Exploring AlphaFold for Protein Structure Prediction

Understanding metrics, applications, and comparisons with other tools.

Introduction to Protein Structural Analysis

- Proteins are the functional building blocks of life, and their functions are determined by their 3D structures.

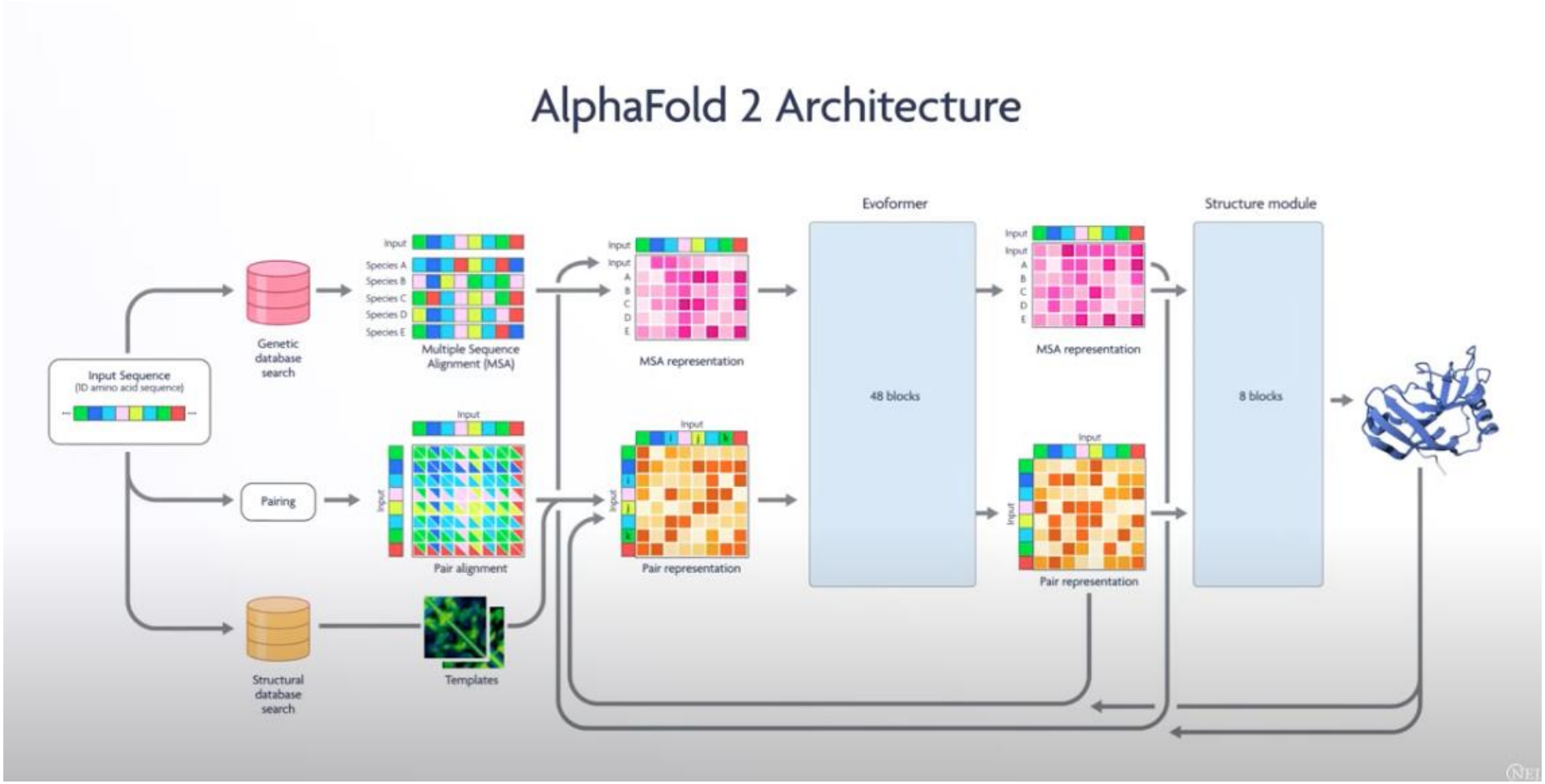
- Understanding protein structures aids in:
- - Drug design.
- - Functional studies.
- - Understanding molecular mechanisms.

- Key structural levels:
- - Primary: Amino acid sequence.
- - Secondary: Alpha helices and beta sheets.
- - Tertiary: 3D arrangement of residues.
- - Quaternary: Multi-chain complexes.

Tools for Protein Structure Prediction

- Before AlphaFold, popular tools included:
- - Homology modeling (e.g., SWISS-MODEL): Relies on template structures.
- - Threading methods (e.g., I-TASSER): Combines template alignment and fragment assembly.
- - Ab initio methods (e.g., Rosetta): Predicts structure from sequence using energy minimization.

- Challenges with these tools:
- - Dependency on templates.
- - Limited accuracy for novel folds.
- - High computational costs.



December 2024

AlphaFold 3 upgrade enables the prediction of other types of biomolecular systems

What Makes AlphaFold Different?

- Key innovations of AlphaFold:
- - End-to-end deep learning model.
- - Predicts residue distances and orientations directly.
- - Excels in novel fold prediction without relying on templates.

- Advantages:
- - High accuracy rivaling experimental methods.
- - Confidence metrics (pLDDT, PAE) for result interpretation.
- - Multimer mode for predicting protein-protein interactions.

Understanding AlphaFold Metrics

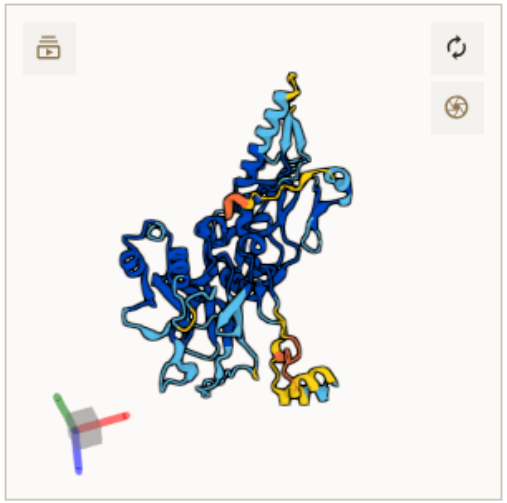
- AlphaFold Metrics:
- - **pLDDT (Predicted Local Distance Difference Test)**: Residue-level confidence (0–100 scale).
- - **PAE (Predicted Aligned Error)**: Quantifies the error between residues in multi-domain proteins or complexes.
- - **pTM (Predicted TM-score)**: Global confidence in the fold's accuracy (0–1 scale).
- - **ipTM (Inter-chain pTM)**: Confidence for inter-chain interactions in multimers.

- Interpretation:
- - High pLDDT (>70): Reliable regions.
- - High ipTM (>0.8): Strong inter-chain confidence.

Which structure has the best pLDDT score?

C-type lectin-like domain family (AF-Q8IZS7-F1)

May function in mediating immune cell-cell interactions



Apical membrane antigen 1 (AF-Q3S2X4-F1)

Involved in parasite invasion of erythrocytes

Chloramphenicol acetyltransferase (AF-P36883-F1)

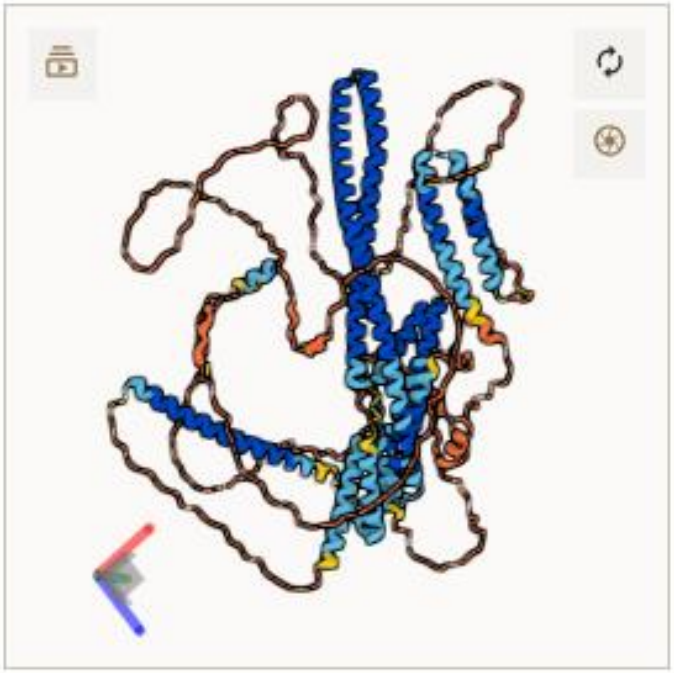
Responsible for resistance to the antibiotic chloramphenicol

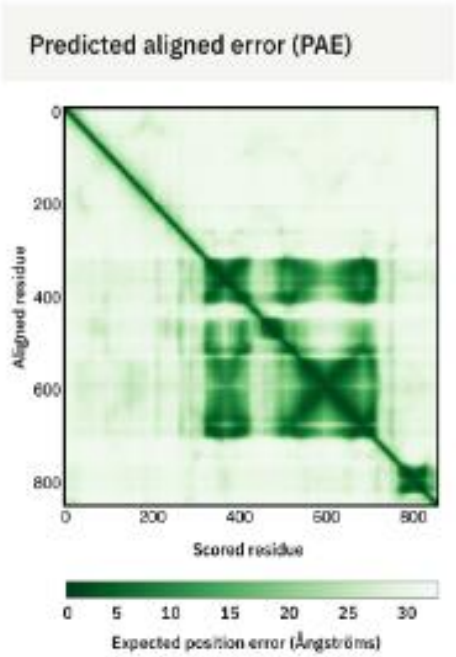
Very high (pLDDT > 90) High (90 > pLDDT > 70) Low (70 > pLDDT > 50) Very low (pLDDT < 50)

Understanding AlphaFold Metrics

- AlphaFold Metrics:
- - **pLDDT (Predicted Local Distance Difference Test)**: Residue-level confidence (0–100 scale).
- - **PAE (Predicted Aligned Error)**: Quantifies the error between residues in multi-domain proteins or complexes.
- - **pTM (Predicted TM-score)**: Global confidence in the fold's accuracy (0–1 scale).
- - **ipTM (Inter-chain pTM)**: Confidence for inter-chain interactions in multimers.

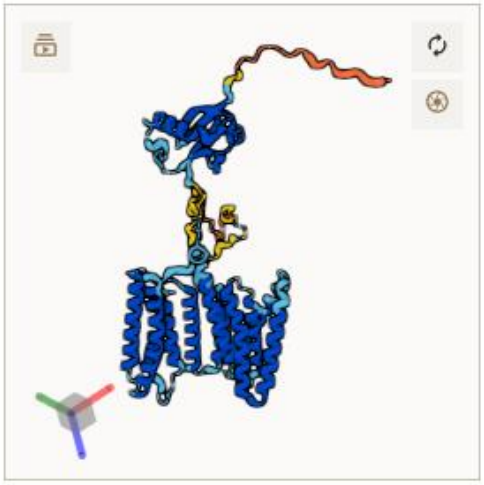
- Interpretation:
- - High pLDDT (>70): Reliable regions.
- - High ipTM (>0.8): Strong inter-chain confidence.

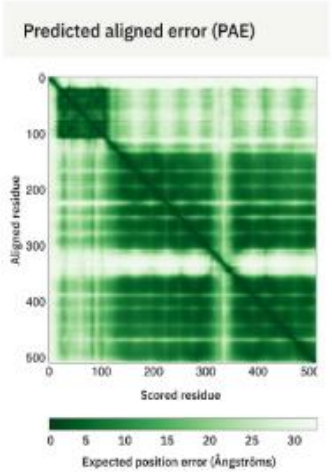




Disrupted in schizophrenia 1 protein (AF-Q9NRI5-F1)

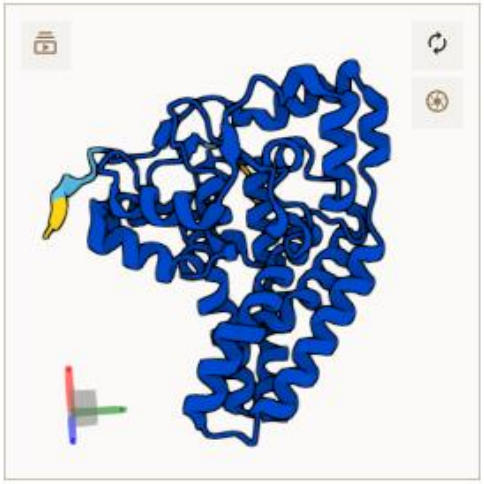
Involved in the regulation of multiple aspects of embryonic and adult neurogenesis

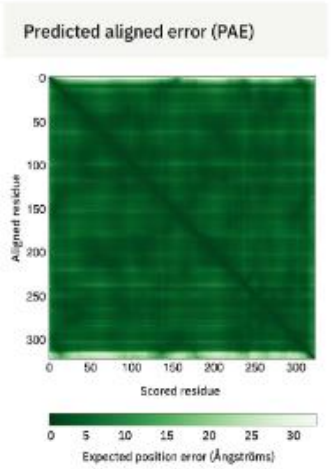




ESX-1 secretion system protein EccD1 (AF-P9WNQ7-F1)

Component of the ESX-1 type VII secretion system, which actively transports various virulence factors into host cells during the course of infection





Sphingolipid delta(4)-desaturase DES1 (AF-O15121-F1)

Has been linked to leukodystrophy. However, no structural or model information for this protein is available.

Comparison Table: AlphaFold vs Other Tools

Feature	AlphaFold	I-TASSER	Rosetta
Template Dependency	Minimal or None	High	Moderate
Novel Fold Prediction	Excellent	Moderate	Limited
Metrics	pLDDT, PAE, pTM, ipTM	C-score, TM-score	Energy Score
Protein-Protein Interactions	Yes (AlphaFold-Multimer)	Limited	Limited
Computational Cost	Efficient	Moderate	High

Applications and Validation

- Applications:
- - Drug discovery.
- - Understanding molecular interactions.
- - Designing protein complexes.

- Validation methods:
- - Cryo-EM or X-ray Crystallography for high-resolution validation.
- - Size-Exclusion Chromatography for oligomerization studies.