

Multiple sequence alignment in AlphaFold2

Multiple sequence alignment

- Allows for alignment of more than 2 sequences at once
 - 3 sequences → 3D dynamic programming table, not feasible (exponential)
 - # cells in DP table with k strings, each of length n → $(n+1)^k$
 - # operations per cell: $2^k - 1$
 - Total runtime: $O((2n)^k)$
- Heuristic example: progressive alignment = align all pairs, measure how closely related each pair is

Application of MSA: AlphaFold2

Deep learning model, predict protein's 3D conformation from primary structure

- **Levinthal's paradox:** if a protein were to find its native 3D shape by randomly testing every possible conformation, it would take longer than the age of the universe → there needs to be other mechanisms for protein to find its native conformation
- **Lattice model:** grids of amino acids in 2D space, group together based on properties → find optimal tiling of amino acids within grid
- **Monte Carlo methods:** propose random change in conformation, determine probability of occurrence

Why MSA in protein prediction?

- AlphaFold2 uses MSA to align an input protein sequence with other proteins sequences → find **conserved regions** of amino acids between similar proteins of other species
 - Principle of **coevolution**: if two amino acids sit close to each other in the folded 3D protein, a mutation in one forces a matching (compensatory) mutation in the other to maintain structural integrity (want stabilizing interactions)
- **Profile HMM (Hidden Markov Model)** = sequence generating model: trained to output sequences from a protein family using hundreds of sequences of proteins → see likelihood of each output coming from the model
 - 3 states: match or insertion emit amino acid, deletion emits a gap
 - Each arrow = transition
 - Number = probability of transitioning from one state to another
 - At each state, output an amino acid and its probability
 - Training: feed sequences from an MSA
 - Given new sequence → find most likely series of states that generated it (order of states = alignment); multiply probabilities
- AlphaFold2: use reference MSA generated to train profile HMM