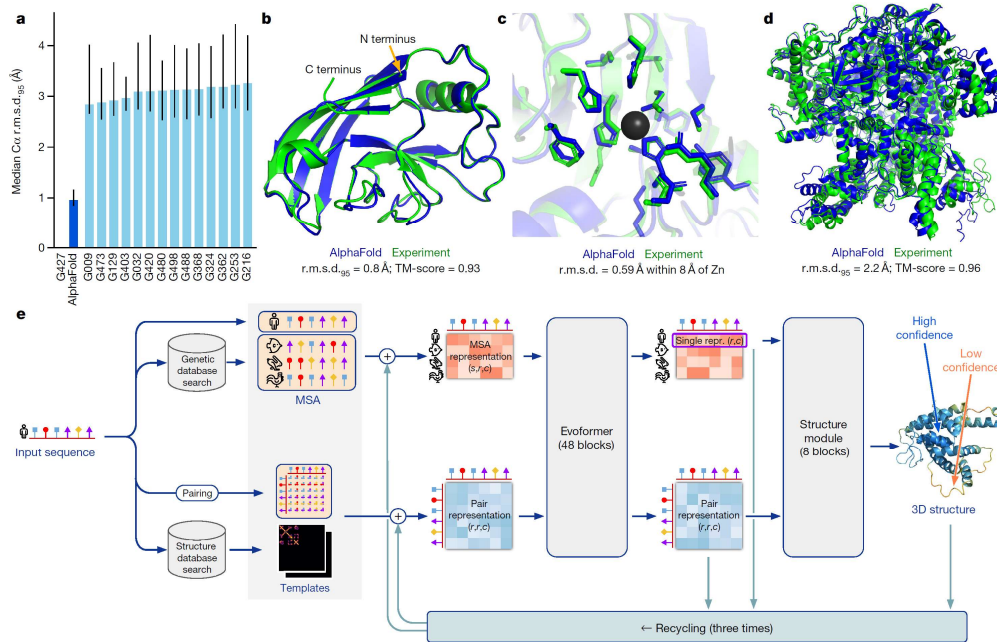
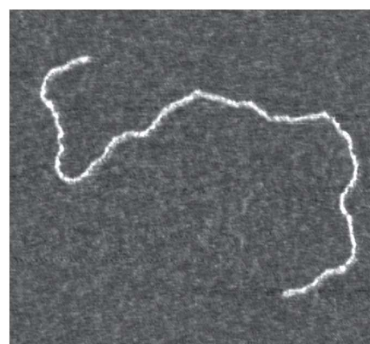
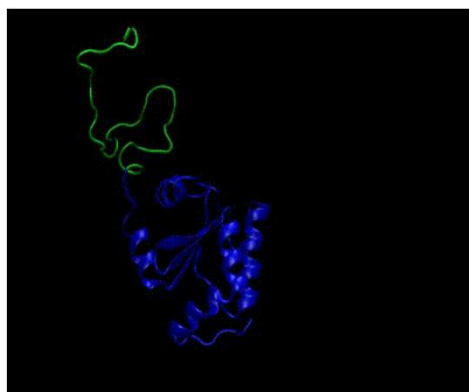


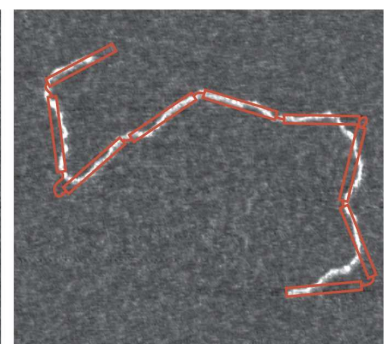
Fig. 1 | AlphaFold produces highly accurate structures. **a**, The performance of AlphaFold on a set of 25 proteins. **b**, An example of a well-predicted zinc-binding site (AlphaFold has accurate side chains). **c**, Another example of a well-predicted zinc-binding site. **d**, A complex of two proteins. **e**, The AlphaFold architecture.

Jumper et al. *Nature* (2021)

Random polymers almost never fold.



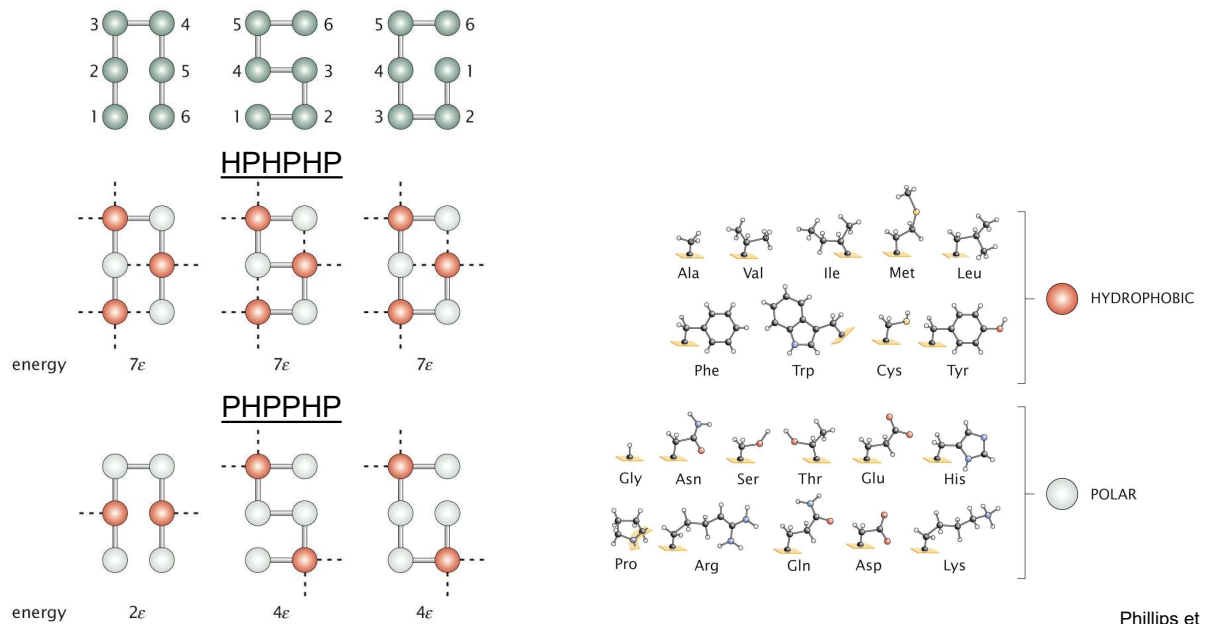
100 nm



Wiggins et al. *Nat Nanotechnol* (2006)

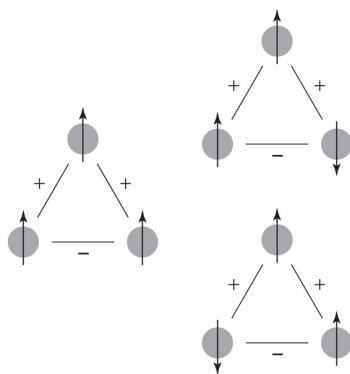
Wikipedia

Proteins as compact polymers with interacting residues

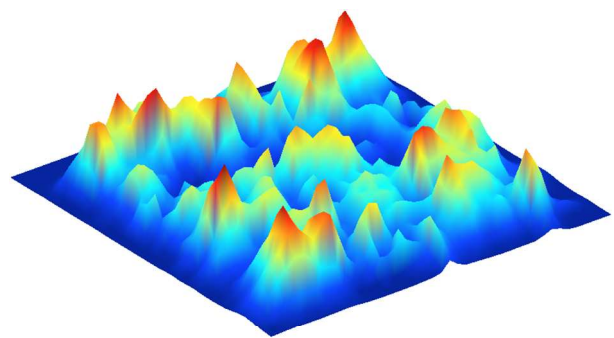


Spin glass will be our model system for **complex systems**.

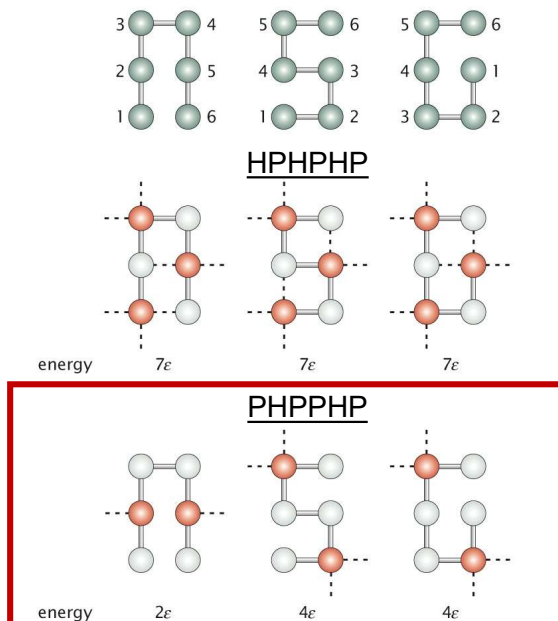
Frustration → Rugged landscapes



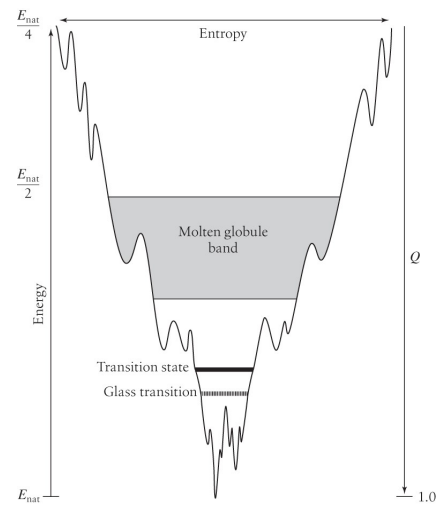
$$E = - \sum_{ij} J_{ij} s_i s_j$$



Proteins as compact polymers with interacting residues

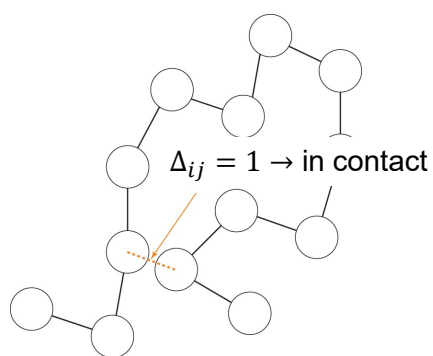


Minimize frustration → Funnel landscape

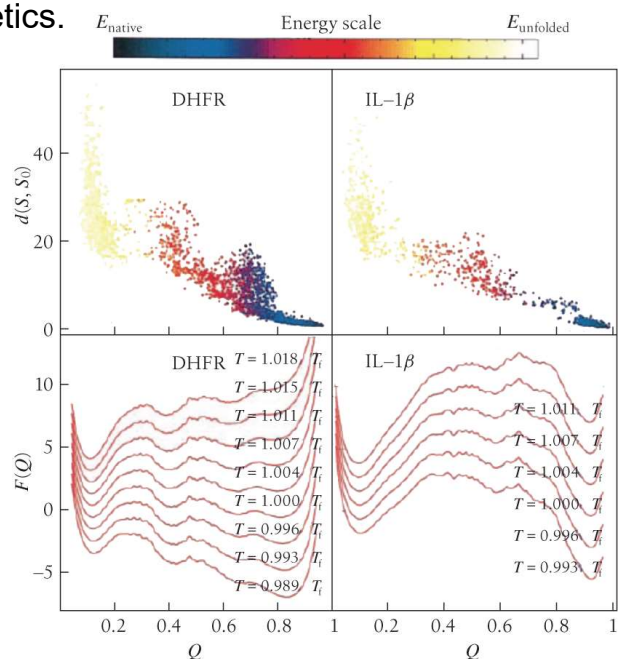


Onuchic et al. *PNAS* (1995)
Phillips et al

Native conformation predicts folding kinetics.

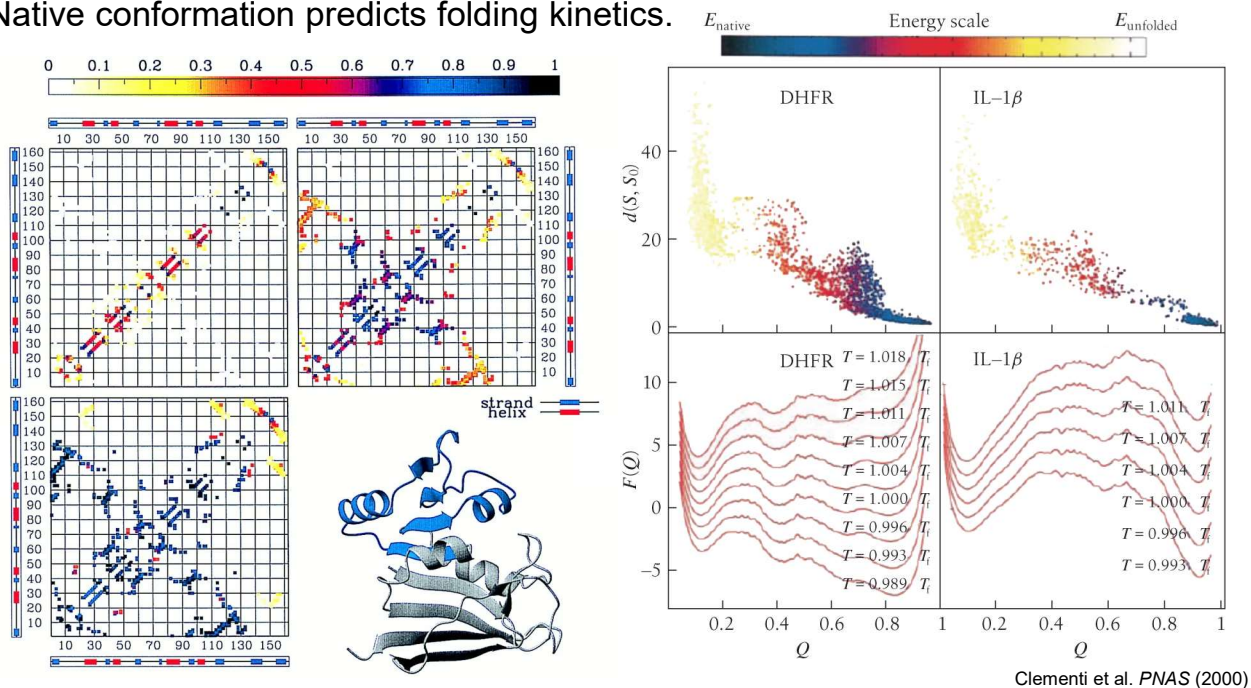


$$E = - \sum_{ij} c_{ij}^{\text{native}} \Delta_{ij} + \dots$$

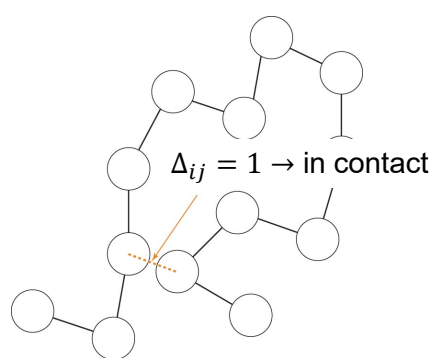


Clementi et al. *PNAS* (2000)

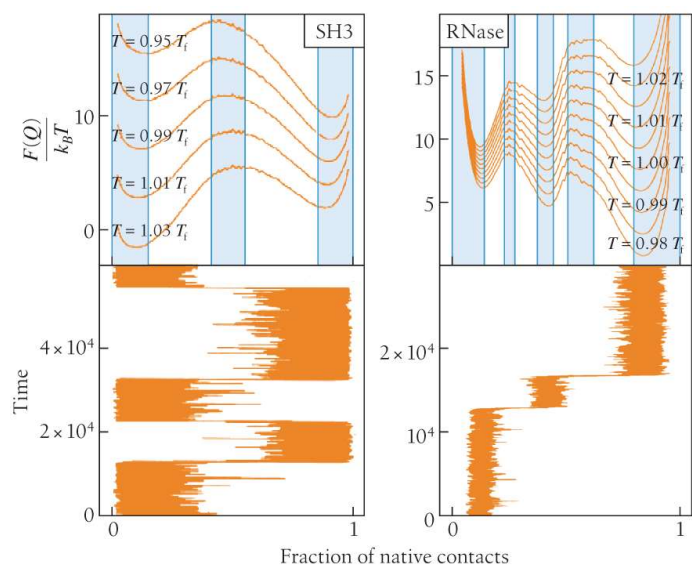
Native conformation predicts folding kinetics.



Native conformation predicts folding kinetics.

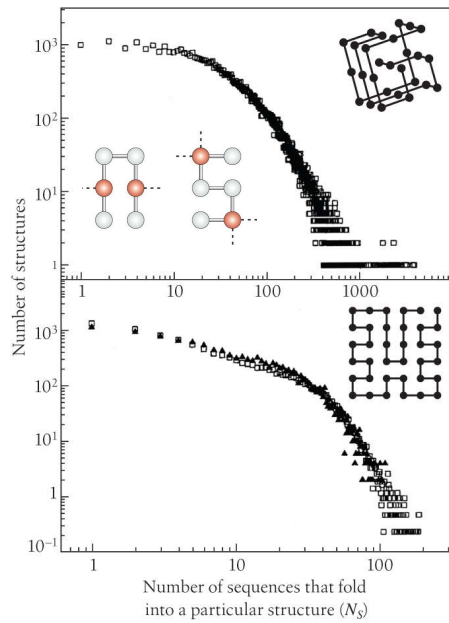


$$E = - \sum_{ij} c_{ij}^{\text{native}} \Delta_{ij} + \dots$$



Clementi et al. *J Mol Biol* (2000)

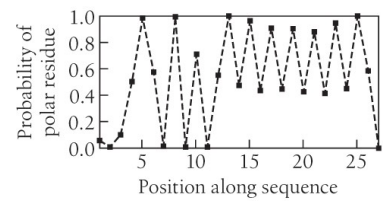
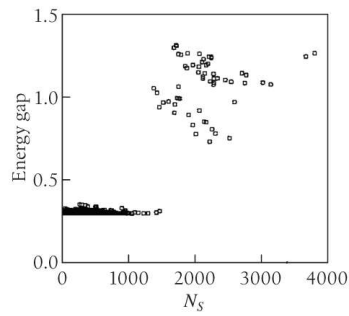
The problem of protein design



Emergence of Preferred Structures in a Simple Model of Protein Folding

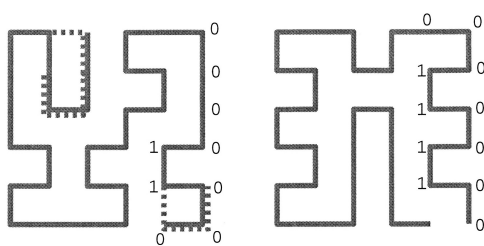
Hao Li, Robert Helling,* Chao Tang,† Ned Wingreen

Protein structures in nature often exhibit a high degree of regularity (for example, secondary structure and tertiary symmetries) that is absent from random compact conformations. With the use of a simple lattice model of protein folding, it was demonstrated that structural regularities are related to high “designability” and evolutionary stability.



Li et al. *Science* (1996)

Designable structures are atypical.



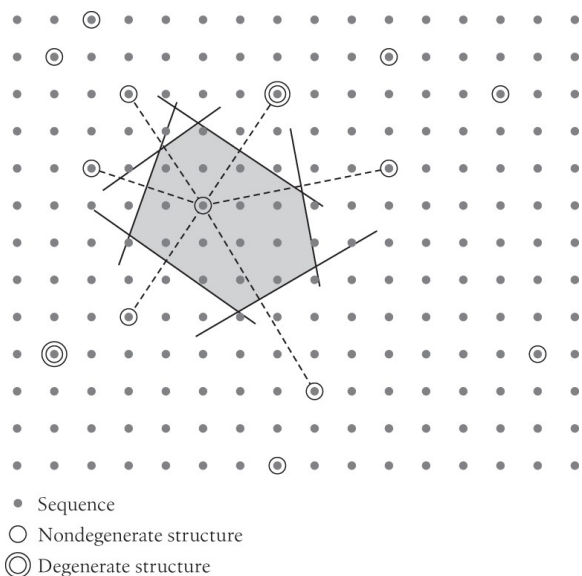
$S_a = 00011000000111111100000110000011110$

$S_b = 001100110000110000110011000011111100$

$$E = \sum_i (s_i - \sigma_i)^2$$

Hydrophobic/
Polar

Core/
Surface



Li et al. *PNAS* (1998)

Figure 1: SCA conservation analysis of the protein structure and sequence.

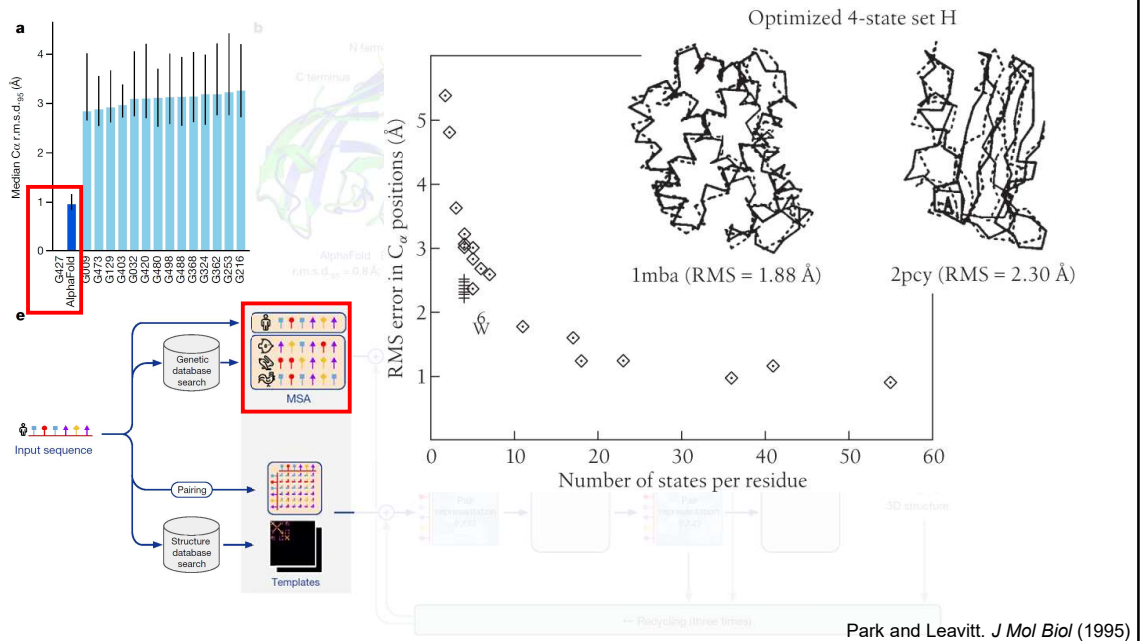
The figure displays the SCA conservation analysis of the protein structure and sequence. The top left shows a 3D ribbon diagram of the protein structure, highlighting residues W7, W30, and Y31. The top right shows a bar chart of SCA conservation scores for residues 1-35, with peaks corresponding to beta-strands $\beta 1$, $\beta 2$, and $\beta 3$. The bottom left shows three heatmaps of SCA conservation scores for the protein structure, comparing Natural sequences, Independent sites model, and + Pairwise correlations. The bottom right shows four pie charts illustrating the distribution of conservation scores for Natural, Coupled conservation, Site-independent conservation, and Random models. A color scale from 0 to 2 is provided for the heatmaps. The legend indicates: Folded (red), Unfolded (blue), Insoluble (yellow), and Unexpressed (grey).

Socolich et al. *Nature* (2005)

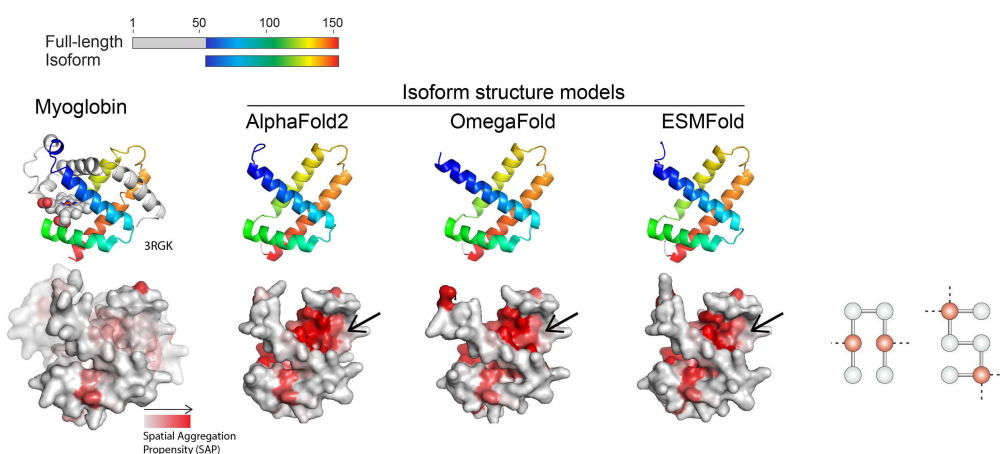
Figure 1: Protein structure and contact map of Spo0F. Left: 3D ribbon diagram of Spo0F with helices $\alpha1$, $\alpha2$, and $\alpha4$. Residues are colored by cluster: red (14-29), green (251-299), and blue (300-309). Right: Contact map (MI vs DI) showing clusters of residues. A dashed red line at MI \approx 0.25 and DI \approx 0.055 separates the clusters. A blue circle highlights a cluster of residues around (MI, DI) = (0.2, 0.06).

Weigt et al. *PNAS* (2009)

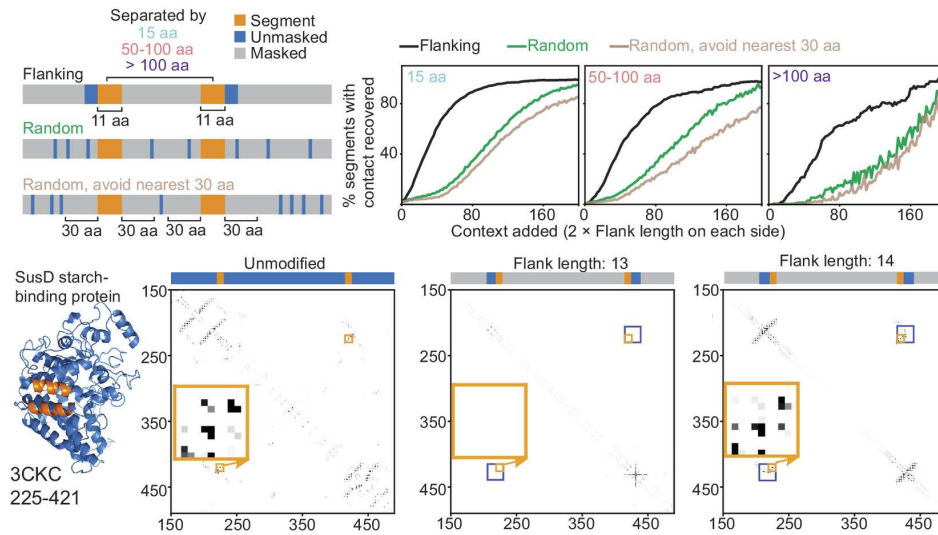
How much information is required to specify structure?



What information is contained in large models?



What information is contained in large models?



Zhang et al. *PNAS* (2024)

Native protein sequences are close to optimal for their structures

Brian Kuhlman and David Baker*

Department of Biochemistry and Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, WA 98195

Edited by William F. DeGrado, University of Pennsylvania School of Medicine, Philadelphia, PA, and approved July 11, 2000 (received for review March 20, 2000)

How large is the volume of sequence space that is compatible with a given protein structure? Starting from random sequences, low

many common features with the energy functions used for protein design. Therefore, it is possible that refinement builds a

progress

A surprising simplicity to protein folding

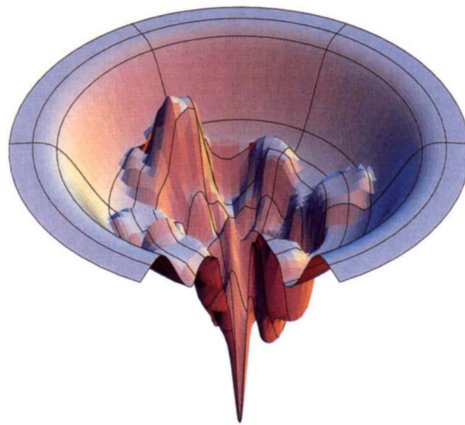
David Baker

Department of Biochemistry, University of Washington, J567 Health Sciences Building, Box 357350, Seattle, Washington 98195, USA

The polypeptide chains that make up proteins have thousands of atoms and hence millions of possible inter-atomic interactions. It might be supposed that the resulting complexity would make prediction of protein structure and protein-folding mechanisms nearly impossible. But the fundamental physics underlying folding may be much simpler than this complexity would lead us to expect: folding rates and mechanisms appear to be largely determined by the topology of the native (folded) state, and new methods have shown great promise in predicting protein-folding mechanisms and the three-dimensional structures of proteins.

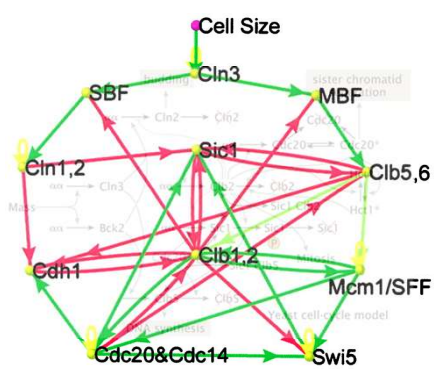
Summary

- Proteins are designable (in theory and in practice).
- Spin glasses are everywhere and frustration is a key feature of complex systems.

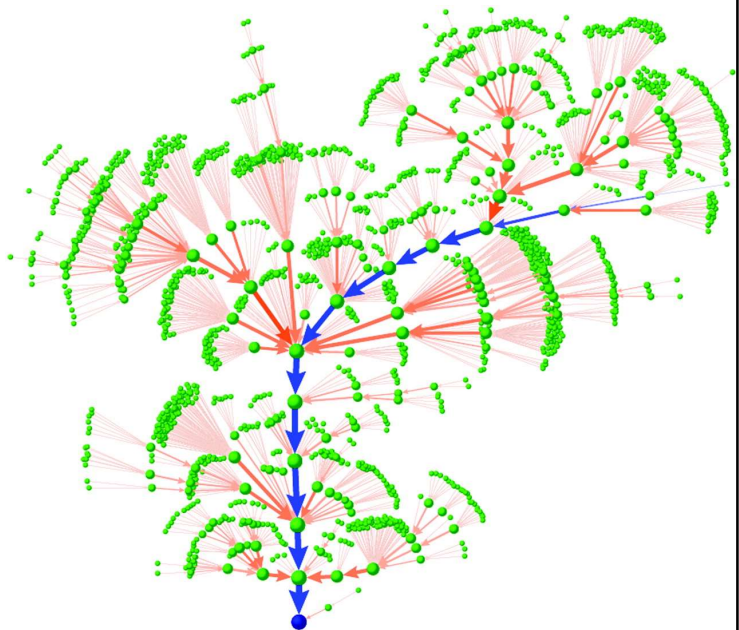


Dill and Chan. *Nat Struct Mol Biol* (1997)

Spin glass for cell cycle

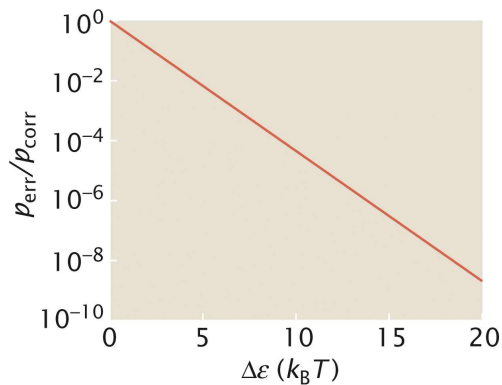


$$s_i(t+1) = \theta \left[\sum_j J_{ij} s_j(t) \right]$$

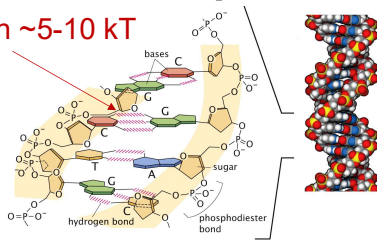


Li et al. *PNAS* (2004)
Phillips et al

Detour: How specific?



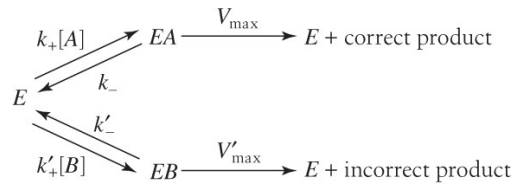
Mismatch ~5-10 kT



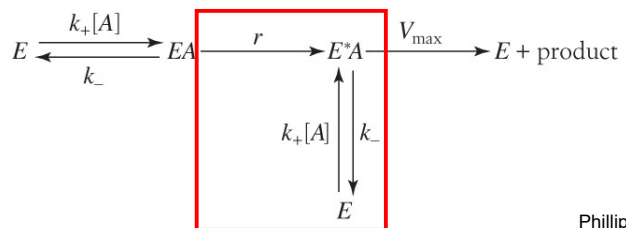
No error correction

A = "correct" substrate

B = "incorrect" substrate

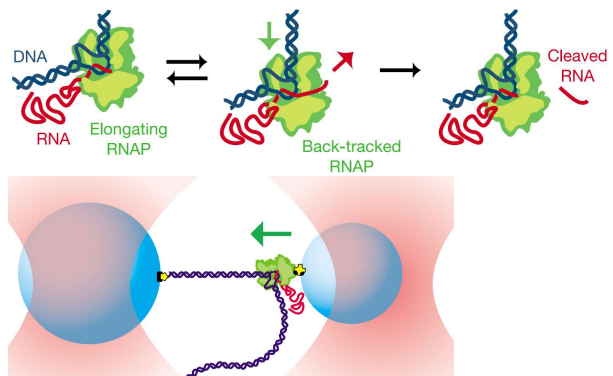
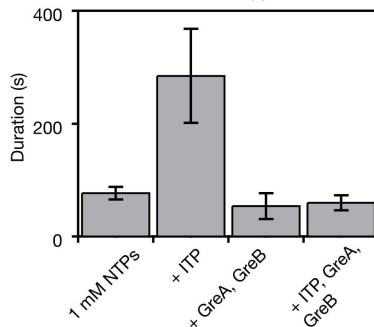
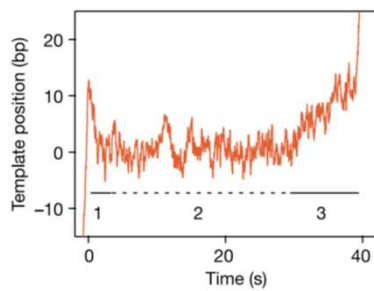


Kinetic proofreading



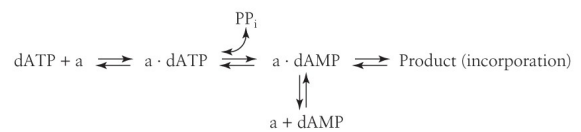
Phillips et al
Bialek

Detour: How specific?



Kinetic proofreading

Replication of DNA, or transcription of RNA, a = DNA template



Shaevitz et al. *Nature* (2003)
Bialek