The Protein Design Problem

- **Protein folding problem**: predict the three-dimensional structure of a protein from its sequence (difficult?)

- **Protein design problem**: given a desired structure, design an amino acid sequence capable of assuming that structure (difficult, too?)

Why?
- Must design a stable protein fold
- Must design the precise orientation of amino acid residues

- Why do we want to de novo design a protein?

1) De novo protein design assists the protein folding problem

2) Being able to design a protein from scratch allows us to better understand the roles of certain residues in proteins of interest

3) We may be able to create artificial enzymes and receptors once de novo protein design is sufficiently sophisticated

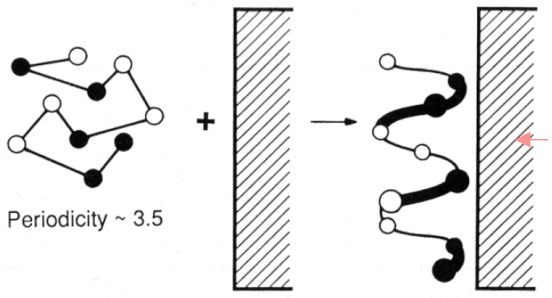
Example 1: Design of a α-helix

- **Minimalist approach**: reduce the 20 proteinogenic amino acids to two (one hydrophobic and one polar)

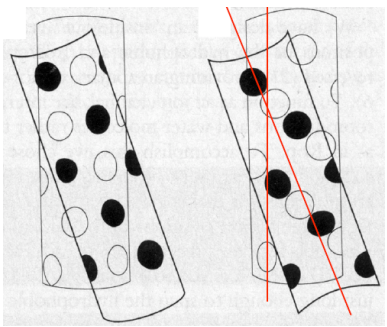
-"**Binary patterning**": place hydrophobic residues on one side of each helix with a periodicity of 3.5 residues, and polar residues at other positions

- Rely on the tendency of **hydrophobic surfaces** to aggregate to drive secondary (helix) and tertiary (4-helix bundle) structure formation

- Consider **side chain packing** between helices: Interhelical angle of 20° packs "knobs into holes"



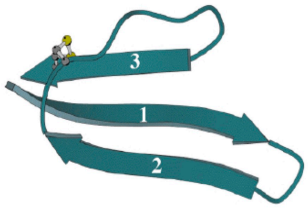
hydrophobic surface (another helix or solvent interface)



Example 2: Design of a β-sheet

- Goal: design a simple, well-folded, β-sheet toxin hand (TH)

- Designing a predominantly β-sheet protein is challenging compared with α-helical protein design



- **Strategies:**

(1) **"tie" the terminal together**:

(2) **"turn" region:**

(3) **overall choice of amino acids:**

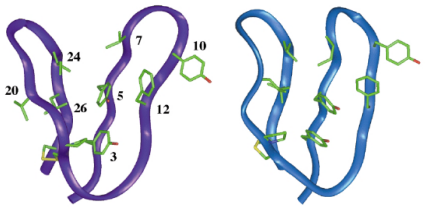
(4) **choice of amino acids in the core**:

(5) **choice of amino acids on the surface of the sheet**:

Nat. Struct. Biol. **2001**, *8,* 535 – 539.

Results:

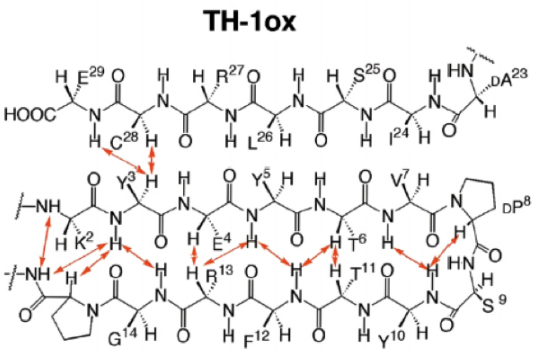
1. T_m = 54-57°C (weak cooperative thermal unfolding)
2. NMR structure shows a hydrophobic core
3. Sedimentation suggests monomer at 750 μM

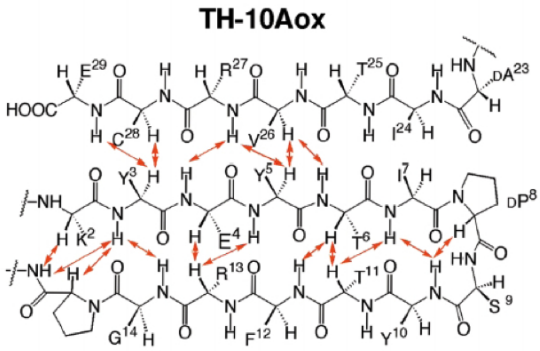


TH-1 TH-10A

Conclusion:

Segments of α-, β-, and defined turn structures can be designed, while non-disordered loop regions remain difficult.



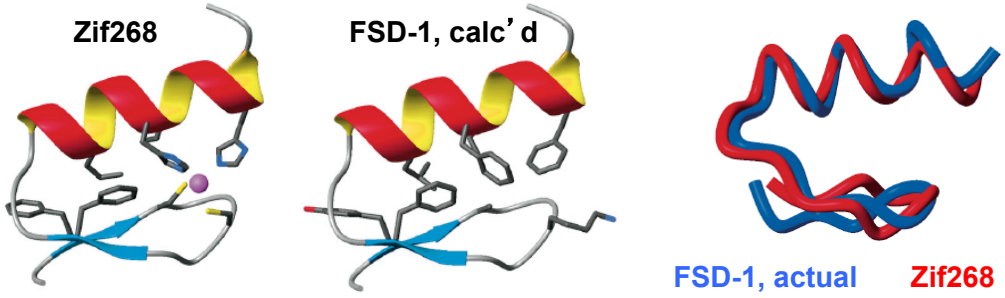


Automated Protein Design

- Can computational approaches design all parts of a protein (core, surface, and boundary)?

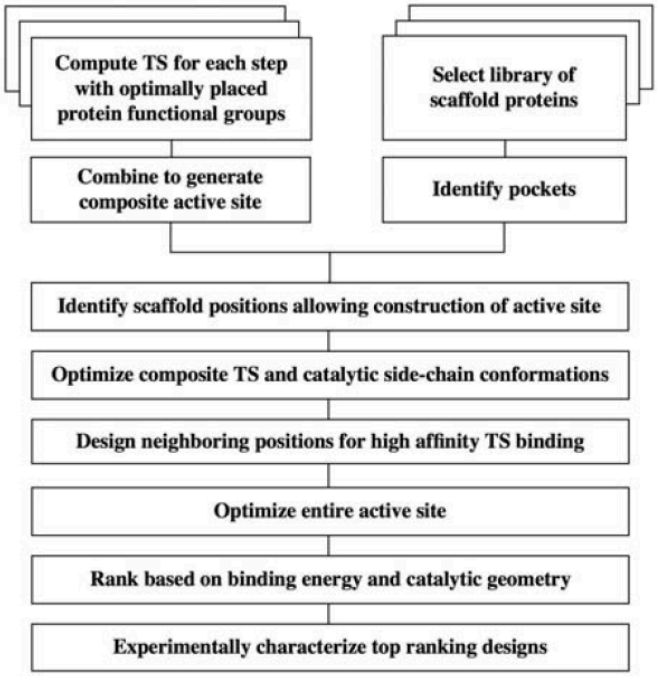
- Ideally, want a general algorithm that solves any inverse protein folding problem (input a structure, output a sequence)

- Strategy: start with a desired backbone; for every side chain, consider all reasonable rotamers of all 20 amino acids at that position and score based on van der Waals, hydrogen bonding, secondary structure propensities, and solvation energetics…



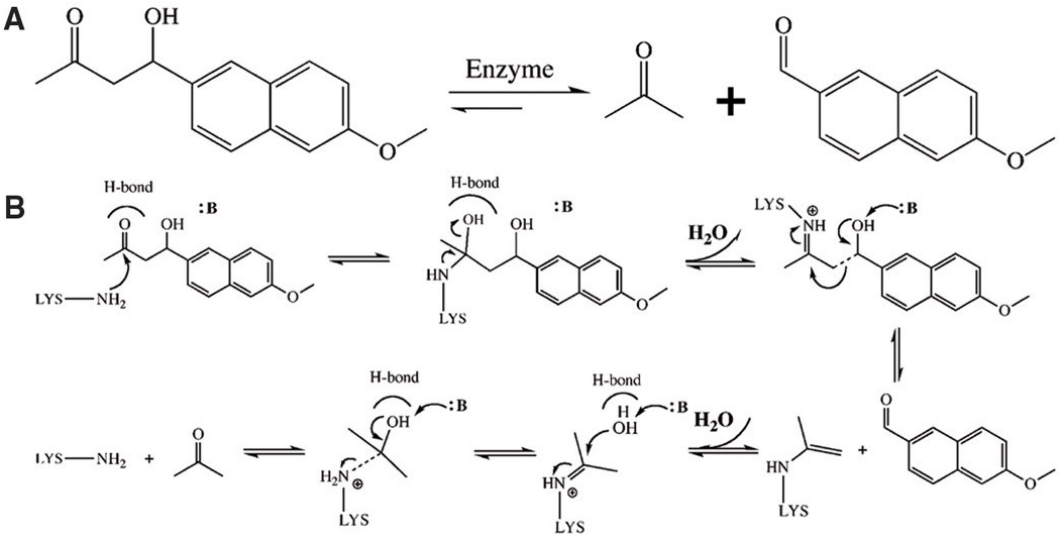
Computational Redesign of Proteins

- Another approach: start with natural protein scaffold, design new function

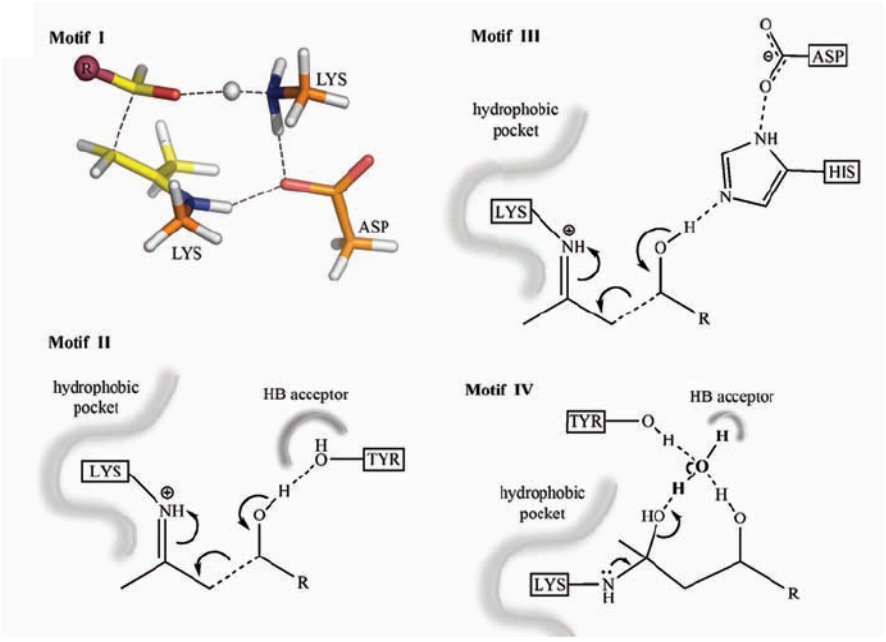


- It is necessary to ensure that the designed active site not only strongly stabilizes the highest energy transition state, but also interacts favorably with the reaction intermediates and other transition states along the reaction pathway

Science **2008**, *319*, 1387-1391.



Science **2008**, *319*, 1387-1391.



1. Why put lysine in a hydrophobic pocket?

2. What is the role of Tyr? Why use Tyr?

Science **2008**, *319*, 1387-1391.

- Early efforts at de novo protein design were partially successful at designing helical proteins, but many produced molten globules

- Minimalist and empirical approaches have recently achieved the design of some native-like proteins, including β-sheet motifs, but some subsequences still need to be copied from natural proteins

- Computational approaches rapidly evaluate many possible designs *in silico* and have achieved some success in designing structure and catalytic function

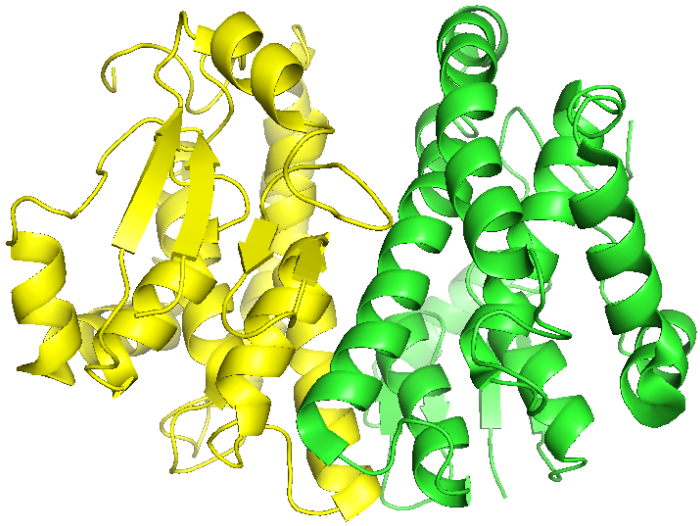
- De novo protein design, especially of catalysts, remains very challenging

- Notable names: Bill DeGrado (UCSF)
Barbara Imperiali (MIT)
Stephen L. Mayo (Caltech)
David Baker (U of Washington)
Kendall N. Houk (UCLA)

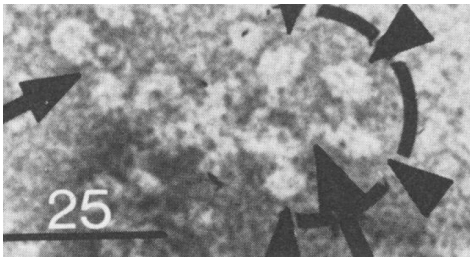
Design/Evolve Protein-Protein Interactions

- Rational design

- Structure-base directed evolution

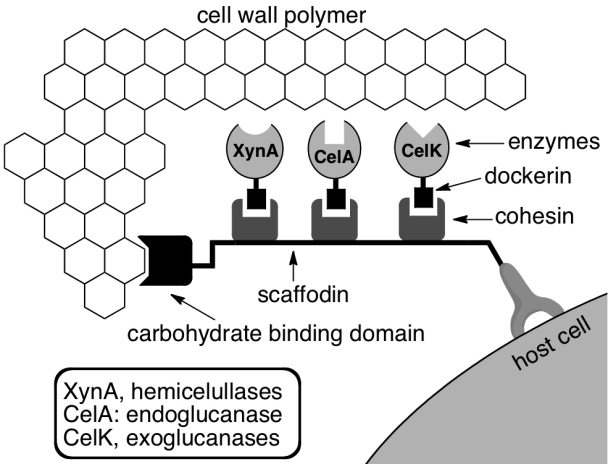


- An real research example in evolving protein-protein interactions

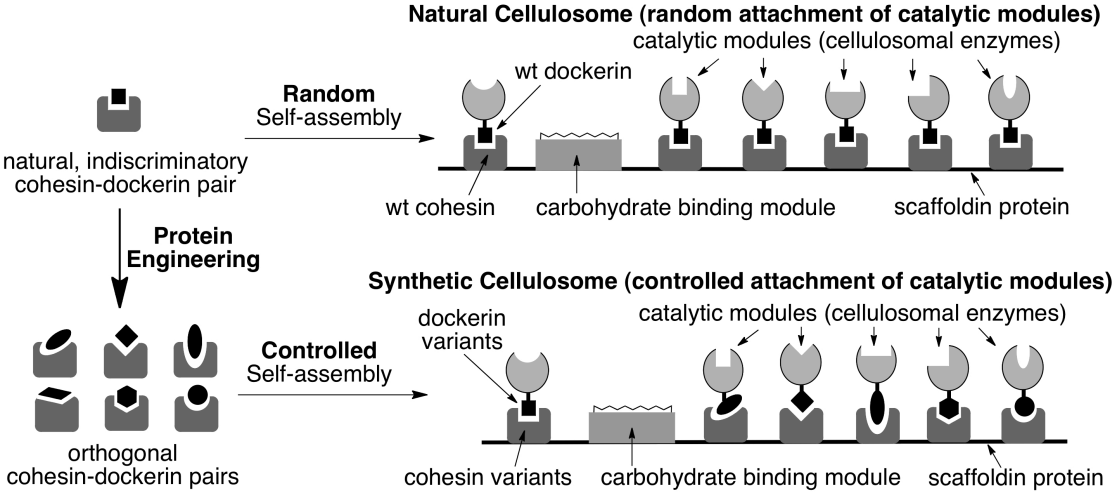


Electron microscopic projection of an artificially flattened cellulosome.

Mayer, Coughlan, Mori, Ljungdahl,
Appl. Environ. Microbiol. **1987**, *53*, 2785.

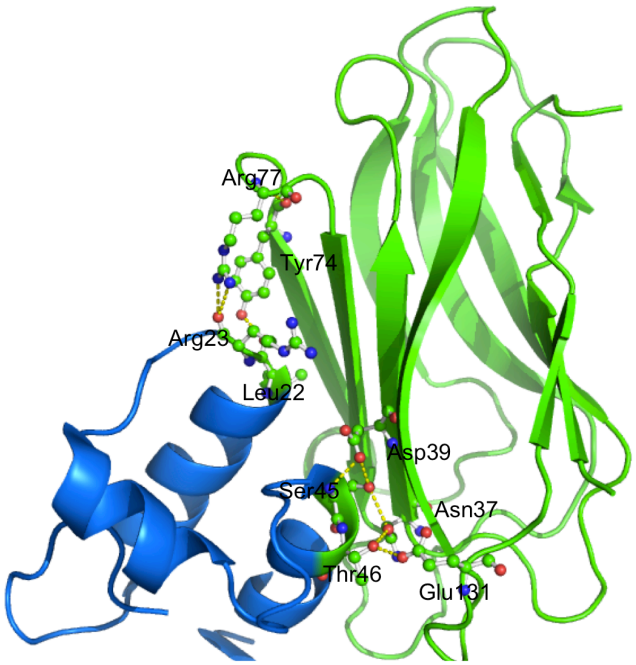


- From random to controlled assembly of Cellulosome

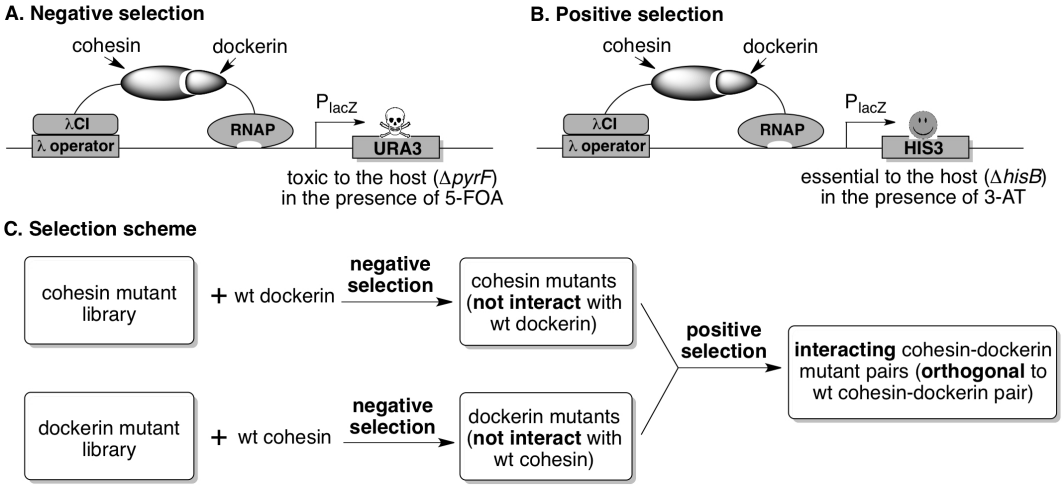


- Cellulosome assembly: based on cohesin - dockerin interaction.
- Nonhomogeneous catalyst: 6.6 x 10^{16} variants within a single species.
- Synergistic action among cellulosomal enzymes.

Construction of Cohesin & Dockerin Libraries

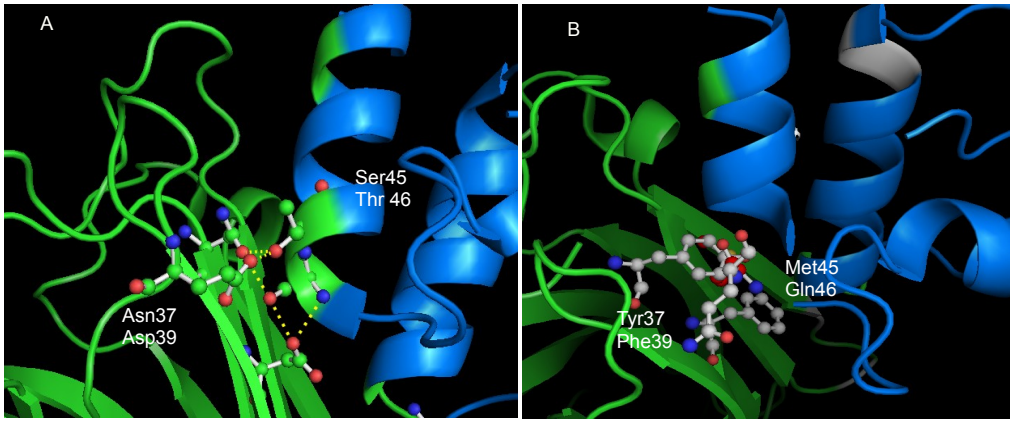


Selection Scheme

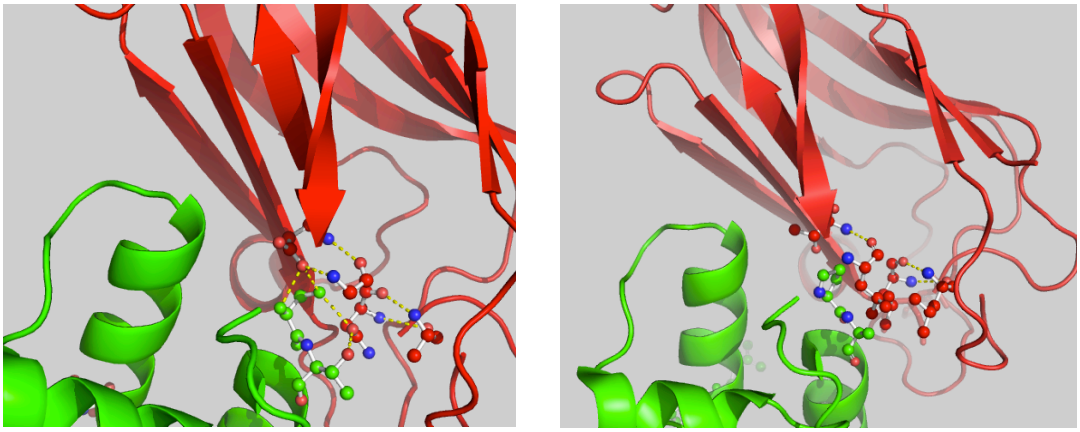


Abbreviation: λcI, bacteriophage λ repressor protein; RNAP, α-subunit of RNA polymerase; P_{lacZ}, the lac promoter; 3-AT, 3-amino-1,2,4-triazole; 5-FOA, 5-fluoroorotic acid.

To visualize mutations that affect protein-protein interactions



To visualize mutations that affect protein-protein interactions



A 2nd example