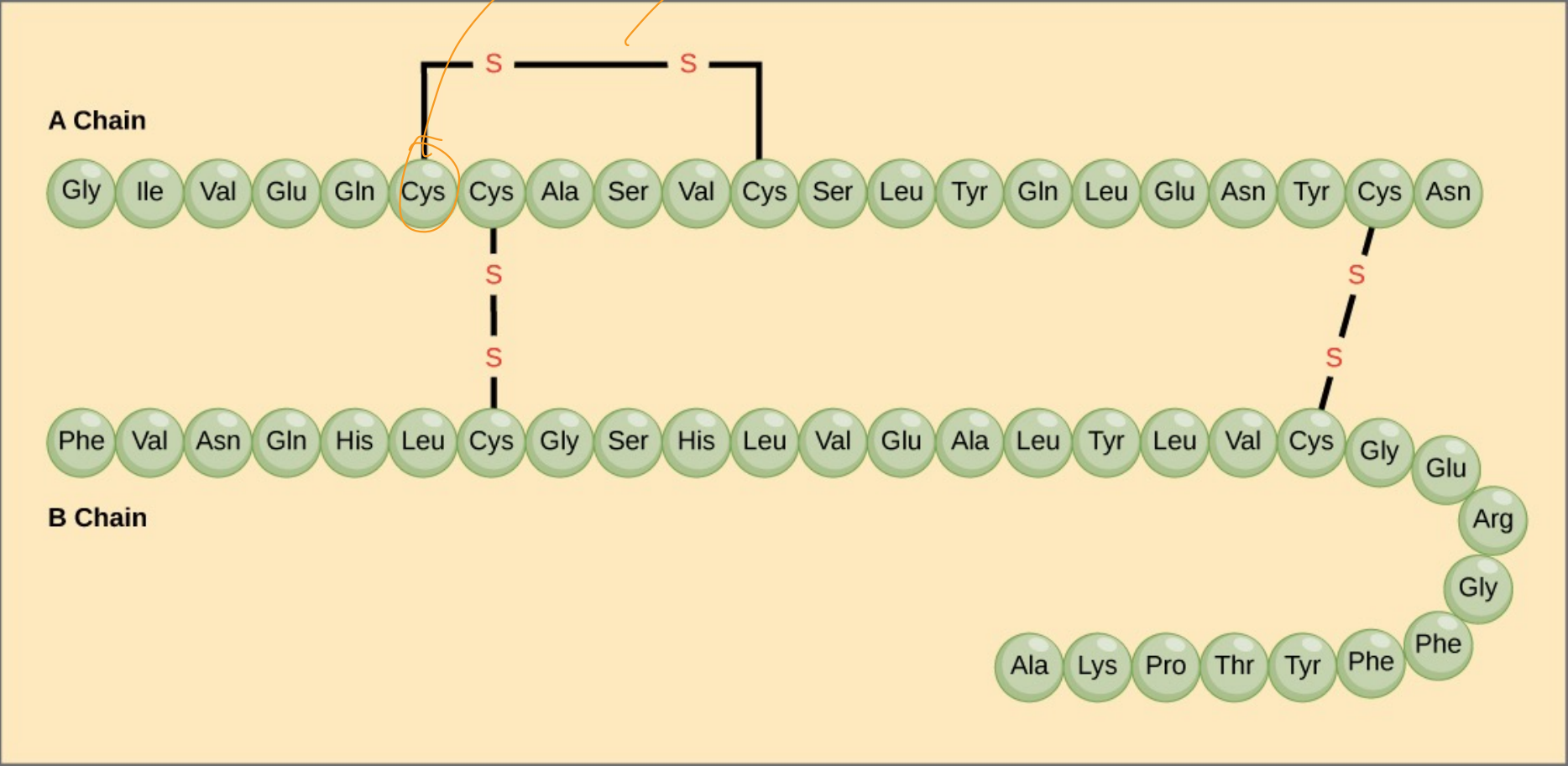


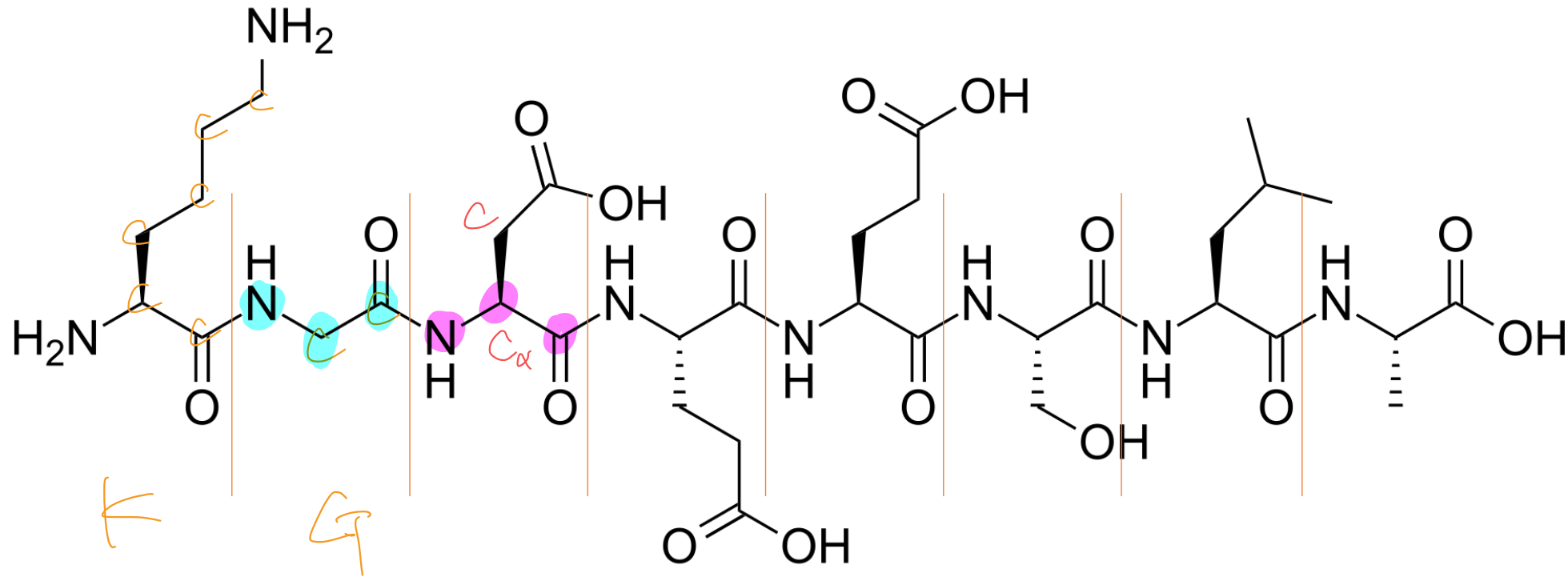
Protein Structures

Primary Structure

disulfide bond

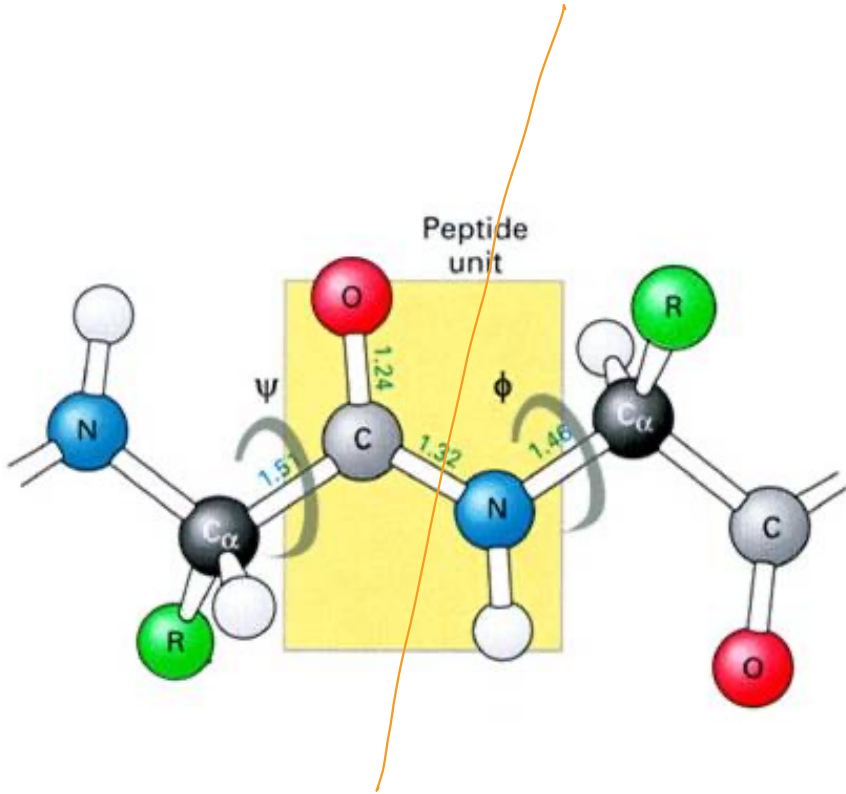
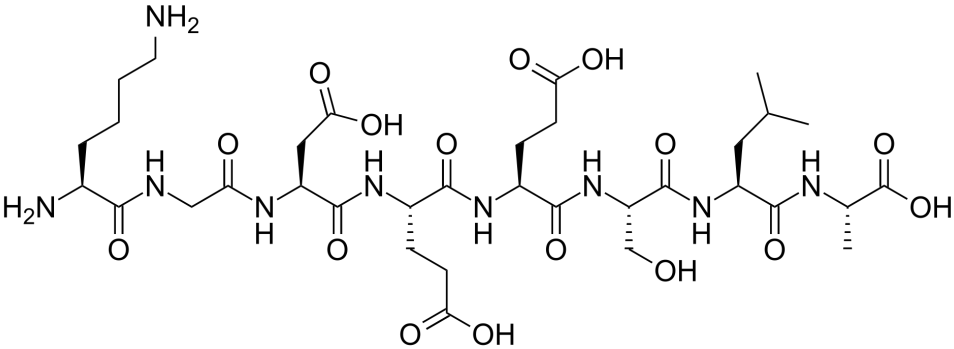


Beefy meaty peptide (delicious peptide)



KGDEESLA

Torsion Angles



Angstrom, Å, 10^{-10} m.

Secondary Structure

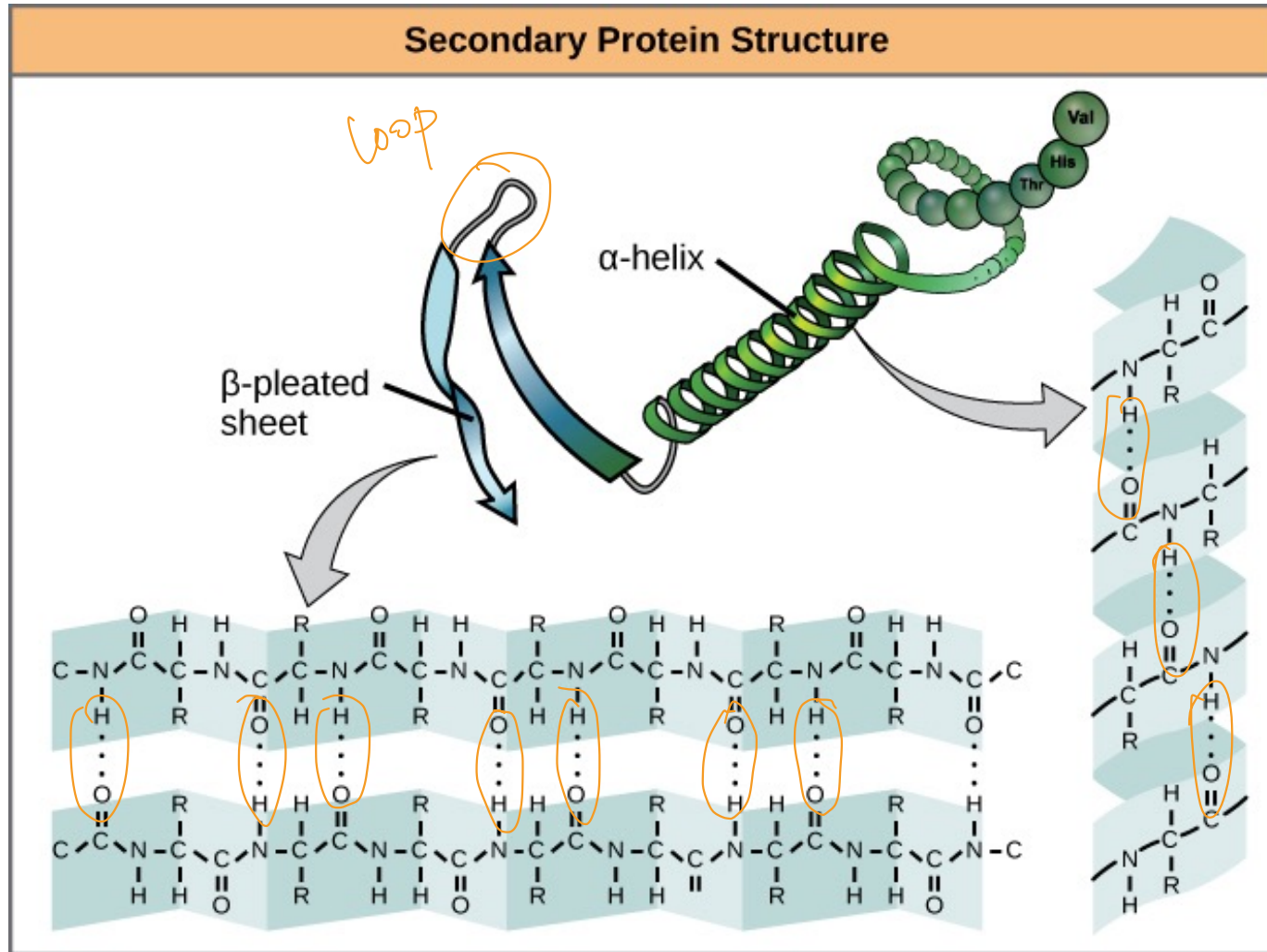
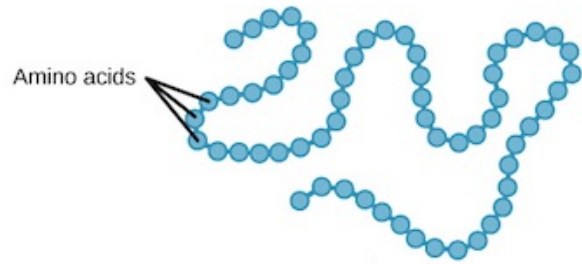
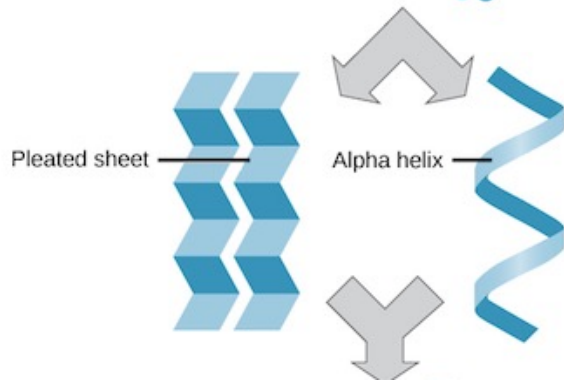


Image credit: OpenStax Biology.

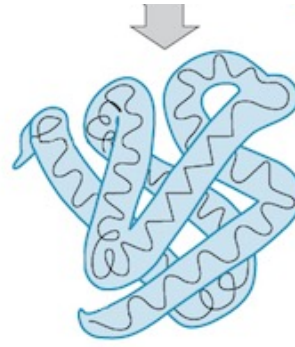
Tertiary and Quaternary Structures



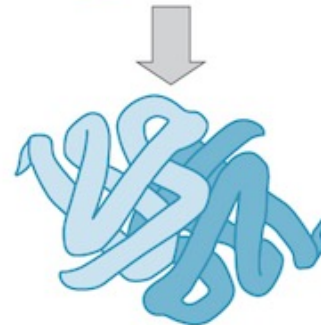
Primary protein structure
sequence of a chain of amino acids



Secondary protein structure
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern



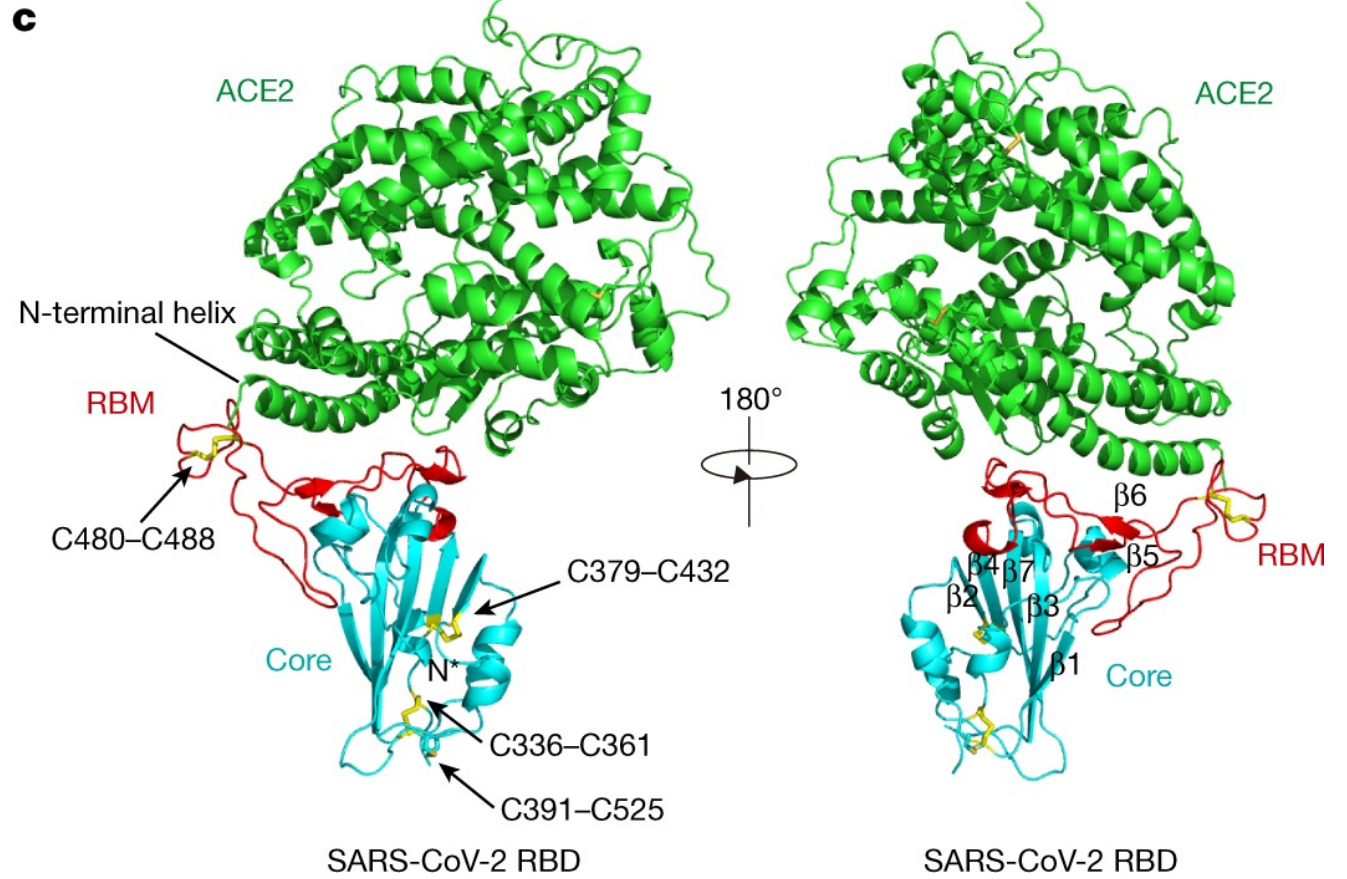
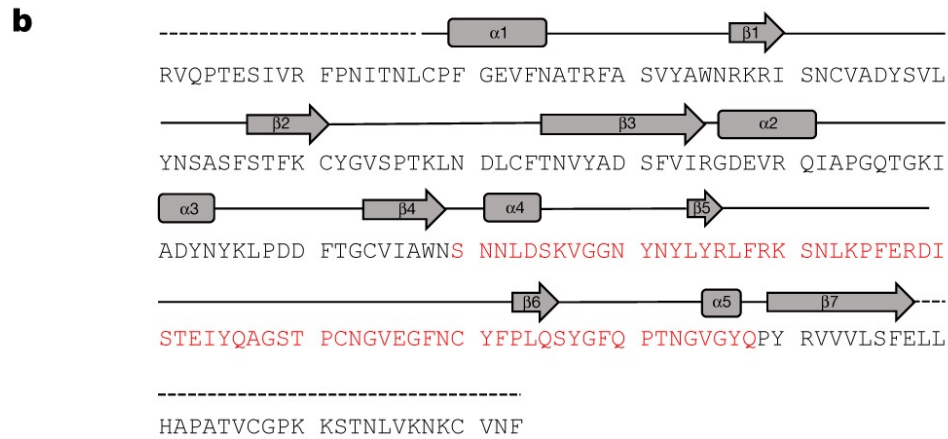
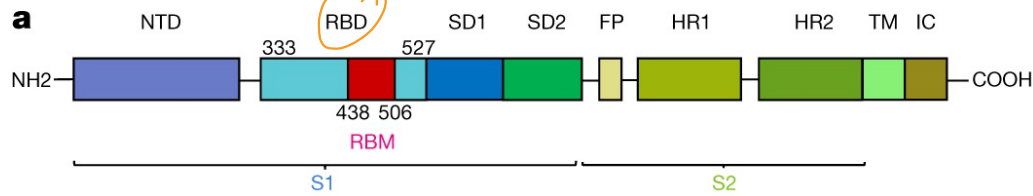
Tertiary protein structure
three-dimensional folding pattern of a protein due to side chain interactions



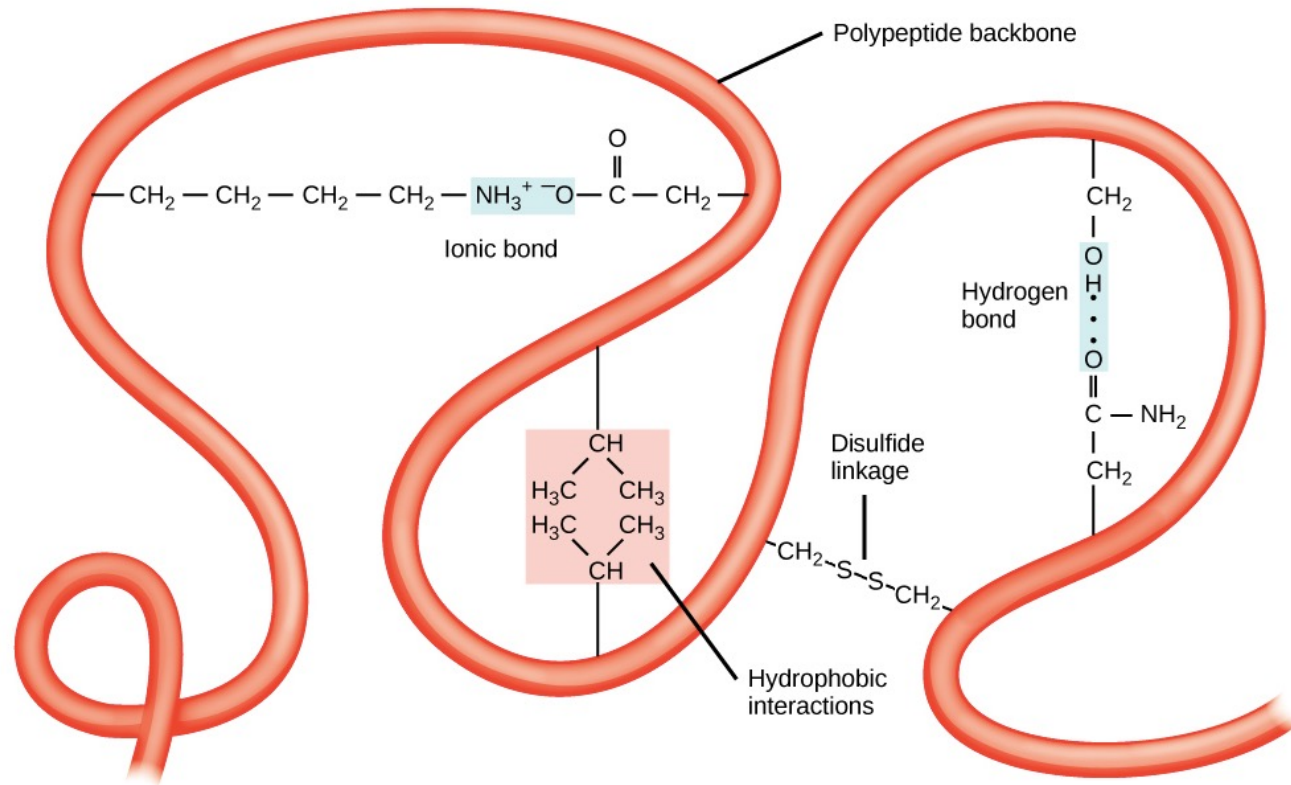
Quaternary protein structure
protein consisting of more than one amino acid chain

SARS-COV2 Spike protein and ACE

RBD Receptor Binding Domain



Free Energy Minimization



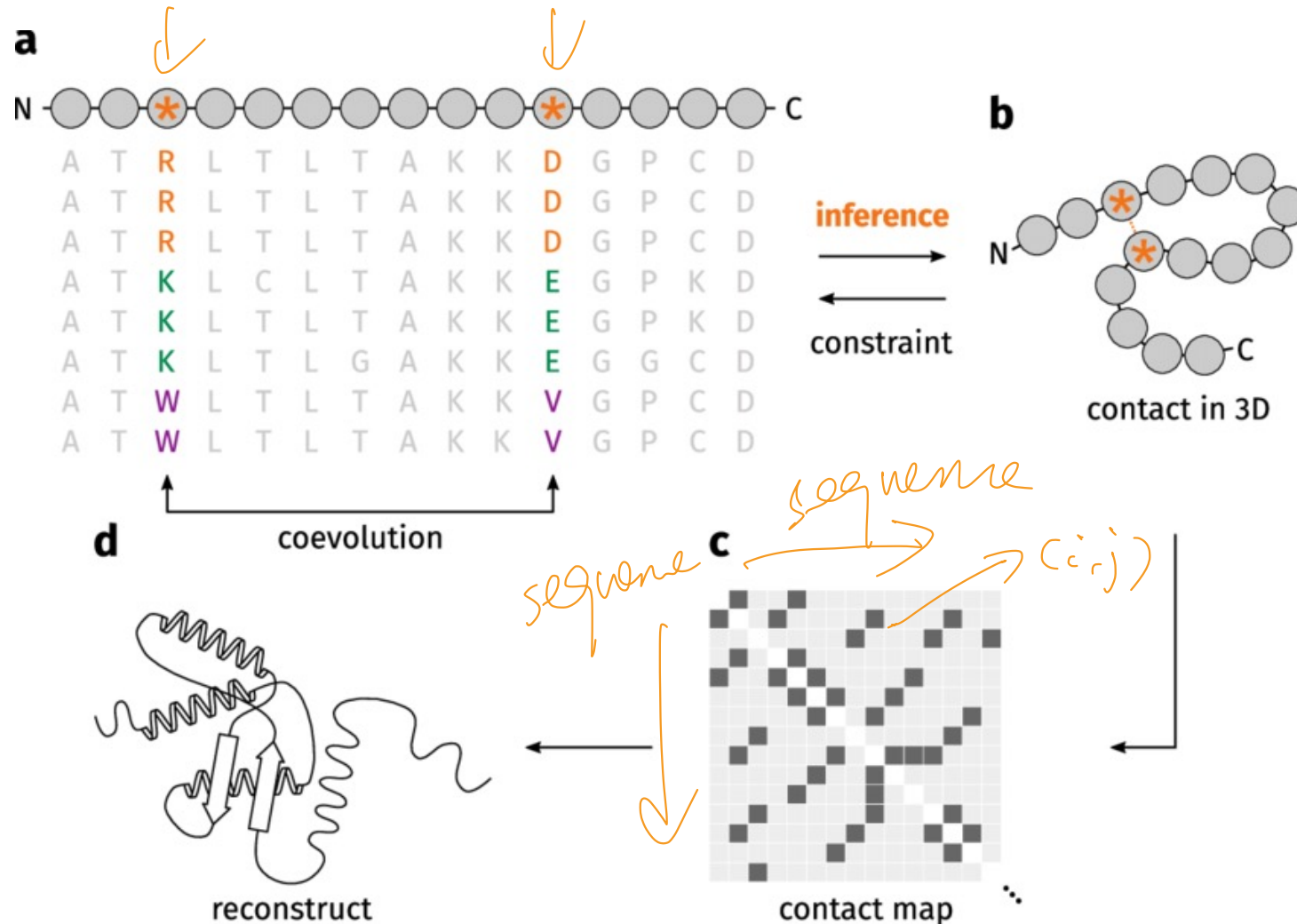
* Free energy function

* Enumerate all torsion

angles to minimize.

- Certain interactions reduce the "free energy".
- It is believed that the natural structure minimizes the free energy.

Additional Information (e.g. Coevolution)



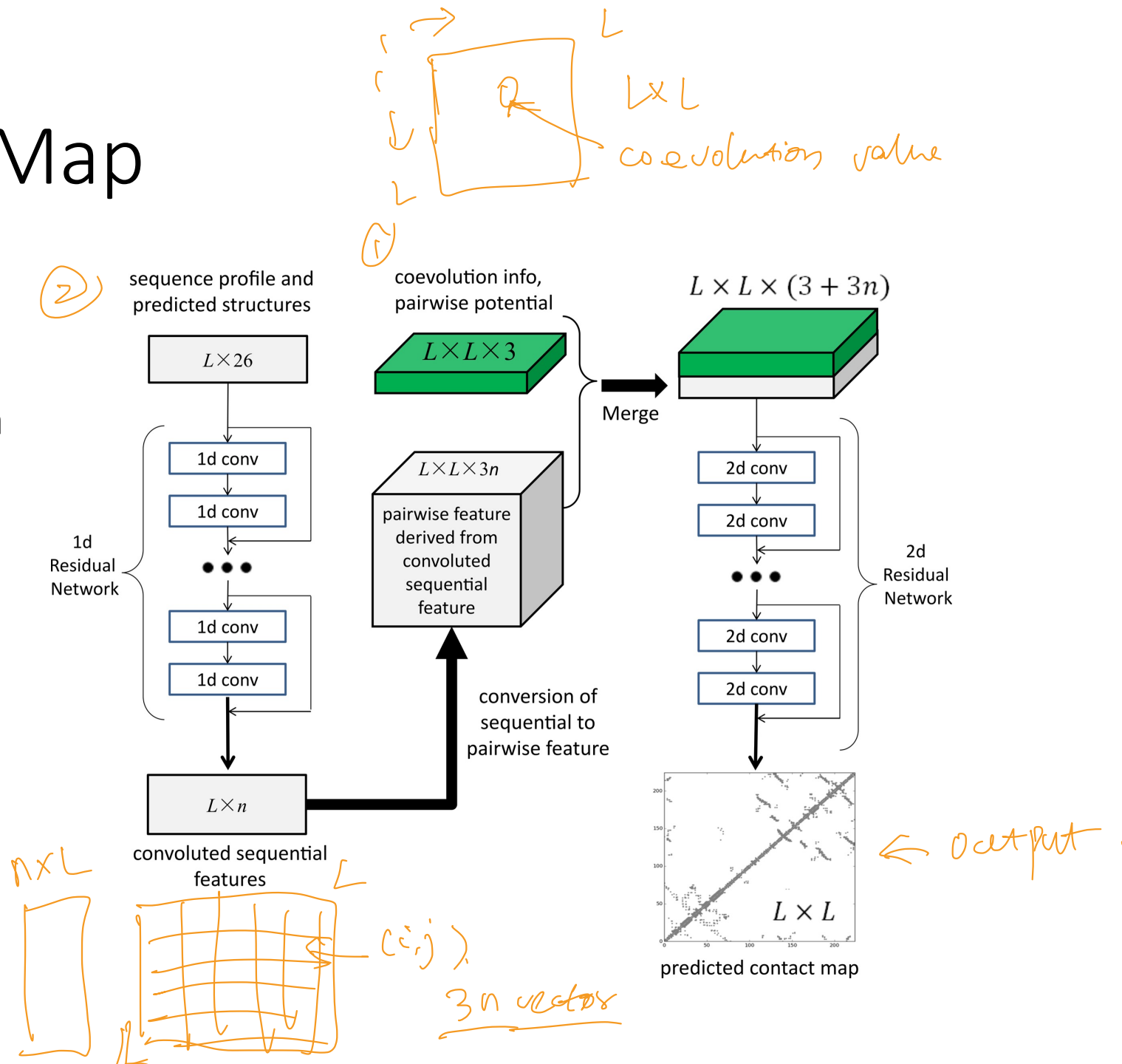
- Additional information can be incorporated in the “energy” function.
- It’s therefore just a scoring function.

Search Approach

- Monte Carlo
 - Sampling of the conformational landscape and minimize energy
- Threading
 - Use existing protein structures as models

Raptor X Contact Map

- Use ResNet to predict the contact map.
- Use some traditional approach to predict structure assisted by contact map.
- Wang S, Sun S, Li Z, Zhang R, Xu J (2017) Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. PLoS Comput Biol 13(1): e1005324. doi:10.1371/journal.pcbi.1005324

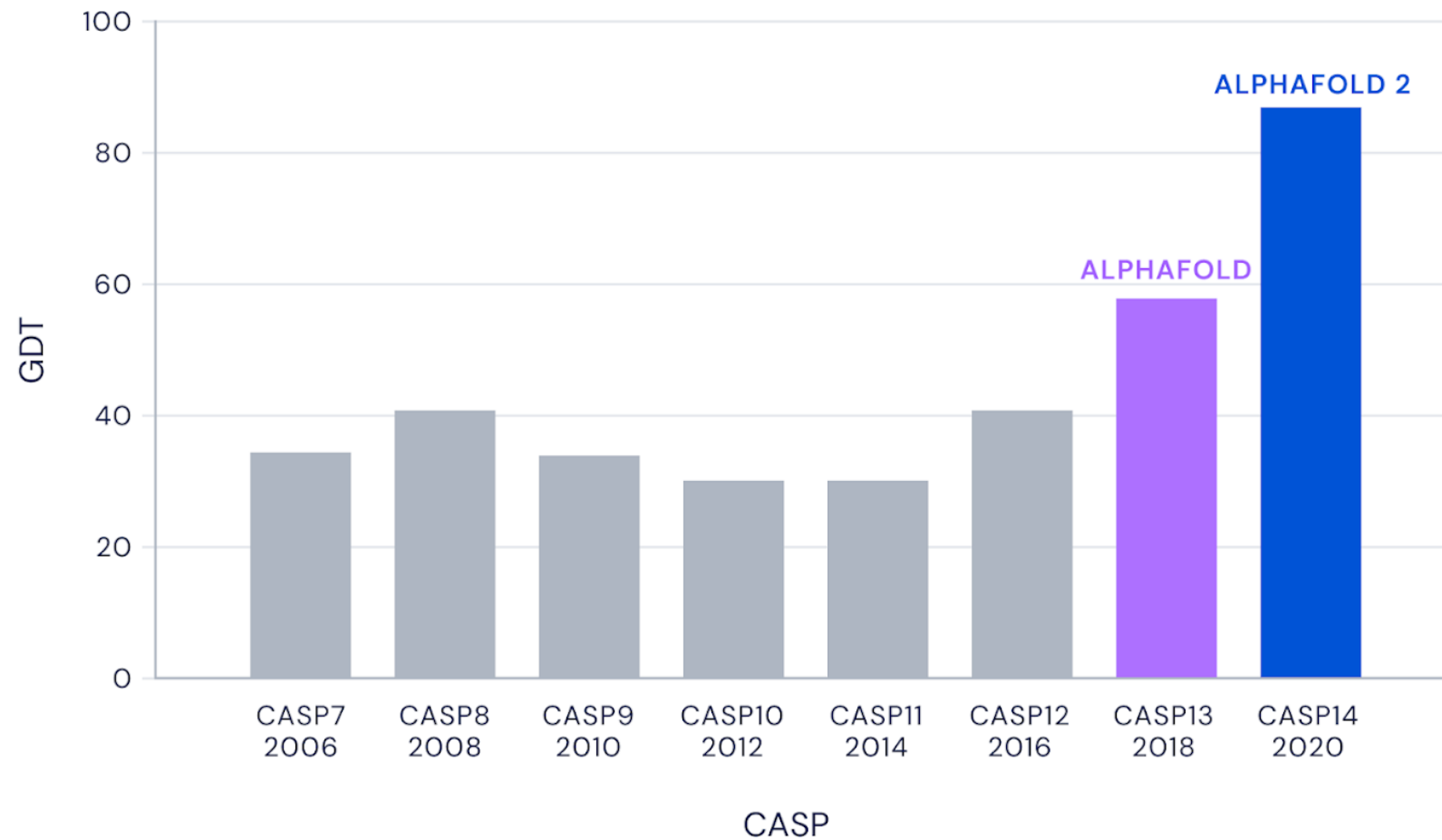


Alpha Fold

- AlphaFold
 - Takes Jinbo Xu's approach + a gradient descent search strategy to minimize the energy.
- AlphaFold2
 - Predict the contact map (2D representation) with neural network.
 - Followed by torsion angle and structure prediction with neural network.
 - New network structures.

CASP: Critical Assessment of Protein Structure Prediction

Median Free-Modelling Accuracy

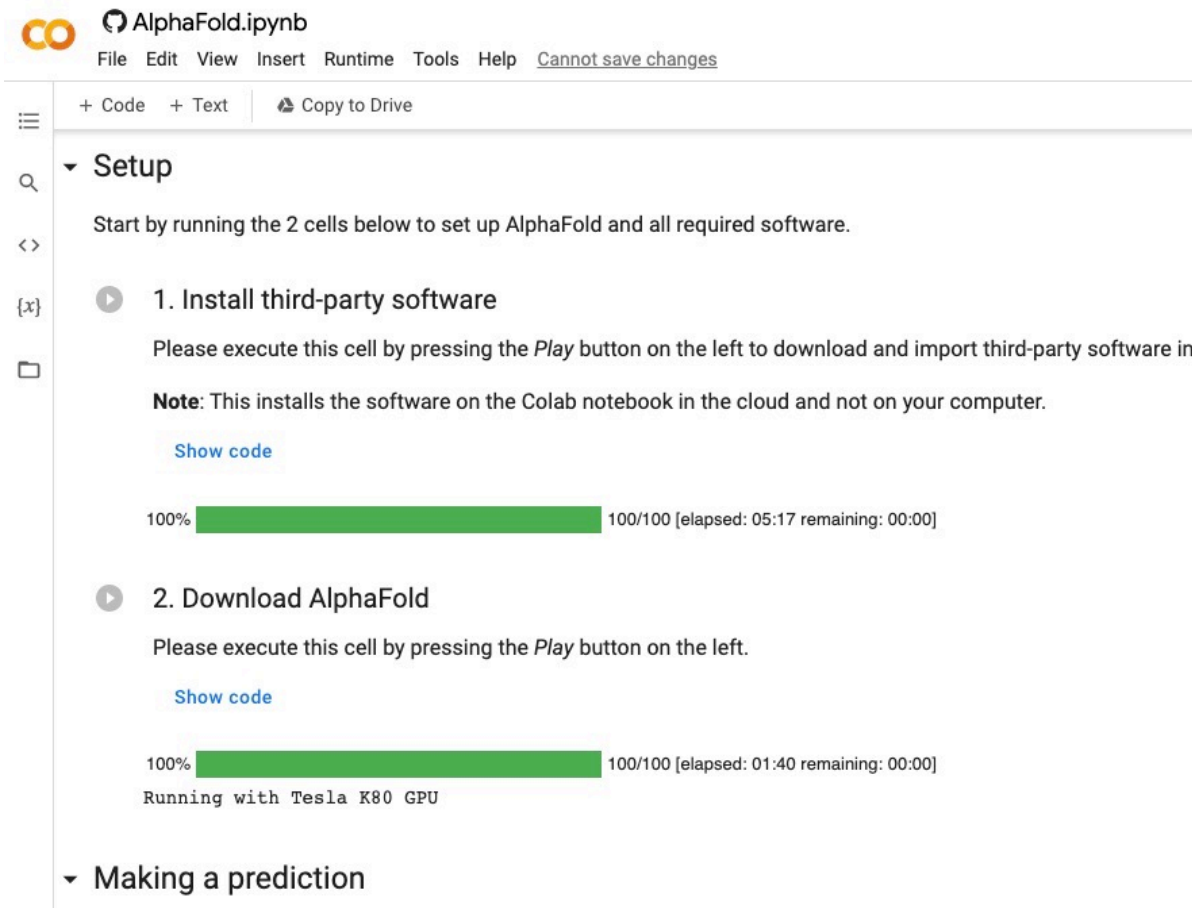


Structure Prediction Software

- Rosetta: David Baker, University of Washington
- RaptorX: Jinbo Xu, Toyota Technological Institute at Chicago
- Alpha Fold: DeepMind
- Many others..



Use Alpha Fold online





AlphaFold.ipynb
File Edit View Insert Runtime Tools Help [Cannot save changes](#)

+ Code + Text Copy to Drive

Setup

Start by running the 2 cells below to set up AlphaFold and all required software.

- 1. Install third-party software**
Please execute this cell by pressing the *Play* button on the left to download and import third-party software in
Note: This installs the software on the Colab notebook in the cloud and not on your computer.
[Show code](#)
100%  100/100 [elapsed: 05:17 remaining: 00:00]
- 2. Download AlphaFold**
Please execute this cell by pressing the *Play* button on the left.
[Show code](#)
100%  100/100 [elapsed: 01:40 remaining: 00:00]
Running with Tesla K80 GPU

Making a prediction

<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>

Use Alpha Fold Online

3. Enter the amino acid sequence(s) to fold

Enter the amino acid sequence(s) to fold:

- If you enter only a single sequence, the monomer model will be used.
- If you enter multiple sequences, the multimer model will be used.

sequence_1: "MAAHKGAEHKAAAEHHEQAAKHHHAAAEHHEKGEHEQAHHADTAYAHKHAEEHAAQAAKHDAEHHAPKPH"

sequence_2: "Insert text here"

sequence_3: "Insert text here"

sequence_4: "Insert text here"

sequence_5: "Insert text here"

sequence_6: "Insert text here"

sequence_7: "Insert text here"

sequence_8: "Insert text here"

[Show code](#)

Using the single-chain model.

4. Search against genetic databases

Once this cell has been executed, you will see statistics about the multiple sequence alignment (MSA) that will be used by AlphaF

[Show code](#)

Getting MSA for sequence 1

Searching mgnify: 100%  147/147 [elapsed: 29:00 remaining: 00:00]

58 unique sequences found in uniref90 for sequence 1
110 unique sequences found in smallbfd for sequence 1
9 unique sequences found in mgnify for sequence 1

5. Run AlphaFold and download prediction

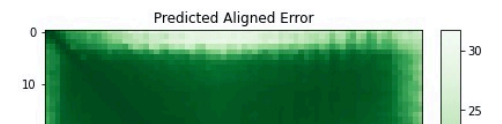
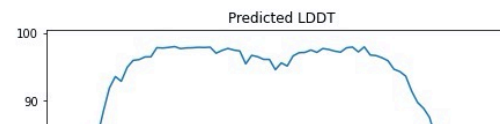
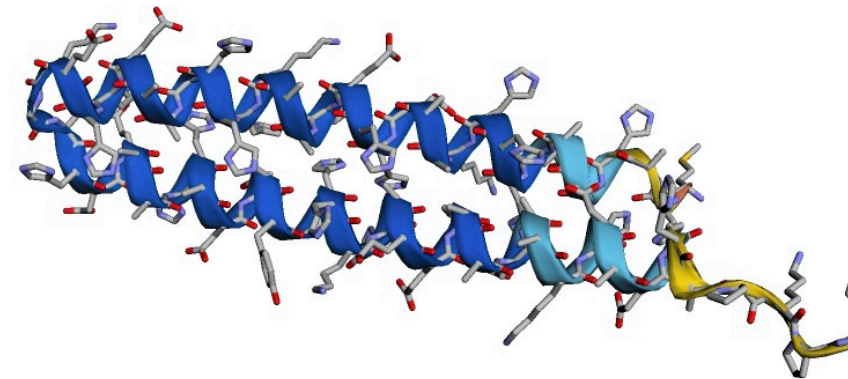
Once this cell has been executed, a zip-archive with the obtained prediction will be automatically downloaded to your computer.

In case you are having issues with the relaxation stage, you can disable it below. Warning: This means that the prediction might have distracting sr

run_relax:

[Show code](#)

AMBER relaxation: 100%  7/7 [elapsed: 31:30 remaining: 00:00]



Protein Structure Resources

- Structure databases
 - PDB (Protein Data Bank)
 - <https://www.rcsb.org/>
 - Alpha Fold Predicted
 - <https://alphafold.ebi.ac.uk/>
- Structure visualization software
 - Pymol

Protein Structure Resources

- Structure databases
 - PDB (Protein Data Bank)
 - <https://www.rcsb.org/>
 - Alpha Fold Predicted
 - <https://alphafold.ebi.ac.uk/>
- Structure visualization software
 - Pymol