iBiology.org Teaching Tools
David Baker's Lecture Part 1:
Introduction to Protein Design

Teaching Tools were prepared by Aditya Anand and Lynn Wang.

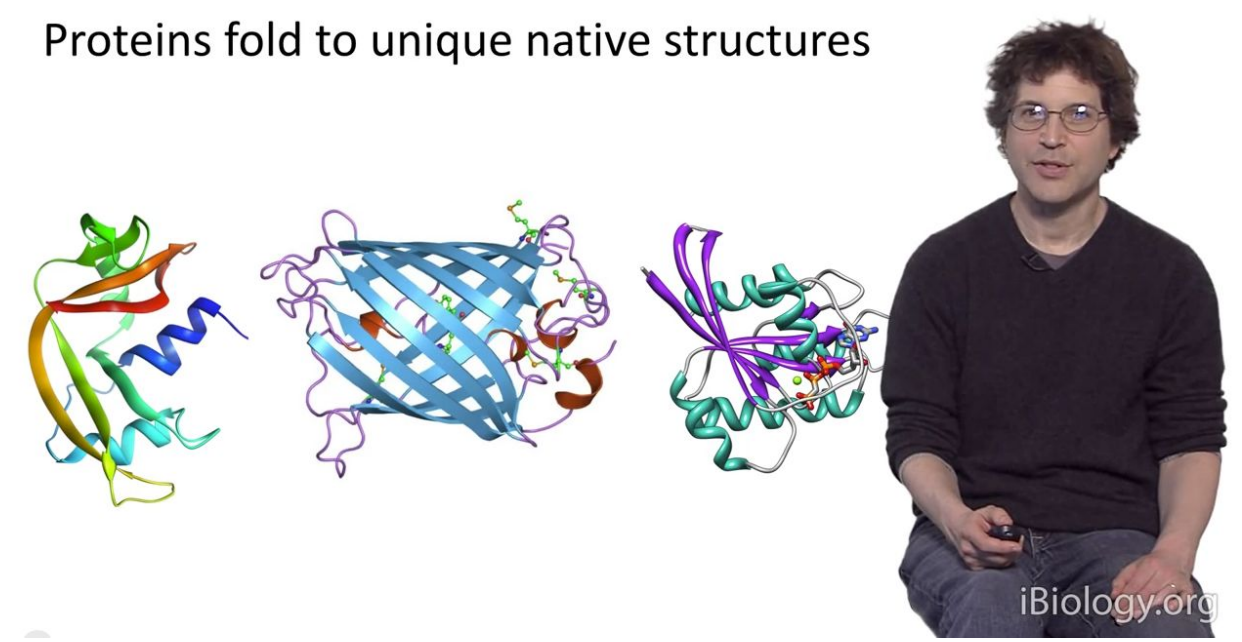
Contents

---

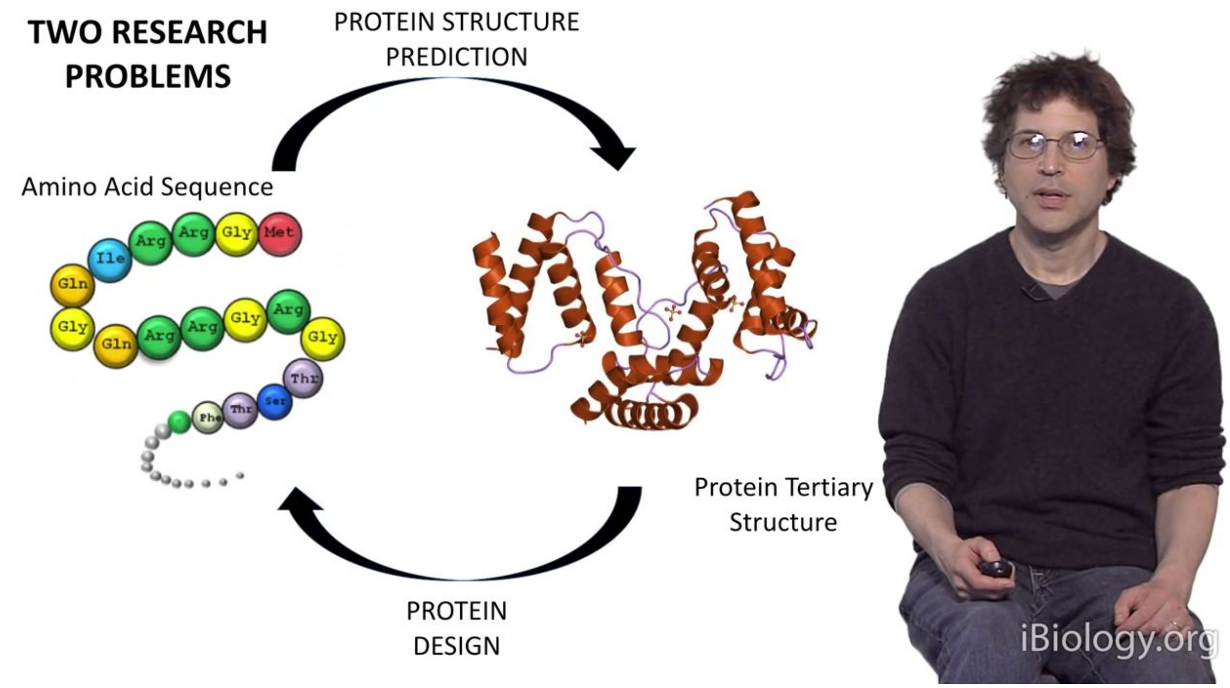
1. Keywords and Terms

Protein design, protein structure prediction, protein engineering, idealized proteins, Rosetta methodology

2. Lecture Notes

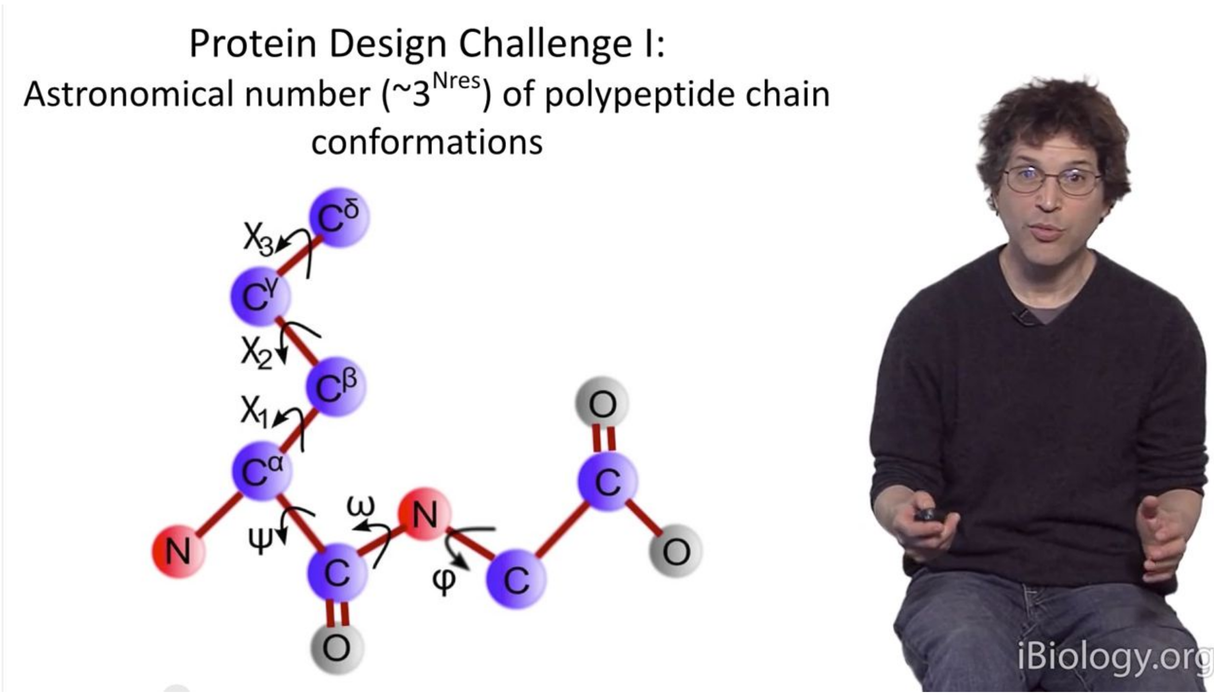


0:21—
Proteins are encoded by genes and fold to form unique native structures. Proteins can fold into many different conformations, with the native state conformation being the lowest energy state.

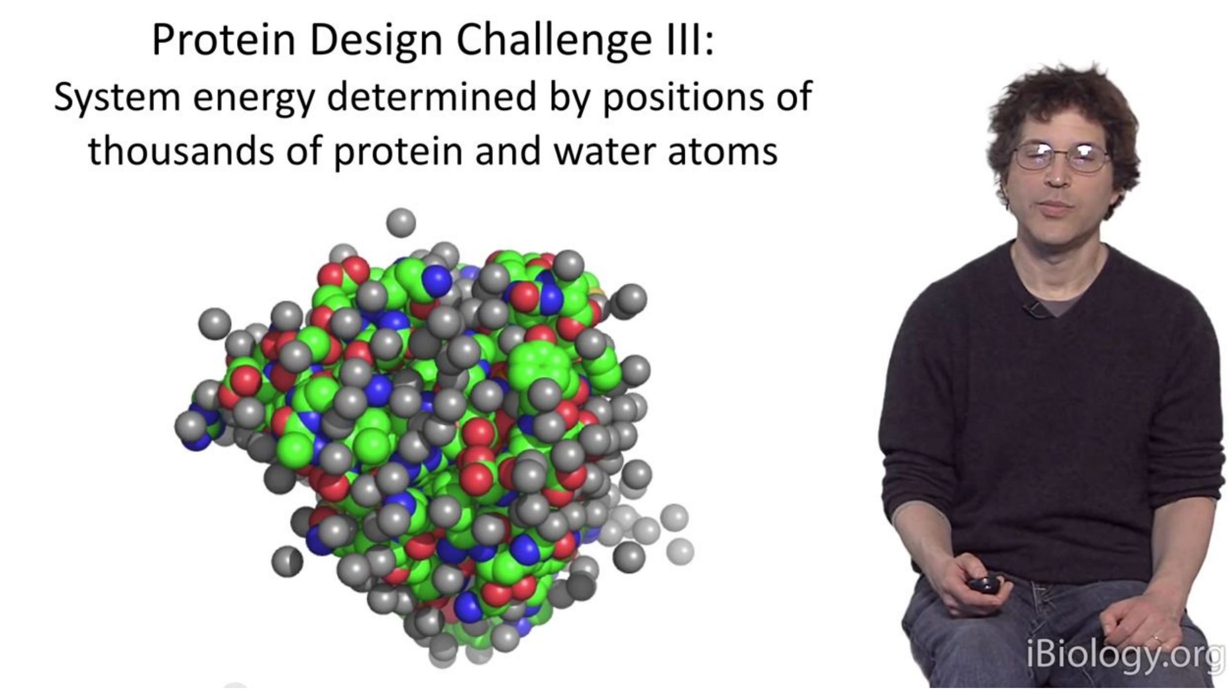


1:09—

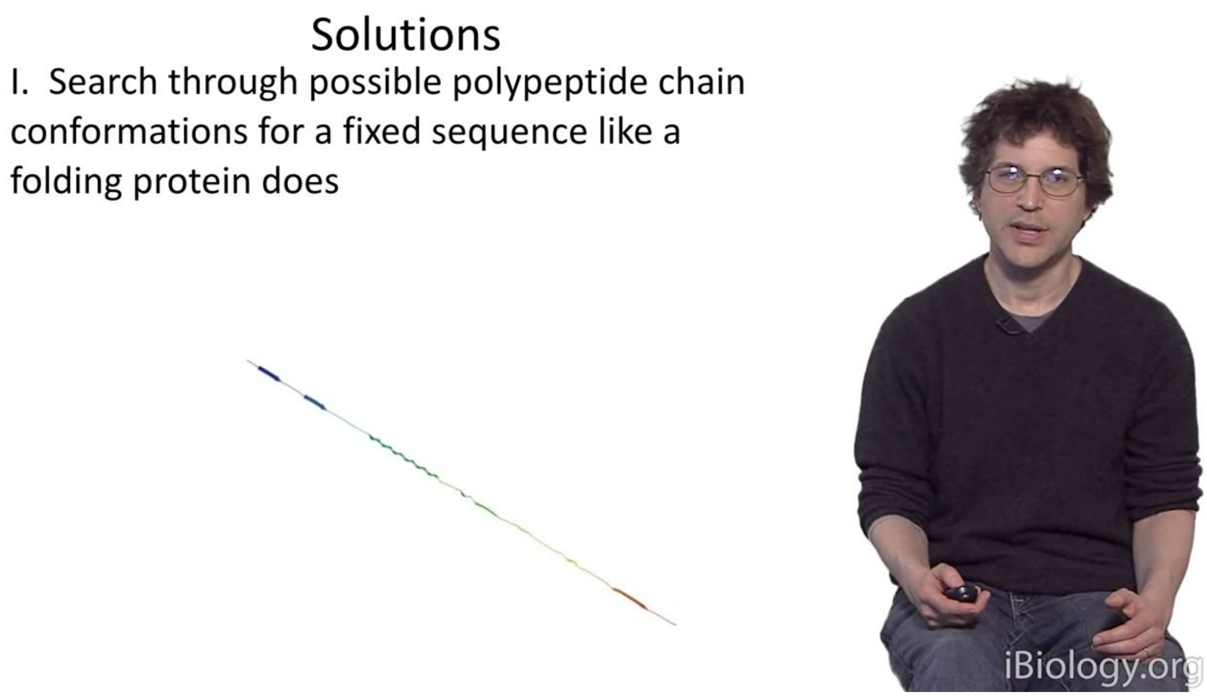
Two main research problems in protein design are protein structure prediction and protein design. With protein structure, one determines the lowest energy structure and function for a protein based on a fixed amino-acid sequence. With protein design, one begins with an ideal protein structure and then determines the amino acid sequence required to make this protein.



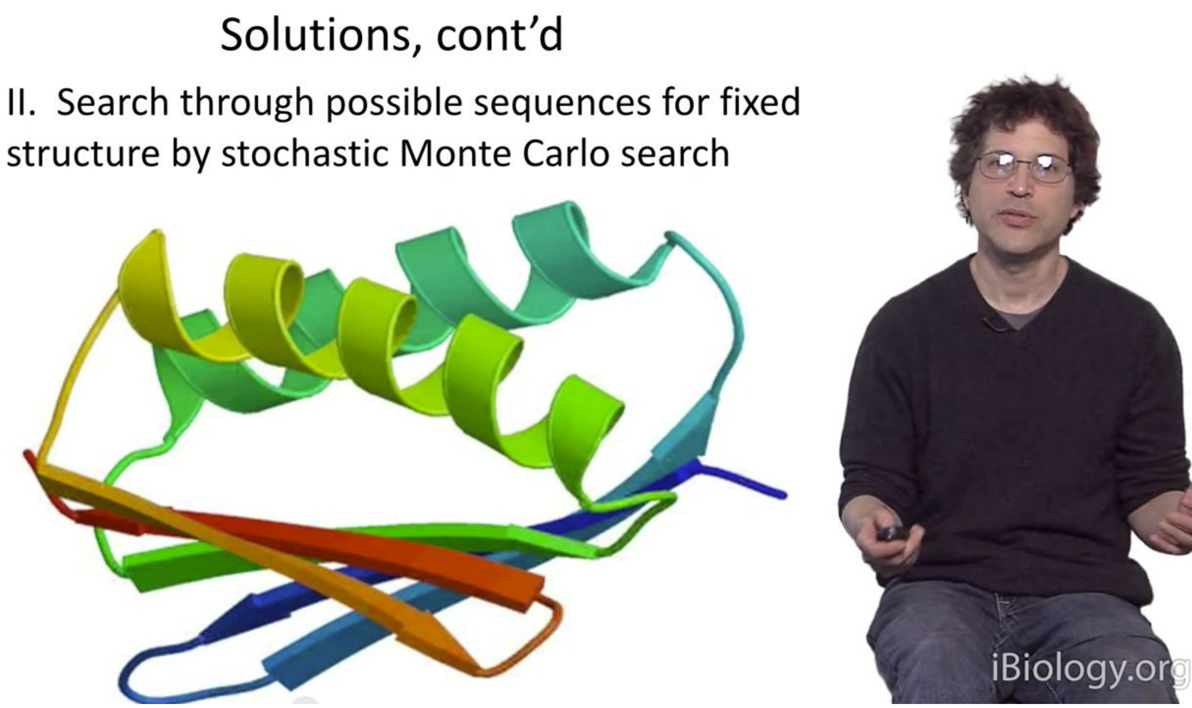
2:47—
Protein design is difficult due to the large number of polypeptide chain conformations and protein sequences.



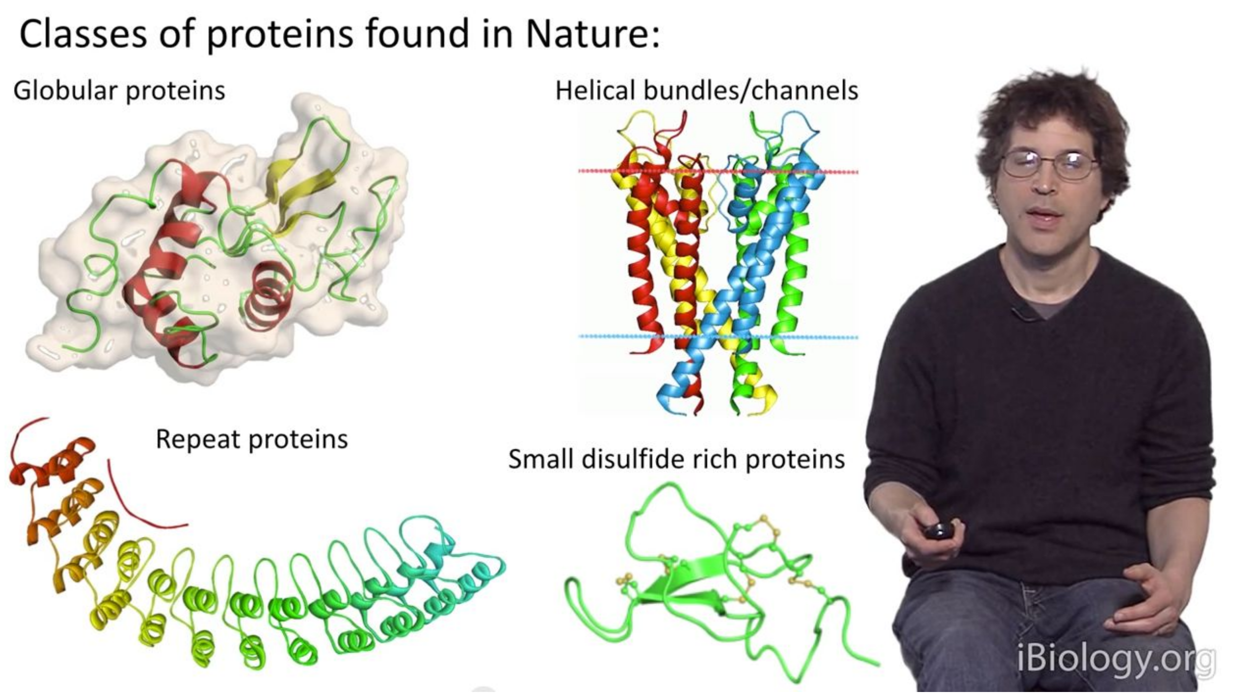
3:48—
Proteins are composed of many atoms and surrounded by many water atoms. The energy of all these atoms needs to be determined for proper protein design.



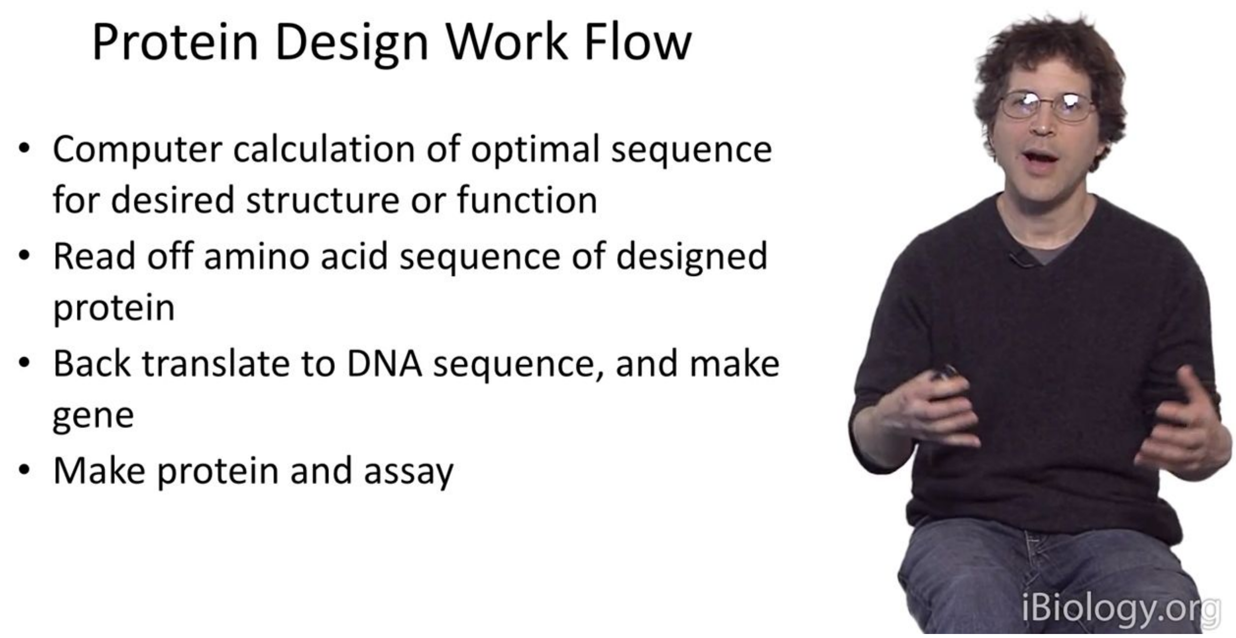
4:21—
The Baker lab uses the Rosetta methodology to mimic protein folding. This method samples through possible protein conformations to identify the lowest energy structure.



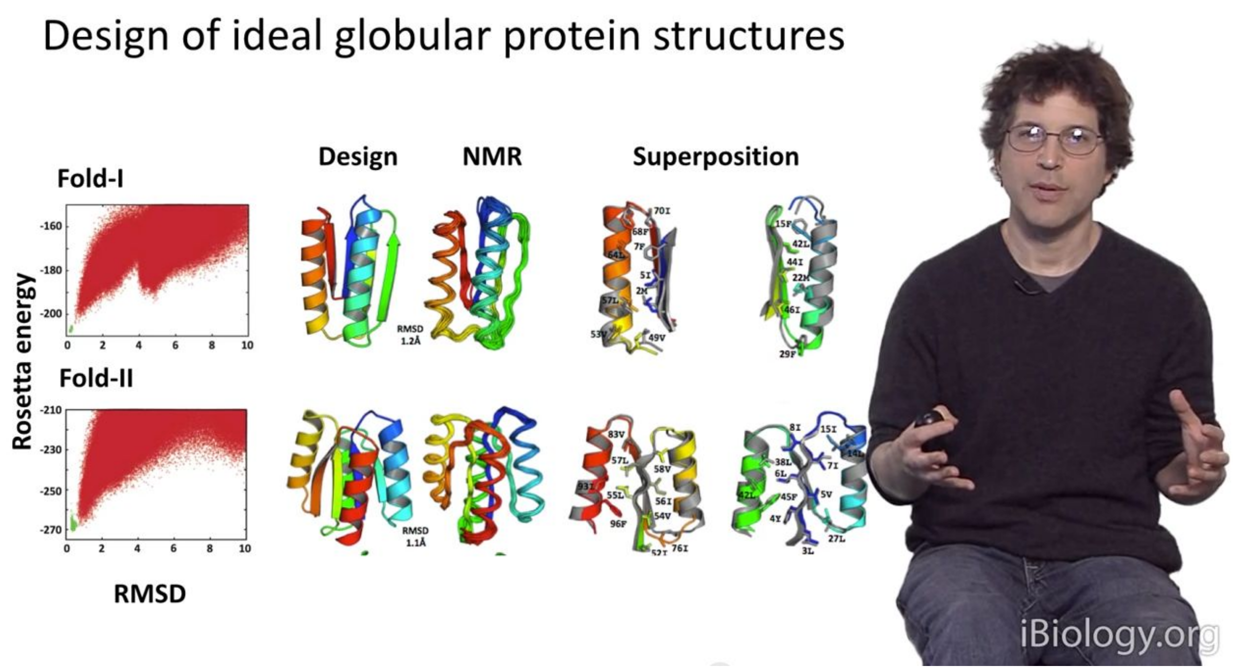
5:39—
The Monte Carlo Search is another method for identifying the lowest energy sequence. It substitutes different amino acids at random locations in a protein structure.



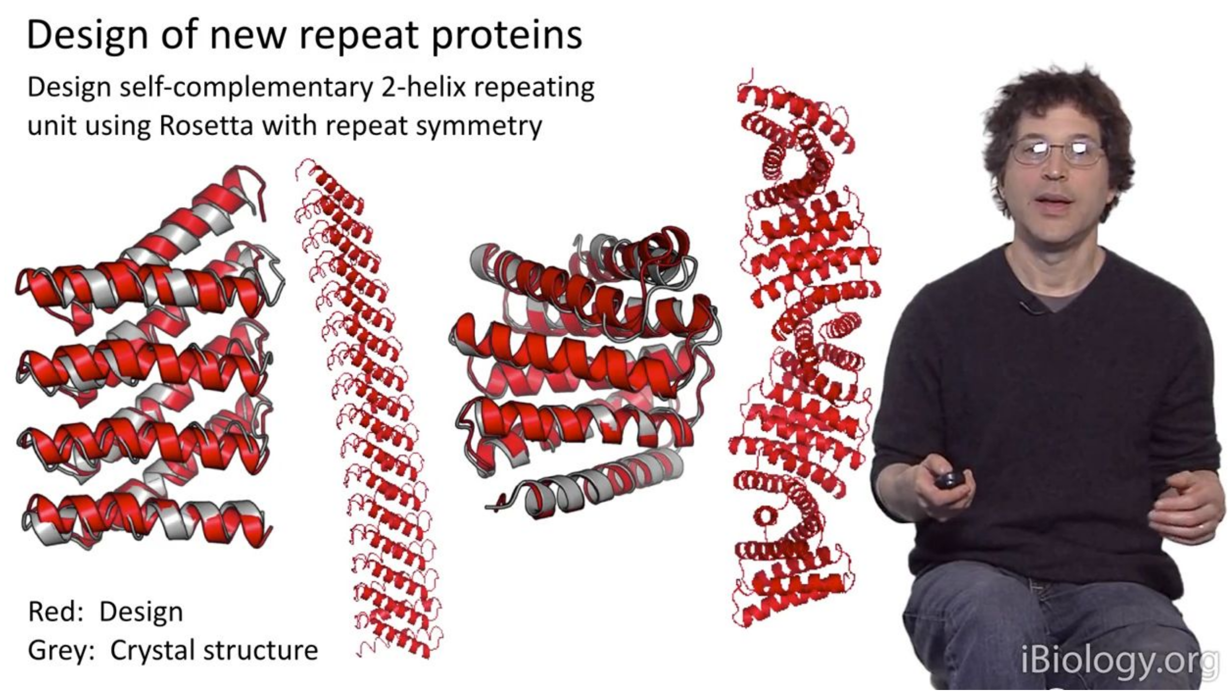
8:15 –
Four classes of proteins are found in nature: globular proteins, helical bundles, repeat proteins, and small disulfide rich proteins. Baker will discuss his lab's efforts to design ideal versions of these proteins.



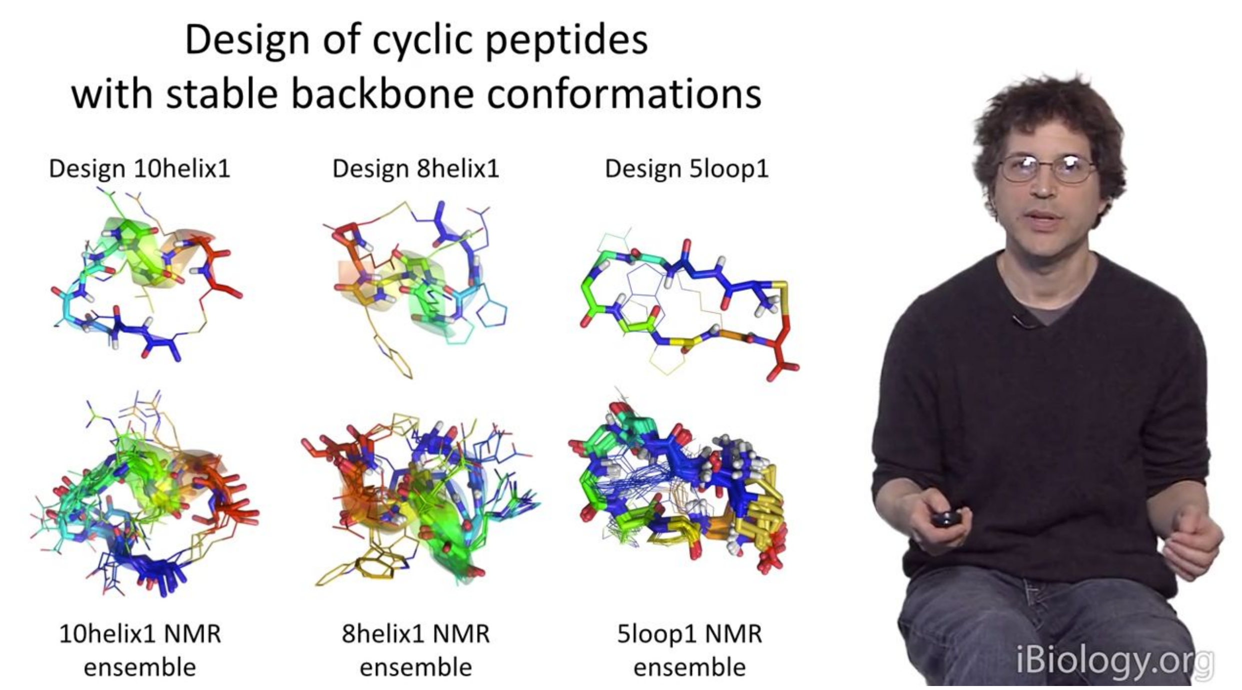
10:31—
The workflow for protein design begins by designing the protein on the computer. A gene is then made encoding the protein. The gene is placed in bacteria, which are allowed to grow a copy of the gene. The protein is then extracted to determine its structure and properties.



12:21-
Once the lab determines the design calculation for the ideal protein, they recruit volunteers to participate in "Rosetta at home", where they predict the structure for a specific sequence. This was done with their work on globular proteins. Each red dot in the left-hand graph represents the results from one volunteer. The ideal protein is then made in the lab and the structure is solved using NMR.



18:12—
Another example includes new repeat proteins designed by the Baker Lab. Please note that the crystal structure (in red) is almost identical to the design (in gray). These proteins could potentially be used in designing nanomaterials.



19:05—
The final class of proteins the Baker Lab studies are small disulfide proteins, which could potentially become the basis for therapeutics.

19:49—
To summarize, Baker explained the protein structure prediction and protein design problems, the approach taken for these problems, and how the lab designs idealized versions of proteins found in nature. These proteins will likely be the basis for designing new world functional proteins for solving modern day problems.

3. Review Questions

1. What are the two main research problems in protein design that Baker discusses?
2. In which state is a protein structure most stable?
3. What are the three challenges associated with protein design?
4. Which methodologies does the Baker lab use for protein design?
5. Which proteins are found in nature?

6. Describe the protein design workflow that the Baker lab follows.

4. Answers to Review Questions

1. Protein structure prediction and protein design.
2. The protein structure is most stable in its lowest energy state.
3. The number of possible polypeptide conformations and amino acid sequences and the energy state of thousands of atoms that make up a protein make protein design a challenging task.
4. The Rosetta methodology, Monte Carlo Search, and modeling methods.
5. Globular proteins, helical bundles, repeat proteins, small disulfide proteins.
6. After the protein is designed on the computer, bacteria are transfected with a vector containing the gene encoding the protein of interest. This creates many copies of the gene that researchers then use to study the protein's structure and properties.

5. Discussion Questions

1. What is the purpose of designing an "idealized" protein?
2. What do all three methods for solving protein design have in common? Why is this important?
3. Can you come up with a few applications of protein design?
4. Think of a specific problem in which an idealized protein could act as a solution.

6. Answers to Discussion Questions

1. Idealized proteins enable scientists to design proteins that have the exact capabilities/functions required. These proteins can then be used to solve many modern day problems. The discussion paper that is part of this lecture series is a good example of an application---vaccine design.

2. All three methods involve calculating the lowest energy state for a given protein structure. This is important because proteins are found naturally at their lowest energy state. Thus, an idealized protein would need to be stable or at its lowest energy state in order to be functional.
3. Possible protein design applications—vaccine design, biosensor design, drug design therapeutics, new research tools/techniques, enzyme design, protein design for specificity and affinity,
4. Possible answer: Since many diseases involve protein-protein interactions, one could design a protein that binds to one of these partners to interrupt the interaction that leads to a disease phenotype.

7. Explain or Teach These Concepts to a Friend

1. Explain the two main research problems of protein design.
2. Explain why proteins fold naturally at their lowest energy state.
3. Explain how the Rosetta methodology works. For extra credit, go to the website
4. http://boinc.bakerlab.org/rosetta/, where you can help the Baker lab determine the 3D shape of proteins yourselves.

8. Questions for Discussion Paper

Discussion Paper:

Correia, B. E. Proof of principle for epitope-focused vaccine design. 2014. Nature. 507: 201–206.

1. How does epitope-focused vaccine design work? What are the advantages of this method compared to traditional vaccine design methods?
2. Briefly describe the computational method of protein design discussed in this paper (Fold From Loops). How many FFL designs were ultimately chosen for filtering and human-guided optimization?

3. The authors then did an immunological evaluation of specific FFL designs they optimized. What evidence did the authors have to support and/or go against the clinical relevance of their designed FFL scaffolds?
4. What evidence did the authors have from their antibody characterization that their designed scaffolds can “re-elicit” neutralizing antibodies?
5. Do you think there is enough evidence supporting the efficacy of FFL scaffolds for use as a vaccine against RSV? If no, which additional experiments are required? Comment on the feasibility of this approach for developing vaccines against other viruses, i.e. HIV or Ebola.

9. Answers to Questions for Discussion Paper

1. Epitope-focused vaccine design involves designing an immunogen for a specific binding area of a virus—an epitope. The binding of the immunogen would lead to the induction of protective neutralizing antibodies. Many viruses have resisted vaccines designed by traditional methods even though such viruses have identified vulnerable epitopes. This is due to the variability amongst different strains of the same virus and thus difficulty in identifying a specific immunogen to target the epitope. The advantage of epitope-focused vaccine design would be the ability to design an ideal immunogen protein that would overcome this setback.
2. Once the epitope-antibody complex has been identified for the virus of interest, the Fold Form Loops (FFL) method selects the functional motif of the epitope and matches it up with its target topology. Next, backbone conformations are designed for the epitope that match the target topology. Low-energy amino-acids are then selected for the backbone conformation of interest. Finally, a series of filtering and optimization are carried out to identify the best designs that will be selected for use. (Also, refer to Figure 1)
3. The authors tested the binding of human sera from six RSV-positive patients and found that three sera reacted to one of the designed epitopes, FFL_001. This suggests that FFL_001 is clinically relevant for RSV. (Also, refer to Figure 3)
4. The authors analyzed the monoclonal antibodies generated by an immunized macaque with the FFL_001 epitope and found that two monoclonal antibodies (17-HD9 and 31-HG7) generated have high affinity for the scaffolding protein of

FFL_001. They also saw that the two antibodies target the same helix-turn-helix compared to mota and pali (two control RSV antibodies). (also, refer to Figure 4)

5. This is an open-ended question up to the students. Possible experiments the students propose would be to take this vaccine to the clinical stage for testing in humans. Likewise, further experiments could be done in animals as well. The epitope-focused approach to designing vaccines would be especially suitable for diseases such as HIV, as HIV is highly antigenically variable. Ideal proteins would thus allow researchers to design a protein that will bind specifically to the epitope of a virus, ensuring that an immune response is elicited.