(In silico) model building

BCH394P/BCH364C
Systems Biology & Bioinformatics
Caitie McCafferty
clmccafferty@utexas.edu

Today’s agenda

• Sequence to structure
• Types of in silico modeling
  • Comparative modeling (homology modeling)
  • Evolutionary coupling modeling
  • Ab initio modeling with neural networks
Anfinsen’s dogma

In physiological conditions, a globular protein’s native structure is determined by the protein’s amino acid sequence.

Protein folding problem

1. What is the folding code?
2. What is the folding mechanism?
3. Can we predict a native protein structure from its primary, amino acid sequence?
Why might we need to build computational models?

Other considerations: modeling and resolutions based on your need
As with any experiment, keep your intended application in mind...
• do you want to examine a ligand binding site (high resolution)
• or maybe a residue neighborhood (medium resolution)
• or domain boundary definition or even topology classification (low resolution)

CASP competition to evaluate structure prediction
Comparative Modeling: some numbers

Comparative modeling workflow

- Sequence-based methods (BLAST)
- Profile-based methods (HMM)
- Target sequence is aligned to templates using PSSM or HMM approaches
- Copy coordinates from template to assemble the backbone of the model
- Energy functions to simulate loops
- Side-chain modeling using rotamer libraries
- Energy minimization to resolve clashes

I-TASSER for comparative modeling


“ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments”
Comparative modeling limitations

Continued reading on comparative modeling


Evolutionary coupling-based modeling

“The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated.”

- Sewall Wright

Finding evolutionary covariation

Local statistical models i.e. mutual information

- Assume pairs of residues are statistically independent of other pairs of residues
- Issue: cooperative interaction exist and are crucial in folding

Global statistical models i.e. maximum entropy

- Correlated pairs of residues are dependent on each other
- Given all pair correlations, which best explain all the others (going from correlation to causation)

Detecting evolutionary couplings

- Sequence of target protein
- Build multiple sequence alignment for target sequence
- Calculate co-occurrence frequencies for all pairs of columns for all amino acids
- Derive causative correlations (predicted contacts) by using global probability model for sequences (see Fig. 2b)
- Use predicted contacts as residue-residue distance constraints
- Generate approximate three-dimensional coordinates using distance geometry algorithm
- Refine three-dimensional coordinates by molecular dynamics (simulated annealing)
- Rank generated models to select most plausible prediction
Maximum entropy models for selecting global residue contacts


EVcouplings

Welcome to EVolutionary Couplings!
The Evolutionary Couplings server provides functional and structural information about proteins derived from the evolutionary sequence record using methods from statistical physics.

Webserver and resources
Evolutionary couplings can be used to predict many interesting aspects about protein and RNA molecules from sequence alone. Here is what we have worked on so far:

- EVcouplings server
- Modeler
- EVcomplex submission
- 3Dseq

EVcouplings server
Complete evolutionary couplings from sequence alignments will predict 3D structure for your protein of interest. This webserver allows to run homology
- Docking, Evolution, Enrich, and EVcomplex jobs.

Modeler
Precomputed evolutionary couplings and 3D models for thousands of experimentally uncharacterized proteins.

EVcomplex submission
Predict interesting residues in protein complexes from sequence covariation for your complex of interest.

3Dseq
Data for in-vitro experimental evolution.
Evolutionary modeling limitations

Continued reading on EC modeling


How one scientist coped when AI beat him at his life's work

A Harvard biologist on his journey from melancholy to acceptance.

By Sian Samuel | Feb 15, 2016, 4:10pm EST

A protein bound to a receptor is shown in the Molecular Graphics Lab in California. | Ann Johansson/Cordis via Getty Images
Best team in CASP performance over the years

Median Free-Modelling Accuracy

Global distance test: fraction of protein that is correctly predicted


How much better did AF2 do over other methods?

97 total targets

https://moalquraishi.wordpress.com
AlphaFold2 at CASP14

This is exciting but how does it work?!

Input sequence is searched against genetic databases to construct a MSA and structure databases to find template structures. This is then used to build a pair representation and an MSA representation.

But what is a pair representation?!?

- Distogram is used to map 2D pairwise distances
- Distogram are independent of translations and rotations, so no need to align structures (much faster)


This is exciting, but how does the Evoformer work????

Notice the input and output of the Evoformer are the same
In this process the MSA and pair representations are refined through exchanging information with each other

Does anyone remember the triangle inequality theorem?
We need to make sure predicted distances are actually compatible using triangles!

The end result is an MSA and pair representation that could actually be a real model!

This is exciting, but how does the structure module work?!

- End-to-end folding
- Protein backbone is represented as a gas of rigid bodies that move independently of each other
- Iterates through these steps to update backbone and build side chains

Okay we can predict protein structure with high accuracy... now what?
Proteome-wide structure prediction

![Human muscle protein titin predicted by AlphaFold, 34350 residues.](image)

27

AlphaFold Multimer for protein complexes

(a) A2B2G2 heteromer

(b) A3B3 heteromer

(c) Protein-protein complex

(d) A2B2 heteromer

Combining structure prediction with other structure determination methods

**Ab initio** NN modeling limitations

Continued reading on **Ab initio** modeling

Modeling servers

• Comparative modeling
  • https://zhanglab.ccmb.med.umich.edu/I-TASSER/

• Evolutionary modeling
  • https://evcouplings.org/job

• AlphaFold2 modeling
  • https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb

Thank you!

Questions?