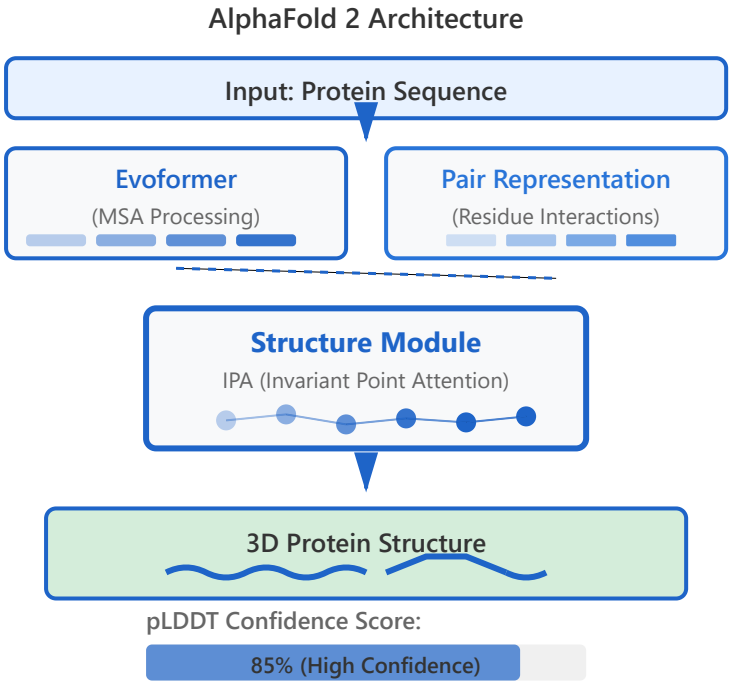
AlphaFold Revolution



Architecture innovations
Evoformer + Structure module

MSA processing
Evolutionary information extraction

Structure module
IPA: SE(3)-equivariant attention

Confidence metrics
pLDDT per-residue scores

Database impact
200M+ structures predicted

1. Architecture Innovations

AlphaFold 2 introduced a revolutionary neural network architecture that combines the **Evoformer** module for processing evolutionary information with the **Structure**

Module for generating 3D coordinates. This two-stage pipeline represents a paradigm shift in protein structure prediction.

Evoformer Block

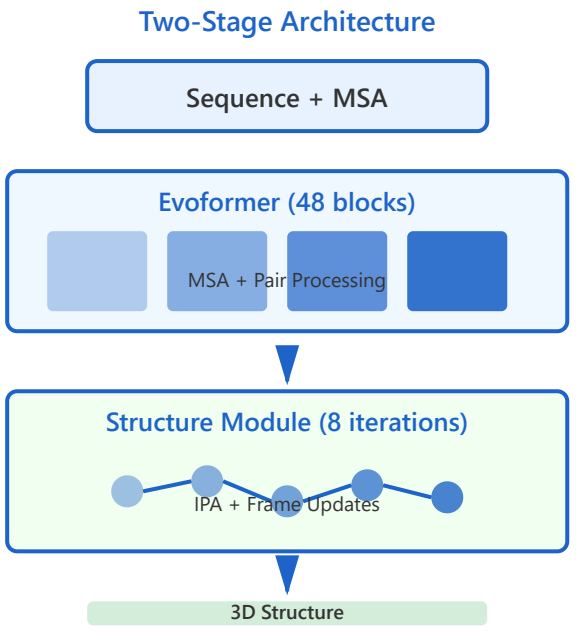
Processes multiple sequence alignments (MSA) and pairwise residue representations through 48 stacked blocks. Each block contains row/column attention mechanisms and transition layers that refine evolutionary patterns.

End-to-End Differentiable

Unlike traditional template-based methods, AlphaFold 2 is trained end-to-end, allowing gradient flow from 3D structure prediction back to sequence processing, enabling sophisticated feature learning.

Iterative Refinement

The structure module operates iteratively, refining the predicted structure through multiple cycles while maintaining geometric consistency through SE(3)-equivariant operations.



2. MSA Processing - Evolutionary Information

The **Multiple Sequence Alignment (MSA)** is crucial for AlphaFold's success. By analyzing thousands of related protein sequences from different species, the model extracts

evolutionary constraints that reveal which residues co-evolve, indicating structural proximity.

Evolutionary Co-variation

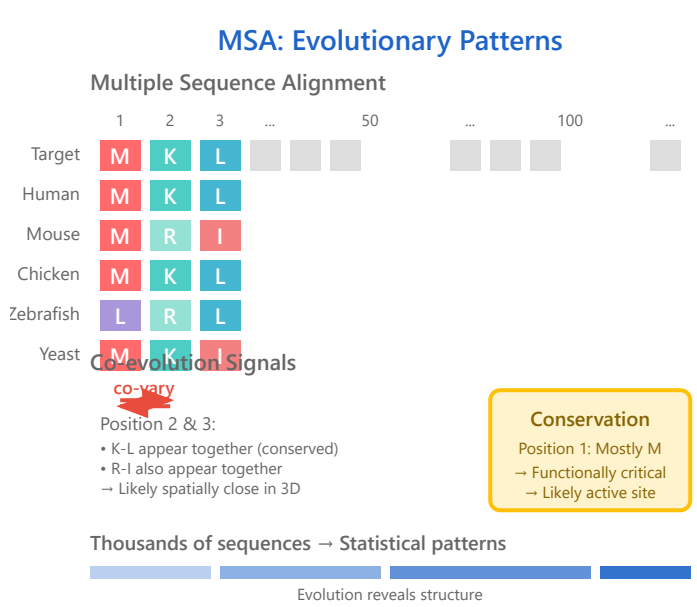
When two positions in a protein consistently mutate together across species, they are likely in close spatial proximity. AlphaFold learns these co-evolution patterns through MSA row and column attention.

MSA Representation

Each MSA row represents a homologous sequence. The Evoformer processes this matrix with specialized attention mechanisms that communicate both within sequences (row) and across positions (column).

Database Search

AlphaFold searches large databases (UniRef90, BFD, MGnify) to find homologous sequences, typically gathering thousands of related proteins to build a comprehensive evolutionary profile.



3. Structure Module - IPA & SE(3)-Equivariance

The Structure Module is AlphaFold's breakthrough component, featuring **Invariant Point Attention (IPA)**. This mechanism operates directly in 3D space while maintaining SE(3)-equivariance, meaning it respects rotations and translations of the protein structure.

Invariant Point Attention (IPA)

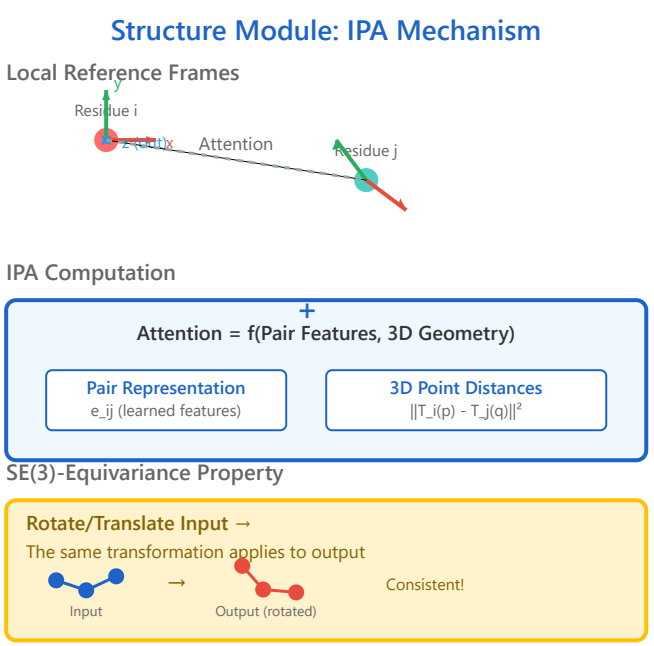
IPA computes attention in both the pair representation space and 3D coordinate space simultaneously. It measures geometric distances between points on local frames, making it rotation/translation invariant.

Local Reference Frames

Each residue has a local coordinate frame (backbone atoms N, Cα, C). The structure module updates both frame orientations and translations iteratively, building the full 3D structure progressively.

Geometric Reasoning

Unlike previous methods that predict distance matrices, AlphaFold directly generates 3D coordinates. This enables natural modeling of chirality, angles, and other geometric constraints inherent to protein structures.



4. Confidence Metrics - pLDDT Scores

AlphaFold provides **per-residue confidence scores (pLDDT)** that indicate how reliable each predicted atom position is. These scores are crucial for researchers to assess which parts of the structure are trustworthy and which regions might be disordered or incorrectly predicted.

pLDDT Definition

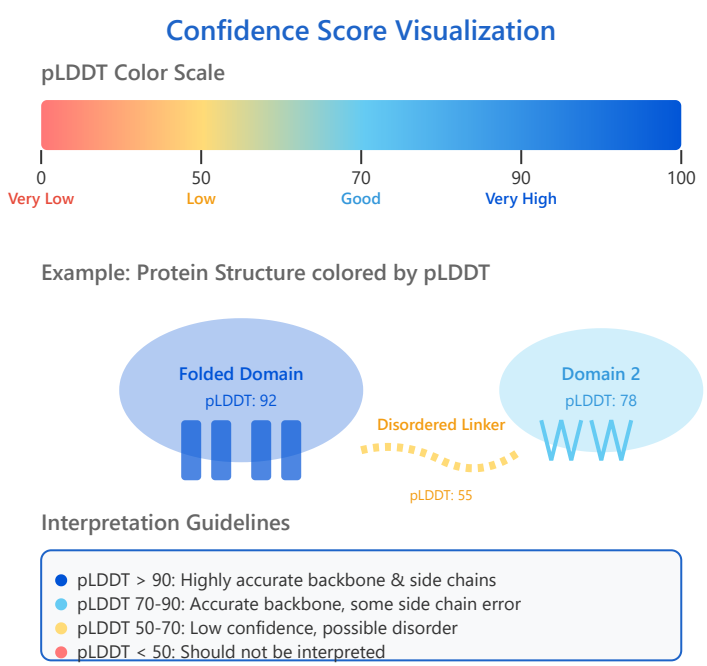
Predicted Local Distance Difference Test (pLDDT) scores range from 0-100, predicting the expected accuracy of Cα atom positions. Scores >90 indicate very high confidence, 70-90 indicate good confidence, 50-70 indicate low confidence, and <50 indicate very low confidence.

Interpretation Guide

High pLDDT regions (blue) typically represent well-folded domains with strong evolutionary constraints. Low pLDDT regions (yellow/red) often correspond to disordered regions, linkers, or areas with insufficient evolutionary information.

PAE (Predicted Aligned Error)

AlphaFold also provides PAE matrices showing confidence in relative positions between residues. This is especially useful for multi-domain proteins to assess domain-domain orientations.



5. Database Impact - 200M+ Structures

AlphaFold has transformed structural biology by predicting **over 200 million protein structures**, covering nearly every known protein. The AlphaFold Protein Structure Database (AlphaFold DB) provides free access to these predictions, democratizing structural data for researchers worldwide.

Coverage Scale

Before AlphaFold: ~170,000 experimentally determined structures in PDB (50 years of work). After AlphaFold: 200+ million predicted

structures covering most of UniProt, representing a 1000x increase in structural knowledge.

Research Acceleration

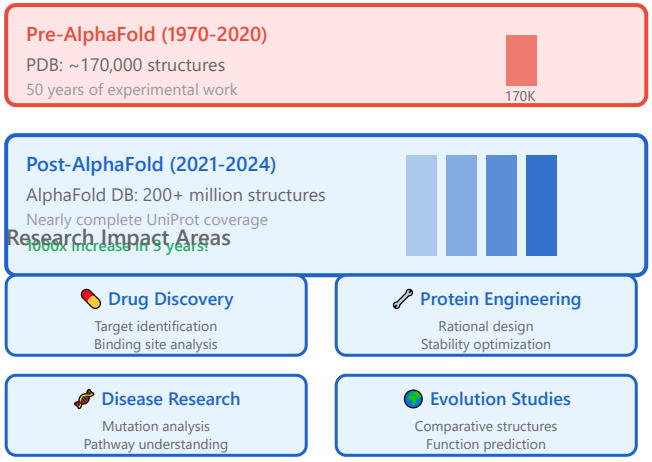
Researchers can now instantly access predicted structures instead of waiting months/years for experimental determination. This has accelerated drug discovery, protein engineering, and fundamental biology research across all fields.

Organism Coverage

AlphaFold DB includes proteomes from model organisms (human, mouse, E. coli, yeast), plants, parasites, and environmental microbes. This enables comparative structural biology and evolutionary studies at unprecedented scale.

Database Growth & Impact

Structural Coverage Timeline



Research Impact Areas