

Lecture 9: 10/15/2026

Multiple Sequence Alignment in AlphaFold 2

Why Multiple Sequence Alignment?

- Previously we only aligned 2 sequences
 - ↳ Good, but one comparison provides limited info (multiple alignments are more clear to see)

Conceptual Example with 3 sequences:

We need to make a 3D DP table $\Rightarrow$ More difficult because time and space complexity increase exponentially, and recurrence relation will need to check more adjacent cells

Naive MSA is NOT scalable:

$$\text{Number of cells} = (n+1)^k$$

$$\text{Number of operations} = (2^k - 1) \text{ adjacent cells}$$

MSA heuristics

- Naive implementations are impractical
- Progressive alignment is the viable alternative

MSA Application: AlphaFold 2

- Deep Learning Model that predicts protein folding from the protein's AA sequence

Historical approaches to protein folding predictions:

- | | |
|-------------------------|----------------------|
| - X-ray crystallography | - Lattice models |
| - Protein engineering | - Molecular Dynamics |
| - NMR | - Monte Carlo |

Today: MSA

MSA in protein folding prediction

- MSA finds conserved regions of AA's between similar proteins of other species
- Principle of coevolution: If 2 AA's sit close to each other in the folded 3D protein, a mutation in one often forces a matching mutation in the other to maintain structural stability

MSA in AlphaFold 2

- A profile Hidden Markov Model is trained to output sequences from a protein family using hundreds of sequences of proteins.

AlphaFold2 process:

- 1) queries a genetic database across species to generate a reference MSA
- 2) the reference MSA is used to train a profile HMM