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

Minchao Fang[1, 2, †], Chentong Wang[1, 2, †], Jungang Shi[1, 2, †], Fangbai Lian[1, 2, †], Qihan Jin[1, 2], Zhe Wang[1, 2], Yanzhe Zhang[1, 2], Zhanyuan Cui[1, 2], YanJun Wang[1, 2], Yitao Ke[1, 2], Qingzheng Han[1, 2], Longxing Cao[1, 2, *]

[1] State Key Laboratory of Gene Expression, School of Life Sciences, Westlake University, Hangzhou, China

[2] Artificial Intelligence Drug Design Core Laboratory, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China

*Corresponding author: caolongxing@westlake.edu.cn

† These authors are co-first authors.

Abstract

Deep learning has revolutionized biomolecular modeling, enabling the prediction of diverse structures with atomic accuracy. However, leveraging the atomic-level precision of the structure prediction model for de novo design remains challenging. Here, we present HalluDesign, a general all-atom framework for protein optimization and de novo design, which iteratively update protein structure and sequence. HalluDesign harnesses the inherent hallucination capabilities of AlphaFold3-style structure prediction models and enables fine-tune free, forward-pass only sequence-structure co-optimization. Structure conditioning at different noise level in the structure prediction stage allows precise control over the sampling space, facilitating tasks from local and global protein optimization to de novo design. We demonstrate the versatility of this approach by optimizing suboptimal structures, rescuing previously unsuccessful designs, designing new biomolecular interactions and generating new protein structures from scratch. Experimental characterization of binder design spanning small molecule, metal ion, protein, and antibody design of phosphorylation-specific peptide revealed high design success rates and excellent structural accuracy. Together, our comprehensive computational and experimental results highlight the broad utility of this framework. We anticipate that HalluDesign will further unlock the modeling and design potential of AlphaFold3-like models, enabling the systematic creation of complex proteins for a wide range of biotechnological applications.

Introduction

Deep learning–based neural networks for protein structure prediction have revolutionized the field, enabling the computational modeling of protein structures with atomic accuracy primarily using sequence information[1-4]. Nevertheless, harnessing the atomic-level precision of these state-of-the-art models into a robust framework for de novo design—rather than just prediction—remains a significant challenge[5]. Recent progress in protein design has leveraged the pretrained structure prediction models[6-9], which can be broadly classified into two categories: gradient-based backpropagation methods and fine-tuning approaches. Gradient-based methods iteratively update sequence and structure compatibility by minimizing network loss[10-12]; however, they typically require high computational costs and often fail to

converge due to the rugged sequence–structure sampling space. In contrast, the fine-tuning approach enable models to generate protein backbones from random noise with forward pass only[13-16]. However, a large number of scaffolds and subsequent sequence sampling are necessary in order to identify sufficient candidates that pass in silico structure prediction filters such as AlphaFold2 or AlphaFold3, prior to experimental characterization.

Recently, AlphaFold3 has extended structure prediction beyond proteins to include small molecules, nucleic acids, ions, and covalent modifications, achieving unprecedented accuracy[17,18]. This breakthrough, along with the various open-source reproductions it inspired[19-21], has spurred a new wave of adaptation and fine-tuning for the modeling and design of diverse biomolecules[22-25]. The AlphaFold3 architecture is primarily composed of the pairformer module, which extracts residue/atom-level and pairwise representations from input features, and a diffusion module, which generates all-atom structures from Gaussian noise, guided by the information extracted from the pairformer. This generative diffusion process inherently leads to "hallucination" – the generation of plausible-looking structures even in the low-confidence regions (Fig. 1a)[17,18]. While this property is an undesirable side-effect for structure prediction, we hypothesized that it could be repurposed for structure optimization and de novo protein design, without requiring fine-tuning or computationally intensive backpropagation. Here, we introduce HalluDesign—a hallucination-driven, forward-pass-only framework that exploits the structure module of AlphaFold3-style models to iteratively optimize and design both protein sequences and structures.

HalluDesign methodology

The real protein sequence–structure landscape is highly rugged[26-28], making it challenging to find optimal solutions through iterative traversing the sequence and structure space. We reasoned that diffusion-based structure prediction models inherently smooth this rugged landscape, which is often observed as 'hallucination' effect (Fig. 1b). To harness this property for optimization and de novo design, we developed HalluDesign, an iterative framework designed to co-optimize protein sequence and structure (Fig. 1c), alternating between a structure update stage and a sequence design stage.

This iterative framework begins with the structure update stage, where the diffusion module of an AlphaFold3-style model leverages its 'hallucination' capabilities to refine the protein structure, guiding it towards a more ideal state conditioned on an input sequence. This process allows plausible conformations to be generated even for suboptimal proteins that lack stable folds and interactions. In the subsequent sequence design stage, an inverse folding deep learning model, such as ProteinMPNN or LigandMPNN[29,30], generates a new amino acid sequence highly compatible with the updated structure. This new sequence is then fed back into the next structure update iteration, serving to better condition the diffusion module and ensure the model operates with a more compatible sequence and structure pair. Crucially, the entire process involves only forward passes through the structure prediction and sequence design models, requiring neither model fine-tuning nor computationally intensive loss backpropagation. The built-in confidence metrics of the structure prediction model, such as pLDDT or ipTM, further provide intrinsic, real-time indicators of the optimization's progress (Supplementary Figs. S8).

The structure update stage provides a unified and controllable strategy by utilizing a truncated diffusion approach. Instead of initializing the diffusion process of the structure generation module of AlphaFold3 with random Gaussian noise, a user-defined noise level is applied to the initial or a previous iteration's

structure. This noised structure then serves as the input for the reverse diffusion process to generate a clean, updated structure. This method offers precise control over the structural exploration by truncating the diffusion trajectory at various timesteps. Shorter trajectories, corresponding to lower noise levels, prompt local refinement, which is well-suited for optimization tasks requiring the preservation of overall topology. Conversely, longer trajectories, corresponding to higher noise levels, enable broader exploration of structure space to escape local minima (Fig. 2a).

Following structure update, the complementary sequence design stage is performed, which is substantially more computationally efficient than the structure update step. First, we employ efficient inverse folding models, such as ProteinMPNN or LigandMPNN, to sample a set of candidate sequences for the updated protein backbone. These sequences are then ranked by CoDP (Contrastive-learning-based Distogram Prediction Model), a fast and lightweight tool we developed to evaluate sequence compatibility with the input structure through distogram prediction and contrastive learning[1,3,4,31,32]. The top-ranked sequence is then selected and serves as the conditioned sequence input for the next structure update stage. The CoDP model was trained on structures from the AlphaFold Database (AFDB)[32]. For each structure, multiple sequences are sampled using ProteinMPNN. CoDP was then trained to distinguish these synthetic sequences, learning to identify those that yield high-confidence predictions in AlphaFold2 and remain structurally consistent with the input backbone, from those sequences that do not.

HalluDesign for protein optimization

While state-of-the-art generative models typically use AlphaFold3 as an external 'oracle' to select high-quality candidates from a large set of designs, we hypothesized that HalluDesign could instead actively refine these existing designs. We first applied HalluDesign to optimize monomeric scaffolds generated by Chroma[33]. Both structural and confidence metrics improved steadily across iterative cycles (Supplementary Figs. S1). We observed distinct optimization behaviors by tuning the diffusion trajectory length: shorter trajectories (e.g., 50 steps) preserved the input topology but converged more slowly, making them ideal for refinement. Longer trajectories (100–200 steps) explored broader structural landscapes and reached high-confidence solutions rapidly, typically within 6–7 cycles (Supplementary Figs. S1). This trade-off was more pronounced for large proteins (>300 aa), which required longer trajectories or additional cycles to accommodate the substantial conformational changes needed for convergence.

To further test the generality of HalluDesign's optimization capabilities, we applied it to scaffolds generated by RSO[34], a state-of-the-art backpropagation-based AlphaFold2 design method. We selected RSO as a challenging test case because it is recognized as one of the most performant methods for monomer generation, particularly for large proteins. Remarkably, our forward-pass-only approach further improved their computational metrics, suggesting that the conformational landscape of AlphaFold3, which HalluDesign leverages, is smoother and more amenable to optimization than the AF2 landscape traversed by gradient-based methods (Supplementary Figs. S2). In contrast, we observed that RFdiffusion partial diffusion were unable to refine these same RSO-generated scaffolds[16,35]. This highlights that protein structure landscape is more accurately modeled by AF3 than that of RFdiffusion, underscoring the necessity of leveraging AF3 as the foundational model for optimization.

Having demonstrated HalluDesign's efficacy on monomers, we next tested its ability to refine biomolecular interactions. We first applied HalluDesign to protein binders generated by RFdiffusion from their published PPI benchmark set, which were under experimental evaluation and already exhibit high confidence metrics. Remarkably, HalluDesign further improved both structural accuracy and confidence across all targets (Fig. 2c). Similarly, structure conditioning with shorter diffusion trajectories achieved comparable optimization performance to longer ones, while better preserving structural integrity. Then we leveraged the all-atom capabilities of the underlying AF3 model by optimizing small-molecule binders. We generated initial binders for a diverse set of ligands using the RFdiffusion all-atom pipeline[36], which shows modest model quality by AF3 metrics. Applying HalluDesign significantly improved the computational metrics across all targets. This enhancement was particularly pronounced for large, polar ligands with many rotatable bonds (Supplementary Figs. S5-7). We attribute this success to HalluDesign's ability to co-optimize both the protein structure and the ligand conformation simultaneously—a distinct advantage over the fixed-ligand approach used by the original generative mode which is absent in the initial, fixed-ligand generation process.

HalluDesign for scaffold-based de novo protein design

Beyond refining existing structures, we hypothesized that HalluDesign could design new interaction with targets using existing protein scaffolds that do not naturally interact with targets. To test this, we performed de novo design of small molecule binders on two distinct topologies. First, we used the nuclear transport factor 2 (NTF2) scaffold[7,37], a compact and engineerable fold known to accommodate diverse ligands. Second, to verify the generality of this approach, we used a large set of pseudocycle scaffolds with pockets aligned along their pseudo-symmetric axis (Fig. 3a)[38]. For both scaffold types, our approach involved placing a randomly selected ligand conformer at the center of the protein—a strategy we term "black hole initialization"—and applying the HalluDesign co-optimization loop (Supplementary Figs. S9-10). Moreover, we truncated the diffusion step to 150 unless otherwise specified for design, as this configuration provided similar optimization performance to the default 200-step diffusion trajectory while reducing computational consumption.

Despite the initial ligands placements for both NTF2 and pseudocycle scaffolds lacking any established protein-ligand interactions, HalluDesign simultaneously co-optimized ligand poses, protein side chains, and protein-ligand interactions, converging on plausible complex conformations with high in-silico success rates. This success rate was inversely proportional to ligand size and polarity: for small, hydrophobic ligands, the rate exceeded 60% (Supplementary Figs. S9), while it was considerably lower for larger, polar targets. This direct co-optimization approach bypasses the need for initial docking steps commonly used with previous methods[39]. Furthermore, HalluDesign achieved significantly higher in silico success rates on these tasks compared to both diffusion-based models like RFdiffusion all-atom and backpropagation methods like BoltzDesign1.

Antibody and nanobody design remain challenging due to flexible loop-dominated interfaces, and we hypothesized that HalluDesign could leverage AlphaFold3's strong performance on antibody–protein complexes to facilitate CDR loop reconstruction. The antibody framework is an ideal predefined scaffold platform for targeting a vast diversity of molecular interfaces task, as it provides a highly conserved and general scaffold that presents its hypervariable loops (CDRs) (Fig. 3b). Using Herceptin as a representative scFv[15,40], HalluDesign successfully rebuilt and optimized CDR loops from noisy initial

structures and polyglycine inputs, generating antibodies that formed specific interactions with PDL1 and phosphorylated peptides—demonstrating its capability for structure-targeting and modification-specific antibody design. Both structural accuracy and confidence metrics improved progressively over iterative optimization cycles (Supplementary Figs. S11-12).

HalluDesign for from scratch protein design

We next explored another challenging application: de novo protein design from scratch. This general strategy initiates the designed protein with pure Gaussian noise as its coordinates and a random protein sequence, without any structure input (Fig. 3c). After a single prediction pass generates an initial, plausible conformation, this structure is then fed into our standard HalluDesign cycles. We first applied this approach to design cyclic peptide binders against protein targets. This application involved two modifications to the general strategy: the target protein structure was provided as a template, and we modified the peptide's position index in AF3 to enforce a head-to-tail cyclization[8,41,42] (Supplementary Figs. S13). Leveraging AF3's unique ability to model covalent modifications, this framework also allowed us to incorporate non-natural amino acids (NCAAs)—either to form covalent bonds with a target cysteine or to engineer side-chain-to-side-chain cyclization which broader explore the design space[40,43] (Supplementary Figs. S14).

However, when this de novo strategy was applied to proteins substantially larger than peptides, the AlphaFold3-based framework failed. The optimization process consistently collapsed into low-complexity conformations, such as a single long helix. We hypothesized this is because the random starting point lies too far from the learned manifold of natural proteins, preventing the optimization from finding a convergence path (Fig. 3d). To overcome this, we substituted the AF3 engine with Protenix, an AF3-style architecture built on protein language model embeddings[20]. We reason that the rich contextual information from the language model effectively smooths this rugged landscape, allowing Protenix to find a convergence path even from such highly 'out-of-manifold' starting points.

This capability for unconditional structure generation lays a solid foundation for overcoming the dependency on predefined scaffolds in designing new molecular interactions. By providing a target molecule as input—such as a protein, ligand, or nucleic acid—the Protenix based HalluDesign can directly leverage language model to interpret even random initial sequences, exploring plausible structural space and identifying conformations capable of forming possible interactions with the target. Consequently, Protenix-based HalluDesign enable the *de novo* design of diverse biomolecular binders from scratch, harnessing the intrinsic sequence–structure knowledge embedded within language model–powered structure prediction frameworks.

Experimental validation of protein-ligand interactions

We designed binders for a range of key endogenous small molecules characterized by multiple rotatable bonds and polar groups, including cyclic adenosine monophosphate (cAMP), adenosine triphosphate (ATP), and adrenaline (AD). In addition, we targeted small molecules with relatively large, rigid, and hydrophobic features, such as cholic acid (CA) and the HIV drug dolutegravir (DTG). We also developed binders for the amino acid tryptophan (TRP) and extended our approach to metal ion ligand zinc (Zn^{2+}).

These targets representing a broad spectrum of ligand sizes and chemical properties (Fig. 4a). NTF2 scaffolds were used for ATP, CA, DTG, and AD, while pseudocycle scaffolds or hybrid NTF2 scaffold were employed for cAMP, TRP and Zn^{2+}. We initialized each protein-ligand complex structure using the "black hole initialization" strategy and applied HalluDesign to generate the final models. Unlike conventional high-throughput design approaches that require screening large numbers of candidates[38,44], HalluDesign achieved binder discovery in a low-throughput setting. For each small-molecule target, fewer than 16 designed proteins were experimentally characterized, most of which were successfully expressed in *E. coli*. ITC measurements confirmed one or more binders for every target. Mutational analysis of the binding pockets confirmed that ligand binding occurs at the intended site and that binding strength is highly dependent on the engineered interactions. Cross-binding assays demonstrated high specificity; for example, the TRP binder did not bind to the other aromatic amino acids, tyrosine or phenylalanine, highlighting the platform's potential for designing amino acid-specific binders. Similarly, the AD binder exhibited only minimal binding to its metabolic product, norepinephrine (NE), which differs from AD by a single methyl group, underscoring the fine specificity of our designs.

We successfully obtained protein-ligand complex crystal structures for cAMP, AD and Zn^{2+} (Fig. 4b). The cAMP-binder complex was resolved at 3.1 Å, with clear electron density for cAMP and an overall structure closely matching the design model (Cα RMSD of 0.74 Å for the protein and 1.07 Å for the ligand using the protein as reference). The sidechain configurations of pocket residues were also accurately recapitulated. Notably, all seven hydrogen bonds predicted in the design model between the protein and cAMP were observed in the crystal structure. The AD binder complex was resolved at 2.4 Å resolution, showing good overall agreement with the design model (Cα RMSD = 0.65 Å). In both the design model and the crystal structure, AD forms strong hydrogen-bonding interactions with protein binder. A leucine residue faces the methyl group, which explains the higher specificity for AD compared to NE. The zinc binder complex was resolved at 1.6 Å, with overall agreement with the design model (Cα RMSD of 0.81 Å). Zinc displayed tetrahedral coordination geometry in both the design model and the crystal structure, with all four coordinating residues precisely positioned. We also obtained apo structures for the DTG and TRP binders at 2.0 Å resolution, respectively, also in close atomic agreement with the design models (Cα RMSD of 0.31 Å and 0.81 Å). Pocket residue sidechain orientations in these structures mirrored those in the predicted models, further demonstrating atomic-level accuracy. The high experimental success rate, pocket point mutation analysis, and the atomic agreement between the crystal structures and design models showcase the versatility and accuracy of our platform for creating specific and efficient protein-ligand interactions.

Experimental validation of protein-protein interactions

We next applied HalluDesign to rescue previously unsuccessful designs in the RFdiffusion protein–protein interaction (PPI) binder design task. In these experiments, we truncated the AlphaFold3 diffusion trajectory and conditioned structure prediction on either the initial complex structure or intermediate predictions from earlier optimization iterations. The resulting design models exhibited improved computational metrics for the binders, along with lower RMSD deviation from the original input structures. We selected 16 rescued designs each for PD-L1 and IL7RA for experimental validation, with each candidate derived from a previously failed design (Fig. 5a). Biolayer interferometry (BLI) assays demonstrated that 12 out of 16 PD-L1 binders successfully bound their target, while 8 out of 16 IL7RA binders showed binding to IL7RA. Overall, these designs displayed low nanomolar affinities for their

respective targets. The high success rate in these rescue experiments underscores the efficiency of the HalluDesign approach in optimizing protein–protein interactions.

Targeting amyloid β (Aβ) with high specificity remains a crucial strategy in Alzheimer's disease (AD) therapy[45]. Phosphorylation at Ser8 (pS8-Aβ) promotes toxic oligomer formation and accelerates aggregation, so designing antibodies that selectively recognize this pathogenic form could enable targeted intervention against the most neurotoxic Aβ species[46]. Using HalluDesign, we generated new CDR loops for the Herceptin scFv to interact specifically with phosphorylated Aβ. Unlike conventional methods that rely on predefined peptide positions or predefined target peptide second structure[16,47], HalluDesign models free sampling of peptide conformations. Experimental results demonstrated specific binding to the phosphorylated form of Aβ, with no detectable interaction with native Aβ (Fig. 5b). These findings indicate that the phosphate group is critical for antibody-peptide interaction. This is consistent with our design model, where adjacent Arg residues near the phosphorylation site function as an electrostatic switch, ensuring specific binding to the only phosphorylated form of Aβ.

We then applied Protenix-based HalluDesign to the de novo design of protein binders with biomedical relevance—specifically targeting PD-L1 and FGFR2, which are key proteins in immunotherapy and cancer development[48,49]. Starting from random binder sequences, Protenix-based HalluDesign produced models with strong computational metrics. The resulting binders interacted favorably with the beta sheet face of PD-L1 and FGFR2, demonstrating both high shape and chemical complementarity to the targets. For each target, we experimentally tested 16 de novo designed binders and observed high success rates: 8 out of 16 PD-L1 binders showed binding signals in BLI experiments, with the strongest affinity reaching 32 pM, and 13 out of 16 FGFR2 binders showed binding, with the best achieving digital nanomolar affinity (Fig. 5c). The impressive success rates and high binding affinities achieved by HalluDesign strongly demonstrate its power for de novo binder protein generation, highlighting its promise for developing binding reagents against a range of biomolecular targets.

Conclusions

In summary, we present HalluDesign, a general all-atom framework that fundamentally reframes the utility of state-of-the-art structure prediction models. Instead of employing AlphaFold3-style models as passive, external 'oracles' to filter candidate designs, HalluDesign leverages the hallucination effect in their diffusion module for iterative sequence-structure optimization. Our approach smooths the rugged sequence-structure landscape, enabling rapid convergence without costly backpropagation or model fine-tuning. Furthermore, its variable noise conditioning enables a wide range of applications, from acting as a 'plug-in' tool to optimize any existing design to de novo design from scratch.

We have demonstrated the broad versatility and atomic-level accuracy of this approach, achieving high experimental success rates in tasks ranging from optimizing existing designs to the de novo design of protein binders, small-molecule binders, and phosphorylation-specific antibodies. As structure prediction models continue to evolve, achieving higher accuracy across a wider range of molecular modalities, the performance of this hallucination-driven design strategy will be further enhanced, opening a wider exploration space for protein design based on these powerful prediction models.

Code Availability

We used the following publicly available code repositories in this work:
LigandMPNN (https://github.com/dauparas/LigandMPNN),
AlphaFold3 (https://github.com/google-deepmind/alphafold3),
Protenix (https://github.com/bytedance/Protenix),
Chai-1 (https://github.com/chaidiscovery/chai-lab), and
Boltz1 (https://github.com/jwohlwend/boltz).

All data and source code developed in this study are available at
CoDP (https://github.com/MinchaoFang/CoDP) and
HalluDesign (https://github.com/MinchaoFang/HalluDesign).

Acknowledgements

We appreciate the members of Cao's labs for their valuable discussions. We extend our special thanks to Dang Lab in Westlake University for valuable discussion about covalent binding cyclic peptide.

References

1 Du, Z. *et al.* The trRosetta server for fast and accurate protein structure prediction. *Nature protocols* **16**, 5634-5651 (2021).

2 Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *biorxiv*, 2021.2010. 2004.463034 (2021).

3 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *nature* **596**, 583-589 (2021).

4 Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023).

5 Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547-552 (2021).

6 Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387-394 (2022).

7 Yeh, A. H.-W. *et al.* De novo design of luciferases using deep learning. *Nature* **614**, 774-780 (2023).

8 Li, Q., Daumiller, D. & Bryant, P. RareFold: Structure prediction and design of proteins with noncanonical amino acids. *bioRxiv*, 2025.2005. 2019.654846 (2025).

9 Wicky, B. I. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56-61 (2022).

10 Norn, C. *et al.* Protein sequence design by conformational landscape optimization. *Proceedings of the National Academy of Sciences* **118**, e2017228118 (2021).

11 Frank, C., Schiwietz, D., Fuß, L., Ovchinnikov, S. & Dietz, H. Alphafold2 refinement improves designability of large de novo proteins. *bioRxiv*, 2024.2011. 2021.624687 (2024).

12 Pacesa, M. *et al.* BindCraft: one-shot design of functional protein binders. *bioRxiv*, 2024.2009. 2030.615802 (2024).

13 Liu, Y. *et al.* De novo protein design with a denoising diffusion network independent of pretrained structure prediction models. *Nature Methods* **21**, 2107-2116 (2024).

14 Wang, C. *et al.* Proteus: exploring protein structure generation for enhanced designability and efficiency. *bioRxiv*, 2024.2002. 2010.579791 (2024).

15 Bennett, N. R. *et al.* Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*, 1-11 (2025).

16 Liu, C. *et al.* Diffusing protein binders to intrinsically disordered proteins. *Nature*, 1-9 (2025).

17 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).

18 Desai, D. *et al.* Review of AlphaFold 3: transformative advances in drug design and therapeutics. *Cureus* **16** (2024).

19 team, C. D. *et al.* Chai-1: Decoding the molecular interactions of life. *BioRxiv*, 2024.2010. 2010.615955 (2024).

20 Team, B. A. A. S. *et al.* Protenix-advancing structure prediction through a comprehensive AlphaFold3 reproduction. *BioRxiv*, 2025.2001. 2008.631967 (2025).

21 Wohlwend, J. *et al.* Boltz-1 democratizing biomolecular interaction modeling. *BioRxiv*, 2024.2011. 2019.624167 (2025).

22 Cho, Y., Pacesa, M., Zhang, Z., Correia, B. & Ovchinnikov, S. BoltzDesign1: Inverting All-Atom Structure Prediction Model for Generalized Biomolecular Binder Design. *bioRxiv*, 2025.2004. 2006.647261 (2025).

23 Ren, M. *et al.* PXDesign: Fast, Modular, and Accurate De Novo Design of Protein Binders. *bioRxiv*, 2025.2008. 2015.670450 (2025).

24 Team, C. D. *et al.* Zero-shot antibody design in a 24-well plate. *bioRxiv*, 2025.2007. 2005.663018 (2025).

25 Zhang, O. *et al.* ODesign: A World Model for Biomolecular Interaction Design. *arXiv preprint arXiv:2510.22304* (2025).

26 Leaver-Fay, A. *et al.* in *Methods in enzymology* Vol. 523 109-143 (Elsevier, 2013).

27 Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation* **13**, 3031-3048 (2017).

28 Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nature reviews molecular cell biology* **20**, 681-697 (2019).

29 Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).

30 Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. *Nature Methods*, 1-7 (2025).

31 Hadsell, R., Chopra, S. & LeCun, Y. in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06).* 1735-1742 (IEEE).

32 Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **50**, D439-D444 (2022).

33 Ingraham, J. B. *et al.* Illuminating protein space with a programmable generative model. *Nature* **623**, 1070-1078 (2023).

34 Frank, C. *et al.* Scalable protein design using optimization in a relaxed sequence space. *Science* **386**, 439-445 (2024).

35 Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023).

36 Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).

37 Basanta, B. *et al.* An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proceedings of the National Academy of Sciences* **117**, 22135-22145 (2020).

38 An, L. *et al.* Binding and sensing diverse small molecules using shape-complementary pseudocycles. *Science* **385**, 276-282 (2024).

39 Fry, B., Slaw, K. & Polizzi, N. F. Zero-shot design of drug-binding proteins via neural selection-expansion. *bioRxiv*, 2025.2004. 2022.649862 (2025).

40 Zhu, C. *et al.* Epitope-directed antibody elicitation by genetically encoded chemical cross-linking reactivity in the antigen. *ACS Central Science* **9**, 1229-1240 (2023).

41 Li, Q., Vlachos, E. N. & Bryant, P. Design of linear and cyclic peptide binders from protein sequence information. *Communications Chemistry* **8**, 211 (2025).

42 Rettie, S. A. *et al.* Cyclic peptide structure prediction and design using AlphaFold2. *Nature Communications* **16**, 4730 (2025).

43 Wang, X. S. *et al.* A genetically encoded, phage-displayed cyclic-peptide library. *Angewandte Chemie* **131**, 16051-16056 (2019).

44 Lee, G. R. *et al.* Small-molecule binding and sensing with a designed protein family. *bioRxiv*, 2023.2011. 2001.565201 (2023).

45 Pinheiro, L. & Faustino, C. Therapeutic strategies targeting amyloid-β in Alzheimer's disease. *Current Alzheimer Research* **16**, 418-452 (2019).

46 Kumar, S. & Walter, J. Phosphorylation of amyloid beta (Aβ) peptides–A trigger for formation of toxic aggregates in Alzheimer's disease. *Aging (Albany NY)* **3**, 803 (2011).

47 Wu, K. *et al.* Design of intrinsically disordered region binding proteins. *Science* **389**, eadr8063 (2025).

48 Li, P. *et al.* FGFR2 promotes expression of PD-L1 in colorectal cancer via the JAK/STAT3 signaling pathway. *The Journal of Immunology* **202**, 3065-3075 (2019).

49 Palakurthi, S. *et al.* The combined effect of FGFR inhibition and PD-1 blockade promotes tumor-intrinsic induction of antitumor immunity. *Cancer immunology research* **7**, 1457-1471 (2019).

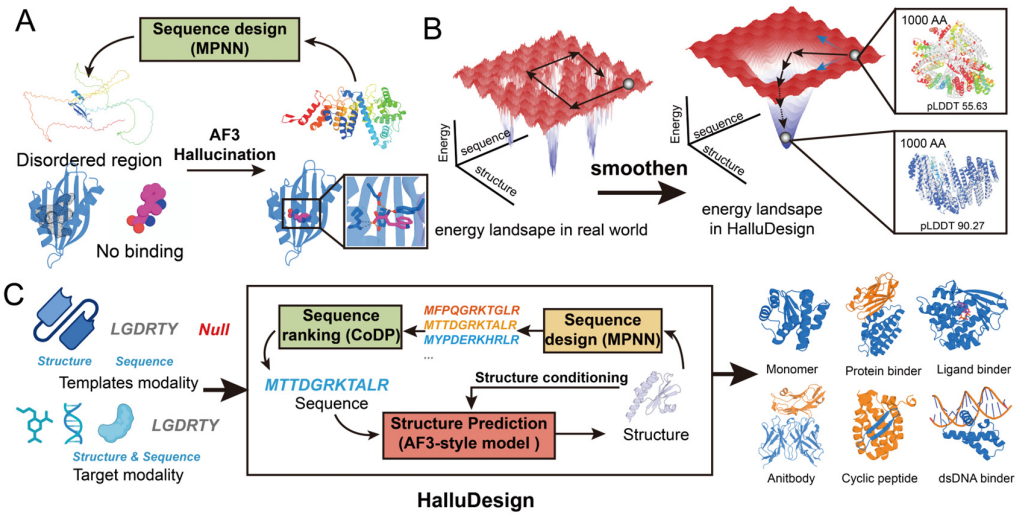


Figure 1:

Overview of the HalluDesign framework illustrating its principles, architecture, optimization strategies, and design pipeline. A-C

(A) Structure Hallucination serves as a potent structural optimization tool that can "hallucinate" ligand binding sites and artificially stabilize protein folds, thereby enabling more accurate and efficient design processes.

(B) By smoothing the sequence – structure landscape, our approach minimizes the risk of getting trapped in local minima. This allows for the optimization of up to 1000 low-quality Chroma backbones into high-quality, refined structures.

(C) Overall Architecture of the HalluDesign Framework: The system is versatile, capable of handling a wide range of biomolecular targets. It can optionally integrate structure or template data, such as antibody frameworks or ligand-binding topologies. In the absence of templates, HalluDesign is fully capable of generating proteins from scratch.

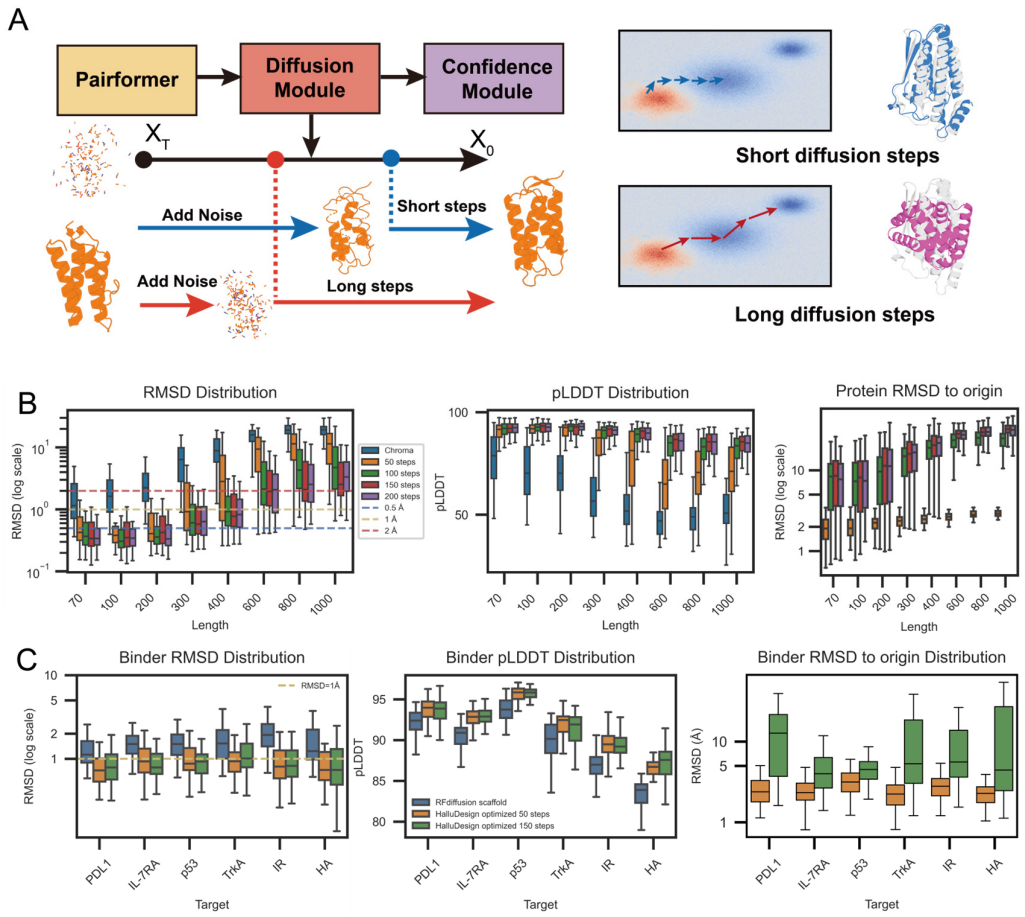


Figure 2:

Different steps can serve different purpose. A-C

(A) Illustration of the effect of diffusion trajectory length on optimization. Longer trajectories explore a broader structural space and produce larger conformational changes, whereas shorter trajectories focus on local refinement and better preserve the original structure.

(B) Final RMSD, PLDDT, and RMSD-to-origin metrics across different optimization steps. For low-quality backbones such as *Chroma*, increasing the number of optimization steps yields a markedly improved optimized state.

(C) Final binder RMSD, binder PLDDT, and binder RMSD-to-origin metrics across different optimization steps. For high-quality Binder complex backbones from RFdiffusion Supplementary File, increasing the number of optimization steps provides no substantial improvement, while causing a larger deviation from the original structure. .

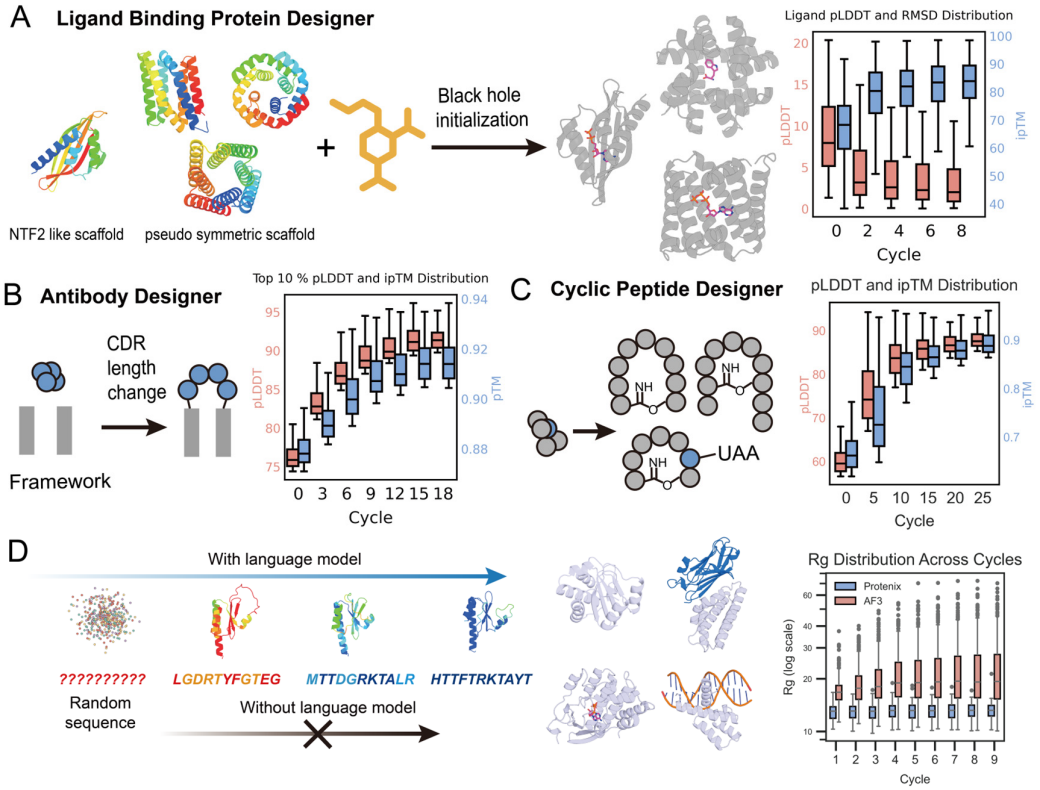


Figure 3

HalluDesign act as predefined scaffold designer. A-D

(A) HalluDesign utilizes scaffolds derived from NTF2 scaffold from previous study or our generated pseudocycle scaffolds. Ligands are centered within the protein structure with random rotations, a process we call Black Hole Initialization. Our model excels in handling ligands with flexible conformations. The ligand pLDDT distribution, RMSD distribution, protein pLDDT, and ipTM scores change with different cycles, providing insight into the refinement of structure quality during the design process

(B) Our method is particularly effective at rebuilding the CDR regions of antibodies with modified lengths. As design cycles progress, both the pLDDT and pTM values consistently increase, indicating improvements in structural accuracy and stability.

(C) Our method is also effective at building cyclic peptides with head-to-tail and side-to-side topologies. As the design cycles progress, both pLDDT and ipTM values consistently increase, indicating improvements in model confidence and interface quality.

(D) Methodology for designing protein and all bimolecular binder from scratch.

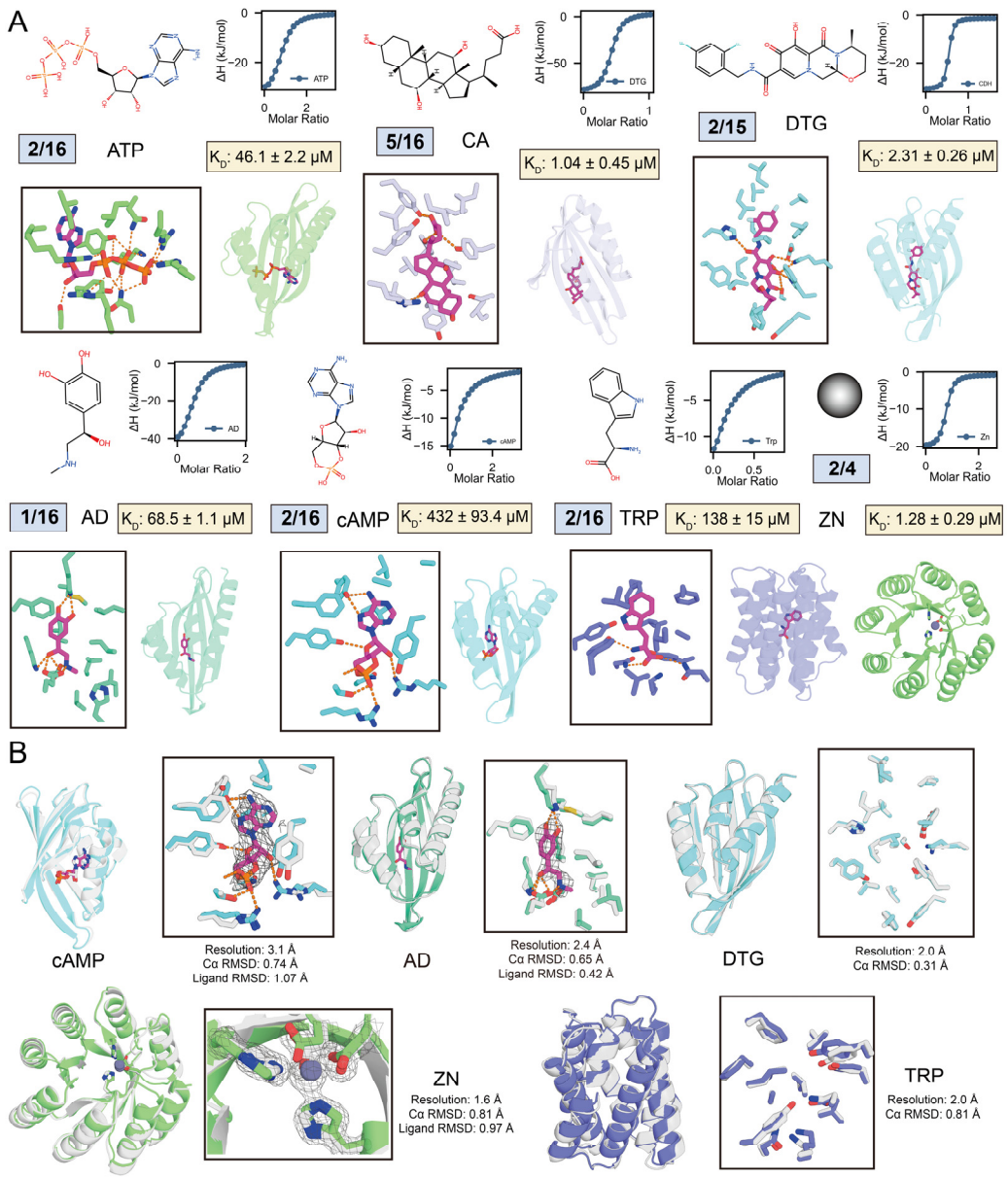


Figure 4:

Ligand binder design. A-B

(A) The design of ligand binders using HalluDesign includes various ligands such as DTG, CDH, ATP, and cAMP, which are bound to the NTF2 scaffold. For TRP and ZN binding, pseudocycle scaffolds are used.

(B) The solved structures are aligned with the corresponding designed models to validate the accuracy and reliability of the design process. White represents the solved protein structure, and the ligand is shown using the electron density map.

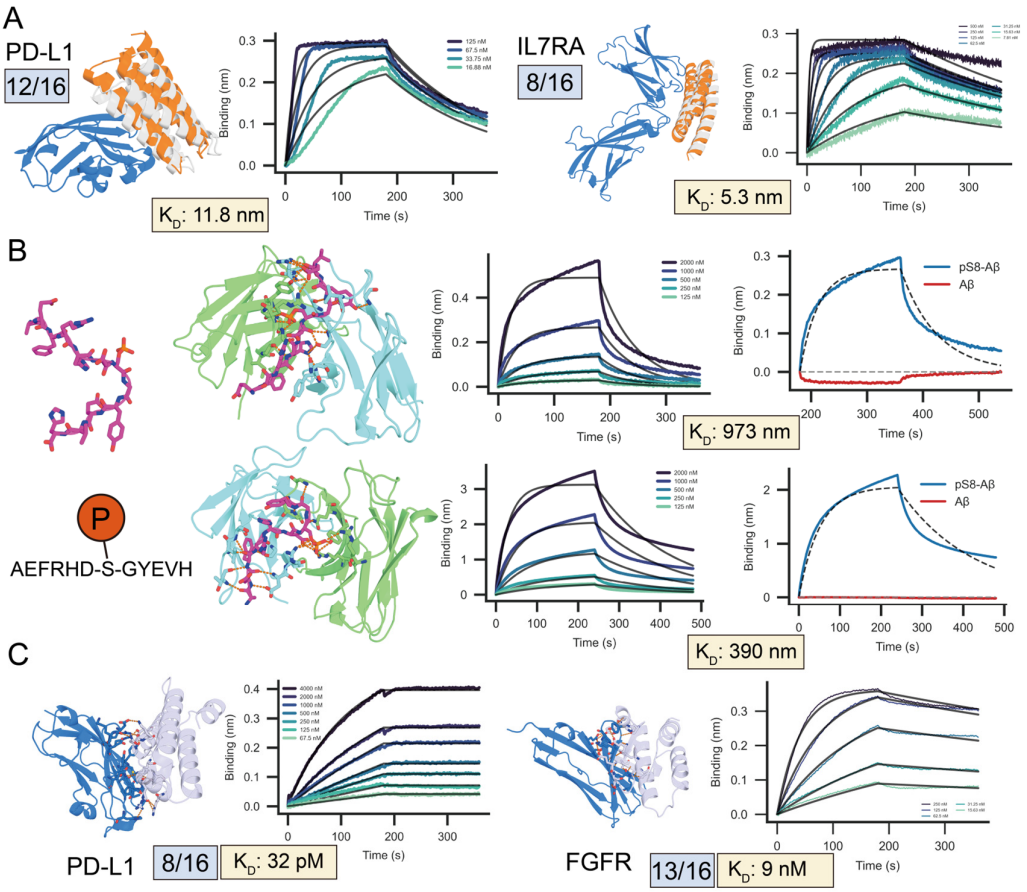


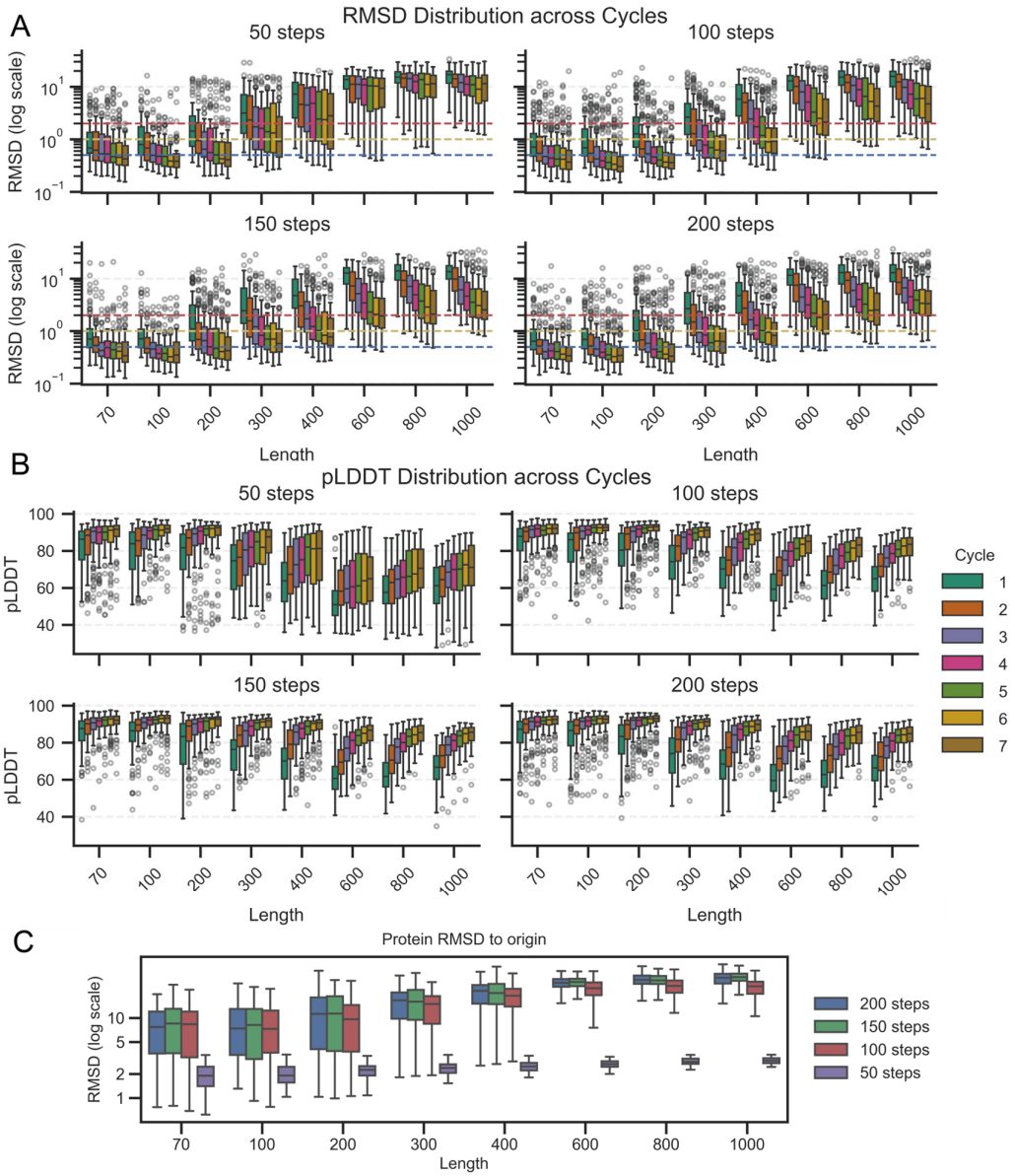
Figure 5:

PPI interactions Optimization and Antibody design. A-C

(A) Optimized Failed Scaffold from RFdiffusion for PDL1: RFdiffusion reports a success rate of 12/95 for PDL1, whereas our method achieves a higher success rate of 12/16. Optimized Failed Scaffold from RFdiffusion for IL7RA: RFdiffusion reports a success rate of 32/96 for IL7RA, while our method achieves a higher success rate of 8/16. White represents the original design, blue indicates the target protein, and orange denotes the optimized structure.

(B) BLI binding affinities are measured for two designed antibodies targeting the phosphorylated Serine (S) at site 8 of Aβ. The binding affinities are also compared for both non-phosphorylated Aβ and phosphorylated Aβ at site 8, with measurements taken at a concentration of 1000 nM. Green and blue represents the two domains of designed ScFv, red indicates the target peptide.

(C) BLI binding affinity measurement for PD-L1 and FGFR best binder which we designed from scratch. White represents the designed complex, blue indicates the target protein.

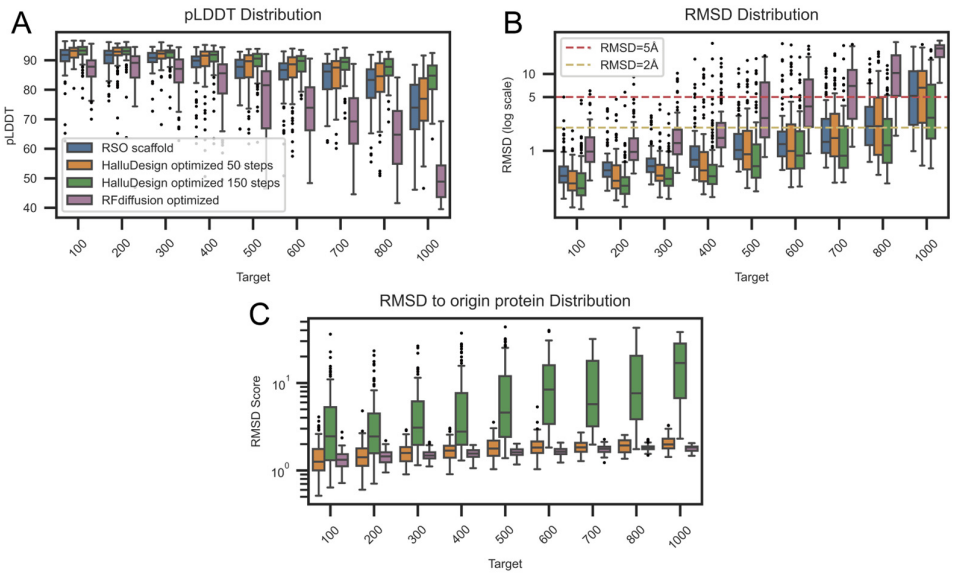


Supplementary Information Figure 1: Effects of different optimization steps on Chroma backbone monomer refinement. A-C

(A) RMSD curves of Chroma backbone optimization under different step settings.

(B) pLDDT curves of Chroma backbone optimization under different step settings.

(C) RMSD to the original structure under different levels of perturbation during Chroma backbone optimization.



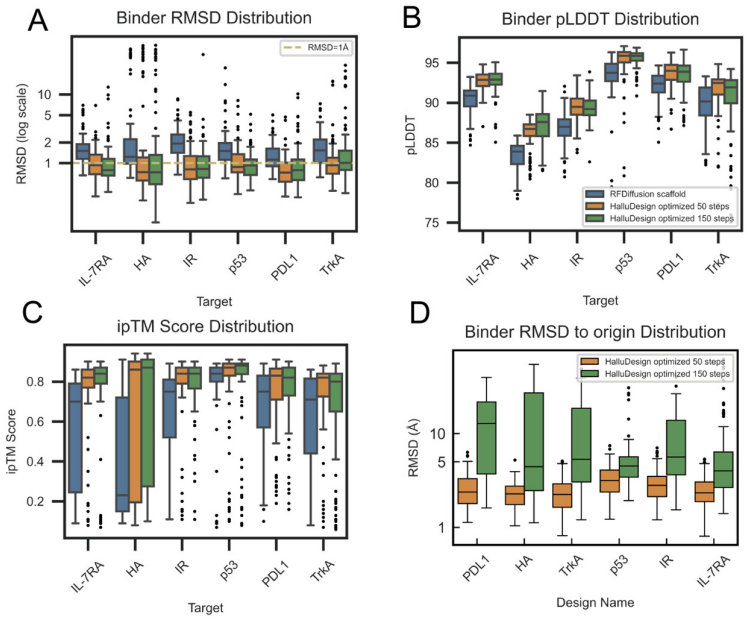
Supplementary Information Figure 2: Benchmark of HalluDesign optimization, origin scaffold and RFdiffusion de-optimization on the RSO monomer. A-C

Comparison between HalluDesign AF3-based optimization and the de-optimization effect of RFdiffusion partial diffusion on the same RSO scaffold. For the RFdiffusion partial diffusion task, RFD was applied with 12 steps, corresponding to a total diffusion process of 50 steps, while HalluDesign was applied with 50 steps and 150 steps, corresponding to a total diffusion process of 200 steps. .

(A) pLDDT scores of the original RSO scaffold, RFdiffusion partially diffusion de-optimized scaffold, and HalluDesign-optimized scaffolds under different step settings.

(B) RMSD values of the original RSO scaffold, RFdiffusion partially diffusion de-optimized scaffold, and HalluDesign-optimized scaffolds under different step settings.

(C) RMSD relative to the original RSO scaffold for the RFdiffusion partially diffusion de-optimized scaffold and HalluDesign-optimized scaffolds under different step settings.



Supplementary Information Figure 3: Benchmark of HalluDesign optimization on RFdiffusion PPI scaffolds from the supplementary dataset. A-D

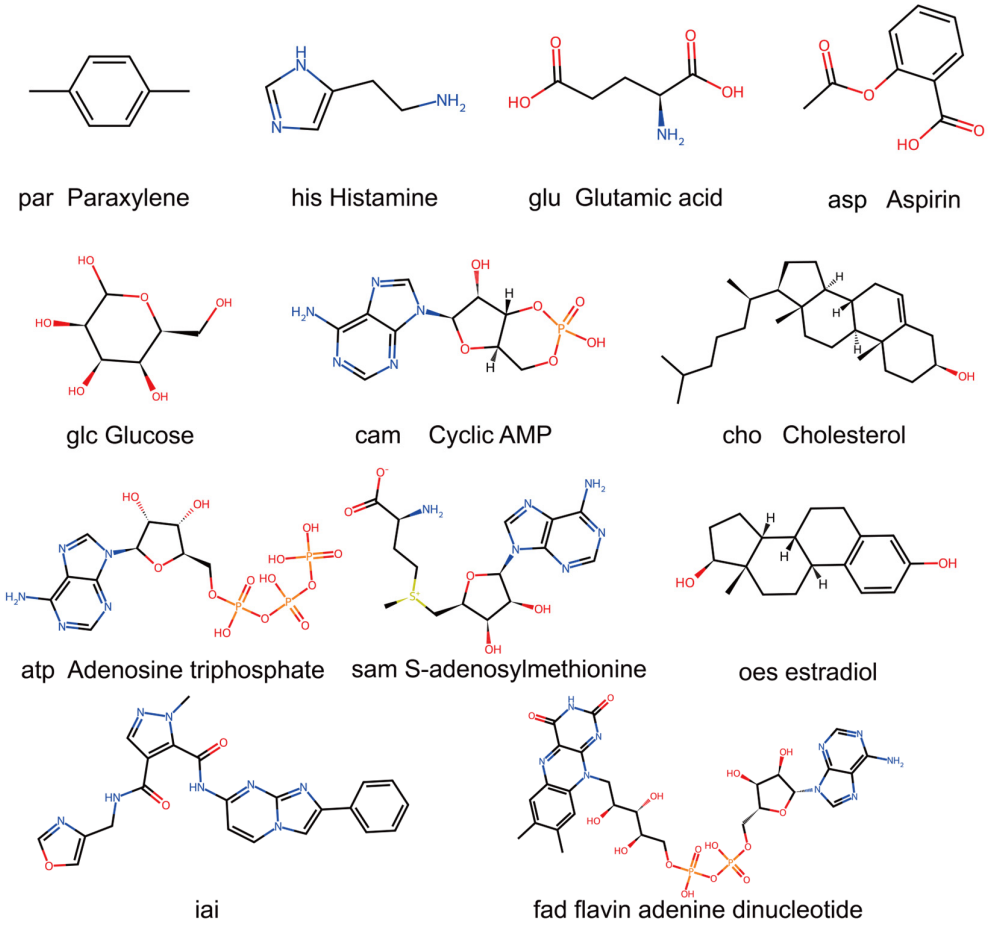
HalluDesign AF3-based optimization was performed on protein binder scaffolds generated by RFdiffusion. These scaffolds were previously experimentally validated and demonstrated strong *in silico* performance metrics. The results indicate that increasing the number of optimization steps leads to substantial structural deviations of the optimized scaffolds from their original counterparts.

(A) Binder RMSD (aligned by target) curves of RFdiffusion scaffolds under different optimization step settings.

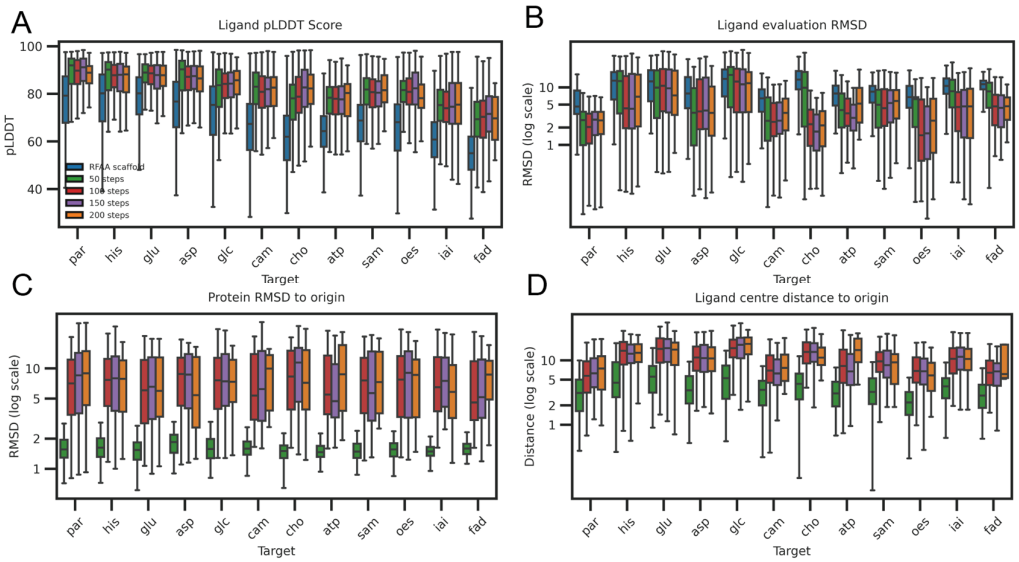
(B) Binder pLDDT curves of RFdiffusion scaffold optimization under different step settings.

(C) AlphaFold-Multimer ipTM curves of usion scaffold optimization under different step settings.

(D) Binder RMSD (aligned by target) between the original RFdiffusion scaffolds and the optimized scaffolds under different optimization step settings.



Supplementary Information Figure 4: All Ligand structures, common names, and abbreviations in our ligand benchmark.



Supplementary Information Figure 5: Benchmark of ligand-binding protein scaffold optimization under different step settings. A-D

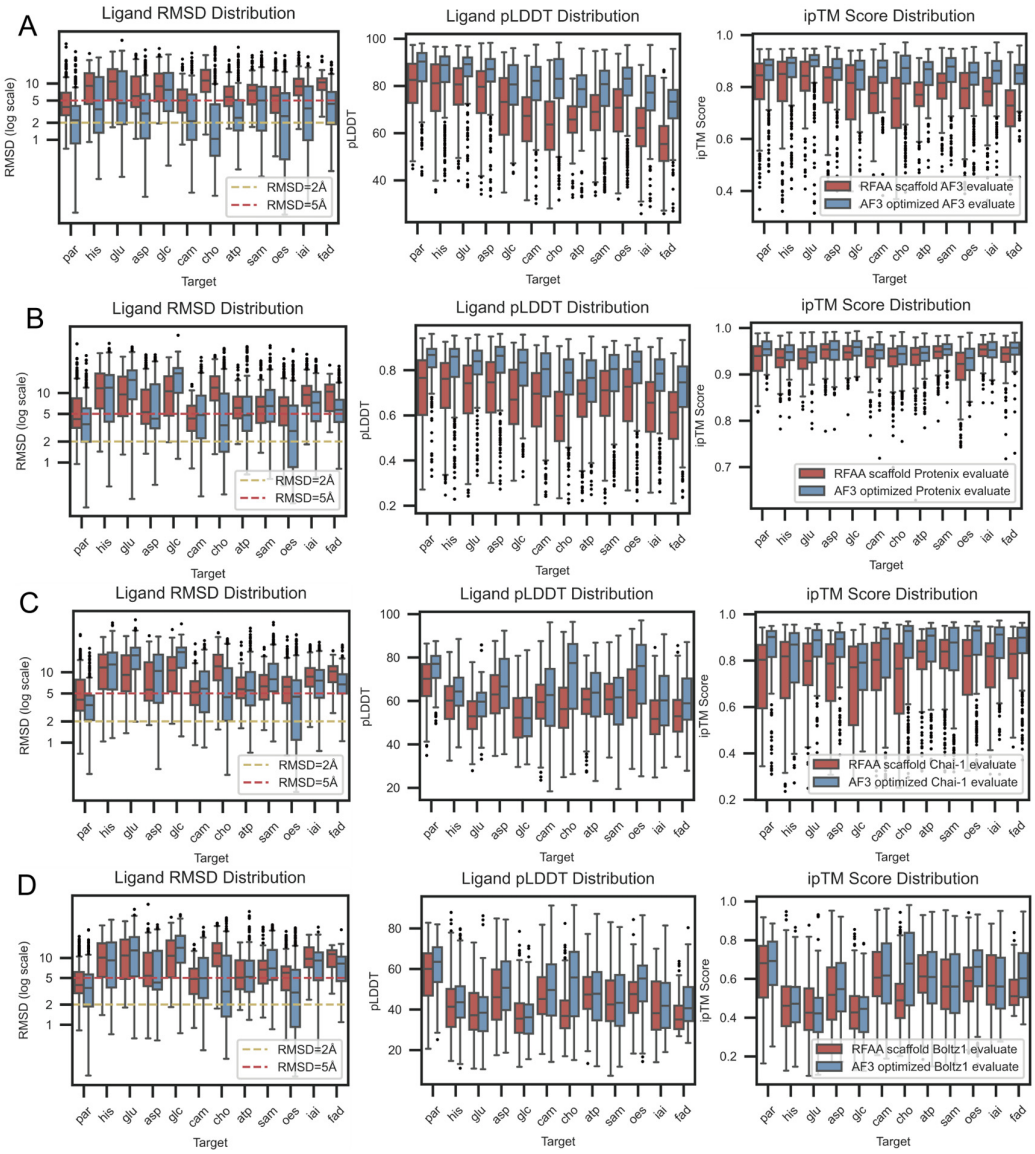
Seven optimization cycles were performed, with both optimization and evaluation conducted using the AF3 framework. The metrics from the final cycle were used for plotting.

(A) Effect of different HalluDesign step settings on the RMSD of twelve ligand-binding proteins relative to the original scaffold.

(B) Effect of different HalluDesign step settings on the center-of-mass distance of twelve ligands relative to the original scaffold.

(C) Effect of different HalluDesign step settings on the pLDDT scores of twelve ligand-binding proteins.

(D) Effect of different HalluDesign step settings on the evaluation RMSD metrics of twelve ligand-binding proteins.



Supplementary Information Figure 6: Benchmark of HalluDesign AF3 optimization on protein–ligand complexes using different structure predictors (AF3, Protenix, Chai-1, and Boltz1), A-D

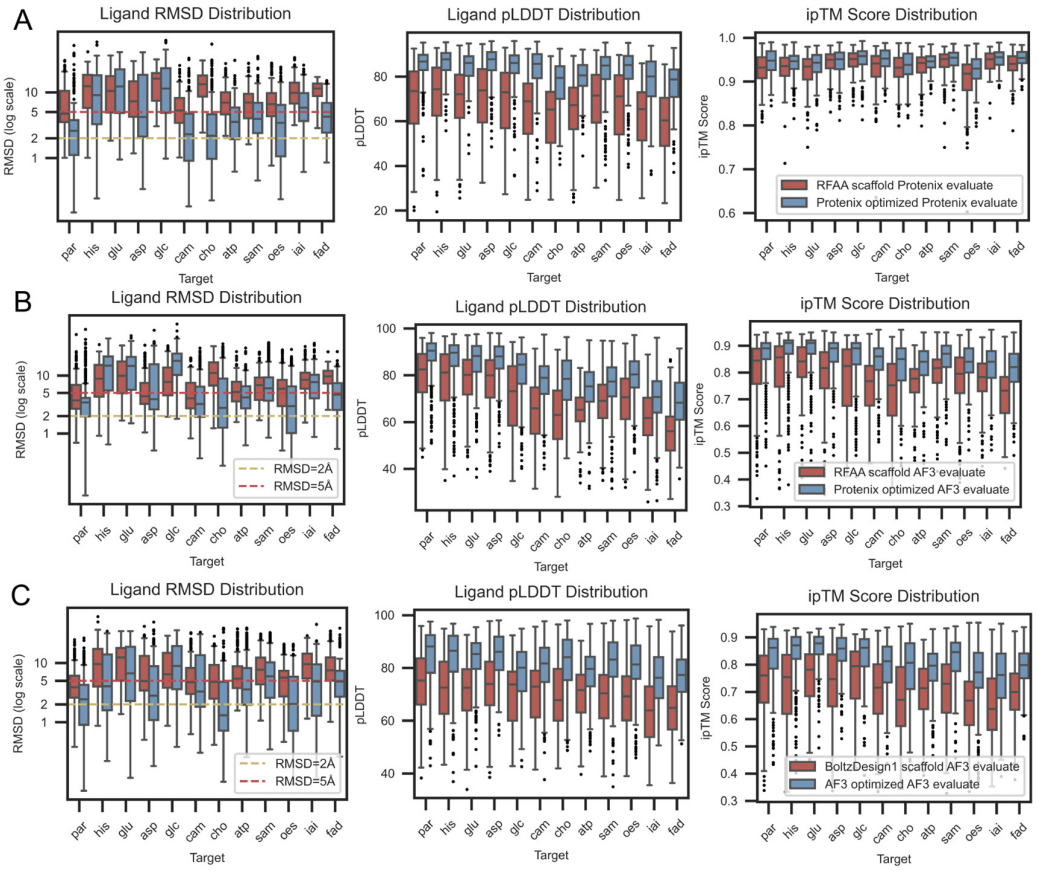
Red line indicates 5 Å RMSD, and yellow line indicates 2 Å RMSD.

(A) Optimization with HalluDesign AF3, evaluated by AF3 without MSA or template inputs.

(B) Optimization with HalluDesign AF3, evaluated by Protenix without MSA but with ESM embeddings.

(C) Optimization with HalluDesign AF3, evaluated by Chai-1 without MSA or template inputs.

(D) Optimization with HalluDesign AF3, evaluated by Boltz1 without MSA or template inputs.

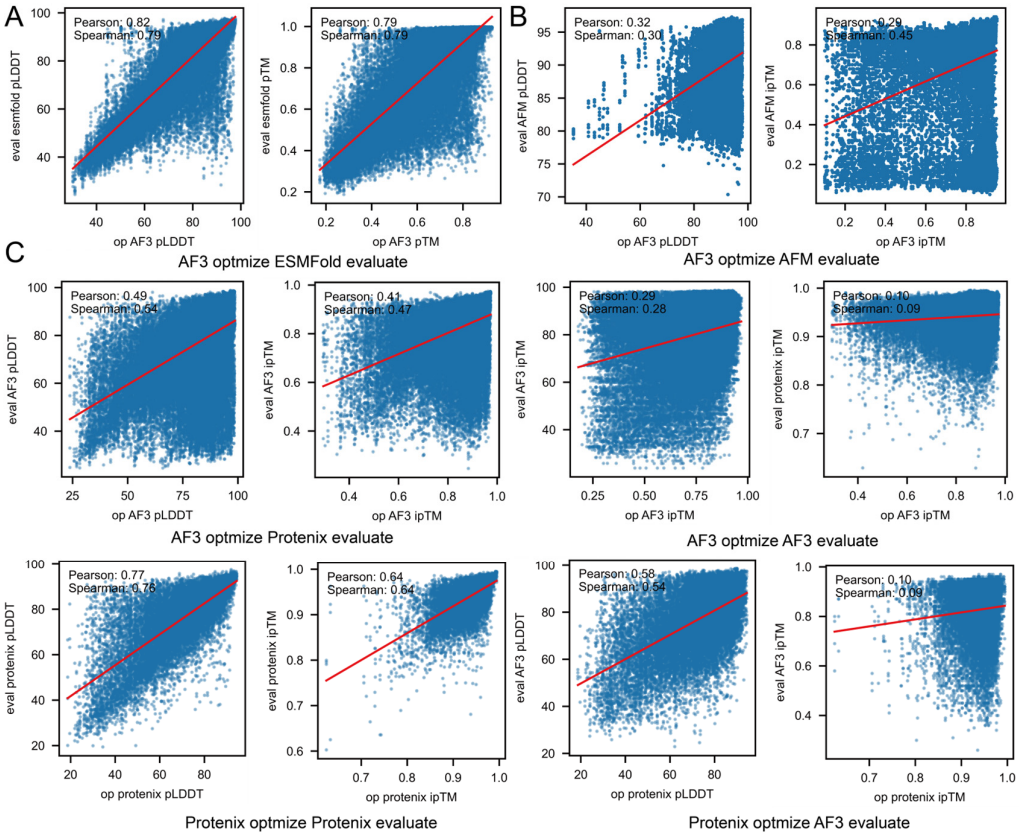


Supplementary Information Figure 7: Benchmark of HalluDesign optimization on protein–ligand complexes using different optimizer and different source scaffold. A-C

(A) Optimization with HalluDesign Protenix, evaluated by Protenix without MSA but with ESM embeddings.

(B) Optimization with HalluDesign Protenix, evaluated by AF3 without MSA or template inputs.

(C) Optimization with HalluDesign AF3 on BoltzDesign1 generated ligand-binding scaffolds, evaluated by AF3 without MSA or template inputs.

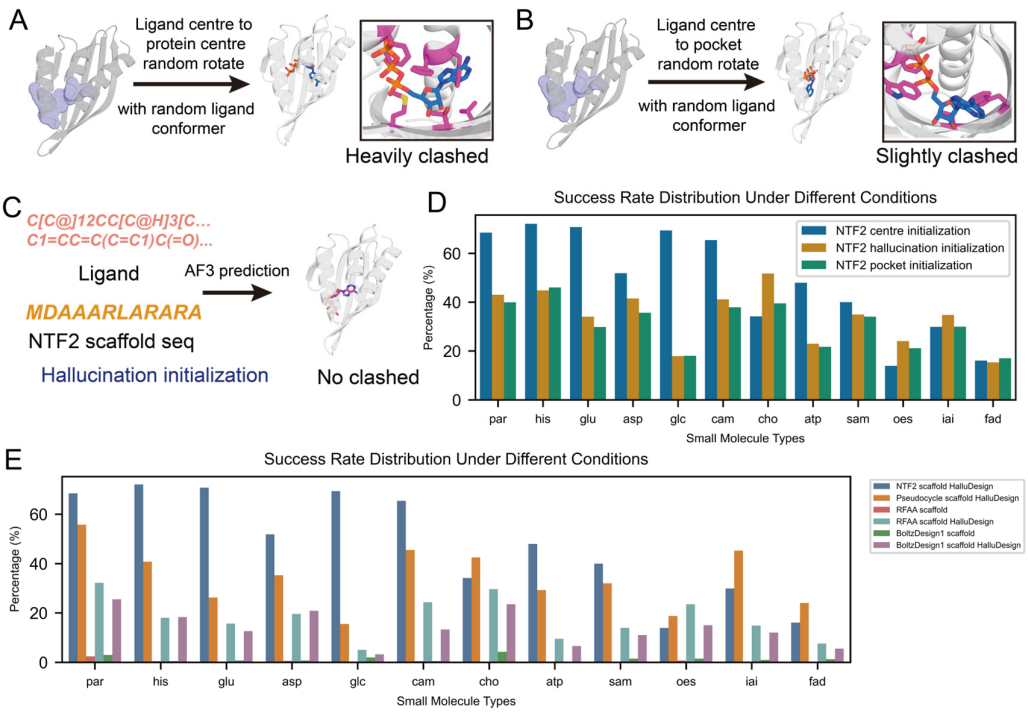


Supplementary Information Figure 8: Consistency between optimizer confidence and evaluator confidence. A-C

(A) Consistency between monomer designs optimized by AF3 and their confidence as evaluated by ESMFold, data from Chroma monomer optimization.

(B) Consistency between protein–binder complex designs optimized by AF3 and confidence scores from AlphaFold-Multimer evaluation. data from RFdiffusion binder scaffold 50 steps optimization.

(C) Consistency between ligand–protein complex designs optimized with AF3 or Protenix and confidence scores from AF3 or Protenix evaluation. data from RFAA scaffold 150 steps optimization.



Supplementary Information Figure 9: Scaffold-based ligand-binding protein design. A-E

The success rate is defined as: protein RMSD ≤ 1.5 Å, protein pLDDT ≥ 85, ligand RMSD ≤ 2 Å, ligand pLDDT ≥ 85, and ipTM ≥ 0.85.

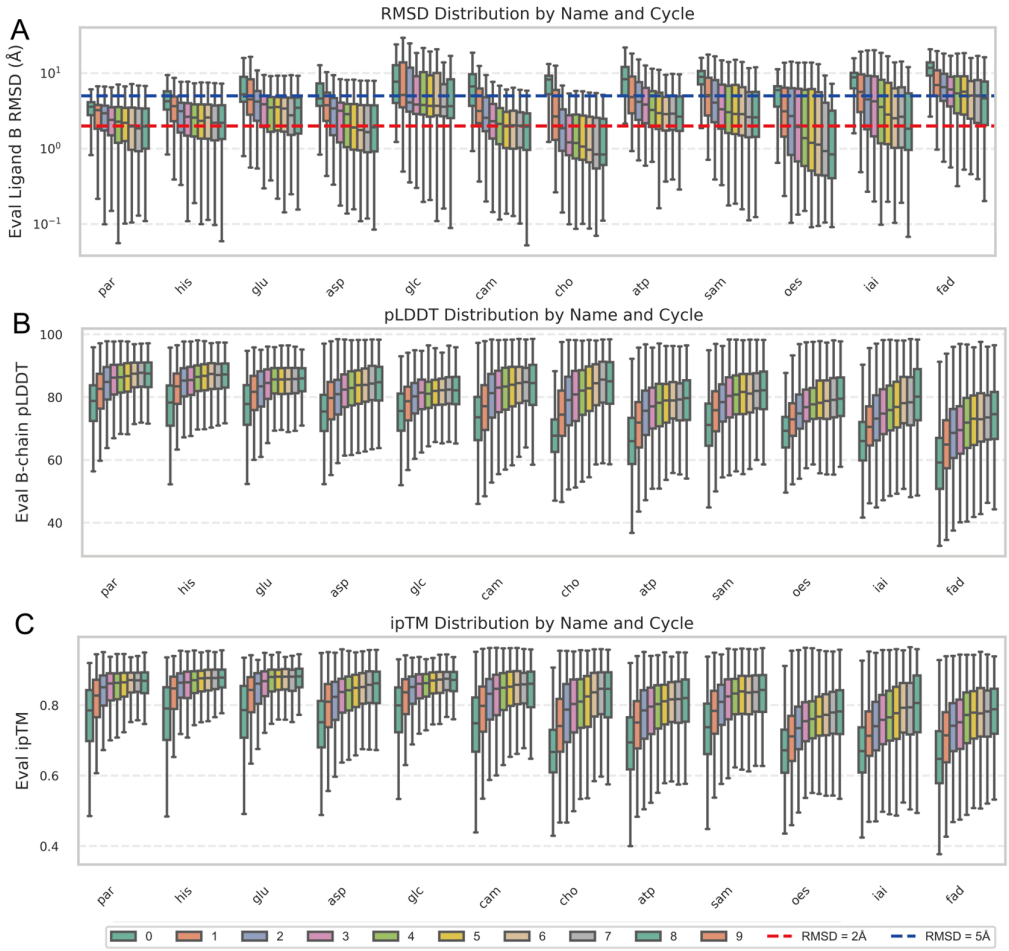
(A) Illustration of ligand center initialization strategies. Ligands are centered on the protein mass center and randomly rotated.

(B) Illustration of ligand pocket initialization strategies. Binding pockets are defined using *Fpocket*, with ligands centered on the pocket's spatial center and randomly rotated.

(C) Illustration of ligand hallucination initialization strategies. AF3 prediction is directly performed, allowing the model to determine the optimal ligand-binding position.

(D) Optimization of NTF2 scaffolds using HalluDesign AF3, with both optimization and evaluation conducted under different ligand initialization strategies by AF3 without MSA or template inputs.

(E) Optimization of different ligand binder scaffold source.

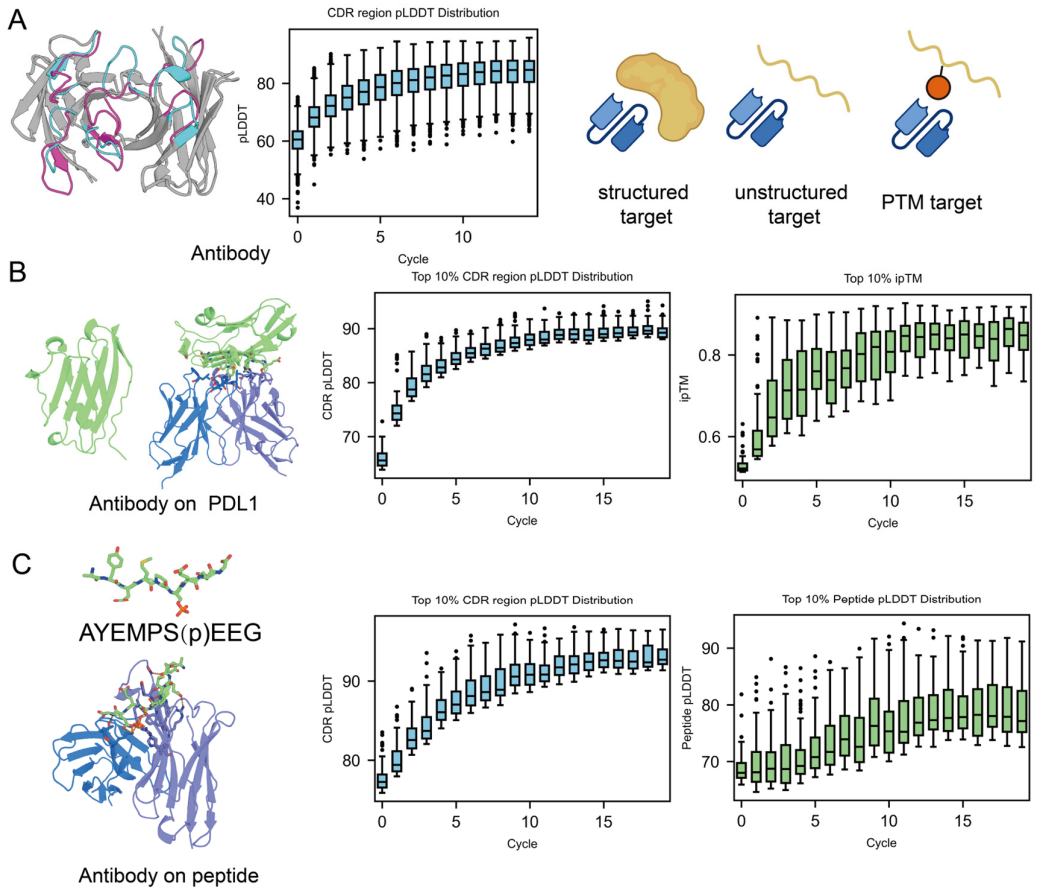


Supplementary Information Figure 10: Scaffold-based ligand-binding protein design metrics across optimization cycles for our small molecular dataset. A-C

(A) Ligand evaluation RMSD trend across optimization cycles under the ligand center initialization strategy for the NTF2 scaffold, designed and optimized using AF3.

(B) Ligand evaluation ipTM trend across optimization cycles under the ligand center initialization strategy for the NTF2 scaffold, designed and optimized using AF3.

(C) Ligand pLDDT trend across optimization cycles under the ligand center initialization strategy for the NTF2 scaffold, designed and optimized using AF3.

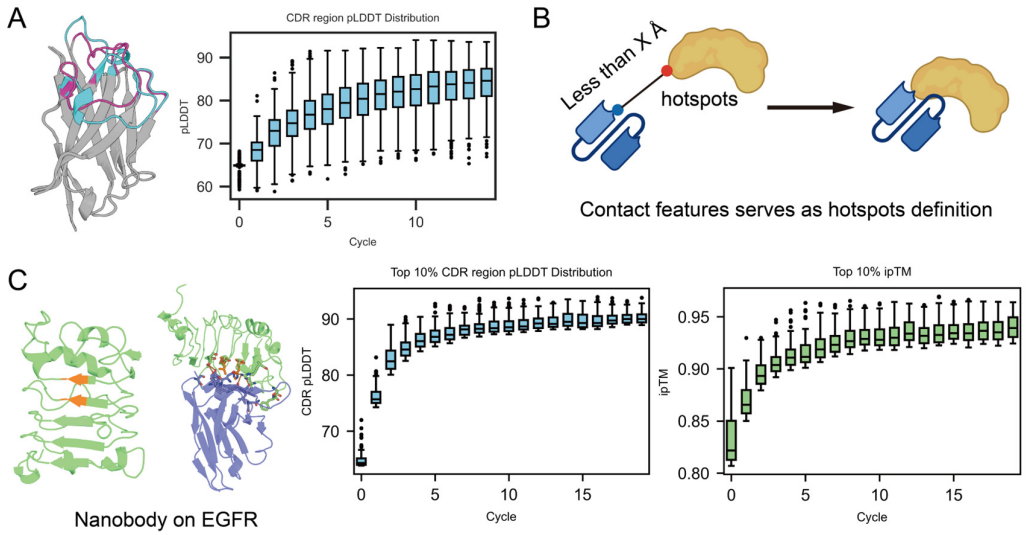


Supplementary Information Figure 11: Framework-fixed, CDR-region–redesigned antibodies utilizing AlphaFold 3. A-C

(A) Example of CDR-region redesign in an antibody. CDR residues are centered and perturbed with coordinate noise, and initialized as glycine sequences.

(B) Design of antibodies targeting PD-L1, optimized and evaluated using AF3.

(C) Optimization curves of antibodies targeting a phosphorylated peptide, optimized and evaluated using AF3.

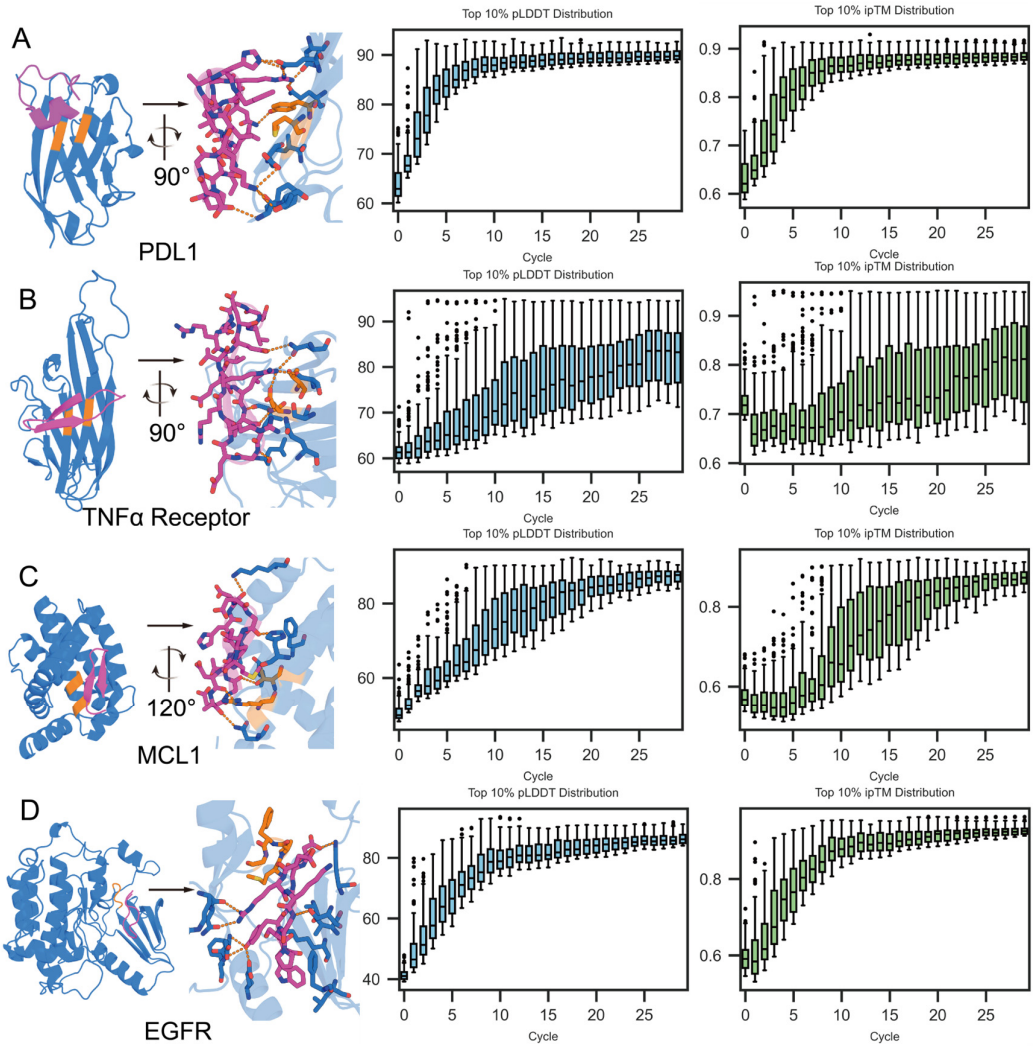


Supplementary Information Figure 12: Framework-fixed, CDR-region–redesigned nanobody utilizing Protenix. A-C

(A) Example of CDR-region redesign in a nanobody. CDR residues are spatially centered and perturbed with coordinate noise, with glycine sequences used for initialization.

(B) In the Protenix-based HalluDesign workflow, contact distance features are leveraged to identify and define hotspot regions.

(C) Nanobody designs targeting EGFR were generated, optimized, and evaluated using Protenix.



Supplementary Information Figure 13: Cyclic peptide design. A-D

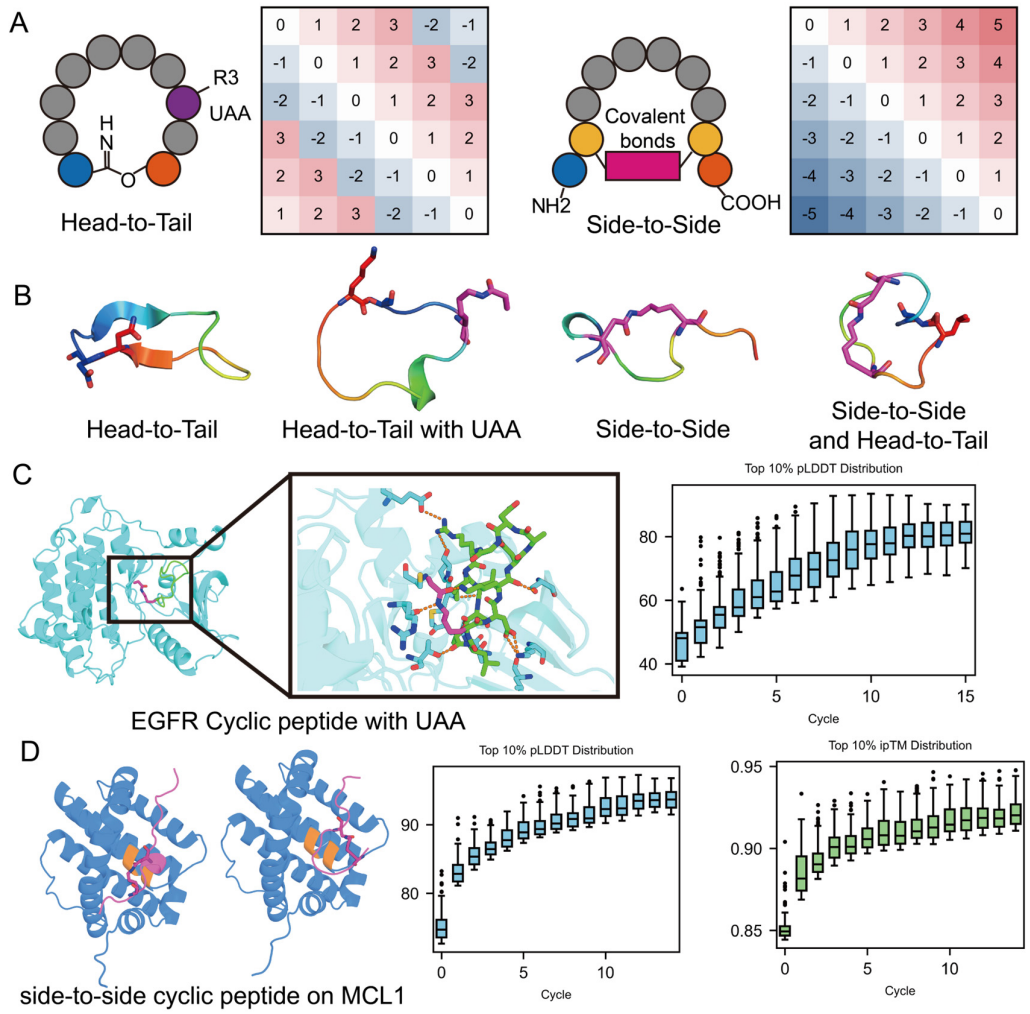
Binding hotspots are shown in orange, cyclic peptides in red, and target proteins in blue. The optimization curves display the top 10% of cyclic peptide pLDDT and ipTM scores.

(A) Example design and optimization curve of head-to-tail cyclic peptides targeting PDL1

(B) Example design and optimization curve of head-to-tail cyclic peptides targeting TNFα

(C) Example design and optimization curve of head-to-tail cyclic peptides targeting MCL1

(D) Example design and optimization curve of head-to-tail cyclic peptides targeting EGFR



Supplementary Information Figure 14: Cyclic peptide design with UAA. A-D

(A) Schematic illustration of position embeddings: head-to-tail cyclic peptides with cyclic positional encoding versus side-to-side cyclic peptides with linear positional encoding.

(B) Representative examples of different cyclic peptide architectures that can be designed within our framework. Side-to-side cyclic peptide design was performed with Protenix. This approach can be directly applied without code modification to cases where a peptide simultaneously forms two covalent bonds with a small molecule, representing side-chain–to–side-chain linkages that mimic peptide bonds.

(C) Design of a head-to-tail cyclic peptide incorporating the unnatural amino acid AcrK to covalently target Cys797 in the EGFR intracellular domain. Optimized with Protenix and evaluated by AF3. Cyclic peptide pLDDT = 88.65 and warhead NCAA pLDDT = 86.4.

(D) Optimization benchmark and schematic illustration of side-to-side cyclic peptide design against the apoptosis regulator MCL1. Optimization and evaluation were both performed with Protenix. However, Protenix tends to suffer from over-hallucination.