High-Affinity Protein Binder Design via Flow Matching and In Silico Maturation

Qilin Yu[a,†], Liangyue Guo[a,†], Xiayan Qin[a,†], Xikun Huang[a,†], Baihui Tian[a], Hongzhun Wang[a], Yu Liu[a], Yunzhi Lang[a], Di Wang[a], Zhouhanyu Shen[a], Jie Lin[a], and Mingchen Chen[a,*]

[a]Changping Laboratory

Abstract

The de novo design of high-affinity protein binders remains a central challenge in protein engineering and therapeutic discovery. While deep generative models have advanced backbone generation and interface design, achieving picomolar or nanomolar binding affinities typically demands extensive experimental screening or iterative in vitro maturation. A general computational approach for the direct production of high-affinity binders has remained elusive. Here, we introduce an integrated framework that synergizes PPIFlow, a flow-matching-based generative model, with a novel in silico maturation strategy. PPIFlow employs a pairformer architecture to explicitly reason over pairwise geometric and chemical interactions, modeling protein backbone rigid-body transformations as continuous flows. To bridge the affinity gap, we implement a dedicated in silico affinity maturation stage that combines interface rotamer enrichment with partial flow refinement to optimize energetic packing. This pipeline is further accelerated by AF3Score, a score-only adaptation of AlphaFold3 that enables high-fidelity and computationally efficient candidate prioritization. Across a diverse set of therapeutic targets, this synergistic approach consistently produces picomolar and nanomolar affinity binders without experimental affinity maturation. Notably, the framework proves highly effective for the de novo design of single-domain antibodies (VHHs), producing sub-nanomolar binders across multiple targets. These results establish that coupling robust backbone generation with focused in silico maturation renders the purely computational design of high-affinity binders feasible.

[*]† These authors contributed equally to this work.
[†]* Corresponding author: mingchenchencpl.ac.cn

1 Introduction

Protein–protein interactions (PPIs) underpin nearly all cellular processes and represent a primary class of therapeutic targets. The ability to design proteins that bind specific epitopes with high affinity and precision would revolutionize applications in pathway modulation, diagnostics, and biologic drug development [1]. Traditional binder discovery, relying on immunization [2], library screening [3, 4], and directed evolution [5, 6], remains resource-intensive and often offers limited control over the targeted epitope or the resulting molecular geometry.

Deep learning has recently catalyzed a paradigm shift in *de novo* protein design. Generative approaches utilizing structure hallucination [7, 8], inverse folding [1, 9], and diffusion models [10, 11] have enabled the creation of novel folds and functional binders. Notable frameworks such as RFdiffusion and BindCraft have significantly improved backbone sampling and diversity, and design of binders [10, 12, 13, 14]. However, a critical bottleneck remains: most computationally designed binders exhibit only modest affinities, typically requiring extensive experimental screening or iterative *in vitro* maturation to achieve therapeutic relevance.

Achieving high-affinity binding is fundamentally more challenging than generating plausible interfaces. High-affinity PPIs require exacting geometric complementarity, optimized side-chain packing, and cooperative interactions that are difficult for backbone-centric generative models to capture [14]. Antibodies, which naturally achieve strong and specific protein–protein interactions, therefore represent a particularly stringent benchmark for *de novo* design. While recent advances in *de novo* antibody design, such as RFantibody and Germinal, have expanded the accessible space for complementarity-determining regions (CDRs) [11, 15], generation of high-affinity binders and single-domain antibodies (VHHs) with high success rate and without the need for library-based directed evolution remains an elusive goal.

Flow matching has recently emerged as a powerful alternative to diffusion, offering superior training stability and efficient sampling by learning continuous mappings between simple distributions and com-

plex structures [16, 17, 18, 19]. Here, we introduce **PPIFlow**, a flow-matching–based framework and integrated design workflow for the *de novo* generation of high-affinity binders targeting precise epitopes (Fig. 1a). PPIFlow integrates a pairformer module to explicitly model interface contacts and utilizes a multi-stage training scheme for robust generalization across diverse protein scaffolds. To achieve *in silico* affinity maturation, we introduce three key innovations: (i) **interface rotamer enrichment** to identify energetically optimal residues; (ii) **partial flow refinement** to optimize interface packing by perturbing and regenerating existing designs; and (iii) **AF3Score**, an efficient score-only adaption of AlphaFold3 (Fig. 1c) for high-confidence complex prioritization[20].

Using this integrated platform, we demonstrate unprecedented performance, successfully generating pM-affinity binders for six out of seven diverse therapeutic targets and pM to nM-affinity VHHs for seven out of eight targets. These results establish PPIFlow as a reliable platform for the *de novo* design of protein and antibody therapeutics with epitope precision.

2 Results

2.1 A flow matching based protein generation model

PPIFlow models protein binder design as a continuous rigid-body flow. We represent protein backbones as collections of rigid residue frames in SE(3) [21] (Fig. 1a). Unlike diffusion-based approaches [10], PPIFlow employs flow matching to directly learn the translational and rotational velocity fields that transport random distributions to physically realistic structures [16, 18]. This formulation provides a geometrically consistent parameterization of rigid-body dynamics, enabling stable training and precise conditioning on interface constraints.

The pairformer module jointly updates residue-level and pairwise embeddings to explicitly capture structured geometric relationships across binding interfaces [22]. These representations are decoded into frame updates via Invariant Point Attention (IPA) [21], balancing generation accuracy with parameter efficiency (Fig. 2c).

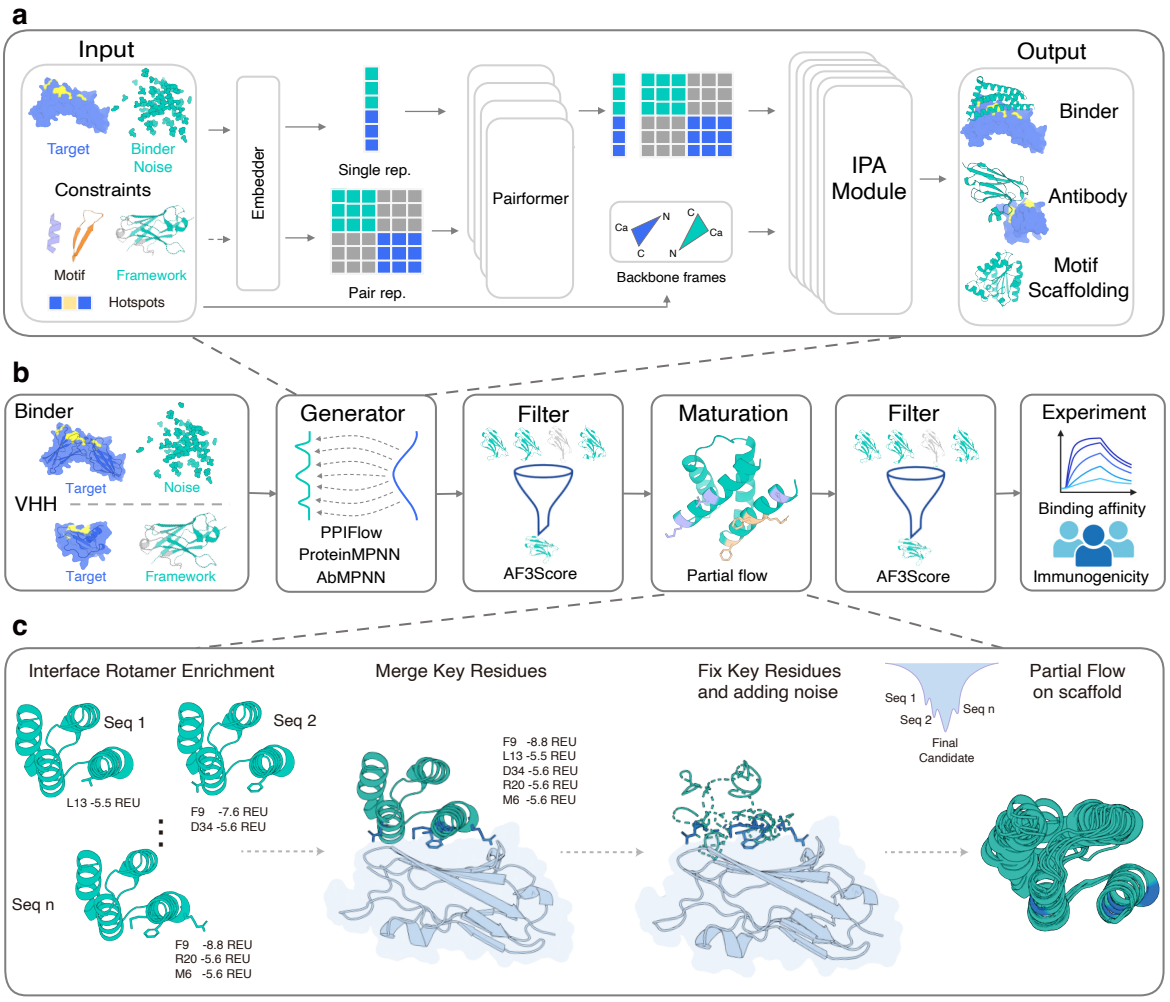

Figure 1: PPIFlow model architecture and high-affinity design workflow. a, Overview of the PPIFlow model framework. b, End-to-end design pipeline. PPIFlow-generated backbones undergo sequence design via ProteinMPNN (monomers, motifs, binders) or AbMPNN (VHHs, scFvs). High-confidence candidates are identified via AF3Score filtering and prioritized for *in silico* maturation and template-free AlphaFold3 validation. c, *In silico* maturation strategy: (1) interface rotamer enrichment to identify energetically favorable residues; (2) selection and merging of key interface residues; (3) fixation of critical residues with concomitant noise addition to neighboring backbone regions; and (4) partial flow-based refinement of unconstrained regions to optimize interfacial packing.

Training follows a staged curriculum that adapts the model from general backbone generation to constrained PPI design (Fig. 2a). Optimization relies on a flow-matching objective [17] supplemented by auxiliary geometric loss terms to promote interface stability. We validated these design choices through ablation studies, and demonstrated that both staged curriculum and geometric constraints are critical for achieving robust binder generation (Fig. 2b).

2.2 PPIFlow generates high-designability proteins

PPIFlow performs robustly across five generative tasks. For monomers (up to 700 residues), the model achieved >90% design success (scRMSD $<$ 2 Å vs. ESMFold) while maintaining structural novelty and diversity (Fig. 2d; Fig. S1) [23, 24, 10, 18, 13]. In motif scaffolding, PPIFlow outperformed state-of-the-art methods on MotifBench (score 42.38 vs. 28.6 for RFdiffusion) [25], solving challenging multi-segment cases (e.g., 4JHW, 4XOJ) where prior approaches failed (Fig. S2) [10, 26].

We further validated PPIFlow on minibinder design (75/120/160/200 residues) against 15 diverse targets. The model surpassed recent open-source methods like RFdiffusion, BindCraft and BoltzGen in design success, novelty, and diversity (Fig. 2e and Fig. S3) [12, 27]. Notably, PPIFlow shows a pronounced advantage for larger binders, as reflected by both AF3Score-based designability and Rosetta-derived interface energetics. This length-dependent improvement likely stems from the multi-stage training strategy combined with PPIFlow's outstanding monomer designability (Fig. S3).

In VHH and antibody design, PPIFlow generated diverse CDRs while preserving input frameworks (median RMSD 0.18 Å) (Fig. S4, S5). The model targeted epitopes precisely (median hotspot coverage > 0.9) (Fig. 2f), yielding complexes with well-packed, high-quality interactions (Fig. 2g).

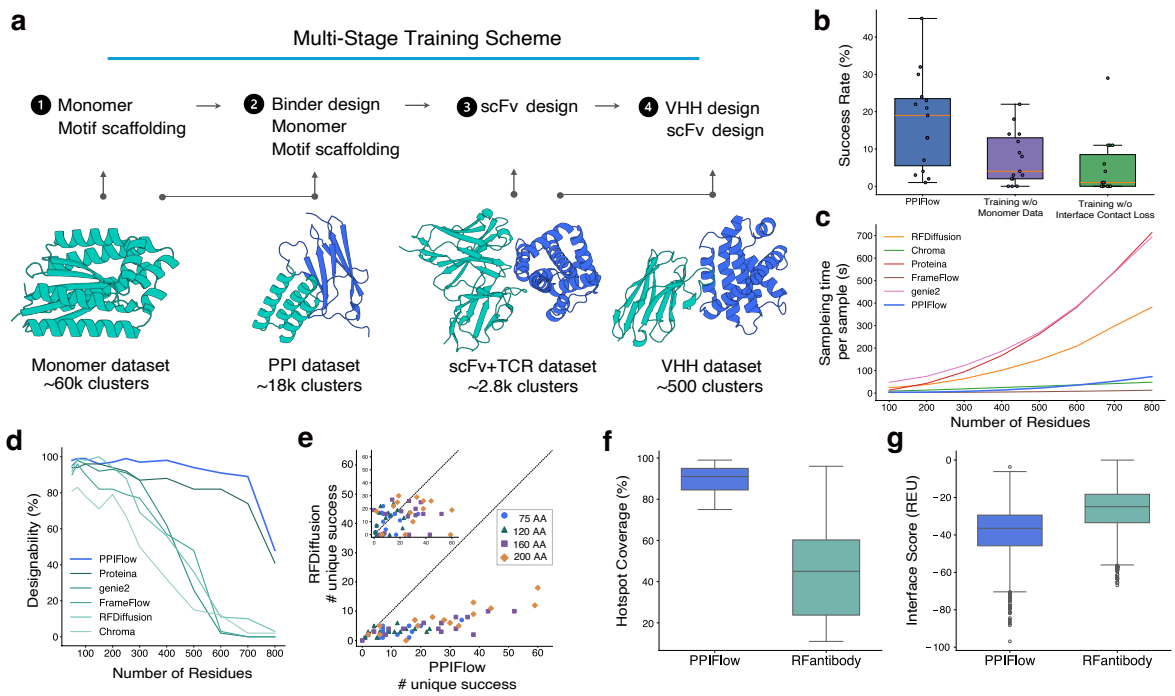

Figure 2: Training curriculum and *in silico* performance benchmarks. a, Multi-stage training scheme of PPIFlow. The foundation model is first trained on a monomeric protein dataset for structure recovery and scaffolding. Subsequent fine-tuning stages progressively adapt the model to protein–protein interactions (PPI), scFv generation, and VHH design, leveraging both general and antibody-specific datasets to preserve framework integrity. b, Ablation analysis of the training objective, highlighting the necessity of monomer dataset integration and interface contact loss for optimal binder design. c, Inference efficiency comparison of PPIFlow against representative generative architectures. d, Designability benchmarks for monomers. e, Designability benchmarks for binders, inset: comparison between PPIFlow (x-axis) & BindCraft (y-axis). f–g, VHH design quality metrics, including predicted hotspot coverage (**f**) and Rosetta interface scores (**g**) for generated antibody–antigen complexes.

2.3 High-affinity design via interface rotamer enrichment and partial flow

While generative models have streamlined binder design [28], achieving picomolar affinity typically requires experimental optimization. To bridge this gap, we developed a pipeline integrating interface rotamer enrichment with partial flow refinement (Fig. 1c). In the **initial generation stage**, PPIFlow generates backbones targeting a predefined epitope (Fig. 1b). We design sequences using ProteinMPNN [29] or AbMPNN [30] and filter candidates via AF3Score to retain structurally plausible designs.

During maturation, we identify energetic "anchor" residues (< -5 Rosetta Energy Units) on the interface [31]. We fix these rotamers while perturbing the backbone to an intermediate flow state ($t = 0.6$). Crucially, this plasticity allows the backbone to escape local minima and resolve steric clashes

imposed by the anchors, reshaping the binding interface rather than merely repacking side chains [32]. We then regenerate perturbed regions and redesign unconstrained sequences ($T = 0.1$). This active remodeling accommodates enriched residues and significantly improves binding compatibility (Fig. S3).

Matured designs are re-evaluated via AF3Score, Rosetta Interface Analyzer, and template-free AlphaFold3. We prioritize the top 30 candidates, selected by a composite score (AF3 ipTM $\times$ 100 $-$ Rosetta interface score), for experimental validation.

2.4 *De novo* design of high-affinity mini-binders

To experimentally validate the pipeline, we designed *de novo* binders against seven therapeutically relevant targets: IL7RA, IFNAR2, IL17A, PD-L1, TRKA, PDGFR, and VEGFA (Fig. 3a). Spanning diverse structural classes and functional roles in oncology and autoimmune disease, these targets provide a stringent benchmark. We characterized 30 candidates per target; all 210 expressed successfully, with 76 (36.2%) binders showing affinities better than 1 μM by Bio-Layer Interferometry (BLI) (Fig. 3b,c; Fig. S6; Table S7), spanning diverse structures and sequences (Table S6,S7). Notably, we identified picomolar-affinity binders for six of the seven targets without post-design experimental optimization.

Success stems largely from the *in silico* maturation stage. Structural comparisons reveal that interface rotamer enrichment followed by partial flow creates compact, highly connected interfaces featuring specific hydrogen bonds, salt bridges, and hydrophobic interactions formed during refinement (Fig. 3d,e; Fig. S9). Across all targets, this maturation systematically improved folding confidence and interface quality, significantly increasing AF3Score pTM and ipTM values (Fig. 3f). In the case of IFNAR2, interface rotamer enrichment markedly outperformed baseline pipelines, yielding superior interface metrics (Fig. 3g,h).

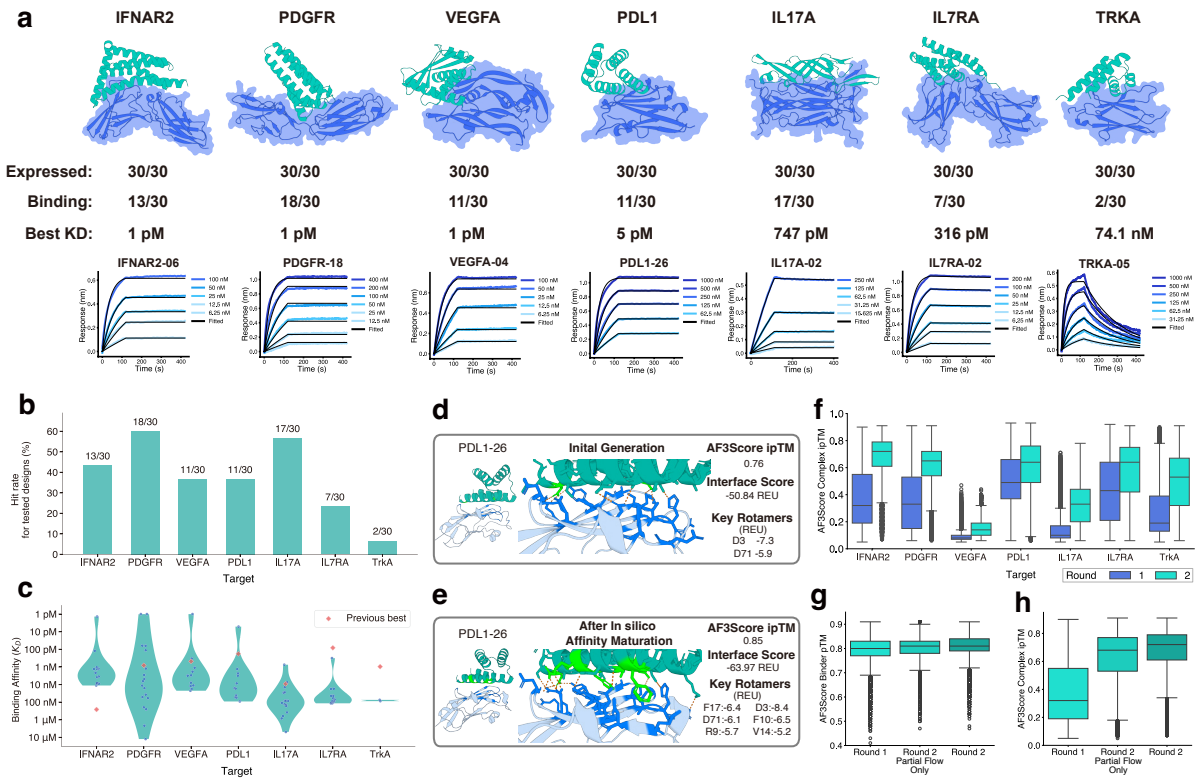

Figure 3: Experimental validation and high-affinity binder generation. a, Structural models of the highest-affinity *de novo* binders for each target. Accompanying data include expression success, binding hit rates, peak affinities, and representative BLI sensorgrams. b–c, Experimental performance across seven therapeutic targets, showing (**b**) hit rates and (**c**) the distribution of binding affinities (violin plots), demonstrating consistent picomolar-to-nanomolar potency. d–e, Illustration of interface side-chain packing after (**d**) initial generation and after (**e**) *in silico* affinity maturation for PDL1-26. f, Comparison of AF3Score profiles between initial Round 1 designs and matured Round 2 designs, illustrating the systematic quality improvement achieved through *in silico* maturation. g–h, Performance comparison of design strategies for IFNAR2. Metrics for AF3Score pTM (**g**) and ipTM (**h**) demonstrate that interface rotamer enrichment combined with partial flow refinement markedly outperforms both Round 1 baselines and partial flow alone.

2.5 *De novo* design of high-affinity VHHs

We extended PPIFlow to the *de novo* generation of single-domain antibodies (VHHs) against eight therapeutic targets: CCL2, HNMT, PDGFR, 1433E, BHRF1, S100A4, EFNA1, and IL13. To prioritize developability, we sampled frameworks from five therapeutically validated scaffolds, generating and maturing CDRs via the PPIFlow-VHH pipeline (see Supplementary Information).

BLI characterization of 240 candidates (30 per target) confirmed 100% expression success. We identified binders for seven of the eight targets, with several achieving exceptional potency without optimization. Notably, VHH CCL2-21 reached 250 pM affinity, while binders for HNMT, PDGFR, and 1433E achieved single-digit nM affinities (Fig. 4a). The overall hit rate was 33.8% (81/240), with 27.9% (67/240) showing affinities better than 1 μM, and these successful designs span diverse structures and sequences (Fig. 4b,c; Fig. S7; Table S6,S8).

As with general protein binders, AF3Score analysis confirms that *in silico* maturation, combining interface rotamer enrichment with partial flow, was critical. This stage systematically improved structural confidence and interface quality (Fig. 4f), evidenced by compact side-chain packing and energetically favorable interactions (Fig. 4d,e; Fig. S10). Benchmarks on CCL2 show that interface rotamer enrichment significantly outperforms baselines, yielding higher AF3Score ipTM values and superior experimental outcomes (Fig. 4g,h).

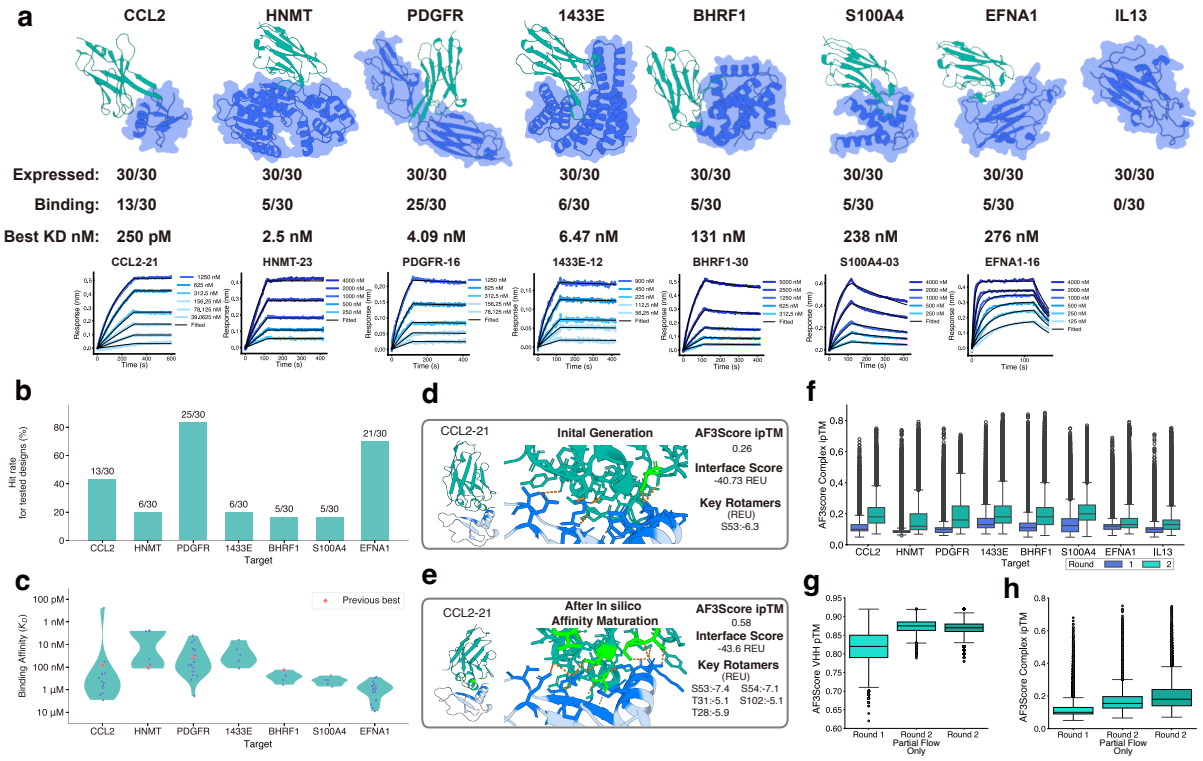

Figure 4: Computational design and experimental validation of high-affinity VHHs. a, Structural models of the highest-affinity *de novo* VHHs for each target, paired with expression data, binding hit rates, peak affinities (K_D), and representative BLI sensorgrams. b–c, Experimental performance across VHH targets, showing (**b**) hit rates and (**c**) affinity distributions (violin plots), highlighting the discovery of sub-nanomolar binders for CCL2, PDGFR, and 1433E. d–e, Illustration of interface side-chain packing after (**d**) initial generation and after (**e**) *in silico* affinity maturation for CCL2-21. f, Comparison of AF3Score profiles between initial (Round 1) and matured (Round 2) VHH designs, demonstrating consistent improvement in predicted interface quality across all targets. g–h, Evaluation of design strategies for the CCL2 target. AF3Score pTM (**g**) and ipTM (**h**) metrics confirm that the full maturation protocol, integrating interface rotamer enrichment with partial flow, significantly outperforms baseline generative and refinement-only pipelines.

3 Discussion

The ability of PPIFlow to generate picomolar affinity binders and VHHs purely *in silico* marks a shift from "design-then-screen and optimize" toward de novo binder generation. Traditionally, bridging the "affinity gap" required massive experimental libraries or directed evolution. Recent retrospective analyses have shown that AlphaFold3 interface confidence scores are strongly predictive of experimental binding success and can therefore serve as an effective structural consistency filter [11]. By incorporating AF3Score, a score-only adaptation of AlphaFold3, PPIFlow bypasses MSA construction and iterative recycling, accelerating evaluation by approximately two orders of magnitude compared to standard inference [20]. This enables rapid, high-fidelity candidate prioritization without the computational burden of traditional protocols.

A central conceptual advance of this work is the integration of global backbone topology sampling with local interface optimization. While flow matching explores diverse architectures, interface rotamer enrichment acts as a computational analogue of affinity maturation. This hybrid approach resolves the tension between broad fold exploration and the stringent geometric requirements of high-affinity binding, addressing the observation that high in silico designability alone does not guarantee experimental success [28, 27].

Although PPIFlow outperforms existing methods such as BindCraft on backbone feasibility metrics, its superior experimental hit rates also stem from the pipeline's synergy. Furthermore, PPIFlow significantly reduces the computational burden associated with binder design. Unlike protocols such as BindCraft that typically necessitate computationally expensive full AlphaFold predictions for filtering, our workflow leverages AF3Score to rank complexes directly, accelerating candidate prioritization by approximately 100-fold. Unlike standard generation methods, PPIFlow robustly accommodate the constraints imposed during "Interface Rotamer Enrichment", allowing the subsequent "Partial Flow" stage to refine packing without breaking the fold. Thus, it is not the generator in isolation, but the holistic integration of pairformer, energy-guided anchoring, and flow-based refinement that bridges the gap between

computational plausibility and picomolar potency.

Limitations remain. PPIFlow currently operates as a backbone-only generator, meaning side-chain packing is handled post-generation [33]. While our maturation pipeline mitigates this, a fully atomistic generative framework could more natively capture the highly specific hydrogen-bonding networks and hydrophobic packing required for certain complex interfaces. We also observed target-dependent variation in success rates. For example, IL13 remained challenging, in such cases, the interface rotamer enrichment stage may fail to identify stable energetic anchors ($<$ -5 REU), preventing effective partial flow refinement. Consistent with this, VHHs generated against IL13 showed systematically lower confidence scores (Fig. 4f).

Finally, the current model is limited to protein–protein complexes; extending the framework to nucleic acids, glycans, and small molecules is essential. Nevertheless, PPIFlow establishes a generalizable platform for rapid, purely computational binder design, significantly reducing reliance on experimental iteration.

4 Data and materials availability

The code and relevant data for PPIFlow will be available at https://github.com/Mingchenchen/PPIFlow. AF3Score is available at https://github.com/Mingchenchen/AF3Score: A Score-Only Adaptation of AlphaFold3 for Biomolecular Structure Evaluation. All relavant data will be available at Zenodo.

5 Acknowledgments

This work was funded by Changping Laboratory.

References

[1] Longxing Cao et al. “Design of protein-binding proteins from the target structure alone”. In: *Nature* 605.7910 (2022), pp. 551–560.

[2] Georges Köhler and Cesar Milstein. “Continuous cultures of fused cells secreting antibody of predefined specificity”. In: *nature* 256.5517 (1975), pp. 495–497.

[3] George P Smith. “Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface”. In: *Science* 228.4705 (1985), pp. 1315–1317.

[4] Greg Winter et al. “Making antibodies by phage display technology”. In: *Annual review of immunology* 12.1 (1994), pp. 433–455.

[5] Frances H Arnold. “Directed evolution: bringing new chemistry to life”. In: *Angewandte Chemie (International Ed. in English)* 57.16 (2017), p. 4143.

[6] Frances H Arnold. “Design by directed evolution”. In: *Accounts of chemical research* 31.3 (1998), pp. 125–131.

[7] Ivan Anishchenko et al. “De novo protein design by deep network hallucination”. In: *Nature* 600.7889 (2021), pp. 547–552.

[8] Jue Wang et al. “Scaffolding protein functional sites using deep learning”. In: *Science* 377.6604 (2022), pp. 387–394.

[9] Milong Ren et al. “Accurate and robust protein sequence design with CarbonDesign”. In: *Nature Machine Intelligence* 6.5 (2024), pp. 536–547.

[10] Joseph L Watson et al. “De novo design of protein structure and function with RFdiffusion”. In: *Nature* (2023), pp. 1–3.

[11] Nathaniel R Bennett et al. “Atomically accurate de novo design of antibodies with RFdiffusion”. In: *Nature* (2025), pp. 1–11.

[12] Martin Pacesa et al. “One-shot design of functional protein binders with BindCraft”. In: *Nature* 646.8084 (2025), pp. 483–492.

[13] John B Ingraham et al. “Illuminating protein space with a programmable generative model”. In: *Nature* 623.7989 (2023), pp. 1070–1078.

[14] Pablo Gainza et al. “De novo design of protein interactions with learned surface fingerprints”. In: *Nature* 617.7959 (2023), pp. 176–184.

[15] Luis S Mille-Fragoso et al. “Efficient generation of epitope-targeted de novo antibodies with Germinal”. In: *bioRxiv* (2025).

[16] Yaron Lipman et al. “Flow matching for generative modeling”. In: *International Conference on Learning Representations* (2023).

[17] Ricky T. Q. Chen and Yaron Lipman. “Flow Matching on General Geometries”. In: *International Conference on Learning Representations* (2024).

[18] Jason Yim et al. “Fast protein backbone generation with SE(3) flow matching”. In: *arXiv preprint arXiv:2310.05297* (2023).

[19] Avishek Joey Bose et al. “Se (3)-stochastic flow matching for protein backbone generation”. In: *arXiv preprint arXiv:2310.02391* (2023).

[20] Yu Liu et al. “Af3score: A score-only adaptation of alphafold3 for biomolecular structure evaluation”. In: *Journal of Chemical Information and Modeling* 65.15 (2025), pp. 8207–8214.

[21] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *nature* 596.7873 (2021), pp. 583–589.

[22] Josh Abramson et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* 630.8016 (2024), pp. 493–500.

[23] Yeqing Lin et al. “Out of many, one: Designing and scaffolding proteins at the scale of the structural universe with genie 2”. In: *arXiv preprint arXiv:2405.15489* (2024).

[24] Tomas Geffner et al. “Proteina: Scaling flow-based protein structure generative models”. In: *arXiv preprint arXiv:2503.00710* (2025).

[25] Zhuoqi Zheng et al. “MotifBench: A standardized protein design benchmark for motif-scaffolding problems”. In: *arXiv preprint arXiv:2502.12479* (2025).

[26] Tianyu Lu et al. “Conditional protein structure generation with protpardelle-1c”. In: *bioRxiv* (2025).

[27] Hannes Stark et al. “Boltzgen: Toward universal binder design”. In: *bioRxiv* (2025), pp. 2025–11.

[28] Nathaniel R Bennett et al. “Improving de novo protein binder design with deep learning”. In: *Nature Communications* 14.1 (2023), p. 2625.

[29] Justas Dauparas et al. “Robust deep learning–based protein sequence design using ProteinMPNN”. In: *Science* 378.6615 (2022), pp. 49–56.

[30] Frédéric A Dreyer et al. “Inverse folding for antibody sequence design using deep learning”. In: *arXiv preprint arXiv:2310.19513* (2023).

[31] Rebecca F Alford et al. “The Rosetta all-atom energy function for macromolecular modeling and design”. In: *Journal of chemical theory and computation* 13.6 (2017), pp. 3031–3048.

[32] Matthias Glögl et al. “Target-conditioned diffusion generates potent TNFR superfamily antagonists and agonists”. In: *Science* 386.6726 (2024), pp. 1154–1161.

[33] Jin Sub Lee and Philip M Kim. “FlowPacker: protein side-chain packing with torsional flow matching”. In: *Bioinformatics* 41.3 (2025), btaf010.