Introduction:
Modern Protein Design

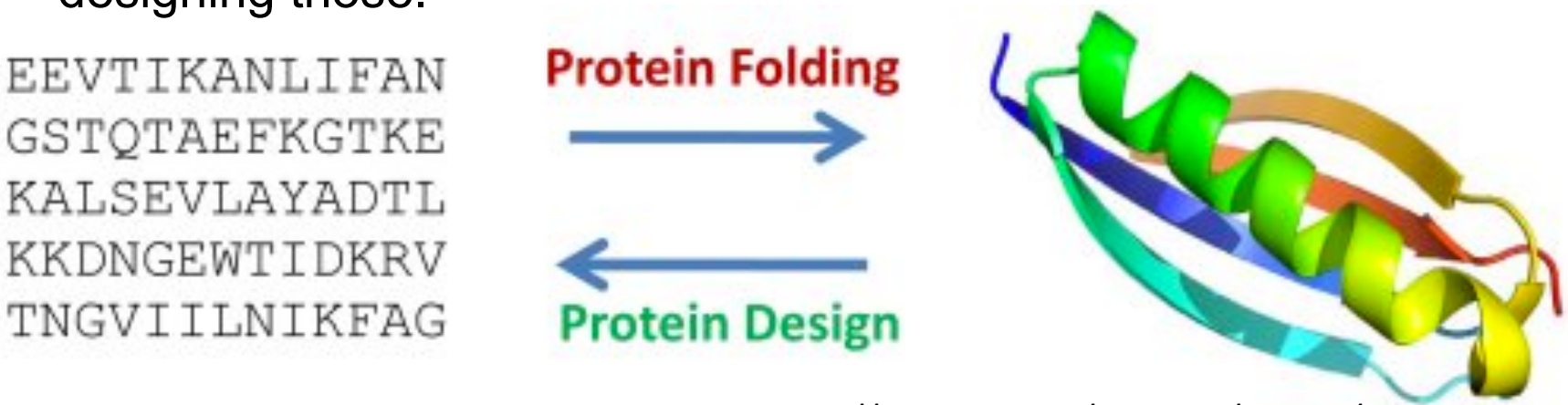
CS/CME/Biophys/BMI 371

Jan. 23, 2018

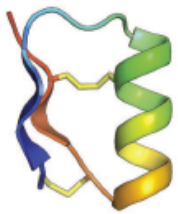
Ron Dror

Protein design: problem definition

- Given the desired three dimensional structure of a protein, design an amino acid sequence that will assume that structure.
  - In many cases, we want the protein to have a particular *function*. We then start by hypothesizing structures that may have that function, and designing those.



http://www.riken.jp/zhangiru/images/sequence_protein.jpg



Bhardwaj et al., *Nature* 2016

Bale et al., *Science* 2016

“Direct” approach doesn’t work

- “Direct” approach to protein design:
  - Given a target structure, search over all possible protein sequences
  - For each protein sequence, predict its structure, and compare to the target structure
  - Choose the best match
- This doesn’t work because:
  - It’s computationally infeasible
  - We’re not that good at predicting protein structure

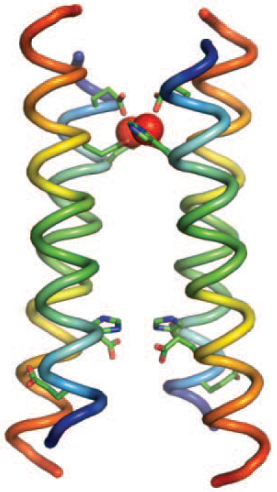
Practical protein design approaches dramatically simplify the problem by making a few assumptions

1. Assume the backbone geometry is fixed
2. Assume each amino acid can only take on a finite number of geometries (*rotamers*)
3. Assume that what we want to do is to minimize the energy of the target structure
   - Simply ignore all the other possible structures that we want to avoid

Each of these assumptions can be relaxed a bit

Papers for Thursday push the boundaries of traditional protein design

- Designing a protein that transports zinc ions through the cell membrane



Joh et al., Science 2014

- Massively parallel design
  - It's now possible to design tens of thousands of proteins, then synthesize and test them in parallel
  - One paper focuses on design of proteins for stability, to gain insight into how they fold
  - Another paper focuses on design of small proteins as drugs—that is proteins designed to bind a target protein

Background material

- Protein design slides from CS/CME/BioE/Biophys/ BMI 279:
  - http://web.stanford.edu/class/cs279/lectures/lecture6.pdf
- "Computer-Based Design of Novel Protein Structures"
  - http://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.35.040405.102046
- "The coming of age of *de novo* protein design"
  - https://cs371.stanford.edu/2018_papers/protein_design/additional/HuangBoyken_DeNovoDesign_Nature2016.pdf

Suggestions for presentations

- No more than 40 minutes total presentation time across all presenters. Even shorter is better.
  - Most presenters spend at least one minute per slide, so you generally want no more than 40 slides total (across all presenters). **Aim for 36 or fewer!**
- You need to explain the main idea of the paper
  - Often the figures in the paper are not sufficient to do this. You may need to make your own.
  - You should read the paper's supplementary information. You'll usually need to read other related papers as well.
- If you can include a demo or video, it's usually worthwhile