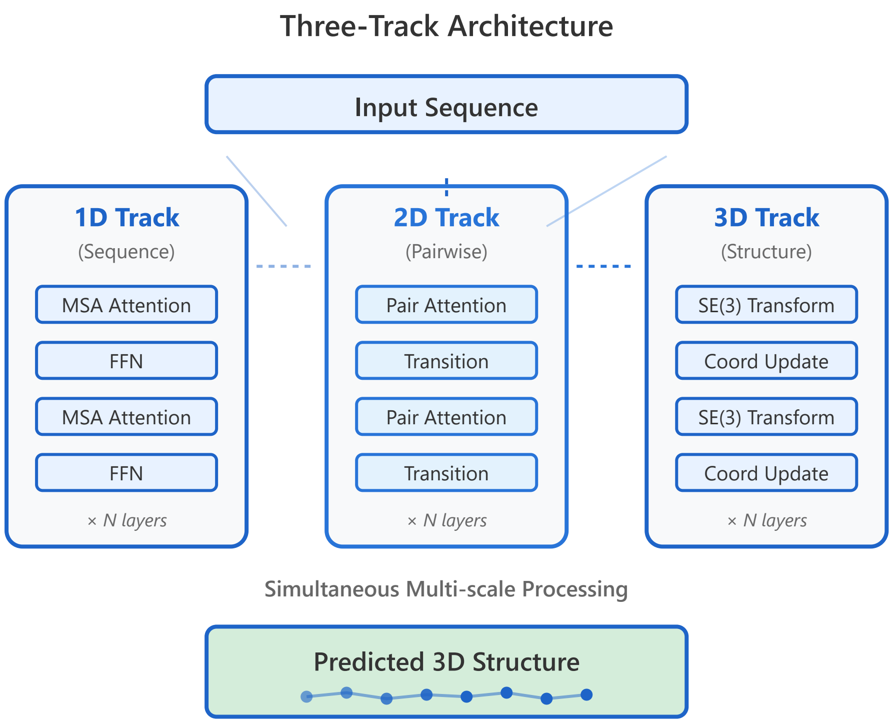
RoseTTAFold: Accurate Protein Structure Prediction



Three-track architecture
1D, 2D, 3D parallel processing

End-to-end learning
Direct structure prediction

Complex prediction
Protein-protein interactions

Speed advantages
Faster than AlphaFold2

Applications
Structure, function, design

1 Three-Track Architecture

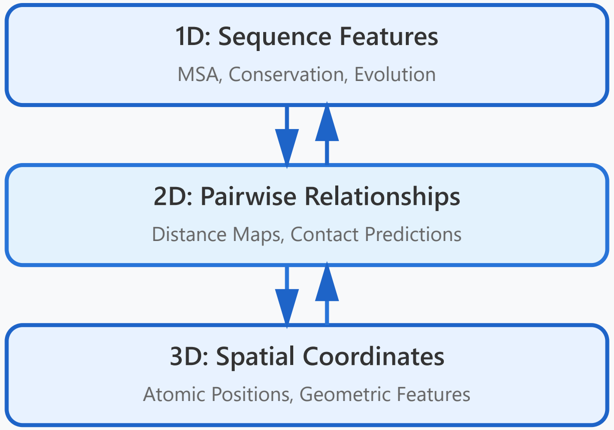
RoseTTAFold employs a unique three-track neural network architecture that simultaneously processes protein information at three different scales: 1D (sequence), 2D (pairwise distances), and 3D (coordinates). This parallel processing approach allows the model to capture multi-scale patterns and relationships that are crucial for accurate structure prediction.

- **1D Track (Sequence):** Processes multiple sequence alignments (MSA) to capture evolutionary information and identify conserved residues across homologous proteins
- **2D Track (Pairwise):** Models residue-residue relationships and distance constraints, capturing local and long-range interactions between amino acids
- **3D Track (Structure):** Directly operates on 3D coordinates using SE(3)-equivariant transformations, ensuring geometric consistency
- **Information Exchange:** The three tracks communicate bidirectionally at each layer, allowing features to flow between different representations

Key Innovation

Unlike traditional methods that process information sequentially, RoseTTAFold's parallel architecture enables simultaneous refinement across all three levels, leading to more coherent and accurate predictions.

Information Flow



Continuous Information Exchange

2 End-to-End Learning

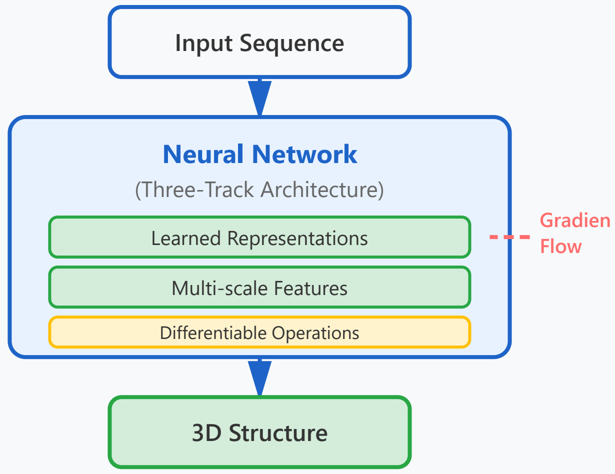
RoseTTAFold implements a fully differentiable end-to-end learning framework that directly maps from protein sequences to 3D structures without requiring intermediate steps or template-based modeling. This approach enables the network to learn complex structure-function relationships directly from data.

- **Direct Prediction:** Eliminates the need for fragment assembly or template-based modeling, which were standard in earlier methods
- **Gradient Flow:** Backpropagation flows through all three tracks simultaneously, allowing for holistic optimization
- **Learned Representations:** The model automatically learns relevant features at each level rather than relying on hand-crafted features
- **Structure Module:** Final layers convert learned representations directly into 3D atomic coordinates with associated confidence scores

Advantage

End-to-end learning allows the model to optimize for the final structural output rather than intermediate objectives, resulting in more accurate and physically realistic predictions.

End-to-End Pipeline



No intermediate templates or fragments

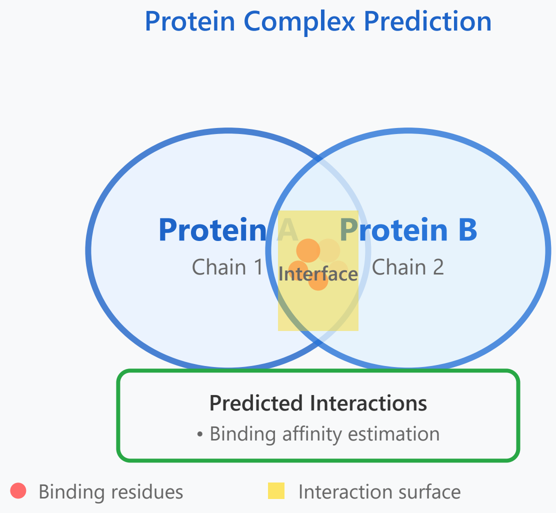
3 Complex Prediction

One of RoseTTAFold's most powerful capabilities is predicting protein-protein interaction complexes and multi-chain assemblies. By treating multiple chains simultaneously during inference, the model can capture inter-chain contacts, binding interfaces, and quaternary structure arrangements.

- **Multi-chain Support:** Processes multiple protein chains simultaneously, capturing inter-molecular interactions
- **Interface Prediction:** Accurately identifies binding sites and interaction surfaces between protein partners
- **Oligomer Assembly:** Can model homo- and hetero-oligomeric structures, including antibody-antigen complexes
- **Functional Insights:** Complex structures reveal mechanisms of protein function, regulation, and signaling pathways

Application Example

RoseTTAFold has successfully predicted structures of antibody-antigen complexes, enzyme-substrate interactions, and large multi-protein assemblies like the SARS-CoV-2 spike protein-ACE2 receptor complex.



Multi-chain simultaneous prediction

4 Speed Advantages

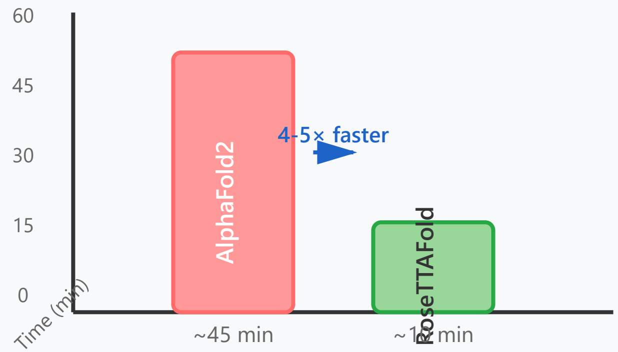
RoseTTAFold achieves significantly faster prediction times compared to AlphaFold2, making it practical for large-scale structural genomics projects and real-time applications. The speed improvement comes from architectural optimizations and efficient implementation without sacrificing accuracy.

- **Computational Efficiency:** Requires fewer computational resources and GPU memory compared to AlphaFold2
- **Faster Inference:** Typical predictions complete in minutes rather than hours for medium-sized proteins
- **Scalable Architecture:** Can process multiple proteins in parallel batches efficiently
- **Resource Accessibility:** Lower computational requirements make it accessible to more researchers without high-end computing clusters

Performance Comparison

For a 300-residue protein, RoseTTAFold typically completes prediction in 5-10 minutes on a single GPU, while maintaining accuracy comparable to AlphaFold2. This speed enables high-throughput screening of entire proteomes.

Computational Performance

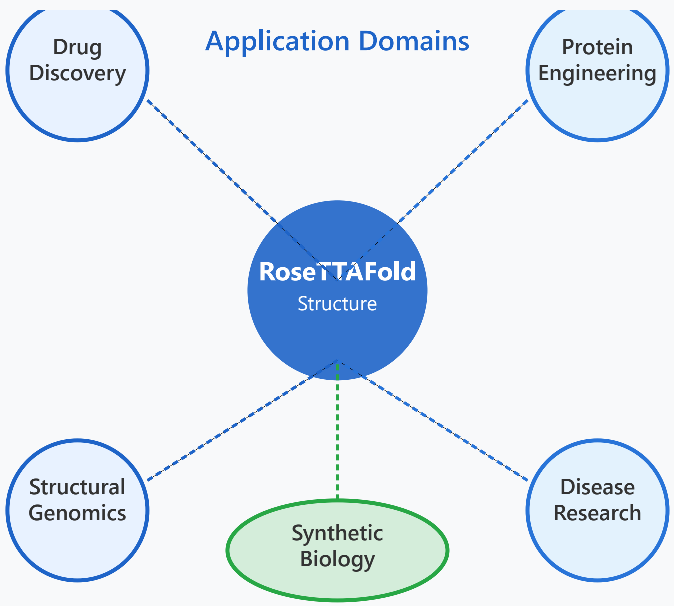


Resource Requirements
- Single GPU sufficient
- Lower memory footprint

For 300-residue protein on single GPU

5 Applications

RoseTTAFold's capabilities extend across multiple domains in structural biology, drug discovery, and protein engineering. Its accuracy, speed, and ability to handle complex structures make it a versatile tool for both fundamental research and practical applications.



- **Drug Discovery:** Identifying binding pockets, predicting drug-target interactions, and virtual screening for therapeutic candidates
- **Protein Engineering:** Guiding rational design of proteins with enhanced stability, altered specificity, or novel functions
- **Structural Genomics:** Large-scale prediction of protein structures for entire genomes, filling gaps in structural databases
- **Disease Research:** Understanding structural basis of genetic diseases, identifying pathogenic variants, and designing therapeutic interventions
- **Synthetic Biology:** Designing novel protein folds, creating artificial enzymes, and engineering biosynthetic pathways

Real-world Impact

RoseTTAFold has been used to predict structures of orphan proteins, design COVID-19 therapeutic candidates, engineer novel enzymes for industrial applications, and accelerate vaccine development by modeling antibody-antigen interactions.