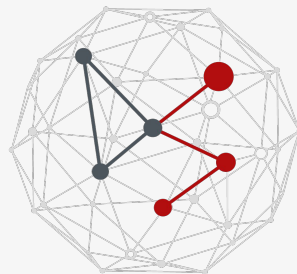


1222-2022
800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

  DIPARTIMENTO
MATEMATICA



DATA SCIENCE
UNIVERSITY OF PADOVA

AlphaFold-2

Master of Science in Data Science

Damiano Piovesan



References

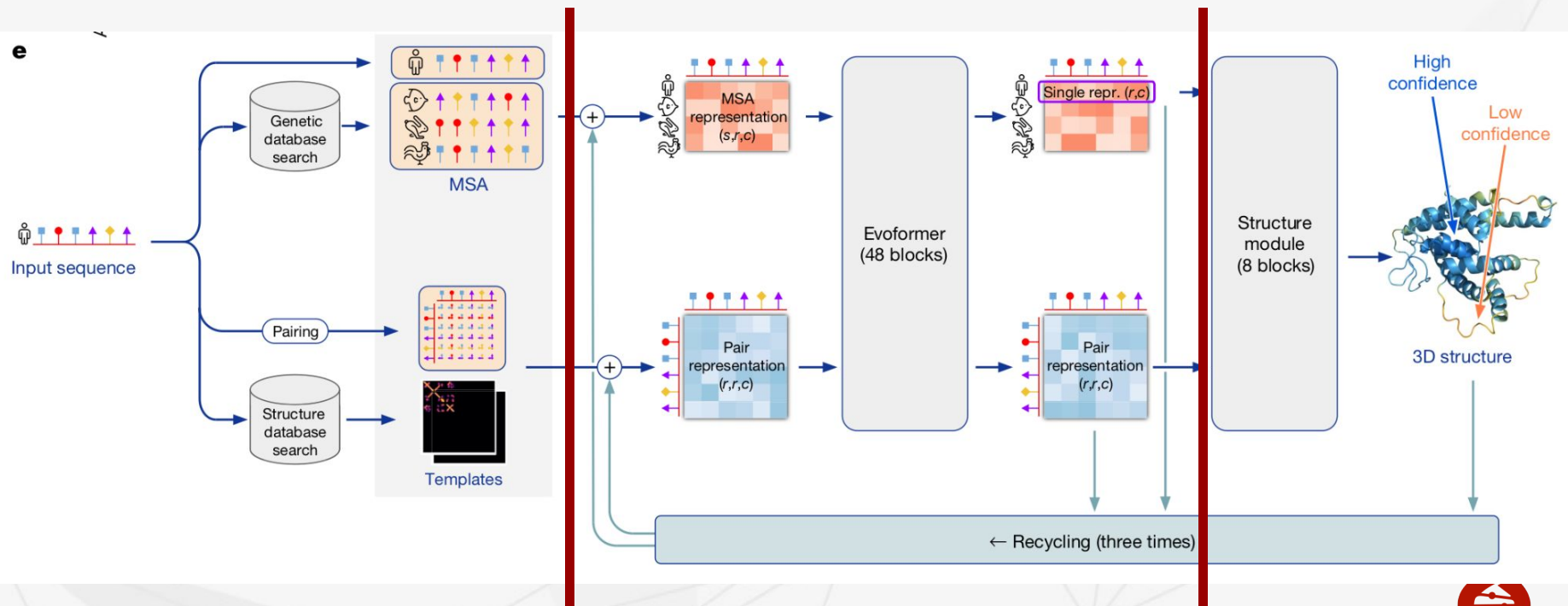
- BLOG - Oxford Protein Informatics Group

<https://www.blopig.com/blog/2021/07/alphafold-2-is-here-whats-behind-the-structure-prediction-miracle/>

- ColabFold
- Nature paper

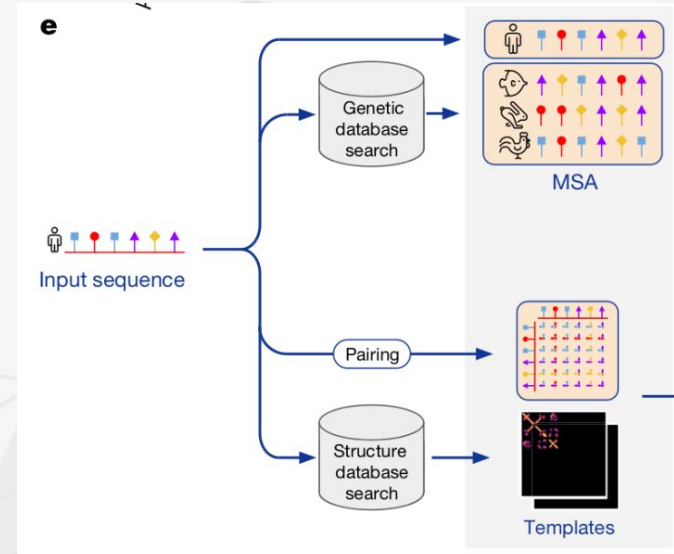


End-to-end network



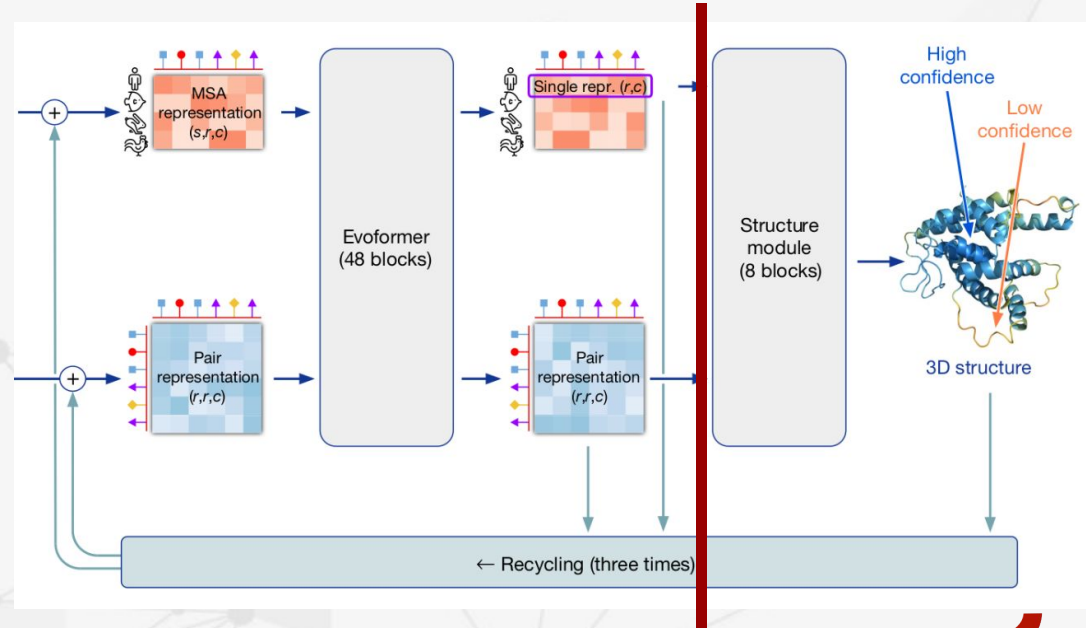
Input

- MSA (Multiple-sequence-alignments)
 - Identify parts that are more likely to mutate and
 - Identify correlations (rows and columns)
- Pair-representation (contacts)
 - Templates (Proteins with a similar structure)



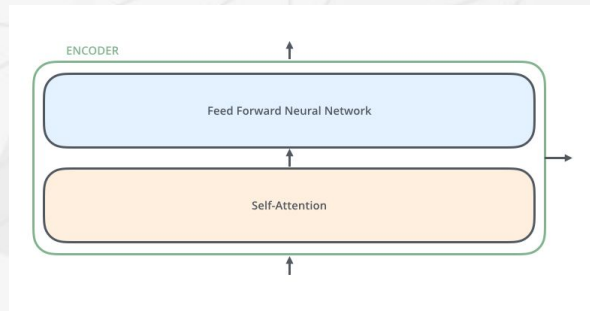
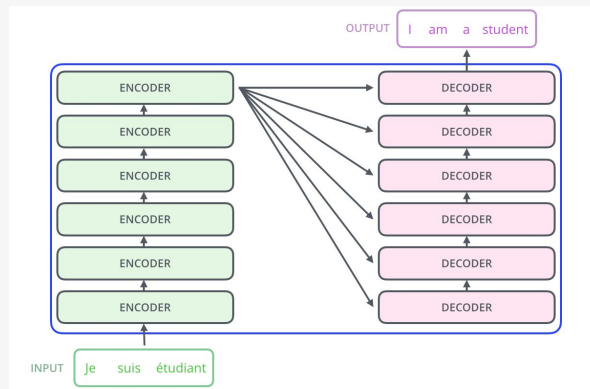
Transformer and structure module

- Evoformer
 - Refine MSA and Pair-representation models by exchanging information between them
- Structure module
 - Generate a final structure (coordinates) in a single step
- Iteration (MSA + Pair + Pred. structure)



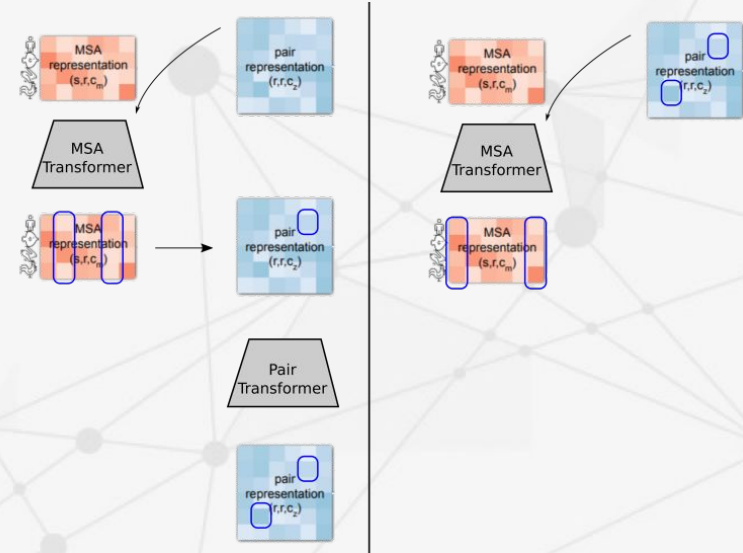
Evoformer

- Variation of the “transformer” deep learning architecture
- Invented by Google Brain
- Based on “Attention” → Identify which parts of the input are more important
- Attention matrix → quadratic memory cost



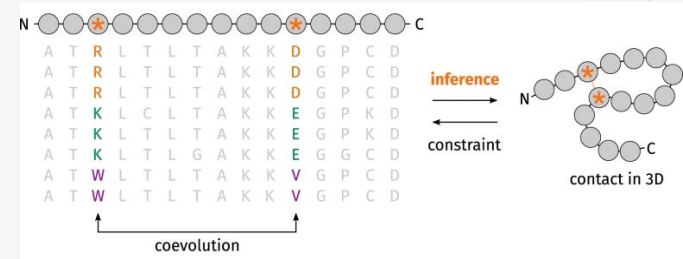
Evoformer

- Embeddings
 - Transform discrete variables into a continuous space
 - “Vanilla dense neural networks”
- Cycle between structural hypothesis and the MSA model
 - Residues A and B are correlated according to the MSA
 - Looking at the structure model it seems that also C and D are close
 - Search for correlation of C and D in the MSA
 - Repeat...



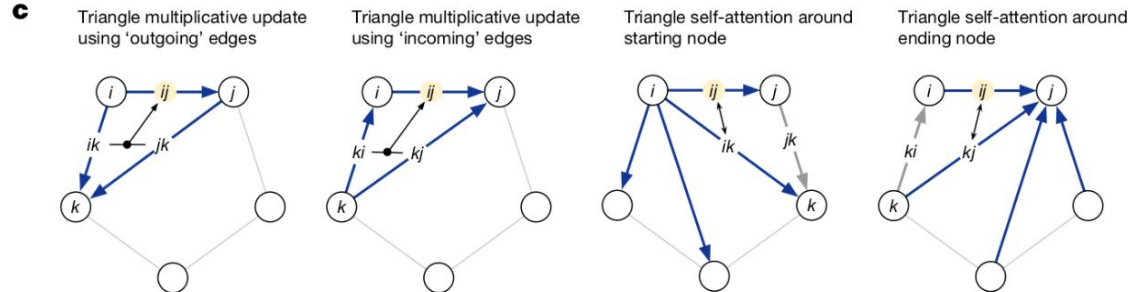
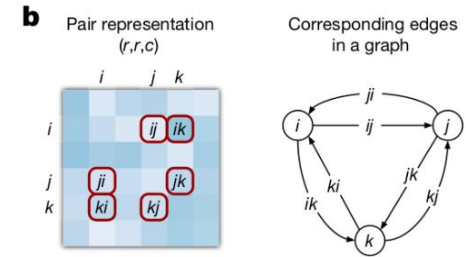
MSA transformer

- “Column-wise”
 - Attention in the vertical direction
 - Which sequences are more informative
- “Row-wise”
 - Attention in the horizontal direction
 - Which pairs of amino acids are more related
 - Incorporates information from the “pair representation”



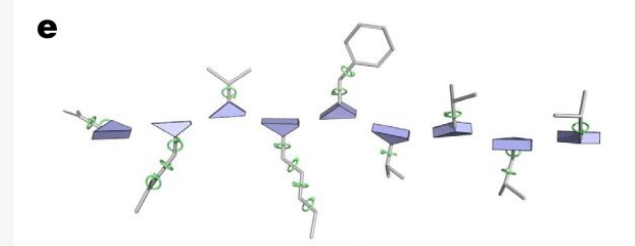
Pair transformer

- Attention is arranged in terms of triangles of residues
- Enforce the triangle inequality
- Triangle multiplicative update, infer the third “missing edge”
- Self-attention



Structure module

- Consider the protein as a “residue gas”, amino acids are independent to each other
- Amino acids modelled as triangles
- Produce a set of “affine matrices” (4x4) that displace and rotate the residues in space
- Prioritize the orientation that optimize side-chains placement
- Peptide bonds geometry is completely unconstrained
- Post-prediction relaxation. Gradient descent in Amber force field



ColabFold

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

