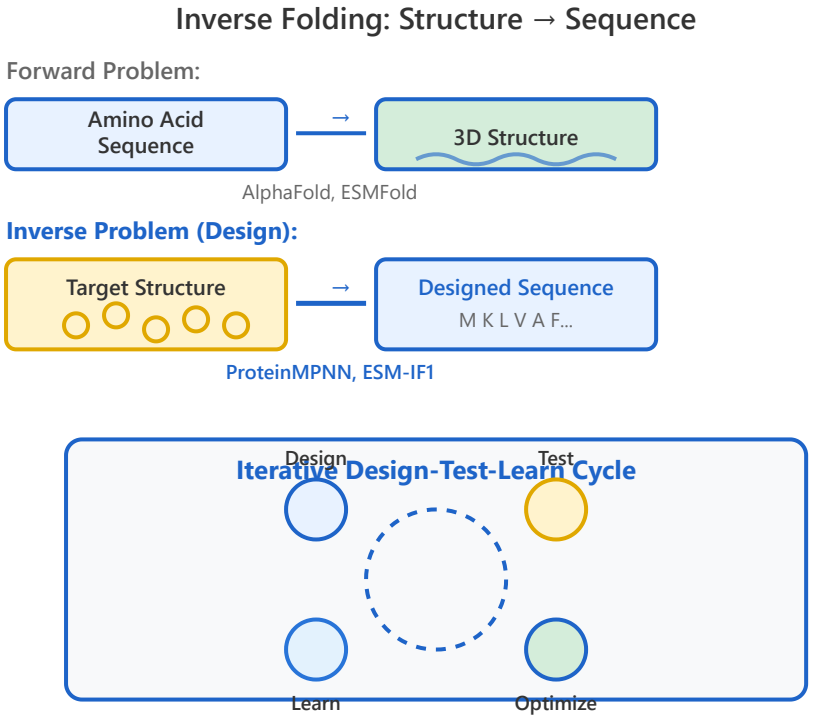
Protein Design



Inverse folding
Structure → sequence prediction

Scaffold design
De novo backbone generation

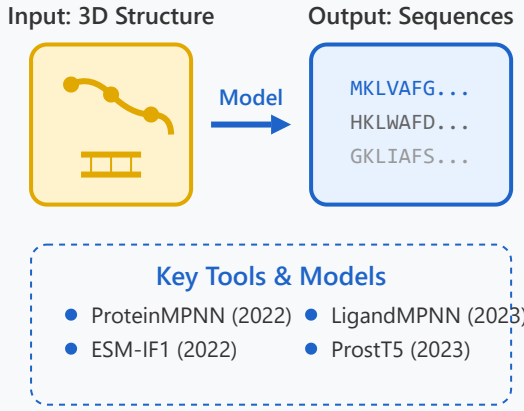
Interface design
Protein-protein interactions

De novo binders
Target-specific protein design

Stability optimization
Thermostability enhancement

Detailed Method Descriptions

Inverse Folding: Structure → Sequence Prediction



Inverse folding is the fundamental problem in computational protein design where we predict amino acid sequences that will fold into a desired three-dimensional structure. Unlike forward folding (structure prediction from sequence), inverse folding solves the reverse problem: given a target structure, what sequences could produce it?

Modern deep learning models like ProteinMPNN and ESM-IF1 have revolutionized this field by learning from vast databases of protein structures. These models encode the geometric constraints of the backbone structure and predict sequences that satisfy both structural and biochemical requirements.

Key Features:

- Graph neural networks encode backbone geometry and residue environments
- Generates multiple diverse sequences for the same structure
- Considers rotamer preferences and side-chain packing
- Can condition on specific residue constraints or motifs
- High success rates (60-90%) in experimental validation

Applications:

- Redesigning protein scaffolds for improved stability
- Engineering enzyme active sites

- Creating novel protein folds
- Protein humanization for therapeutics

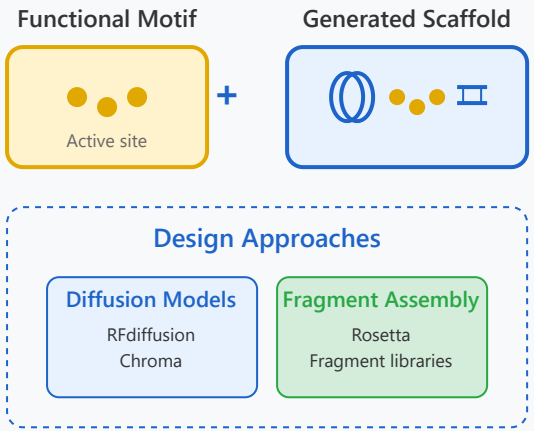
Example:

Given a TIM barrel structure, ProteinMPNN can generate 100+ diverse sequences that all fold into the same barrel topology, enabling exploration of sequence space while maintaining function.

2 Scaffold Design: De Novo Backbone Generation

Scaffold design involves creating entirely new protein backbones that can support specific functional motifs or binding sites. Rather than modifying existing proteins, this approach generates novel three-dimensional architectures from scratch that position key residues in precise geometric arrangements.

Recent breakthroughs using diffusion models (like RFdiffusion) have dramatically improved scaffold design. These models learn to generate protein backbones by reversing a noise-adding process, allowing them to create diverse, designable structures that incorporate functional constraints while maintaining protein-like geometry and secondary structure composition.



Key Features:

- Generates completely novel protein folds not found in nature
- Can incorporate motif constraints (binding sites, active sites)
- Controls secondary structure composition (helices, sheets, loops)
- Symmetry constraints for multi-subunit assemblies
- Iterative refinement through structure prediction validation

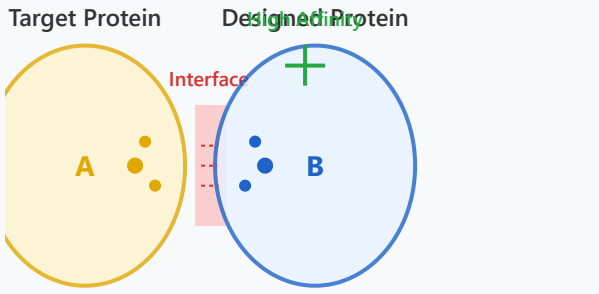
Applications:

- Creating novel enzyme scaffolds with tailored active sites
- Designing protein binders with specific epitope recognition
- Building protein cages and nanomaterials
- Engineering vaccines with optimized antigen presentation

Example:

RFdiffusion was used to design novel protein binders against SARS-CoV-2 spike protein by generating scaffolds that position key binding residues to match the ACE2 receptor interface, resulting in nanomolar-affinity binders.

3 Interface Design: Protein-Protein Interactions



Key Design Factors

- Shape complementarity
- Hydrogen bonding network
- Electrostatic compatibility
- Hydrophobic packing

Interface design focuses on engineering the interaction surfaces between proteins to create or enhance protein-protein interactions. This involves optimizing the complementarity between two protein surfaces through shape, electrostatics, and chemical interactions to achieve high-affinity binding.

Computational approaches model the interface region at atomic detail, considering backbone conformational changes, side-chain rotamers, and the balance between binding affinity and specificity. Modern machine learning methods can predict interface residues and suggest mutations that improve binding while maintaining specificity.

Key Features:

- Shape complementarity scoring using surface geometry analysis
- Electrostatic potential matching across interfaces
- Hydrogen bond network optimization
- Buried surface area maximization
- Hot spot residue identification and enhancement
- Specificity design to avoid off-target interactions

Applications:

- Antibody-antigen interface optimization for therapeutics
- Protein complex stabilization in structural biology
- Designing protein inhibitors for disease targets
- Engineering synthetic signaling pathways
- Creating self-assembling protein materials

Example:

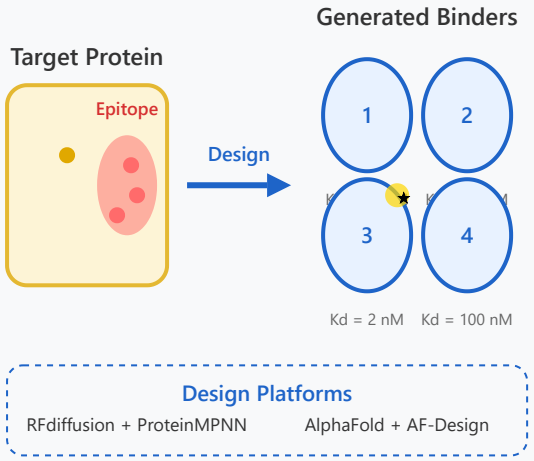
Redesigning the interface between an antibody and a viral protein to increase binding affinity from micromolar to picomolar range through computational optimization of 5-10 key interface residues.

4 De Novo Binders: Target-Specific Protein Design

De novo binder design creates entirely new proteins from scratch that bind to specific target molecules with high affinity and specificity. Unlike antibodies or natural binding proteins, these are computationally designed proteins optimized for a particular binding task.

This approach combines scaffold generation with interface design, using diffusion models to create backbones that present binding residues in optimal geometric arrangements. The process typically involves generating thousands of

candidates, screening them computationally, and experimentally validating the top designs, achieving success rates of 10-50% for nanomolar binders.



Key Features:

- Target-conditioned generation of binding proteins
- Multiple scaffold topologies (helical bundles, repeat proteins, mini-proteins)
- Epitope-specific targeting on protein surfaces
- Integration with AlphaFold for structure prediction validation
- Rapid design-test cycles (weeks instead of months)
- Generates highly specific binders with minimal off-target binding

Applications:

- Therapeutic protein development (antibody alternatives)
- Diagnostic tools and biosensors
- Research tools for target validation
- Blocking viral entry (e.g., COVID-19 inhibitors)
- Creating new signaling molecules in synthetic biology

Example:

Researchers at University of Washington used RFdiffusion to design mini-protein binders against the SARS-CoV-2 receptor binding domain, achieving picomolar affinity binders that neutralize the virus in vitro within 3 weeks of design.

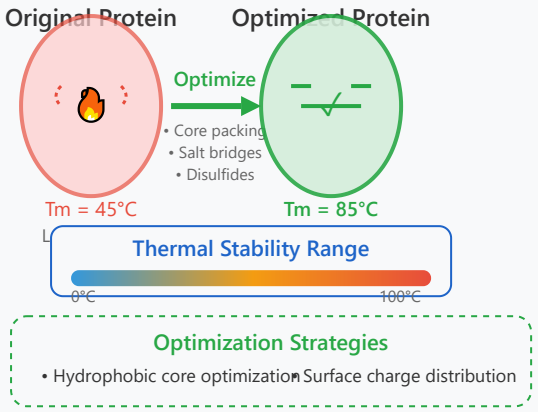
5 Stability Optimization: Thermostability Enhancement

Stability optimization focuses on enhancing protein thermostability and resistance to denaturation through rational design and computational prediction. This involves identifying and modifying residues that contribute to protein unfolding, improving hydrophobic core packing, introducing stabilizing interactions, and removing destabilizing elements.

Machine learning models can now predict the effects of mutations on stability (ΔΔG) with high accuracy, enabling systematic exploration of stability-enhancing variants. This is particularly valuable for therapeutic proteins, industrial enzymes, and any application requiring proteins to function under harsh conditions.

Key Features:

- ΔΔG prediction for mutation effects on folding stability
- Core residue packing optimization using rotamer libraries
- Introduction of disulfide bonds for covalent stabilization
- Salt bridge network design for electrostatic stabilization
- Loop rigidification and proline substitutions



- Removal of thermolabile residues (asparagine, glutamine deamidation sites)
- Consensus design from homologous sequences

Applications:

- Industrial enzyme optimization for high-temperature processes
- Therapeutic protein formulation stability
- Vaccine antigen stabilization
- Biosensor proteins for harsh environments
- Extending shelf-life of protein-based products

Example:

Engineering a thermostable variant of T4 lysozyme by introducing 5 key mutations identified through computational stability prediction, increasing melting temperature from 42°C to 72°C while maintaining full enzymatic activity.