Novelty Assessment Report

Paper: Scaling Atomistic Protein Binder Design with Generative Pretraining and Test-Time Compute
PDF URL: https://openreview.net/pdf?id=qmCpJtFZra
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Protein interaction modeling is central to protein design, which has been transformed by machine learning with broad applications in drug discovery and beyond. In this landscape, structure-based de novo binder design is most often cast as either conditional generative modeling or sequence optimization via structure predictors ("hallucination"). We argue that this is a false dichotomy and propose Complexa, a novel fully atomistic binder generation method unifying both paradigms. We extend recent flow-based latent protein generation architecture and leverage the domain-domain interactions of monomeric computationally predicted protein structures to construct Teddymer, a new large-scale dataset of synthetic binder-target pairs for pretraining. Combined with high-quality experimental multimers, this enables training a strong base model. We then perform inference-time optimization with this generative prior, unifying the strengths of previously distinct generative and hallucination methods. Complexa sets a new state of the art in computational binder design benchmarks: it delivers markedly higher in-silico success rates than existing generative approaches, and our novel test-time optimization strategies greatly outperform previous hallucination methods under normalized compute budgets. We further demonstrate explicit interface hydrogen bond optimization, fold class-guided binder generation, and extensions to small molecule targets and enzyme design tasks, again surpassing prior methods. Code, models and new data will be publicly released.

Core Task Landscape

This paper addresses: **De Novo Protein Binder Design**

A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Generative Model Architectures and Training Paradigms**
- **Inference-Time Optimization and Filtering Strategies**
- **Target-Specific Design Applications and Validation**
- **Computational Pipelines and Workflow Automation**
- **Specialized Binding Modalities and Constraints**
- **Experimental Screening and High-Throughput Validation**
- **Antibody-Specific Design Methods**
- **Template-Based and Motif-Driven Design**
- **Small Molecule and Ligand Binding Design**
- **Reviews, Perspectives, and Methodological Surveys**

Complete Taxonomy Tree

- De Novo Protein Binder Design Survey Taxonomy
- Generative Model Architectures and Training Paradigms
  - Diffusion and Flow-Based Generative Models ★ (3 papers)
  - [0] Scaling Atomistic Protein Binder Design with Generative Pretraining and Test-Time Compute (Anon et al., 2026) View paper
  - [7] De novo design of protein structure and function with RFdiffusion (Joseph L. Watson, 2023) View paper
  - [23] Robust and Reliable de novo Protein Design: A Flow-Matching-Based Protein Generative Model Achieves Remarkably High Success Rates (Junyu Yan, 2025) View paper
  - Protein Language Models for Binder Generation (2 papers)
  - [28] Prot42: a Novel Family of Protein Language Models for Target-aware Protein Binder Generation (Tekin, 2025) View paper
  - [30] De novo design of peptide binders to conformationally diverse targets with contrastive language modeling (Suhaas Bhat, 2024) View paper
  - AlphaFold-Based Hallucination and Inversion (4 papers)
  - [2] De novo design of high-affinity protein binders with AlphaProteo (Zambaldi, 2024) View paper
  - [25] One-shot design of functional protein binders with BindCraft (Martin Pacesa, 2025) View paper
  - [38] Boltzdesign1: Inverting All-Atom Structure Prediction Model for Generalized Biomolecular Binder Design (Yehlin Cho, 2025) View paper
  - [49] AlphaDesign: a de novo protein design framework based on AlphaFold (Michael Jendrusch, 2025) View paper
  - Hybrid and Multi-Scale Frameworks (2 papers)
  - [35] Latent-X: An Atom-level Frontier Model for De Novo Protein Binder Design (Bridgland, 2025) View paper
  - [50] De Novo Design of Protein-Binding Peptides by Quantum Computing (Lars Meuser, 2025) View paper
- Inference-Time Optimization and Filtering Strategies
  - Confidence-Based Filtering and Ranking (2 papers)
  - [13] Pxdesign: Fast, modular, and accurate de novo design of protein binders (Milong Ren, 2025) View paper

- [26] Predicting experimental success in de novo binder design: a meta-analysis of 3,766 experimentally characterised binders (Max D. Overath, 2025) View paper
- Test-Time Compute and Iterative Refinement (1 papers)
- [8] Improving de novo protein binder design with deep learning (Nathaniel R. Bennett, 2022) View paper

- Target-Specific Design Applications and Validation
  - Therapeutic Target Binders
  - Viral Protein Binders (4 papers)
    - [14] De novo design of mini-protein binders broadly neutralizing Clostridioides difficile toxin B variants (Xinchen Lv, 2024) View paper
    - [22] De novo design of picomolar SARS-CoV-2 miniprotein inhibitors (Longxing Cao, 2020) View paper
    - [27] De novo design mini-binder proteins targeting the glycoproteins D to inhibit PRV replication in PK15 cells (Lin Wei, 2025) View paper
    - [32] De novo-designed protein binders neutralize snake toxins (Anahita Bishop, 2025) View paper
  - Cytokine and Signaling Protein Binders (2 papers)
    - [15] De novo design of miniprotein antagonists of cytokine storm inducers (Buwei Huang, 2024) View paper
    - [18] De novo design of mini-binder proteins against IL-2 receptor Î² chain (Ke Ming, 2024) View paper
  - Peptide Hormone Binders (2 papers)
    - [6] De novo design of high-affinity binders of bioactive helical peptides (Susana VÃ¡zquez Torres, 2024) View paper
    - [24] De novo design of high-affinity protein binders to bioactive helical peptides (Susana VÃ¡zquez Torres, 2022) View paper
  - Tool Protein and Affinity Tag Binders (2 papers)
  - [19] De novo design protein binders for MBP and GST tags. (Jinlong Zhou, 2025) View paper
  - [34] Deep learningâ□□driven protein binder design for crop improvement (MS Iqbal, 2025) View paper
- Computational Pipelines and Workflow Automation (2 papers)
  - [11] De novo design of protein binders as functional therapeutics (Huang, 2023) View paper
  - [29] Automated and modular protein binder design with BinderFlow (Carlos ChacÃ³n-SÃ¡nchez, 2025) View paper
- Specialized Binding Modalities and Constraints
  - Peptide Binder Design (2 papers)
  - [16] De novo design of modular peptide-binding proteins by superhelical matching (Kejia Wu, 2023) View paper
  - [42] Target-Specific De Novo Peptide Binder Design with DiffPepBuilder (Fanhao Wang, 2024) View paper
  - Mirror-Image and D-Protein Binders (1 papers)
  - [31] Designing de novo D-protein binders (Haiyan Liu, 2024) View paper
  - Membrane Protein Analogues and Soluble Mimics (1 papers)
  - [1] Computational design of soluble and functional membrane protein analogues (Casper A. Goverde, 2024) View paper
- Experimental Screening and High-Throughput Validation (1 papers)
  - [20] Massively parallel de novo protein design for targeted therapeutics (Jaime Jahncke, 2017) View paper
- Antibody-Specific Design Methods (1 papers)
  - [37] Atomically accurate de novo design of antibodies with RFdiffusion (Nathaniel R Bennett, 2024) View paper
- Template-Based and Motif-Driven Design (4 papers)
  - [17] Design of protein-binding proteins from the target structure alone (Longxing Cao, 2022) View paper
  - [21] Bottom-up de novo design of functional proteins with complex structural features (Che Yang, 2021) View paper
  - [36] De novo design of buttressed loops for sculpting protein functions (Hanlun Jiang, 2024) View paper
  - [43] Robust de novo design of protein binding proteins from target structural information alone (Longxing Cao, 2021) View paper
- Small Molecule and Ligand Binding Design (4 papers)
  - [40] Prospective de novo drug design with deep interactome learning (Atz, 2024) View paper
  - [41] Computational design of ligand-binding proteins with high affinity and selectivity (Christine E. Tinberg, 2013) View paper
  - [46] Assay2mol: large language model-based drug design using bioassay context (Yifan Deng, 2025) View paper
  - [48] New computational protein design methods for de novo small molecule binding sites (James E. Lucas, 2020) View paper
- Reviews, Perspectives, and Methodological Surveys (11 papers)
  - [3] De novo protein designâ□□From new structures to programmable functions (Tanja Kortemme, 2024) View paper
  - [4] Code to complex: AI-driven de novo binder design (Daniel R. Fox, 2025) View paper
  - [5] De novo protein design, a retrospective (Ivan V. Korendovych, 2020) View paper
  - [9] The coming of age of de novo protein design (Po-Ssu Huang, 2016) View paper
  - [10] Computational design of novel proteinâ□□protein interactionsâ□□An overview on methodological approaches and applications (Anthony Marchand, 2022) View paper
  - [12] Recent advances in de novo protein design: Principles, methods, and applications (Xingjie Pan, 2021) View paper
  - [33] â□¦ electrochemical biosensor with S protein binding affinity for COVID-19 detection: Integrating computational design with experimental validation of S protein binding â□¦ (R Khan, 2024) View paper
  - [39] Computational design of affinity and specificity at proteinâ□□protein interfaces (John Karanicolas, 2009) View paper
  - [44] Essentials of de novo protein design: Methods and applications (Enrique Marcos, 2018) View paper
  - [45] Scratch That? De Novo Antibody Design Enters the AI Drug Discovery Toolbox (Fay Lin, 2025) View paper
  - [47] Computational design and experimental optimization of protein binders with prospects for biomedical applications (Alessandro Bonadio, 2021) View paper

Narrative

Core task: de novo protein binder design. The field has matured from early computational and experimental efforts into a rich ecosystem organized around several complementary strategies. At the highest level, the taxonomy distinguishes generative model architectures—including diffusion and flow-based frameworks that learn to sample novel binder structures—from inference-time optimization methods that refine candidates through energy-based filtering or iterative search. Parallel branches address target-specific applications (designing binders for particular therapeutic or diagnostic targets), computational pipelines that automate multi-step workflows, and specialized modalities such as antibody-specific methods, template-based approaches, and small-molecule binding sites. Additional branches capture experimental screening platforms and broader reviews that synthesize methodological advances. Representative works like RFdiffusion[7] exemplify the power of diffusion models, while AlphaProteo[2] and related efforts demonstrate how large-scale training can yield high-affinity binders across diverse targets.

Within the generative model branch, diffusion and flow-based methods have become particularly active, balancing the need for structural realism with computational efficiency. RFdiffusion[7] pioneered the application of denoising diffusion to protein backbone generation, enabling flexible hotspot-constrained design, whereas Flow Matching Design[23] explores continuous normalizing flows for smoother sampling trajectories. Scaling Atomistic Binder[0] sits squarely in this diffusion-and-flow cluster, emphasizing atomistic resolution and scalable training regimes that push beyond earlier coarse-grained or backbone-only models. Compared to RFdiffusion[7], which focuses on backbone geometry, Scaling Atomistic Binder[0] incorporates finer chemical detail to improve predicted binding interfaces. In contrast to Flow Matching Design[23], which prioritizes flow-matching dynamics, Scaling Atomistic Binder[0] leverages diffusion with explicit atom-level representations. These distinctions reflect ongoing exploration of how best to encode physical constraints and sampling efficiency in generative frameworks, a central question as the field moves toward clinically validated therapeutics and high-throughput experimental validation.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. De novo design of protein structure and function with RFdiffusion

Authors: Joseph L. Watson, David Juergens, N. Bennett, Brian L. Trippe, Jason Yim, et al. (38 authors total) | **Year/Venue**: 2023 • Nature | **URL**: View paper

Abstract

There has been considerable recent progress in designing new proteins using deep-learning methods^ 1 â□□ 9 . Despite this progress, a general deep-learning framework for protein design that enables solution of a wide range of design challenges, including de novo binder design and design of higher-order symmetric architectures, has yet to be described. Diffusion models^ 10 , 11 have had considerable success in image and language generative modelling but limited success when applied to protein mod...

Relationship Analysis

Both papers belong to the Diffusion and Flow-Based Generative Models category, using probabilistic denoising processes to generate protein structures for binder design. They overlap in their core approach of applying diffusion/flow-based generative modeling to de novo protein binder design conditioned on target structures. The key differences are that the original paper (Complexa) uses flow matching with a partially latent representation, introduces a large-scale synthetic training dataset (Teddymer), and uniquely combines generative modeling with inference-time optimization strategies (beam search, MCTS), while RFdiffusion uses diffusion models based on the RoseTTAFold architecture without the inference-time compute scaling paradigm.

---

2. Robust and Reliable de novo Protein Design: A Flow-Matching-Based Protein Generative Model Achieves Remarkably High Success Rates

Authors: Junyu Yan, Zibo Cui, Wenqing Yan, Yuhang Chen, Mengchen Pu, et al. (7 authors total) | **Year/Venue**: 2025 • bioRxiv | **URL**: View paper

Abstract

Generative models have achieved significant progress in the field of protein design, particularly in the generation of tertiary structures. However, they still face several challenges, such as balancing the designability and diversity of the generated structures, generating specified structures under highly constrained conditions, and the most challenging aspect, direct functional design such as motifs and binders. We present OriginFlow, an efficient protein generative model that combines Stocha...

Relationship Analysis

Both papers belong to the Diffusion and Flow-Based Generative Models category, employing flow matching techniques for de novo protein binder design. They overlap in their use of flow-based architectures for generating protein binders conditioned on targets, evaluation using structure prediction confidence metrics (ipAE, pLDDT), and validation on similar protein targets. However, Complexa distinguishes itself through its novel Teddymer synthetic training dataset derived from domain-domain interactions, unified framework combining generative modeling with inference-time optimization (beam search, MCTS), and partially latent representation building on La-Proteína, while OriginFlow focuses on combining SDE and flow matching models with emphasis on wet-lab validation achieving 90% success rates across expression, solubility, and affinity metrics.

Contributions Analysis

Overall novelty summary. The paper proposes Complexa, a method unifying generative modeling and hallucination-based optimization for protein binder design. It sits within the 'Diffusion and Flow-Based Generative Models' leaf, which contains only three papers total, indicating a relatively focused but not overcrowded research direction. The taxonomy shows this leaf is one of four under 'Generative Model Architectures and Training Paradigms', suggesting the field has diversified into multiple architectural paradigms rather than concentrating heavily in any single approach.

The taxonomy reveals neighboring leaves include 'Protein Language Models for Binder Generation' (two papers), 'AlphaFold-Based Hallucination and Inversion' (four papers), and 'Hybrid and Multi-Scale Frameworks' (two papers). Complexa's claim to unify generative and hallucination paradigms positions it at the boundary between the diffusion-based leaf and the AlphaFold hallucination branch. The 'Inference-Time Optimization and Filtering Strategies' branch (three papers across two leaves) is also relevant, as Complexa incorporates test-time optimization. This cross-cutting positioning suggests the work bridges previously distinct methodological clusters.

Among thirty candidates examined, the analysis identified one refutable pair for the Teddymer dataset contribution, while the unification framework and Complexa architecture showed no clear refutations across ten candidates each. The Teddymer dataset—synthetic binder-target pairs from domain-domain interactions—appears to have some overlap with prior synthetic training data efforts. The core methodological contributions (unifying generative and hallucination, latent target conditioning with test-time optimization) appear more distinctive within the limited search scope, though the analysis does not claim exhaustive coverage of all relevant prior work.

Based on the top-thirty semantic matches and taxonomy structure, Complexa appears to occupy a relatively novel position by explicitly bridging generative and hallucination paradigms. The limited search scope means potentially relevant work outside these candidates may exist. The taxonomy's modest leaf size (three papers) and the cross-branch positioning suggest the work addresses a recognized gap, though the Teddymer dataset shows some precedent in synthetic training data construction.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Unifying generative modeling and hallucination methods for binder design

Description: The authors introduce a unified framework that integrates flow-based generative modeling with inference-time optimization, bridging the gap between conditional generation approaches and structure predictor-based hallucination methods that were previously treated as separate paradigms in protein binder design.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Pxdesign: Fast, modular, and accurate de novo design of protein binders

URL: View paper

Brief Assessment

Pxdesign[13] develops separate diffusion-based (pxdesign-d) and hallucination-based (pxdesign-h) approaches as distinct methods, rather than unifying them into a single framework as the original paper proposes.

---

2. Fold-Conditioned De Novo Binder Design via AlphaFold2-Multimer Hallucination

URL: View paper

Brief Assessment

Fold-Conditioned Design[53] focuses exclusively on hallucination-based methods using AlphaFold2-Multimer with contact map similarity loss, without incorporating generative modeling components or flow-based architectures.

---

3. HalluDesign: Protein Optimization and de novo Design via Iterative Structure Hallucination and Sequence design

URL: View paper

Brief Assessment

HalluDesign[59] focuses exclusively on hallucination-based optimization using AlphaFold3-style models without training generative components. The original paper's novelty lies in combining flow-based generative pretraining with inference-time optimization, which HalluDesign[59] does not address.

---

4. Design of proteins presenting discontinuous functional sites using deep learning

URL: View paper

Brief Assessment

Discontinuous Sites[56] focuses on scaffolding discontinuous binding interfaces using gradient-based hallucination with TRRosetta, not on unifying generative modeling with hallucination methods as a general framework.

---

5. Protein Hunter: exploiting structure hallucination within diffusion for protein design

URL: View paper

Brief Assessment

Protein Hunter[54] uses hallucination within diffusion models for structure prediction, not a unified framework combining flow-based generative modeling with inference-time optimization as described in the original paper.

---

6. Data and AI-driven synthetic binding protein discovery

URL: View paper

Brief Assessment

Synthetic Binding Discovery[52] focuses on data-driven approaches and AI methods for synthetic binding protein discovery. The provided candidate text fragments are too sparse to establish whether it addresses the unification of generative modeling with hallucination methods as a methodological framework.

---

7. Diï¬usion Models for Protein Structure Design: From Backbone Generation to Atomic-Resolution Enzyme Design

URL: View paper

Brief Assessment

Diffusion Enzyme Design[57] focuses on enzyme design applications using diffusion models for protein structure generation. The candidate does not address the unification of generative modeling with hallucination/inference-time optimization methods for binder design.

---

8. BindEnergyCraft: Casting Protein Structure Predictors as Energy-Based Models for Binder Design

URL: View paper

Brief Assessment

BindEnergyCraft[55] focuses on improving hallucination-based methods by reinterpreting structure predictor confidence outputs as energy-based models, rather than unifying generative and hallucination paradigms as distinct approaches to binder design.

---

9. Casting Protein Structure Predictors as Energy-BasedModels for Binder Design and Scoring

URL: View paper

Brief Assessment

Energy-Based Scoring[58] focuses on casting structure predictors as energy-based models for scoring and hallucination-based optimization, not on unifying flow-based generative modeling with inference-time optimization as proposed in the original paper.

---

10. Scaffolding protein functional sites using deep learning

URL: View paper

Brief Assessment

Scaffolding Functional Sites[51] focuses on scaffolding functional residues using constrained hallucination and inpainting methods, not on unifying generative modeling with hallucination approaches as a general framework for binder design.

Contribution 2: Teddymer dataset of synthetic protein dimers

Description: The authors construct a large-scale dataset of 3.5 million synthetic binder-target pairs by partitioning AlphaFold Database monomers into structural domains using TED annotations and assembling artificial dimers from domain-domain interactions, providing training data for atomistic binder generation.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. ProBID-Net: a deep learning model for proteinâprotein binding interface design

URL: View paper

Prior Art Analysis

ProBID-Net[60] demonstrates that training on synthetic protein dimer datasets derived from domain-domain interactions was already established prior to the original paper's work. The candidate paper explicitly states it was 'trained using natural proteinâprotein complex structures and protein domainâdomain interface structures,' indicating that the approach of using domain-domain interactions to create training data for binder generation was not novel to the original authors. Both papers leverage the same fundamental concept: partitioning proteins into domains and using domain-domain interactions as training data for protein binder design models.

Evidence

Evidence 1 - **Rationale**: Both papers describe using domain-domain interface structures as training data for protein binder design. The candidate paper explicitly mentions training on 'protein domainâdomain interface structures,' demonstrating that this approach predates the original paper's Teddymer dataset. This directly challenges the novelty claim that the authors were first to construct such a dataset from domain-domain interactions. - **Original**: we exploit domain-domain interactions from predicted monomer structures in the alphafold database (afdb). using structural domain annotations from the encyclopedia of domains (ted) (lau et al., 2024b), we partition proteins into domains and assemble artificial protein dimers. after clustering and fi... - **Candidate**: trained using natural proteinâprotein complex structures and protein domainâdomain interface structures, probid-net can discern features from known target protein structures to design specific binding proteins based on their binding sites.

Evidence 2 - **Rationale**: The original paper's core innovation of treating domain-domain interactions as equivalent to binder-target pairs for training is already present in ProBID-Net[60], which explicitly trained on 'protein domainâdomain interface structures' for the same purpose of designing binding proteins. This shows the conceptual approach was not novel. - **Original**: we argue that the biophysical interactions between structural domains of afdb monomers are qualitatively similar to the interactions between chains in multimeric structures (see fig. 3). hence, we propose to split the afdb multi-domain monomers into their individual domains and treat the resulting m... - **Candidate**: trained using natural proteinâprotein complex structures and protein domainâdomain interface structures, probid-net can discern features from known target protein structures to design specific binding proteins based on their binding sites.

---

2. Exploring Artificially Conjugated Ubiquitin Dimers by Means of NMR Spectroscopy and MD Simulations

URL: View paper

Brief Assessment

The candidate paper (Ubiquitin Dimers[68]) studies artificially conjugated ubiquitin dimers using NMR spectroscopy and MD simulations, focusing on experimental characterization of specific protein conjugates rather than constructing large-scale synthetic datasets for machine learning training.

---

3. Target-Specific De Novo Peptide Binder Design with DiffPepBuilder

URL: View paper

Brief Assessment

DiffPepBuilder[42] focuses on peptide binder design using a synthetic dataset (PepPC-F) derived from protein-protein interfaces for peptide-specific applications, not on general protein dimer datasets for atomistic binder generation.

---

4. Design of protein function leaps by directed domain interface evolution

URL: View paper

Brief Assessment

Domain Interface Evolution[64] focuses on experimentally optimizing domain interfaces through directed evolution to enhance binding affinity and specificity, not on creating synthetic datasets for machine learning training.

---

5. Nonnatural proteinâ□□protein interaction-pair design by key residues grafting

URL: View paper

Brief Assessment

Key Residues Grafting[66] focuses on designing protein-protein interactions by grafting key residues onto scaffold proteins, not on constructing large-scale synthetic dimer datasets from domain-domain interactions for training generative models.

---

6. An All-Atom Generative Model for Designing Protein Complexes

URL: View paper

Brief Assessment

All-Atom Generative[65] focuses on multi-chain protein complex generation using PDB biological assemblies and does not describe creating synthetic dimers from domain-domain interactions of AlphaFold Database monomers using TED annotations.

---

7. Synthetic protein switches: design principles and applications

URL: View paper

Brief Assessment

Synthetic Protein Switches[62] focuses on engineered protein switches and their design principles for controlling protein function, not on creating large-scale datasets of synthetic dimers for training atomistic binder generation models.

---

8. Development of artificial antibody against receptor binding domain of SARS-CoV-2 spike protein.

URL: View paper

Brief Assessment

Artificial Antibody[67] focuses on designing affibody molecules against SARS-CoV-2 RBD using in silico docking and cell-free protein synthesis, not on creating large-scale synthetic protein dimer datasets for training generative models.

---

9. Protein domain mimics as modulators of proteinâ□□protein interactions

URL: View paper

Brief Assessment

Domain Mimics[63] focuses on small-molecule protein domain mimics (HBS, OHM, triazolamers) for modulating protein-protein interactions, not on constructing synthetic protein dimer datasets for machine learning training.

---

10. Rational Design and Protein Engineering of {SH2 Domainâ□□â□« Flexible Linkerâ□□â□« Selfâ□□Controlling Peptide} Fusion System With Phosphorylationâ□□Regulated Molecular Switch Functionality

URL: View paper

Brief Assessment

SH2 Fusion System[61] focuses on engineering phosphorylation-regulated molecular switches using SH2 domain fusion proteins, not on creating synthetic protein dimer datasets for machine learning training.

Contribution 3: Complexa framework with latent target conditioning and test-time optimization

Description: The authors develop Complexa, a fully atomistic binder generation method that extends La-Proteína's partially latent flow matching architecture with novel latent target conditioning and implements multiple test-time compute scaling algorithms (beam search, Feynman-Kac steering, Monte Carlo Tree Search) to optimize binder quality during inference.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Manipulating 3D Molecules in a Fixed-Dimensional SE(3)-Equivariant Latent Space

URL: View paper

Brief Assessment

SE3 Equivariant Latent[70] focuses on small molecule manipulation in a fixed-dimensional latent space using VAE and Bayesian flow networks, not protein binder generation with flow matching and test-time optimization algorithms.

---

2. Electron density-based GPT for optimization and suggestion of hostâ□□guest binders

URL: View paper

Brief Assessment

Electron Density GPT[73] focuses on host-guest molecular binding using electron density representations and variational autoencoders, not protein binder design with flow matching and test-time compute scaling algorithms.

---

3. Protein Diffusion Models as Statistical Potentials

URL: View paper

Brief Assessment

Statistical Potentials[77] focuses on using protein diffusion models as statistical potentials for atomistic simulation, operating on latent representation spaces. This differs from Complexa's binder generation framework with latent target conditioning and test-time compute scaling algorithms.

---

4. Bayesian Optimization in the Latent Space of a Variational Autoencoder for the Generation of Selective FLT3 Inhibitors.

URL: View paper

Brief Assessment

Bayesian Latent Optimization[78] focuses on molecular generation using Bayesian optimization in VAE latent space for FLT3 inhibitors, not protein binder design with flow matching and test-time compute scaling algorithms.

---

5. Molecule design by latent prompt transformer

URL: View paper

Brief Assessment

Latent Prompt Transformer[71] focuses on molecule design using latent prompts for conditional generation, not protein binder design with atomistic flow matching and test-time compute scaling algorithms.

---

6. Reinforcement Learning-Inspired Molecular Generation with Latent Space Diffusion and Genetic Algorithm Optimization under Affinity and Similarity Constraints

URL: View paper

Brief Assessment

Reinforcement Molecular Generation[69] focuses on molecular generation with genetic algorithms for drug design, not protein binder generation with flow matching and test-time compute scaling algorithms.

---

7. Target specific peptide design using latent space approximate trajectory collector

URL: View paper

Brief Assessment

Latent Trajectory Collector[72] focuses on peptide design using latent space optimization with CMA-ES, not atomistic protein binder generation with flow matching and test-time compute scaling algorithms like beam search or MCTS.

---

8. Latent molecular optimization for targeted therapeutic design

URL: View paper

Brief Assessment

Latent Molecular Optimization[76] focuses on small molecule drug design using latent chemical embeddings and gradient-based optimization, not protein binder generation with flow matching and test-time compute scaling algorithms.

---

9. Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity

URL: View paper

Brief Assessment

Variational Autoencoder 3D[74] focuses on chemical compound generation using variational autoencoders for molecular structures, not protein binder design with flow matching and test-time optimization.

---

10. Deep Generative Methods for Target Specific Drug Design

URL: View paper

Brief Assessment

Generative Target Design[75] focuses on small-molecule and peptide drug design using latent space sampling with molecular dynamics feedback, not atomistic protein binder generation with flow matching and test-time compute scaling algorithms.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Scaling Atomistic Protein Binder Design with Generative Pretraining and Test-Time Compute View paper
- [1] Computational design of soluble and functional membrane protein analogues View paper

- [2] De novo design of high-affinity protein binders with AlphaProteo View paper
- [3] De novo protein designâ□□From new structures to programmable functions View paper
- [4] Code to complex: AI-driven de novo binder design View paper
- [5] De novo protein design, a retrospective View paper
- [6] De novo design of high-affinity binders of bioactive helical peptides View paper
- [7] De novo design of protein structure and function with RFdiffusion View paper
- [8] Improving de novo protein binder design with deep learning View paper
- [9] The coming of age of de novo protein design View paper
- [10] Computational design of novel proteinâ□□protein interactionsâ□□An overview on methodological approaches and applications View paper
- [11] De novo design of protein binders as functional therapeutics View paper
- [12] Recent advances in de novo protein design: Principles, methods, and applications View paper
- [13] Pxdesign: Fast, modular, and accurate de novo design of protein binders View paper
- [14] De novo design of mini-protein binders broadly neutralizing Clostridioides difficile toxin B variants View paper
- [15] De novo design of miniprotein antagonists of cytokine storm inducers View paper
- [16] De novo design of modular peptide-binding proteins by superhelical matching View paper
- [17] Design of protein-binding proteins from the target structure alone View paper
- [18] De novo design of mini-binder proteins against IL-2 receptor Î² chain View paper
- [19] De novo design protein binders for MBP and GST tags. View paper
- [20] Massively parallel de novo protein design for targeted therapeutics View paper
- [21] Bottom-up de novo design of functional proteins with complex structural features View paper
- [22] De novo design of picomolar SARS-CoV-2 miniprotein inhibitors View paper
- [23] Robust and Reliable de novo Protein Design: A Flow-Matching-Based Protein Generative Model Achieves Remarkably High Success Rates View paper
- [24] De novo design of high-affinity protein binders to bioactive helical peptides View paper
- [25] One-shot design of functional protein binders with BindCraft View paper
- [26] Predicting experimental success in de novo binder design: a meta-analysis of 3,766 experimentally characterised binders View paper
- [27] De novo design mini-binder proteins targeting the glycoproteins D to inhibit PRV replication in PK15 cells View paper
- [28] Prot42: a Novel Family of Protein Language Models for Target-aware Protein Binder Generation View paper
- [29] Automated and modular protein binder design with BinderFlow View paper
- [30] De novo design of peptide binders to conformationally diverse targets with contrastive language modeling View paper
- [31] Designing de novo D-protein binders View paper
- [32] De novo-designed protein binders neutralize snake toxins View paper
- [33] â□¦ electrochemical biosensor with S protein binding affinity for COVID-19 detection: Integrating computational design with experimental validation of S protein binding â□¦ View paper
- [34] Deep learningâ□□driven protein binder design for crop improvement View paper
- [35] Latent-X: An Atom-level Frontier Model for De Novo Protein Binder Design View paper
- [36] De novo design of buttressed loops for sculpting protein functions View paper
- [37] Atomically accurate de novo design of antibodies with RFdiffusion View paper
- [38] Boltzdesign1: Inverting All-Atom Structure Prediction Model for Generalized Biomolecular Binder Design View paper
- [39] Computational design of affinity and specificity at proteinâ□□protein interfaces View paper
- [40] Prospective de novo drug design with deep interactome learning View paper
- [41] Computational design of ligand-binding proteins with high affinity and selectivity View paper
- [42] Target-Specific De Novo Peptide Binder Design with DiffPepBuilder View paper
- [43] Robust de novo design of protein binding proteins from target structural information alone View paper
- [44] Essentials of de novo protein design: Methods and applications View paper
- [45] Scratch That? De Novo Antibody Design Enters the AI Drug Discovery Toolbox View paper
- [46] Assay2mol: large language model-based drug design using bioassay context View paper
- [47] Computational design and experimental optimization of protein binders with prospects for biomedical applications View paper
- [48] New computational protein design methods for de novo small molecule binding sites View paper
- [49] AlphaDesign: a de novo protein design framework based on AlphaFold View paper
- [50] De Novo Design of Protein-Binding Peptides by Quantum Computing View paper
- [51] Scaffolding protein functional sites using deep learning View paper
- [52] Data and AI-driven synthetic binding protein discovery View paper
- [53] Fold-Conditioned De Novo Binder Design via AlphaFold2-Multimer Hallucination View paper
- [54] Protein Hunter: exploiting structure hallucination within diffusion for protein design View paper
- [55] BindEnergyCraft: Casting Protein Structure Predictors as Energy-Based Models for Binder Design View paper
- [56] Design of proteins presenting discontinuous functional sites using deep learning View paper
- [57] Diï¬□usion Models for Protein Structure Design: From Backbone Generation to Atomic-Resolution Enzyme Design View paper
- [58] Casting Protein Structure Predictors as Energy-BasedModels for Binder Design and Scoring View paper
- [59] HalluDesign: Protein Optimization and de novo Design via Iterative Structure Hallucination and Sequence design View paper
- [60] ProBID-Net: a deep learning model for proteinâ□□protein binding interface design View paper
- [61] Rational Design and Protein Engineering of {SH2 Domainâ□□â□« Flexible Linkerâ□□â□« Selfâ□□Controlling Peptide} Fusion System With Phosphorylationâ□□Regulated Molecular Switch Functionality View paper
- [62] Synthetic protein switches: design principles and applications View paper
- [63] Protein domain mimics as modulators of proteinâ□□protein interactions View paper
- [64] Design of protein function leaps by directed domain interface evolution View paper
- [65] An All-Atom Generative Model for Designing Protein Complexes View paper
- [66] Nonnatural proteinâ□□protein interaction-pair design by key residues grafting View paper
- [67] Development of artificial antibody against receptor binding domain of SARS-CoV-2 spike protein. View paper
- [68] Exploring Artificially Conjugated Ubiquitin Dimers by Means of NMR Spectroscopy and MD Simulations View paper

- [69] Reinforcement Learning-Inspired Molecular Generation with Latent Space Diffusion and Genetic Algorithm Optimization under Affinity and Similarity Constraints View paper
- [70] Manipulating 3D Molecules in a Fixed-Dimensional SE(3)-Equivariant Latent Space View paper
- [71] Molecule design by latent prompt transformer View paper
- [72] Target specific peptide design using latent space approximate trajectory collector View paper
- [73] Electron density-based GPT for optimization and suggestion of hostâguest binders View paper
- [74] Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity View paper
- [75] Deep Generative Methods for Target Specific Drug Design View paper
- [76] Latent molecular optimization for targeted therapeutic design View paper
- [77] Protein Diffusion Models as Statistical Potentials View paper
- [78] Bayesian Optimization in the Latent Space of a Variational Autoencoder for the Generation of Selective FLT3 Inhibitors. View paper