Latent Generative Search unlocks *de novo* Design of Untapped Biomolecular Interactions at Scale

Kieran Didi[1,2,*], **Danny Reidenbach**[1,*], **Matthew Penner**[3], **Supriya Ravichandran**[4], **Marshall Case**[4], **Mike Nichols**[4], **Erik Swanson**[4], **Alex Reis**[4], **Maggie Prescott**[4], **Yue Qian**[5], **Dongming Qian**[5], **Jingjing Yang**[5], **Weiji Li**[5], **Le Li**[5], **Daichi Shonai**[6], **Sean Gay**[6], **Bhoomika Basu Mallik**[7], **Ho Yeung Chim**[7], **Liurong Chen**[7], **Miguel Atienza Juantay**[7], **Hubert Klein**[7], **Anna U. Macintyre**[8], **Maxim Secor**[8], **Daniele Granata**[8], **Zhonglin Cao**[1], **Guoqing Zhou**[1], **Tomas Geffner**[1], **Xi Chen**[1], **Micha Livne**[1], **Zuobai Zhang**[1], **Tianjing Zhang**[1], **Kyle Gion**[1], **Michael M. Bronstein**[2,9], **Martin Steinegger**[10], **Kristine Deibler**[8], **Scott Soderling**[6], **Alena Khmelinskaia**[7], **Florian Hollfelder**[3], **Christian Dallago**[1,6], **Emine Kucukbenli**[1], **Arash Vahdat**[1], **Pierce Ogden**[4], **Karsten Kreis**[1]

Abstract: In domains from game-play to language, the most powerful AI systems combine a generative model that learns a latent representation of the solution space, with adaptive search at inference time. This principle has not yet been realized in *de novo* protein design: generative methods produce structures in a single shot without optimizing them, while hallucination methods search over sequences without a learned generative prior to guide them. Finally, both approaches rely on separate models for sequence design. Here, we show that joint sequence and structure generation in a continuous latent space, combined with reward-guided inference-time search, fundamentally changes what protein design can achieve. In a massive-scale screening of over one million protein designs against 127 diverse and challenging targets, our approach, Proteina-Complexa, produced hits for 86 of them, demonstrating unmatched breadth. As part of this experimental campaign, we also carried out the largest experimental head-to-head benchmark of computational binder design methods to date. Proteina-Complexa produced more validated hits than any baseline method, with model generated sequences outperforming all baselines, including post hoc redesign, demonstrating that our underlying co-design of sequence, backbone, and side chains exceeds the quality of commonly used inverse folding models. Across five additional campaigns against challenging therapeutic targets, we achieve a 63.5% hit rate against PDGFR (top $K_D = 93.6\,\text{pM}$), nanomolar Nipah virus binders, nanomolar binders to a muscle-wasting receptor that block myostatin signaling in cells, 40–50% hit rates for kinase mini-protein and peptide binders, and the first *de novo* carbohydrate binders—a target class thought to be inaccessible to current design methods. These results establish latent generative search as a new paradigm for protein design, opening target classes and scales previously out of reach.

Project page: https://research.nvidia.com/labs/genair/proteina-complexa/

1. Introduction

Proteins mediate nearly all molecular interactions in biology. The ability to design proteins that bind specified targets with high affinity and selectivity could transform therapeutics, diagnostics, and synthetic biology [Huang et al., 2016]. Deep learning has made this goal increasingly realistic [Ferruz et al., 2023, Koh et al., 2025, Kortemme, 2024, Listov et al., 2024]. Structure-prediction models like AlphaFold2 [Jumper et al., 2021] and AlphaFold3 [Abramson et al., 2024] provide accurate predictions of target structures. Generative methods such as RFdiffusion [Watson et al., 2023] can produce candidate binder backbones, and hallucination approaches such as BindCraft [Pacesa et al., 2025] optimize sequences using structure-predictor gradients. Together with improved filtering [Bennett et al., 2023], these advances have increased design success rates by orders of magnitude [Cao et al., 2022, Chevalier et al., 2017].

Despite these advances, current protein design methods do not operate in a learned latent space of a pretrained generative model, where reward-guided search can efficiently steer generation toward user-defined objectives, introducing a fundamental limitation. In other AI domains, the combination of learned

[1]NVIDIA, [2]University of Oxford, [3]University of Cambridge, [4]Manifold Bio, [5]Viva Biotech, [6]Duke University, [7]LMU Munich, [8]Novo Nordisk, [9]AITHYRA, [10]Seoul National University

*Equal contribution.
Corresponding author: kkreis@nvidia.com

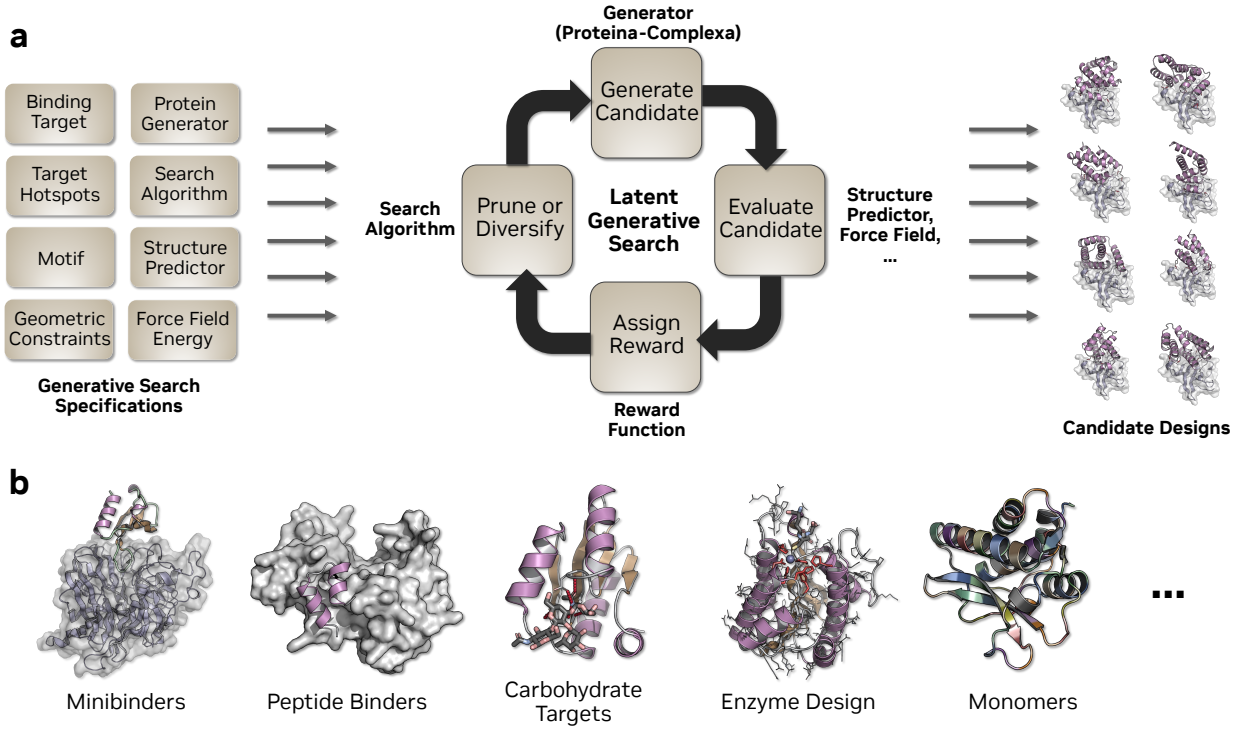


Figure 1 | **Latent Generative Search with Proteina-Complexa. a)** Our latent generative search framework combines several key ingredients: *(i)* A base generative model co-generates candidate protein structures and sequences—we utilize the flow-based Proteina-Complexa model. *(ii)* Generated candidates are evaluated by auxiliary models, for instance via structure prediction models or physical force fields. *(iii)* A reward function assigns the evaluated candidates scores, for example based on folding models' interface confidence predictions or counting the number of interface hydrogen bonds. *(iv)* A search algorithm, such as Beam or Tree Search, selects promising proteins for diversification or prunes others, based on the reward scoring. For Proteina-Complexa, generation is iterative and operates in a fully continuous, partially latent space; see Sec. A. In that case, latent generative search rolls out stochastic generation trajectories, evaluates and scores them, and prunes poorly performing trajectories early on, while diversifying promising ones. **b)** The latent generative search framework is general and allows for various protein design tasks given appropriate input specifications, reward definitions and models. It can be applied, for instance, during monomer generation (Sec. 4), to design minibinders (Sections 5, 6 and 8) and peptide binders (Sec. 7), to address carbohydrate targets (Sec. 9), or when scaffolding catalytic residues in the context of enzyme design [Didi et al., 2026]. Latent generative search combines data-driven learning with compute-driven optimization via inference-time search guided by the pretrained generative prior. This stands in contrast to prior pure generation or pure optimization—*i.e.* hallucination—methods.

representation and search has been transformative. AlphaGo [Silver et al., 2016] first combined learned priors with search for game-play. For image and video generation, continuous latent embeddings and diffusion models enable high-resolution generation, and can be coupled to search strategies for improved performance [Ma et al., 2025, Rombach et al., 2022]. In natural language processing, large language models have been shown to benefit from inference-time scaling to solve harder problems [Snell et al., 2025, Wei et al., 2022]. In contrast, protein structure generative models typically produce structures in a single shot. The quality of these generations frequently falls short without extensive downstream filtering and iterative structural refinement [Frank et al., 2024, Hong and Kortemme, 2024]. Hallucination-based protein methods instead rely on unconstrained optimisation, rendering the search slower, more brittle, and susceptible to non-physical or adversarial sequence–structure solutions [Lisanza et al., 2025]. Both paradigms additionally depend on post hoc sequence redesign – commonly via inverse-folding networks – to achieve competitive experimental success rates [Pacesa et al., 2025]; however, due to limitations in these redesign methods, this procedure often degrades performance for electrostatic and functionally critical interactions [Albanese et al., 2025, Pacesa et al., 2025].

These limitations are most pronounced for therapeutically relevant and difficult targets, such as polar protein

interfaces, carbohydrates, and flexible peptides, where sequence and backbone geometry must be co-optimized to place hydrogen-bonding interactions precisely. We hypothesized that learning a latent space that jointly encodes sequence and structure, and searching that space at inference time, would overcome these limitations.

Here, we introduce a latent generative search framework based on Proteina-Complexa [Didi et al., 2026], realizing this vision, and we experimentally validate it extensively (Fig. 1). Building on La-Proteina's latent representation [Geffner et al., 2026], Proteina-Complexa jointly encodes amino-acid identity and all-atom structure in a latent space, trains an expressive base generative model on large-scale protein structure data, and supports reward-guided inference-time search, combining the strengths of generative and hallucination approaches (Sec. 2). Proteina-Complexa's overarching latent generative search framework is universal – different scoring models, reward objectives, conditioning inputs and search algorithms enable diverse protein design applications.

We validate Proteina-Complexa across comprehensive experiments. In the largest head-to-head benchmark of models to date, with over one million proteins screened against 127 targets by multiplexed phage display, Proteina-Complexa outperforms all competing methods, and co-generated sequences surpass all baselines (Sec. 3). The model co-generates well-folded, hyperstable proteins from 100 to 800 amino acids with sequence quality matching ProteinMPNN, establishing a generative framework that does not require post hoc sequence redesign (Sec. 4). Across five challenging campaigns, we achieve (i) a 63.5% hit rate against PDGFR with picomolar affinity (Sec. 5), (ii) nanomolar binders to a receptor implicated in muscle wasting that block signaling in cells without affinity maturation (Sec. 6), (iii) 40–50% hit rates for kinase mini-protein and peptide binders (Sec. 7), (iv) nanomolar Nipah-virus binders via *de novo* design and scaffold re-engineering (Sec. 8), and (v) the first fully *de novo* carbohydrate binders, a target class previously thought inaccessible (Sec. 9).

2. Latent Generative Search with Proteina-Complexa

We briefly describe the Proteina-Complexa framework [Didi et al., 2026, Geffner et al., 2026]; full technical details are provided in Appendix A.

A core challenge in atomistic protein generation is that the number of atoms varies by amino acid type, complicating joint modeling of sequence and structure. La-Proteina [Geffner et al., 2026] addresses this with a partially latent representation: backbone C_α coordinates are modeled explicitly, while each residue's amino acid identity and remaining atomic coordinates are compressed into a fixed-length continuous latent variable by a variational autoencoder. A flow-based generative model then learns the joint distribution over backbone coordinates and latent variables, so that every generated sample is a complete protein—backbone, side chains, and sequence—with no post hoc sequence redesign required.

Proteina-Complexa [Didi et al., 2026] extends this to binder–target complexes by conditioning on a target structure and optional interface hotspot residues. Training proceeds in stages—pretraining on monomeric proteins, training on Teddymer (a large-scale synthetic dataset of binder–target pairs derived from domain–domain interactions in AlphaFold Database [Varadi et al., 2021] predicted structures), and fine-tuning on experimental multimer structures from the PDB—overcoming the scarcity of experimental complex data that has limited prior approaches.

The computational efficiency of our framework enables reward-guided search at inference time: rather than generating a single sample and accepting or rejecting it, Proteina-Complexa can explore multiple trajectories through the generative process and steer them toward high-quality binders, while operating 30–60× faster than RFdiffusion [Watson et al., 2023], a widely adopted core method for *de novo* protein design. The framework supports beam search [Fernandes et al., 2025], sequential Monte Carlo [Ramesh and Mardani, 2025], and Monte Carlo Tree Search [Singhal et al., 2025, Yoon et al., 2025] (details in Appendix A.4). By the same token, it also supports partial redesign of existing proteins: by noising a known binder and jointly regenerating backbone and sequence, Proteina-Complexa performs codesign-based scaffold re-engineering rather than fixed-backbone sequence optimization, a capability that we apply to the Nipah virus target (Section 8). [file:2]

In practice, we do not optimize against a single proxy for binder quality, but instead search and subsequently filter over combinations of complementary reward terms that capture distinct aspects of the design landscape. These rewards include structure-prediction confidence metrics from tools such as AlphaFold-Multimer [Evans et al., 2022] for protein targets and RosettaFold3 [Corley et al., 2025] for small-molecule targets, protein language model likelihoods that favor sequence plausibility, force-field interaction terms including hydrogen-

bond and energy-based evaluations [Leaver-Fay et al., 2025], and geometric constraints that encourage physically plausible interfaces, such as pocket burial, clash avoidance, and hotspot engagement. Using mixtures of these objectives allows the search procedure to explore a broader and more biologically relevant design space, as different rewards emphasize complementary properties including foldability, interface confidence, sequence naturalness, interaction chemistry, and target-specific shape complementarity. This multi-objective formulation is particularly important for enabling diverse screening libraries, where no single score is sufficient to capture all of the features associated with experimentally successful binders.

3. Massive-Scale Benchmark: Testing 1 Million Binders for 127 Targets

To validate the performance of the Proteina-Complexa framework, a multiplexed *in vitro* campaign was conducted at unprecedented scale, with more than one million designed proteins screened against 127 targets across more than 1,000 design tasks. A campaign of this magnitude demands not only a powerful generative model but also a computationally efficient one: Proteina-Complexa's latent generative framework enables rapid sampling of diverse, high-quality binder candidates, making it uniquely suited to large-scale design exploration where slower structure-based methods would be prohibitively expensive. To our knowledge, this provided the first large-scale experimental framework in which contemporary open binder-design methods, including Proteina-Complexa, RFdiffusion, BindCraft and BoltzGen, could be compared head-to-head in a unified wet-lab setting across both design methodology and sequence re-design strategy. The breadth of the multiplexed readout allowed on-target activity and off-target engagement to be resolved across a broad target panel, yielding a richer view of binding specificity than is accessible in conventional small-scale validation studies. Within this framework, self-generated sequences from Proteina-Complexa gave the strongest overall performance, providing, to our knowledge, the first *in vitro* evidence that end-to-end co-generated sequences can be evaluated at scale and can outperform both leading open design methods and sequence re-design pipelines. Overall, the design campaign required more than 140,000 GPU hours, and the complete details of the sampling procedure, search configuration, and in silico success criteria are provided in Appendix B.

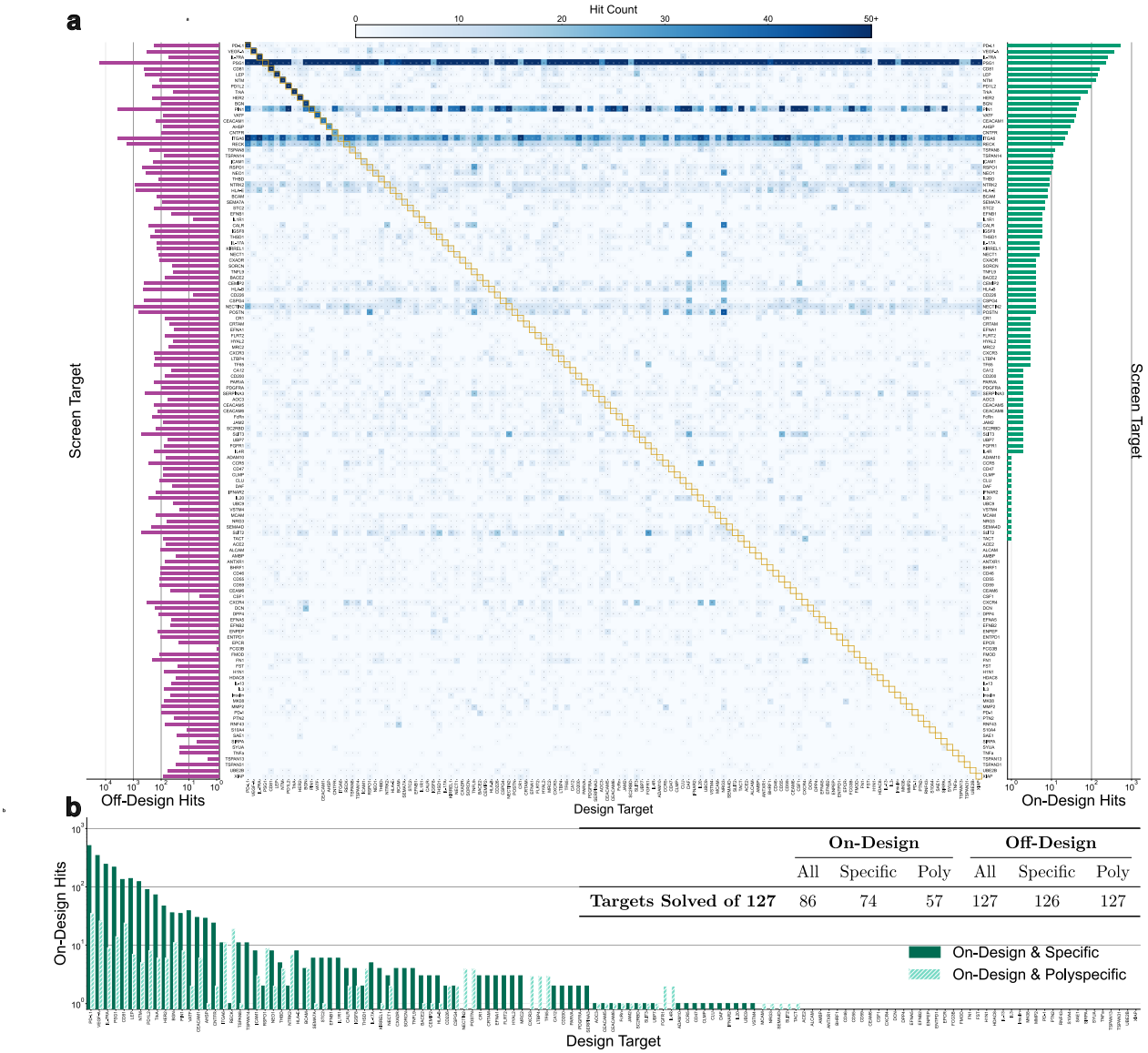
3.1. Co-generated binders cover two-thirds of a 127-target panel

High-throughput screening of 467,176 Proteina-Complexa-generated sequences across the full 127-target panel reveals broad coverage of the target landscape (Fig. 2a). The hit distribution heatmap shows on-design hits concentrated along the diagonal, confirming that designed sequences preferentially bind their intended targets. Proteina-Complexa produced at least one on-design hit for 86 of 127 targets (68%), with 74 of those yielding target-specific binders and 57 yielding poly-specific binders engaging 2–4 targets (Fig. 2b). Off-design binding was also pervasive: 126 of 127 targets had at least one specific off-design hit, indicating that the designed library contains latent binding activity beyond the intended targets. This dual on- and off-design coverage demonstrates that Proteina-Complexa explores a sufficiently diverse structural space to generate functional binders across a large and varied target panel from a single design campaign.

The choice of interface hotspot residues used to condition generation has a substantial effect on experimental success. Across targets with multiple tested hotspot combinations, hit fractions vary widely, and the computationally top-ranked hotspot does not always yield the highest experimental hit rate (Fig. 11). Some targets are robustly solved regardless of hotspot selection, while others are highly sensitive to the conditioning residues, underscoring the value of evaluating multiple hotspot combinations per target rather than relying on a single in silico prediction.

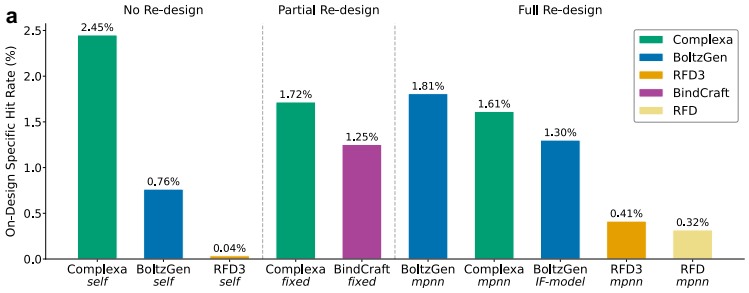
3.2. End-to-end codesign outperforms all baselines and post hoc re-design

To compare Proteina-Complexa against contemporary open methods, we analyzed on-design specific hit rates for the top-ranked in silico hotspot across the 75 targets where at least one method achieved at least one on-design hit (Fig. 3a,b). Proteina-Complexa with self-generated sequences achieved an on-design specific hit rate of 2.45%—more than three times higher than the next-best self-generated baseline (BoltzGen self, 0.76%) and nearly double the best fully redesigned baseline (BoltzGen with ProteinMPNN, 1.81%). In absolute numbers, Proteina-Complexa self-generated sequences produced 691 on-design hits (630 specific) from 25,707 tested sequences, compared to 514 (414 specific) for BoltzGen with ProteinMPNN, 311 (257 specific) for BindCraft, 86 (83 specific) for RFdiffusion3 (RFD3), and 63 (56 specific) for RFdiffusion (RFD).



	On-Design			Off-Design		
	All	Specific	Poly	All	Specific	Poly
Targets Solved of 127	86	74	57	127	126	127

Figure 2 | **Specificity landscape across 127 targets reveals broad target coverage. a)** Heatmap of screening hit counts from Proteina-Complexa designs. Data include 467,176 unique sequences screened across 127 overlapping design/screen targets and side-bar totals. Rows correspond to screen targets and columns to design targets. Color intensity indicates the number of hits in each cell (capped at 50 for visualization), and numeric annotations denote the corresponding counts. Targets are ordered by total on-design hits, defined as *on-design specific* + *on-design polyspecific*. Here, *on-design specific* denotes confident engagement of exactly one target (the design target), whereas *on-design polyspecific* denotes confident engagement of the design target plus additional targets (2–4 total targets). Side panels summarize per-target off-design hits (left) and on-design hits (right) on logarithmic scales. **b)** Per-target on-design composition in the same target order as panel **a**, shown as log-scale bars split into *On-Design & Specific* (darker green) and *On-Design & Polyspecific* (lighter striped green). The embedded summary table reports target-level coverage across categories (targets with at least one hit).



b

Method	On-Design			# Seqs.
	All	Specific	Specificity	
RFD (mpnn)	63	56	88.9%	17,607
RFD3 (mpnn)	86	83	**96.5%**	20,014
RFD3 (self)	16	10	62.5%	28,188
BindCraft (fixed)	311	257	82.6%	20,511
BoltzGen (mpnn)	514	414	80.5%	22,881
BoltzGen (IF-model)	375	300	80.0%	23,063
BoltzGen (self)	244	195	79.9%	25,509
Proteina-Complexa (mpnn)	365	314	86.0%	19,449
Proteina-Complexa (fixed)	374	334	89.3%	19,444
Proteina-Complexa (self)	**691**	**630**	91.2%	25,707

Figure 3 | Baseline comparison of binder design methods on solvable targets. (a) On-design specific hit rate (%) for each method-sequence redesign combination, evaluated on the 75 targets where at least one method achieved an on-design hit. Methods are grouped by redesign strategy: no redesign (native model output), partial redesign (fixed interface), and full redesign (inverse folding or MPNN-generated sequences). (b) On-design hit counts and specificity across methods. Specificity is defined as the ratio of on-design specific hits to total on-design hits, quantifying the fraction of on-design binders that engage exactly one target. Bold indicates best, underline indicates second best per column.

Critically, the sequence design strategy matters as much as the backbone generation method. For Proteina-Complexa, self-generated sequences (691 hits, 2.45%) outperformed both ProteinMPNN-redesigned sequences (365 hits, 1.61%) and fixed-interface redesign (374 hits, 1.72%) applied to the same backbones. This pattern did not hold for other methods: BoltzGen performed best with ProteinMPNN redesign (1.81%) rather than its own self-generated sequences (0.76%), and RFdiffusion3 self-generated sequences yielded only 0.04%. These results establish that Proteina-Complexa is the only method by which co-generated sequences consistently outperform post hoc redesign, providing the first large-scale experimental evidence that end-to-end codesign can eliminate the need for separate inverse-folding models. For further ablations on the impact of sampling temperature on Proteina-Complexa, please see Appendix B.

4. Biophysical Characterization of Designed Monomers

A prerequisite for reliable binder design is that the underlying generative model produces well-folded, stable proteins with high-quality sequences. Many current all-atom generation frameworks are limited to shorter chains and rely on ProteinMPNN for sequence assignment, inherently coupling their output quality to that of the inverse folding model. Proteina-Complexa's partially latent representation, by contrast, co-generates sequence and structure jointly and scales to proteins of several hundred residues. Therefore, we first characterized the biophysical properties of Proteina-Complexa-generated monomers.

We evaluated 48 designed monomeric folds spanning 100 to 800 amino acids (Fig. 4a). For 24, the sequence was co-generated with the backbone by Proteina-Complexa; for the other 24, a ProteinMPNN-designed sequence was assigned to the same Proteina-Complexa backbone, enabling a controlled comparison. Under standard *E. coli* expression conditions, 41 of 48 constructs were solubly expressed and yielded detectable purified protein after affinity purification and size-exclusion chromatography (SEC). For the 38 constructs with no significant aggregation, SEC coupled to multi-angle light scattering (MALS) confirmed the expected monomeric molecular weight (Fig. 4b). Circular dichroism (CD) spectroscopy indicated that proteins were folded and hyperstable, with melting temperatures exceeding 95°C (Fig. 4d). Of these, 23 constructs were produced at high yields ($\geq$50 mg/L). Critically, these results show no dependence on sequence origin: Proteina-Complexa co-generated sequences performed comparably to ProteinMPNN-designed sequences across all metrics, demonstrating that the codesign framework produces sequences of equivalent quality to the current state-of-the-art inverse folding model—while being fully orthogonal to it.

To quantify this comparison directly, we selected four design pairs sharing the same Proteina-Complexa backbone but differing in sequence source (co-generated versus ProteinMPNN), and measured soluble yields in triplicate under identical conditions (Fig. 4b). Overall yields ranged from 91.6 to 343.0 mg/L. For two pairs (600 and 800 amino acids), ProteinMPNN sequences yielded 2–4$\times$ more protein; for one pair (500 amino acids) the two sources performed equivalently; and for the final pair (800 amino acids) the Proteina-Complexa sequence surpassed ProteinMPNN by 1.5$\times$. These results establish Proteina-Complexa as a viable alternative

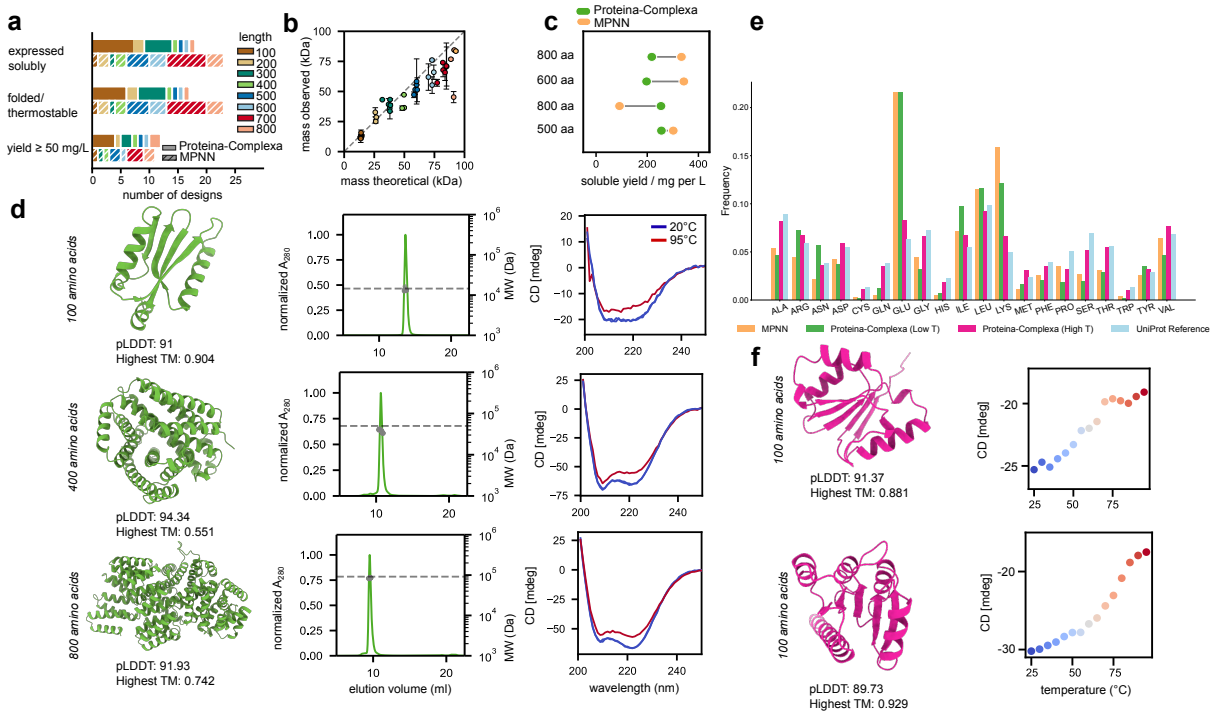


Figure 4 | **Proteina-Complexa-designed monomeric folds and sequences are well expressed and hyperstable. a)** Overall outcome for 48 Proteina-Complexa designs spanning 100–800 amino acids; 24 had sequences co-generated with the backbone. **b)** Expression yield comparison between ProteinMPNN and Proteina-Complexa co-generated sequences for four shared backbones. **c)** For selected folds: design model (left), SEC profile overlaid with MALS-determined molecular weight (middle), and CD spectra at 20°C (blue) and 95°C (red) (right). Average pLDDT from AF2 single-sequence predictions and Foldseek TM-score to the closest known fold are reported. **d)** Theoretical versus MALS-determined molecular weight for the designed set. **e)** Amino acid frequency distributions for ProteinMPNN, Proteina-Complexa at low and high sampling temperatures, and the UniProt reference. **f)** Melting curves visualised via CD spectra from two proteins that were generated with high sampling temperature.

to ProteinMPNN for sequence design on which all previous methods, both generative [Butcher et al., 2025a, Stark et al., 2025, Watson et al., 2023] and hallucination [Cho et al., 2025, Pacesa et al., 2025] rely.

An additional property of the latent generative framework is that sampling temperature provides a continuous knob to modulate the character of generated sequences. At lower sampling temperatures, the model produces sequences with amino acid distributions resembling those of ProteinMPNN—enriched in designability-associated residues and yielding highly soluble, well-expressed proteins (Fig. 4e). At higher temperatures, generated sequences increasingly recapitulate natural amino acid distributions. We posited that such *de novo* sequences (sequence identity <= 40 %), when selected with additional bias towards natural proteins by relying on MSA to be confidently predicted, would display more native-like properties, e.g. lower melting temperatures. Indeed, from a set of 13 selected designs, the two designs that expressed solubly exhibited cooperative melting behaviors under 95 °C (Fig. 4f). When compared to the low temperature generated set, the observed trade-off between engineered stability and natural-like biophysical properties suggests a potential route to explore the space between maximally stable, designed-like proteins and more dynamic, native-like folds—a property that may prove valuable for applications where conformational flexibility or natural sequence composition is desirable, though beyond the scope of the present work.

5. Designing Binders against PDGFR and PD-L1 with Diverse Topologies

As a first case study, we chose two previously studied targets: PD-L1 [Cao et al., 2022, Watson et al., 2023] as well as the polar PDGFR [Fox et al., 2025], which in previous studies was targeted with specific approaches

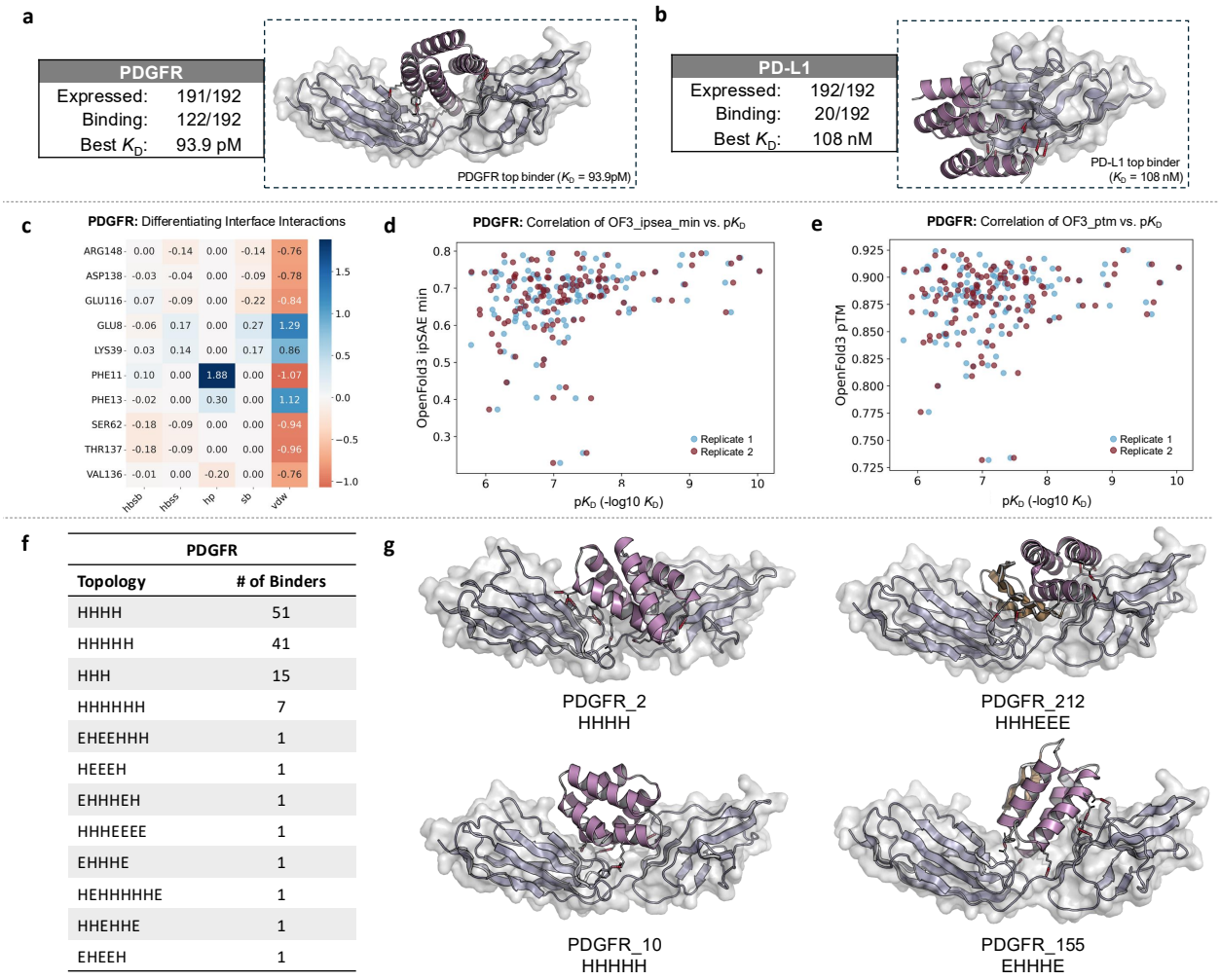


Figure 5 | ***In vitro* results and analysis of *de novo* designed protein binders for PDGFR and PD-L1. a)** Summary of *in vitro* results for PDGFR and predicted structure of top binder. **b)** Summary of *in vitro* results for PD-L1 and predicted structure of top binder. **c)** Difference in average interface interactions between binders and non-binders by residue and interaction type (hbsb = hydrogen bond side-chain to backbone, hbss = side-chain to side-chain, hp = hydrophobic, sb = salt-bridge, vdw = van der Waals). **d)** OpenFold3 ipSAE versus pKd. **e)** OpenFold3 pTM versus pKd. **f)** Secondary structure topological clustering of PDGFR binders. **g)** Predicted complexes of four topologically diverse PDGFR binders.

such as beta-strand pairing due to its polar nature [Sappington et al., 2026]. Interface hotspots were identified by Rosetta alanine scanning of PDB complexes with known binders, retaining positions with `DG_Binding` > 0 (excluding glycines) and selecting the top 9 residues per target. Multiple hotspot combinations were evaluated (Table 3; see Appendix D for detailed analysis): for PDGFR, combinations including core hydrophobic residues (A77, A82, A11) consistently yielded the highest hit rates and affinities, while for PD-L1, multi-residue conditioning improved success rates up to 3-fold over single-residue sets. Binder length was constrained to 55–95 residues.

For each hotspot combination, 1,500 candidates were generated using Proteina-Complexa's standard beam search protocol with AF2 ipAE reward, yielding initial pools of 9,000 for PDGFR and 10,500 for PD-L1. A two-stage filtering pipeline was applied: the first stage filtered on physicochemical properties and confidence metrics (isoelectric point, ESM pseudo-likelihood, minimum ipAE, complex iPTM and pLDDT; details in Appendix D), narrowing the pools to 221 (PDGFR) and 445 (PD-L1) candidates. The second stage validated monomeric structures with OpenFold3 and Boltz-2, requiring pLDDT > 80 and RMSD < 2Å to the Proteina-Complexa complex structure, leaving 196 (PDGFR) and 422 (PD-L1) candidates. From these, 192

designs per target were selected for experimental testing—for PDGFR by decreasing iPTM, for PD-L1 by first retaining 20 non-purely-helical topologies and filling the remaining 172 slots by iPTM.

Of the 192 designs per target, 191 (PDGFR) and 192 (PD-L1) were successfully expressed and purified. SPR identified 122 designs with detectable binding to PDGFR and 20 to PD-L1 (Fig. 5a,b). Affinities ranged from 93.6 pM to 1.34 μM for PDGFR and 108 nM to 3.71 μM for PD-L1. The top PDGFR binder adopts a helical motif complementing a concave groove on the receptor surface, while the top PD-L1 binder presents a flatter interaction face consistent with engaging a protein–protein interaction patch. Overall hit rates were 63.5% for PDGFR—with the most potent binder in the double-digit picomolar range—and 10.4% for PD-L1.

Residue-level interaction analysis of PDGFR (Fig. 5c) reveals that hydrophobic contacts, particularly at hotspot Phe11, are the primary discriminators between binders and non-binders, while salt-bridge and hydrogen-bond signals are weaker. Van der Waals differences further refine the interaction footprint, identifying positions with favorable shape complementarity in binders. Among structure prediction metrics evaluated across multiple methods (OpenFold3, AF2, Boltz-2), only OpenFold3 ipSAE and pTM showed enrichment for stronger binders (Fig. 5d,e): higher values generally correspond to higher pKd, though the correlation is imperfect, suggesting these metrics capture interface quality but miss determinants such as solvation and dynamics.

A key advantage of latent generative search is structural diversity. Topological clustering of PDGFR binders (Fig. 5f) reveals multiple distinct fold families among the hits, including helical insertion motifs that complement the receptor groove as well as alternative scaffolds engaging the surface differently. Four representative complexes (Fig. 5g) illustrate this diversity: all maintain core complementarity to PDGFR but vary in helix orientation, insertion depth, and side-chain engagement. The presence of multiple successful topologies underscores that latent search explores the design space broadly rather than collapsing to a single structural solution—a property critical for early-stage drug discovery, where diverse scaffolds maximize the exploitable chemical space.

6. Designed Binders against ActRIIA block Myostatin Signaling in Cells

Activin receptor type IIA (ActRIIA) has emerged as a compelling therapeutic target in disorders characterized by muscle wasting and lean mass imbalance. It functions as a high-affinity type II receptor for multiple TGF-β superfamily ligands—including myostatin (GDF8), activins, and GDF11—that suppress skeletal muscle growth through Smad2/3 signaling. Clinically, muscle wasting is a pervasive comorbidity across cancer cachexia, chronic organ failure, sarcopenia, and metabolic disease [Rodgers and Ward, 2022]. This is particularly relevant in the context of GLP-1 receptor agonist therapies, where substantial weight loss can be accompanied by unintended lean mass reduction. Blockade of ActRIIA-mediated protein–protein interactions therefore represents an attractive strategy to counteract catabolic signaling. While most therapeutics in development are monoclonal antibodies or ligand traps, non-antibody modalities capable of disrupting endogenous ActRIIA ligand engagement remain scarce.

Proteina-Complexa was employed to generate de novo mini-binders targeting ActRIIA. Seven rounds of design were carried out, varying sampling size, search algorithms (beam search, Feynman–Kac steering, and MCTS), and hotspot conditioning. Interface hotspots were selected from the ActRIIA:Bimagrumab and ActRIIA:Activin A complexes, comprising residues N36, K39, D40, E48, D53, D55, K56, W79, L80, D81, D82, and N84, with spatially clustered subsets used to diversify the designs. Candidates passing filtering criteria were re-ranked by averaged complex ipSAE score, and the top 200, after scaffold-level clustering with FoldSeek to ensure structural diversity, were selected for experimental validation. Sequences were co-generated by Proteina-Complexa rather than redesigned post hoc, reflecting the inherently coupled codesign framework.

Among the 200 selected designs, 192 were successfully expressed and purified, of which 110 showed high yields (>50 mg/L). All 192 mini-binders were analyzed by SPR, identifying 16 designs with detectable ActRIIA binding, 8 of which exhibited sub-μM affinities (Fig. 6 and Fig. 17). Equilibrium dissociation constants (K_{D}) and $R_{\max}$ values, obtained from two independent measurements, were largely consistent across replicates. The tightest binder, design #51, showed a K_{D} of 36 nM. Structure modeling suggests that all identified binders adopt helical conformations and engage the receptor at its protein–protein interaction interface (Fig. 6). Despite converging on a similar binding mode, the sub-μM hit sequences display low mutual similarity, with pairwise sequence identity ranging from 6% (#17 and #31) to 18% (#17 and #104). In general, the campaign

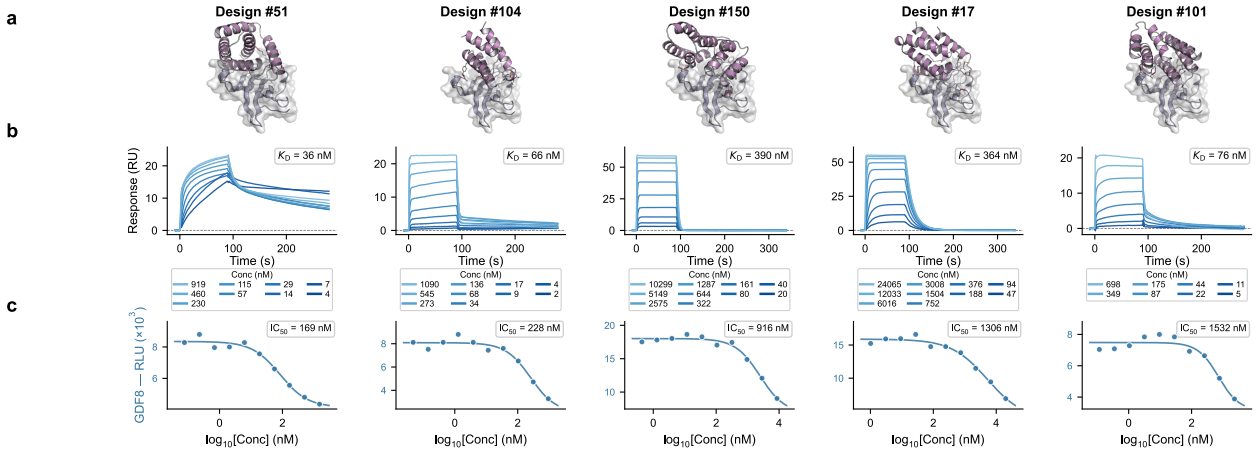


Figure 6 | ***In vitro* results and analysis of de novo designed protein binders for ActRIIA. a)** Predicted structures of binders that achieved nanomolar affinity and functional downstream effects. **b)** SPR sensorgrams of each design with corresponding K_D values. **c)** Efficacy/potency of designed mini binders at blocking activin A or myostatin-induced Smad2/3 response. Designs were tested for their ability to inhibit myostatin (also called GFD8, 200 ng/mL) -induced Smad2/3 response in a CAGA-luciferase reporter gene assay in HEK293T/17 cells.

yielded an 8.3% hit rate from raw computational designs without experimental optimization.

To assess whether the designed binders could functionally interfere with ligand-induced signaling, selected designs were tested in a Smad2/3-responsive luciferase reporter assay using HEK293 cells stimulated with either Activin A (20 ng/mL) or GDF8/myostatin (200 ng/mL). Several designs showed measurable inhibition of ligand-induced signaling (Fig. 6 and Fig. 16). Under Activin A stimulation, designs #150, #17, and #51 exhibited modest inhibition. Under GDF8 stimulation, functional effects were more pronounced: designs #104 and #51 showed IC_{50} values of 227.5 nM and 168.9 nM, respectively, while designs #150, #17, and #101 exhibited IC_{50} values of 916 nM, 1,306 nM, and 1,532 nM. These results are consistent with the biophysical affinity measurements and indicate that the designed binders engage the receptor in an orientation overlapping with the native ligand interface, functionally disrupting Smad2/3 signaling. That computationally designed mini-proteins—with no affinity maturation—already exhibit therapeutically relevant IC_{50} values in a cellular assay underscores the quality of the codesign framework. Experimental details are provided in Sec. E.

7. Designing Peptides and Minibinders against Kinase Targets

Protein kinases govern nearly all biological signaling through phosphorylation of serine, threonine, and tyrosine residues [Manning et al., 2002, Olsen et al., 2006], yet genetically encoded tools for interrogating their catalytic activity remain limited. To address this gap, we designed de novo binders targeting the catalytic domains of two kinases—PAK1 (PAK1c) and CK1δ (CK1δc)—using Proteina-Complexa. Critically, these campaigns span two distinct size regimes: conventional mini-protein binders (49–74 amino acids) for PAK1 and short peptide binders (<31 amino acids) for CK1δ. Designing functional binders in the peptide regime is particularly challenging due to the drastically reduced conformational complexity and binding surface area, and high success rates in this regime would demonstrate that the latent codesign framework generalizes across the full protein-to-peptide size spectrum.

7.1. PAK1 mini-protein binder selection and validation

Group I PAK kinases are central effectors of Rho-family GTPases that regulate cytoskeletal dynamics and cell survival [Chen et al., 2010, Edwards et al., 1999, Kreis and Barnier, 2009, Zhang et al., 2022]. We generated a library of 50 mini-protein binder candidates (49–74 amino acids) targeting PAK1c using Proteina-Complexa with its standard beam-search protocol and an AlphaFold2 interface pAE reward. These candidates were screened using a split-neomycin resistance (split-NeoR) protein-fragment complementation assay in

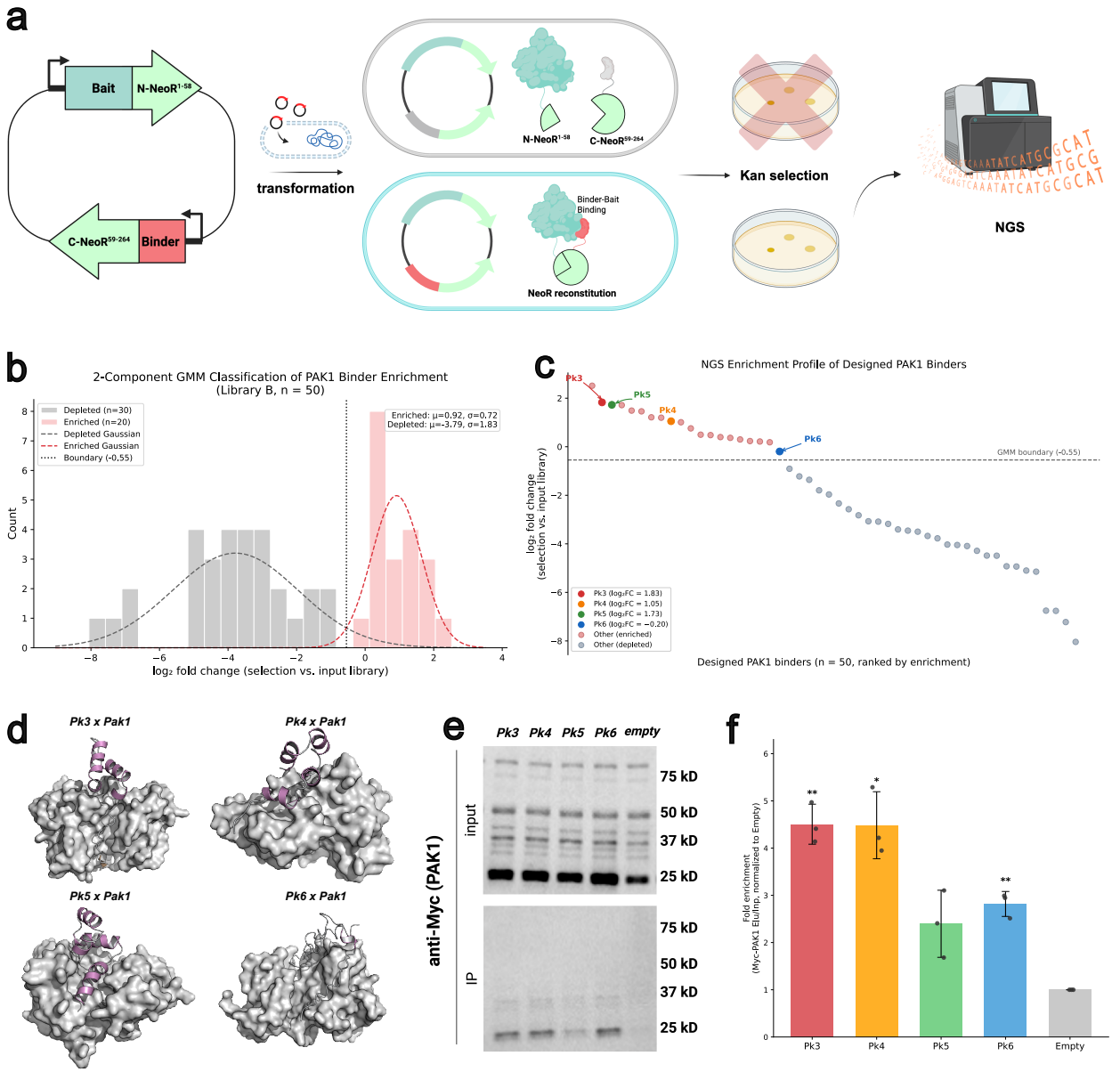


Figure 7 | **Selection and validation of *de novo* designed PAK1 binders. a)** Schematic of the split-NeoR protein-fragment complementation assay: binder candidates fused to C-NeoR and PAK1 bait fused to N-NeoR are co-expressed in *E. coli*; reconstitution of aminoglycoside phosphotransferase activity confers kanamycin resistance, and surviving colonies are analyzed by NGS. **b)** Two-component Gaussian mixture model (GMM) fitted to $\log_2$ fold change values of 50 designed PAK1 binders, classifying an enriched group ($n = 20$, pink) and a depleted group ($n = 30$, gray). Dashed line: GMM decision boundary ($\log_2\text{FC} = -0.55$). **c)** NGS enrichment profile for all 50 designed binders ranked by enrichment; Pk3 (red), Pk4 (orange), Pk5 (green), and Pk6 (blue) are highlighted. **d)** AlphaFold2 Multimer predicted structures of binder–PAK1 complexes (PAK1 in green, binder in magenta). **e)** Representative Western blot of HA-Trap co-immunoprecipitation from HEK293T cells. **f)** Quantification of Myc-PAK1 co-precipitation (mean fold enrichment ± SD, $n = 3$; $^*p < 0.05$, $^{**}p < 0.01$, one-sample t-test vs. 1).

E. coli [He et al., 2025, Paschon et al., 2005], in which binder–target engagement reconstitutes aminoglycoside phosphotransferase activity and confers kanamycin resistance (Fig. 7A). NGS analysis of colonies surviving selection showed that 20 of the 50 designs (40%) were classified as enriched by a two-component Gaussian mixture model (Fig. 7B,C).

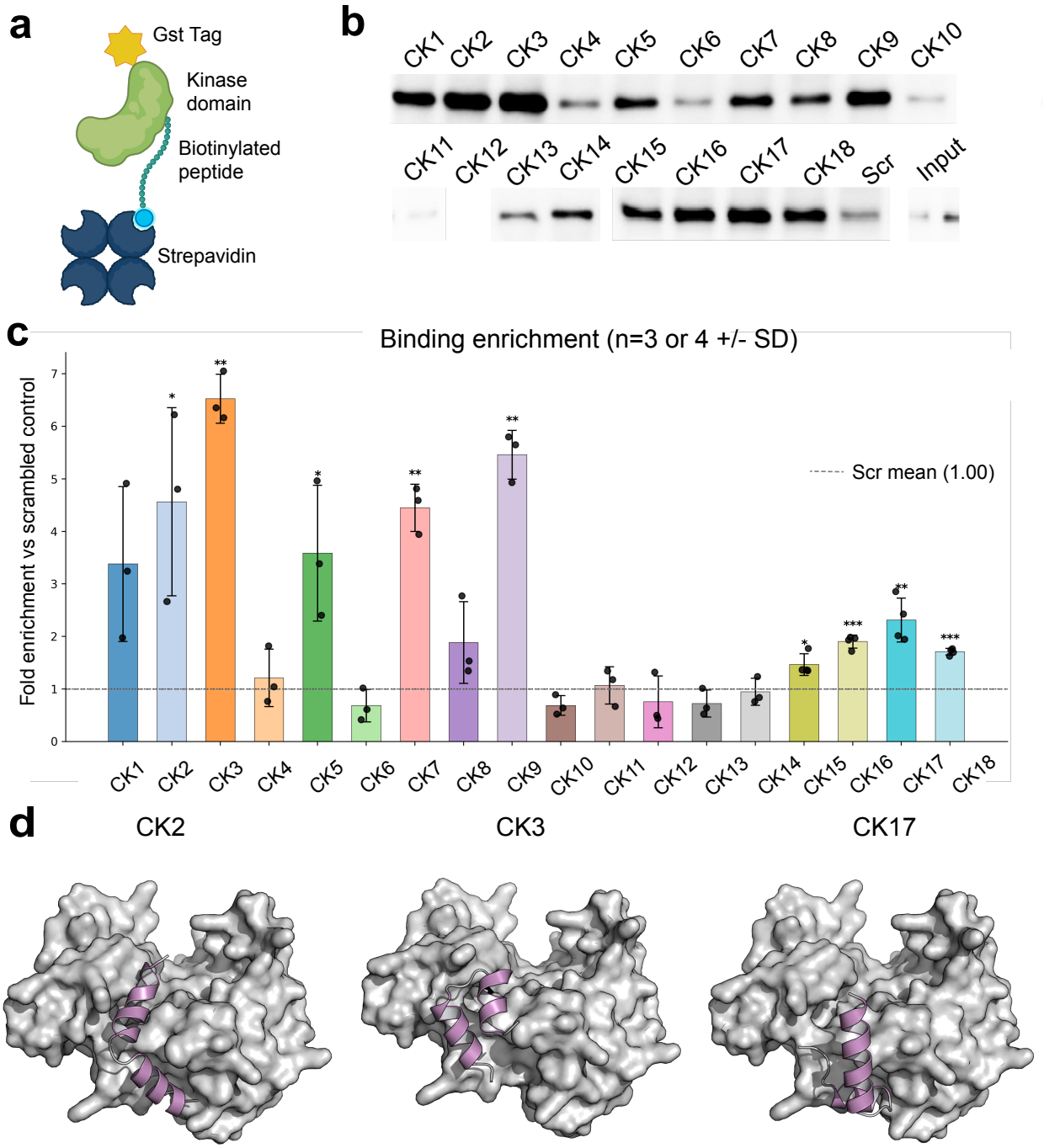


Figure 8 | **Analysis of Proteina-Complexa-designed CK1δc peptide binders by affinity pull-down.** **a)** Schematic of the streptavidin bead pull-down assay: predicted peptide binders are immobilized via a biotinylated residue to magnetic streptavidin beads, then used to capture purified GST-CK1δc. **b)** Representative Western blot (anti-GST) for each binder (CK1–CK18) versus a scrambled peptide control (Scr) and total input. **c)** Quantification of binding enrichment normalized to the scrambled control (dashed line, scr mean = 1.00); bars represent mean ± SD ($n = 3$–4); $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, one-sample one-tailed t-test vs. 1. **d)** Visualization of a subset of successful binders (CK2, CK3, and CK17), binding to CKδc.

Four candidates were selected for mammalian validation: Pk3 and Pk5 (top enriched), Pk4 (mid-ranked), and Pk6 (at the GMM boundary, $\log_2\text{FC} = -0.20$). Co-immunoprecipitation in HEK293T cells confirmed that Pk3 and Pk4 each achieved ∼4.5-fold enrichment over the empty-vector control ($p < 0.01$ and $p < 0.05$, respectively), while Pk6 showed moderate but significant enrichment (2.82 ± 0.26, $p < 0.01$; Fig. 7E,F). AlphaFold2 Multimer ipTM+pTM scores were consistent with these results, with the two strongest binders by co-IP (Pk3 and Pk4) having the highest predicted scores (0.848 and 0.765; Fig. 7D). Notably, Pk6—classified at the bacterial selection boundary—still showed significant co-precipitation in mammalian cells, suggesting the bacterial assay may underestimate binding activity for some candidates.

7.2. CK1δ peptide binder validation

Casein kinase 1 delta (CK1δ) plays critical roles in circadian rhythm regulation, Wnt signaling, and neurodegeneration [Eng et al., 2017, Li et al., 2004, Rosenberg et al., 2015, Xu et al., 2019], making it an attractive target for selective tool development. To test whether Proteina-Complexa can design functional binders in the short-peptide regime, we synthesized 18 computationally derived CK1δc-targeting peptides (<31 amino acids; Table 6) via Fmoc solid-phase synthesis, each bearing a biotinylated lysine for streptavidin pull-down (Fig. 8A). Binding to purified GST-tagged CK1δc was assessed by Western blot.

Of the 18 candidates, 9 (50%) achieved statistically significant enrichment over the scrambled control ($p < 0.05$, one-tailed t-test), with several exceeding 5-fold enrichment (Fig. 8B-D). The strongest binders were CK3 (6.52 ± 0.47, $p = 0.001$), CK9 (5.46 ± 0.46, $p = 0.002$), and CK7 (4.45 ± 0.45, $p = 0.003$). A second tier of binders—CK15, CK16, CK17, and CK18—showed more modest but highly reproducible enrichment (1.5–2.3-fold, $p < 0.01$). That a 50% hit rate is achieved for peptides of fewer than 31 amino acids—a regime where reduced conformational complexity and minimal binding surface area make design particularly challenging—underscores the generality of the latent codesign framework. Beyond validating the model's ability to operate across size regimes, these peptide binders open practical applications as research tools, cell-permeable probes, and starting points for peptidomimetic development that are less accessible to larger protein scaffolds. Experimental details are provided in Appendix F.

8. *De Novo* Design and Binder Re-Engineering for the Nipah Virus

Nipah virus (NiV) is a BSL-4 paramyxovirus with case fatality rates of 40–75% and no approved vaccines or therapeutics, making it a WHO priority pathogen [Moore et al., 2024]. The NiV attachment glycoprotein (NiV-G) mediates host cell entry by engaging the ephrin-B2 and ephrin-B3 receptors, and its receptor-binding site represents a critical epitope for neutralization. Designing binders that engage this specific site—rather than elsewhere on the glycoprotein surface—is essential for downstream antiviral function but poses a considerable challenge, as the receptor-binding pocket is recessed and partially occluded.

We evaluated Proteina-Complexa on this target in two complementary modes as part of the Adaptivy binder competition: fully *de novo* design and structure–sequence redesign of an existing binder scaffold via partial diffusion. For the *de novo* campaign (Fig. 9a), candidates were generated conditional on hotspot residues located at the receptor-binding site and explored using beam search. Search was guided by a multi-objective reward that combined Boltz-2 maximum ipSAE with interface-focused terms for shape complementarity, interface hydrophobicity, and hydrogen-bond satisfaction, thereby enriching for designs with both high structural confidence and chemically plausible binding interfaces. Following downstream filtering and ranking, 14 *de novo* designs were selected for experimental characterization.

For the redesign campaign, we applied Proteina-Complexa's diffuse–denoise protocol (Fig. 9b), in which an input binder structure is partially noised and both backbone and sequence are jointly regenerated, enabling true codesign-based redesign rather than sequence-only optimization. Using this procedure, six redesigned candidates were selected for experimental testing.

All 20 designs were expressed, purified, and characterized by SPR (Fig. 9C). From the 14 fully de novo designs, one binder was identified with a K_D of 56 nM. From the 6 diffuse-denoise redesigns, 5 showed detectable binding, all in the nanomolar range. Predicted structures of all six binders engaging the NiV-G receptor-binding site are shown in Fig. 9B. This illustrates an additional strength of the latent codesign framework: because both structure and sequence are jointly optimized during partial diffusion, the method performs genuine codesign-based scaffold reengineering, in contrast to conventional approaches that fix the

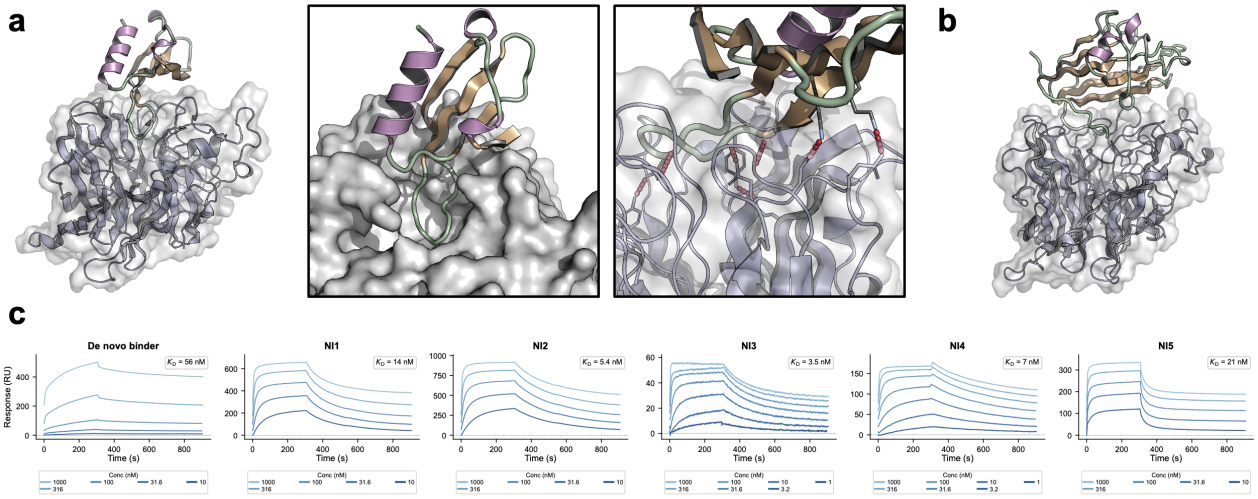


Figure 9 | ***De novo* design and binder re-engineering for the Nipah virus attachment glycoprotein (NiV-G). a)** Schematic of the diffuse-denoise codesign protocol: an existing binder scaffold is partially noised and jointly re-generated in both backbone and sequence, guided by inference-time rewards, performing true structure–sequence codesign-based redesign rather than fixed-backbone sequence optimization. **b)** Predicted structures of designed binder–NiV-G complexes. Left: the de novo binder (K_D = *56* nM). Right: five diffuse-denoise redesigned binders, all with nanomolar affinity. NiV-G is shown in surface representation with the receptor-binding site highlighted. **c)** SPR sensorgrams for all binders with detectable binding, with equilibrium dissociation constants (K_D) annotated.

backbone and redesign only the sequence. The de novo hit at 56 nM demonstrates that Proteina-Complexa can generate functional binders to difficult viral epitopes from scratch, while the redesign results show that the same framework can efficiently diversify existing scaffolds when a starting point is available. Importantly, both modes target the receptor-binding site required for downstream neutralizing activity, positioning these binders as useful starting points for potential therapeutic development.

9. Beyond Protein Targets: De Novo Binder Design for ABO-Transplant Compatibility Carbohydrates

A persistent limitation of *de novo* binder design is the dependence on hydrophobic interface contacts. Across current methods—including RFdiffusion [Watson et al., 2023], BindCraft [Pacesa et al., 2025], RFdiffusion3 [Butcher et al., 2025b], and BoltzGen [Stark et al., 2025]—success rates correlate strongly with epitope hydrophobicity, and all rely on ProteinMPNN [Sumida et al., 2024], which struggles with polar interface residues. Highly polar target sites remain a considerable challenge [Cao et al., 2022]. Carbohydrates represent the extreme case: they are simultaneously small and densely polar, presenting hydroxyl-rich surfaces with no hydrophobic character that models can exploit. Lectins and carbohydrate-binding modules (CBMs) typically bind sugars with millimolar affinity [Lis and Sharon, 1998, Weis and Drickamer, 1996], because displacing the hydration shell of sugar hydroxyls incurs a large desolvation penalty [Dam and Brewer, 2002, Lemieux, 1996, Toone, 1994]. To our knowledge, no computational method has demonstrated *de novo* design of a protein that binds a free carbohydrate with measurable affinity.

The ABO histo-blood group antigens—among the first human glycan systems to be elucidated [Watkins, 2001]—are attractive targets for *de novo* design. These terminal glycan caps are present on diverse cell types, and incompatible antigens cause hyperacute rejection of blood transfusions and solid organ transplants, both of which can be fatal. Binders designed against these antigens could enable custom enzymes that efficiently convert between ABO types for universal organ donation, as concentrating active sites on cell surfaces greatly aids enzymatic conversion [Watkins, 2001].

We selected the blood group B antigen—a terminal trisaccharide (α-D-Gal-(1→3)-[α-L-Fuc-(1→2)]-D-Gal)—as a stringent test (Fig. 10a). Proteina-Complexa was conditioned on the sugar ligand structure with

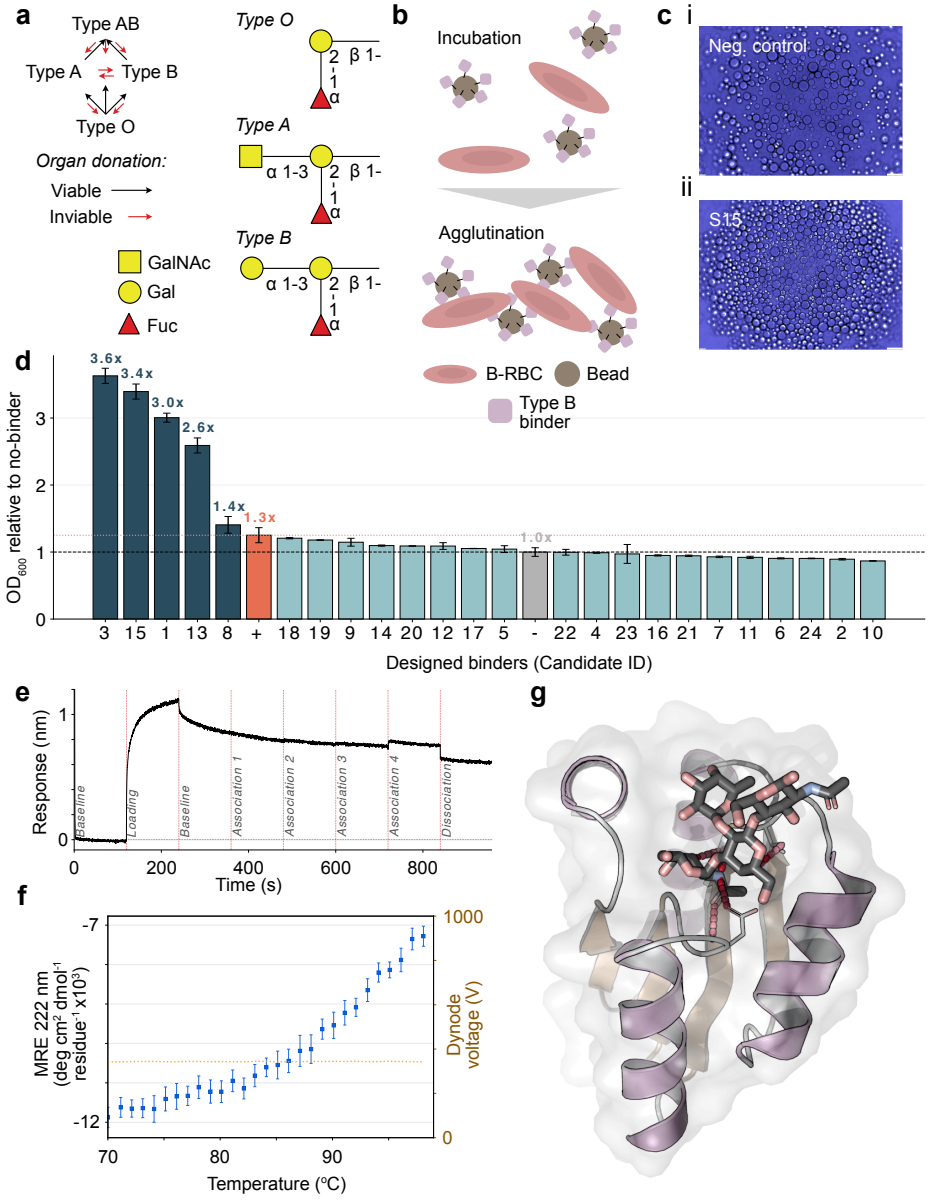


Figure 10 | ***De novo* designed protein binders for the blood group B carbohydrate antigen. a)** The ABO blood group system and its role in transplant compatibility. Coloured arrows indicate the direction and feasibility of organ donation between blood types. Blood group A, B, and O (H) antigen structures are displayed; the A and B antigens differ by a single terminal sugar modification (GalNAc vs. Gal), while the H antigen (type O) lacks the terminal sugar. **b)** Schematic of the modified hemagglutination assay: designed binders immobilized on Ni-NTA beads cross-link blood group B antigens on RBCs, causing agglutination. **c)** Confocal microscopy of (i) the negative control (beads + RBCs, no protein) and (ii) the fastest-appearing agglutination hit (NV15), focused on the region of greatest bead density. **d)** Quantification of agglutination signal for all 24 designs using OD_{600} normalized to the negative control (mean ± SD, $n = 3$ replicate plates). Designs NV1, NV3, NV13, and NV15 show strong signals (2.6–3.6×), substantially exceeding the binding positive control (1.3×). **e)** Biolayer interferometry (BLI) sensorgram for NV15. Reference-subtracted response (nm) vs. time; phase boundaries are indicated. Four sequential association steps at increasing glycan–DNA conjugate concentrations are followed by dissociation, confirming direct and concentration-dependent binding to the carbohydrate target. **f)** Circular dichroism thermal stability analysis of NV15. Mean residue ellipticity (MRE) at 222 nm is plotted as a function of temperature (left axis). Dynode voltage (right axis) is reported to monitor for aggregation, which was not observed. Secondary structure elements remain stable beyond 95 °C, precluding accurate determination of a melting temperature within the accessible range. **g)** Predicted structure of design NV15 (15, 87 aa; cyan cartoon) in complex with the blood group B trisaccharide ligand (pink sticks). Interface side chains are shown as sticks.

pocket burial and hydrogen bond rewards as well as interface confidence metrics from RosettaFold3 (RF3). For these, RF3 was run without ligand templating and the update ligand was fed back to the generative trajectory, allowing the model to reason over different ligand conformations depending on the generated protein structure. After filtering, 24 designs (64–88 amino acids) were expressed in *E. coli*, with most yielding soluble protein at the expected size (Fig. 18).

To assess binding on intact RBCs, we performed a plate-based modified hemagglutination assay in triplicate (Fig. 10b–d). Five designs exhibited strong agglutination signals: NV3 (3.6×), NV15 (3.4×), NV1 (3.0×), NV13 (2.6×) and NV8 (1.4×)—all substantially exceeding the binding positive control (1.3×). Signals were highly reproducible across replicates (SD <0.13 for top hits). Biolayer interferometry on NV15 (87 amino acids) confirmed direct carbohydrate engagement, with fast on- and off-rates characteristic of hydrogen-bond-dominated protein–carbohydrate interactions [Dam and Brewer, 2002] (Fig. 10e). Circular dichroism confirmed that NV15 is well folded with secondary structure stable beyond 95 °C (Fig. 10f). Experimental details are provided in Sec. G.

These results represent, to our knowledge, the first *de novo* computationally designed proteins that bind a free carbohydrate. From a single round of design with no experimental optimization, Proteina-Complexa produced five binders from 24 candidates (>20% hit rate) against a target class that requires simultaneously overcoming the polar interface problem, the small molecule binding pocket problem, and the carbohydrate desolvation penalty—challenges that have prevented previous methods from succeeding at tasks like this one.

10. Conclusions

This work demonstrates that joint sequence–structure codesign with inference-time search in a continuous latent space produces functional protein binders across targets and scales no prior method has reached. Co-generated sequences from Proteina-Complexa consistently outperform competing generative models, optimization baselines, and post hoc sequence redesign strategies in compute-matched comparisons across 127 targets—confirming that eliminating the decoupling of backbone generation from sequence design translates into measurable experimental gains.

Screening over one million synthesized proteins against 127 targets via multiplexed phage display—the first head-to-head comparison of multiple binder design methods in an all-to-all format—yields a dataset of unprecedented scale.

The individual campaigns demonstrate capabilities beyond existing methods: picomolar binders against PDGFR spanning diverse topologies; ActRIIA binders that block myostatin signaling in cells without affinity maturation; nanomolar Nipah virus binders from joint structure–sequence scaffold redesign; mini-protein and peptide binders against two protein kinases with high hit rates in both size regimes, including a 50% success rate for peptides of fewer than 31 amino acids; and functional carbohydrate-binding proteins—to our knowledge, the first computationally designed proteins that bind free carbohydrates. Each result addresses a distinct challenge: structural diversity, downstream therapeutic function, scaffold reengineering, size-regime generality from peptides to proteins, and binding at densely polar interfaces devoid of hydrophobic character.

Important open questions remain. While codesign outperforms sequence redesign on average, the margin varies across targets, and understanding which epitope properties benefit most from joint optimization requires further study.

The success with carbohydrates motivates extension to other non-protein modalities. More broadly, the principles of generation-guided search—combining a learned prior with adaptive inference-time compute—now extend from game playing and language modeling to the design of functional proteins, opening target classes and scales previously out of reach.

Acknowledgments

M.B. is partially supported by the EPSRC Turing AI World-Leading Research Fellowship No. EP/X040062/1 and EPSRC AI Hub No. EP/Y028872/1. M.S. is supported by the National Research Foundation of Korea grants (NRF) (2020M3-A9G7-103933, RS-2021- NR061659, RS-2021- NR056571, RS-2024-00396026 and RS-2020-NR049543), the Novo Nordisk Foundation NNF24SA0092560 and the Creative-Pioneering

Researchers Program. S.S. is supported by the NIH grants NS147032 and MH111684. The Khmelinskaia lab acknowledges funding by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder through the ONE MUNICH Project Munich Multiscale Biofabrication, the LMU Munich Excellence Investment Fund, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Individual Project 558049915 and the Excellence Cluster BioSysteM, under Germany's Excellence Strategy - EXC3092 - 533751719.

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Zidek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 2024. 1

Katherine I Albanese, Sophie Barbe, Shunsuke Tagami, Derek N Woolfson, and Thomas Schiex. Computational protein design. *Nature Reviews Methods Primers*, 5(1):13, 2025. 2

Nathaniel R. Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, Frank DiMaio, Steven De Munck, Savvas N. Savvides, and David Baker. Improving de novo protein binder design with deep learning. *Nature Communications*, 14:2625, 2023. 1

Brian Bushnell, Jonathan Rood, and Esther Singer. Bbmerge–accurate paired shotgun read merging via overlap. *PloS one*, 12(10):e0185056, 2017. 28

Jasper Butcher, Rohith Krishna, Raktim Mitra, Rafael I. Brent, Yanjing Li, Nathaniel Corley, Paul Kim, Jonathan Funk, Simon Mathis, Saman Salike, Aiko Muraishi, Helen Eisenach, Tuscan Rock Thompson, Jie Chen, Yuliya Politanska, Enisha Sehgal, Brian Coventry, Odin Zhang, Bo Qiang, Kieran Didi, Max Kazman, Frank DiMaio, and David Baker. De novo design of all-atom biomolecular interactions with rfdiffusion3. *bioRxiv*, 2025a. doi: 10.1101/2025.09.18.676967. 7, 17

Jasper Butcher, Rohith Krishna, Raktim Mitra, Rafael I. Brent, Yanjing Li, Nathaniel Corley, Paul Kim, Jonathan Funk, Simon Mathis, Saman Salike, Aiko Muraishi, Helen Eisenach, Tuscan Rock Thompson, Jie Chen, Yuliya Politanska, Enisha Sehgal, Brian Coventry, Odin Zhang, Bo Qiang, Kieran Didi, Max Kazman, Frank DiMaio, and David Baker. De novo design of all-atom biomolecular interactions with rfdiffusion3. In *bioRxiv* Butcher et al. [2025a]. doi: 10.1101/2025.09.18.676967. See Butcher et al., 2025. 14

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M Jude, Iva Marković, Rameshwar U Kadam, Koen HG Verschueren, et al. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022. doi: 10.1038/s41586-022-04654-9. 1, 7, 14

Bruce R Carlson, Kyle E Lloyd, Amy Kruszewski, Il Hwan Kim, Ramona M Rodriguiz, Camille Heindel, Melissa Faytell, Serena M Dudek, William C Wetsel, and Scott H Soderling. WRP/srGAP3 facilitates the initiation of spine development by an inverse F-BAR domain, and its loss impairs long-term memory. *J. Neurosci.*, 31(7):2447–2460, 2011. 39

Lulu Y Chen, Christopher S Rex, Alex H Babayan, Eniko A Kramar, Gary Lynch, Christine M Gall, and Julie C Lauterborn. Physiological activation of synaptic Rac>PAK (p-21 activated kinase) signaling is defective in a mouse model of fragile X syndrome. *J. Neurosci.*, 30(33):10977–10984, 2010. 10

Aaron Chevalier, Daniel-Adriano Silva, Gabriel J Rocklin, Derrick R Hicks, Renan Vergara, Patience Murapa, Steffen M Bernard, Lu Zhang, Kwok-Ho Lam, Guorui Yao, et al. Massively parallel de novo protein design for targeted therapeutics. *Nature*, 550(7674):74–79, 2017. 1

Yehlin Cho, Martin Pacesa, Zhidian Zhang, Bruno E. Correia, and Sergey Ovchinnikov. Boltzdesign1: Inverting all-atom structure prediction model for generalized biomolecular binder design. *bioRxiv*, 2025. doi: 10.1101/2025.04.06.647261. 7

Nathaniel Corley, Simon Mathis, Rohith Krishna, Magnus S. Bauer, Tuscan R. Thompson, Woody Ahern, Maxwell W. Kazman, Rafael I. Brent, Kieran Didi, Andrew Kubaney, Lilian McHugh, Arnav Nagle, Andrew Favor, Meghana Kshirsagar, Pascal Sturmfels, Yanjing Li, Jasper Butcher, Bo Qiang, Lars L. Schaaf, Raktim Mitra, Katelyn Campbell, Odin Zhang, Roni Weissman, Ian R. Humphreys, Qian Cong, Jonathan Funk, Shreyash Sonthalia, Pietro Liò, David Baker, and Frank DiMaio. Accelerating biomolecular modeling with atomworks and rf3. *bioRxiv*, 2025. doi: 10.1101/2025.08.14.670328. 3

Tarun K Dam and C Fred Brewer. Thermodynamic studies of lectin–carbohydrate interactions by isothermal titration calorimetry. *Chemical Reviews*, 102(2):387–430, 2002. 14, 16

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022. 42

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M Bronstein, et al. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations*, 2026. 2, 3, 22, 23, 25

Roland L Dunbrack Jr. Rēs ipsae loquunt: Whats wrong with alphafolds iptm score and how to fix it. *bioRxiv*, 2025. 25, 43

Donnal C Edwards, Laura C Sanders, Gary M Bhomick, and Gordon N Gill. Activation of LIM-kinase by Pak1 couples Rac/Cdc42 GTPase signalling to actin cytoskeletal dynamics. *Nat. Cell Biol.*, 1(5):253–259, 1999. 10

Gervais W L Eng, Edison, and David M Virshup. Site-specific phosphorylation of casein kinase 1 delta (CK1delta) regulates its activity towards the circadian regulator PER2. *PLoS One*, 12(5):e0177834, 2017. 13

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with alphafold-multimer. *bioRxiv*, 2022. doi: 10.1101/2021.10.04.463034. 3

Guilherme Fernandes, Vasco Ramos, Regev Cohen, Idan Szpektor, and Jo ao Magalhães. Latent beam diffusion models for decoding image sequences. *arXiv preprint arXiv:2503.20429*, 2025. 3, 24

Noelia Ferruz, Michael Heinzinger, Mehmet Akdel, Alexander Goncearenco, Luca Naef, and Christian Dallago. From sequence to function through structure: Deep learning for protein design. *Computational and Structural Biotechnology Journal*, 21:238–250, 2023. 1

Daniel R Fox, Cyntia Taveneau, Janik Clement, Rhys Grinter, and Gavin J Knott. Code to complex: Ai-driven de novo binder design. *Structure*, 33(10):1631–1642, 2025. 7

Christopher Frank, Dominik Schiwietz, Lara Fuß, Sergey Ovchinnikov, and Hendrik Dietz. Alphafold2 refinement improves designability of large de novo proteins. *bioRxiv*, pages 2024–11, 2024. 2

Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Proteina: Scaling flow-based protein structure generative models. In *International Conference on Learning Representations (ICLR)*, 2025. 22, 23

Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. In *The Fourteenth International Conference on Learning Representations*, 2026. 3, 22, 23

Jingxuan He, Li-Ning Zou, Vidhi Pareek, and Stephen J Benkovic. Multiplex mapping of protein–protein interaction interfaces. *Proc. Natl. Acad. Sci. U.S.A.*, 122(52):e2425774122, 2025. 11, 37

Lu Hong and Tanja Kortemme. An integrative approach to protein sequence design through multiobjective optimization. *PLoS computational biology*, 20(7):e1011953, 2024. 2

Po-Ssu Huang, Scott E Boyken, and David Baker. The coming of age of de novo protein design. *Nature*, 537 (7620):320–327, 2016. 1

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589, 2021. 1

Huan Yee Koh, Yizhen Zheng, Madeleine Yang, Rohit Arora, Geoffrey I Webb, Shirui Pan, Li Li, and George M Church. Ai-driven protein design. *Nature Reviews Bioengineering*, 3(12):1034–1056, 2025. 1

Tanja Kortemme. De novo protein designfrom new structures to programmable functions. *Cell*, 187(3): 526–544, 2024. 1

Clayton W. Kosonocky, Alex M. Abel, Aaron L. Feller, Amanda E. Cifuentes Rieffer, Phillip R. Woolley, Jakub Lála, Daryl R. Barth, Tynan Gardner, Stephen C. Ekker, Andrew D. Ellington, Wesley A. Wierson, and Edward M. Marcotte. Validation and analysis of 12,000 AI-driven CAR-T designs in the Bits to Binders competition. *bioRxiv*, 2026. doi: 10.64898/2026.03.03.709355. preprint. 29

Patricia Kreis and Jean-Vianney Barnier. PAK signalling in neuronal physiology. *Cell Signal.*, 21(3):384–393, 2009. 10

Andrew Leaver-Fay, Jeff Flatten, Alex Ford, Joseph Kleinhenz, Henry Solberg, David Baker, Andrew M. Watkins, Brian Kuhlman, and Frank DiMaio. tmol: a gpu-accelarated, pytorch implementation of rosetta's relax protocol. `https://github.com/uw-ipd/tmol`, 2025. 4

Raymond U Lemieux. How water provides the impetus for molecular recognition in aqueous solution. *Accounts of Chemical Research*, 29(8):373–380, 1996. 14

Guojun Li, Hao Yin, and Jeff Kuret. Casein kinase 1 delta phosphorylates tau and disrupts its binding to microtubules. *J. Biol. Chem.*, 279(16):15938–15945, 2004. 13

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023. 23

Halina Lis and Nathan Sharon. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chemical Reviews*, 98(2):637–674, 1998. 14

Sidney Lyayuga Lisanza, Jacob Merle Gershon, Samuel WK Tipps, Jeremiah Nelson Sims, Lucas Arnoldt, Samuel J Hendel, Miriam K Simma, Ge Liu, Muna Yase, Hongwei Wu, et al. Multistate and functional protein design using rosettafold sequence space diffusion. *Nature biotechnology*, 43(8):1288–1298, 2025. 2

Dina Listov, Casper A Goverde, Bruno E Correia, and Sarel Jacob Fleishman. Opportunities and challenges in design and optimization of protein function. *Nature Reviews Molecular Cell Biology*, 25(8):639–653, 2024. 1

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, et al. Scaling inference time compute for diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2523–2534, 2025. 2

Gerard Manning, David B Whyte, Richard Martinez, Tony Hunter, and Sucha Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934, 2002. 10

Kristine A Moore, Angela J Mehr, Julia T Ostrowsky, Angela K Ulrich, Nicolina M Moua, Petra C Fay, Peter J Hart, Josephine P Golding, Virginia Benassi, Marie-Pierre Preziosi, et al. Measures to prevent and treat nipah virus disease: research priorities for 2024–29. *The Lancet Infectious Diseases*, 24(11):e707–e717, 2024. 13

Jesper V Olsen, Blagoy Blagoev, Florian Gnad, Boris Macek, Chanchal Kumar, Peter Mortensen, and Matthias Mann. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127(3): 635–648, 2006. 10

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, Yehlin Cho, Kourosh H. Ghamary, Laura Vinué, Brahm J. Yachnin, Andrew M. Wollacott, Stephen Buckley, Adrie H. Westphal, Simon Lindhoud, Sandrine Georgeon, Casper A. Goverde, Georgios N. Hatzopoulos, Pierre Gönczy, Yannick D. Muller, Gerald Schwank, Daan C. Swarts, Alex J. Vecchio, Bernard L. Schneider, Sergey Ovchinnikov, and Bruno E. Correia. One-shot design of functional protein binders with bindcraft. *Nature*, 2025. doi: 10.1038/s41586-025-09429-6. 1, 2, 7, 14

David E Paschon, Zankhana S Patel, and Marc Ostermeier. Enhanced catalytic efficiency of aminoglycoside phosphotransferase (3')-IIa achieved through protein fragmentation and reassembly. *J. Mol. Biol.*, 353(1): 26–37, 2005. 11, 37

Vignav Ramesh and Morteza Mardani. Test-time scaling of diffusion models via noise trajectory search. *arXiv preprint arXiv:2506.03164*, 2025. 3

Buel D. Rodgers and Christopher W. Ward. Myostatin/activin receptor ligands in muscle and the development status of attenuating drugs. *Endocrine Reviews*, 43(2):329–365, 2022. 9

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

Lauren H Rosenberg, Marie Lafitte, Victor Quereda, William Grant, Wei Chen, Marjeta Bibian, et al. Therapeutic targeting of casein kinase 1δ in breast cancer. *Sci. Transl. Med.*, 7(318):318ra202, 2015. 13

Isaac Sappington, Martin Toul, David S Lee, Stephanie A Robinson, Inna Goreshnik, Clara McCurdy, Tung Ching Chan, Nic Buchholz, Buwei Huang, Dionne Vafeados, et al. Improved protein binder design using β-pairing targeted rfdiffusion. *Nature Communications*, 2026. 8

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016. 2

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *Forty-second International Conference on Machine Learning*, 2025. 3

Charlie Victor Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling parameters for reasoning. In *The Thirteenth International Conference on Learning Representations*, 2025. 2

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, et al. Boltzgen: Toward universal binder design. *bioRxiv*, 2025. doi: 10.1101/2025.11.20.689494. 7, 14

Kiera H Sumida, Reyes Núñez-Franco, Indrek Kalvet, Samuel J Pellock, Basile IM Wicky, Lukas F Milles, Justas Dauparas, Jue Wang, Yakov Kipnis, Noel Jameson, et al. Improving protein expression, stability, and function with proteinmpnn. *Journal of the American Chemical Society*, 146(3):2054–2061, 2024. 14

Erik Swanson, Michael Nichols, Supriya Ravichandran, and Pierce Ogden. mber: Controllable de novo antibody design with million-scale experimental screening. *bioRxiv*, 2025. 24, 28

Eric J Toone. Structure and energetics of protein–carbohydrate complexes. *Current Opinion in Structural Biology*, 4(5):719–728, 1994. 14

Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, and Sameer Velankar. Alphafold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50:D439–D444, 2021. 3, 23

WM Watkins. The abo blood group system: historical background. *Transfusion medicine*, 11(4), 2001. 14

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with rfdiffusion. *Nature*, 620:1089–1100, 2023. 1, 3, 7, 14

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed H. Chi, Quoc V Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, 2022. 2

William I Weis and Kurt Drickamer. Structural basis of lectin-carbohydrate recognition. *Annual Review of Biochemistry*, 65:441–473, 1996. 14

Pengfei Xu, Chiara Ianes, Fabian Gärtner, Congxing Liu, Timo Burster, Vasiliy Bakulev, Najma Rachidi, and Uwe Knippschild. Structure, regulation, and (patho-)physiological functions of the stress-induced protein kinase CK1 delta (CSNK1D). *Gene*, 715:144005, 2019. 13

Jaesik Yoon, Hyeonseo Cho, Doojin Baek, Yoshua Bengio, and Sungjin Ahn. Monte carlo tree diffusion for system 2 planning. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=XrCbBdycDc. 3

Vinicius Zambaldi, David La, Alexander E. Chu, Harshnira Patani, Amy E. Danson, Tristan O. C. Kwan, Thomas Frerix, Rosalia G. Schneider, David Saxton, Ashok Thillaisundaram, Zachary Wu, Isabel Moraes, Oskar Lange, Eliseo Papa, Gabriella Stanton, Victor Martin, Sukhdeep Singh, Lai H. Wong, Russ Bates, Simon A. Kohl, Josh Abramson, Andrew W. Senior, Yilmaz Alguel, Mary Y. Wu, Irene M. Aspalter, Katie Bentley, David L. V. Bauer, Peter Cherepanov, Demis Hassabis, Pushmeet Kohli, Rob Fergus, and Jue Wang. De novo design of high-affinity protein binders with alphaproteo. *arXiv preprint arXiv:2409.08022*, 2024. 24

Kun Zhang, Yong Wang, Tao Fan, Changqi Zeng, and Zhong Sheng Sun. The p21-activated kinases in neural cytoskeletal remodeling and related neurological disorders. *Protein Cell*, 13(1):6–25, 2022. 10

Appendix

A. Model Architecture and Training Details

This section provides additional technical details on the Proteina-Complexa framework. For full architectural specifications and training recipes, we refer the reader to Proteína [Geffner et al., 2025] for the backbone generation architecture, La-Proteína [Geffner et al., 2026] for the partially latent all-atom extension, and Proteina-Complexa [Didi et al., 2026] for binder–target complex generation and inference-time search.

A.1. Partially latent protein representation

A central challenge in all-atom protein generation is that the number of atoms per residue varies with amino acid identity (e.g., glycine has no side chain while tryptophan has 14 heavy atoms), making it difficult to define a fixed-dimensional generative target. La-Proteína [Geffner et al., 2026] addresses this with a *partially latent* representation that splits each residue into two components:

1. **Explicit backbone coordinates.** The C_α position of each residue is modeled directly in 3D space, preserving the geometric structure needed for interface design.
2. **Latent residue variables.** A variational autoencoder (VAE) compresses the amino acid identity and all remaining atomic coordinates (backbone atoms beyond C_α, plus side-chain atoms) into a continuous latent vector $\mathbf{z}_i \in \mathbb{R}^d$ of fixed dimensionality d for every residue i, regardless of amino acid type.

The VAE is trained to reconstruct full atomic detail from $\mathbf{z}_i$ given the local backbone context. Once trained, the encoder maps any residue to a point in a smooth, continuous latent space where nearby points correspond to chemically similar residues and side-chain conformations. This enables the generative model to reason over sequence and structure jointly without handling variable-length atomic representations.

A.2. Flow-based generative model

Given the partially latent representation, La-Proteína models the joint distribution over backbone coordinates $\mathbf{x} = \{\mathbf{x}_i\}$ and latent variables $\mathbf{z} = \{\mathbf{z}_i\}$ using flow matching [Lipman et al., 2023]. The generative process defines a continuous trajectory from a simple prior distribution (Gaussian noise) to the data distribution:

$$p_t(\mathbf{x}, \mathbf{z}) = \text{interpolation between } p_0(\text{noise}) \text{ and } p_1(\text{data}), \quad t \in [0, 1]. \tag{1}$$

A neural network $v_\theta(\mathbf{x}, \mathbf{z}, t)$ is trained to predict the velocity field that transports samples along this trajectory. At generation time, a sample is drawn from the noise prior and evolved to $t = 1$ by numerically integrating the learned velocity field, yielding a complete protein with backbone coordinates and latent vectors that are decoded into full atomic detail and amino acid sequence.

The architecture underlying the velocity network is a scalable transformer, originally developed in Proteína [Geffner et al., 2025] for backbone generation and extended in La-Proteína to jointly process backbone and latent channels.

A.3. Extension to binder–target complexes

Proteina-Complexa [Didi et al., 2026] extends the monomeric generative model to binder design by conditioning on a fixed target structure. The target's backbone coordinates and (optionally) a set of interface hotspot residues are provided as context; the model generates the binder's backbone and latent variables while keeping the target fixed. Training proceeds in three stages to overcome the scarcity of experimental complex data:

1. **Monomeric pretraining.** The model is first trained on single-chain proteins from the PDB and AlphaFold Database, learning general protein geometry and sequence–structure relationships.
2. **Synthetic complex training.** The model is then trained on Teddymer, a large-scale dataset of synthetic binder–target pairs constructed by extracting domain–domain interactions from computationally predicted monomeric structures in the AlphaFold Database [Varadi et al., 2021]. This exposes the model to diverse protein–protein interfaces.
3. **Experimental complex fine-tuning.** Finally, the model is fine-tuned on experimentally determined multimeric structures from the PDB, aligning the learned representations with real binding interfaces.

A.4. Reward-guided inference-time search

A key advantage of the continuous latent space is that it enables *reward-guided search* at inference time. Rather than generating a single protein and accepting or rejecting it, Proteina-Complexa maintains a population of candidate trajectories during the generative process and steers them toward high-quality solutions using

external reward signals. This is analogous to how tree search enhances game-playing agents or how best-of-N sampling and chain-of-thought improve language models—but applied in a continuous generative space rather than a discrete one.

Multiple search strategies are supported, but in most experiments we use beam search due to its efficiency and simplicity [Fernandes et al., 2025]. At each integration step, the current population of partially generated proteins is expanded by spawning multiple continuations, scored by a reward function, and pruned to retain the top candidates. This focuses computational resources on the most promising regions of the generative landscape. The reward signal is computed from structure-prediction confidence metrics (e.g., ipTM, pLDDT, or ipSAE scores from models such as AlphaFold2 or OpenFold3) evaluated on partially or fully generated complexes. Because rewards are applied during generation rather than as a post hoc filter, the search can redirect trajectories away from poor solutions early, substantially improving the yield of high-quality binders relative to unguided generation.

A.5. Codesign-based scaffold re-engineering

Beyond fully de novo generation, Proteina-Complexa supports partial redesign of existing proteins through a diffuse-denoise protocol. Given an input binder structure, the model applies controlled noise to both backbone coordinates and latent variables, then jointly re-generates both—performing true structure–sequence codesign rather than fixed-backbone sequence optimization. The noise level controls the degree of deviation from the input scaffold: low noise preserves most of the original structure while optimizing local interactions, while higher noise enables more substantial backbone rearrangements. This capability is applied to the Nipah virus target in Sec. 8, where an existing binder scaffold is re-engineered to improve affinity while maintaining the target engagement site.

B. Massive-Scale Benchmark: Extended Analysis

B.1. Designing 1.4 Million Sequences

Following Swanson et al. [2025], Manifold Bio selected 138 design targets to define a diverse and biologically relevant benchmark panel for evaluating Proteina-Complexa alongside recent generative and hallucination-based binder design methods. For targets included in prior in vitro studies (e.g., Zambaldi et al. [2024]), we used hotspot definitions reported in the literature. For the remaining targets, where rigorously validated hotspots were unavailable for more than 93% of cases, Manifold Bio proposed approximately 10 single-residue hotspots per target and paired each hotspot with distance-based target crops, yielding more than 1,200 formal design tasks.

For each target, we set out to screen 10,000 sequences. For targets with only one known hotspot, this budget was doubled to maintain comparable search breadth. Half of the total sequence budget was allocated to a diverse hotspot sweep. The remaining half was split among baseline comparisons with matched compute budgets, Proteina-Complexa sequence-redesign ablations, and Proteina-Complexa sampling-temperature ablations.

For each task, we ran one instance of Proteina-Complexa with standard beam search and AlphaFold2 interface predicted aligned error (ipAE) as the reward, generating approximately 27,000 candidate binders. We filtered this pool to the top 1,000 candidates ranked by pAE. For each hotspot task, we recorded candidates satisfying all of the following criteria: minimum interface pAE < 2, complex pLDDT > 0.8, ipTM > 0.7, and fewer than 5 backbone Cα clashes with the uncropped target (with a clash defined as any atom pair within 1 Å).

For each target, the hotspot with the largest number of candidates passing these criteria was designated the *top hotspot*. This top hotspot was used for all baseline experiments shown in Fig. 2 and Fig. 11. For targets with only one hotspot, we repeated the search twice and retained the best 1,000 sequences from each run. We did not evaluate samples from hotspots that yielded no in silico success.

To screen a structurally diverse binder set for each target, we pooled all generated binders from hotspots with at least one *in silico* successful design and clustered them at the target level by predicted binder structure. We then filled the evaluation budget using a staged selection procedure that cycled across hotspots and

sampled without replacement from these target-level structure clusters. Specifically, we iterated through four ranking regimes in sequence: minimum ipSAE, ipTM, minimum ipAE, and random sampling. Within each regime, the top cluster candidate was selected from previously unselected clusters whenever possible, thereby promoting diversity across both hotspot conditions and predicted binding modes while still prioritizing high-confidence AlphaFold2 predictions. The resulting screen, including both on-design and off-design hits, is shown in Fig. 3.

For the top-hotspot comparison, we ranked the 1,000 retained candidates by minimum ipSAE and selected the top 400 without clustering. The same procedure was applied to sampling-temperature and sequence-redesign ablations using the top hotspot condition. Minimum ipSAE [Dunbrack Jr, 2025] was chosen as the primary selection metric to align with recent in vitro practice and to provide a consistent criterion across methods. For sequence-redesign ablations, we followed Didi et al. [2026], applying soluble ProteinMPNN for both full-sequence redesign and fixed-interface redesign. For temperature ablations, we started from the default backbone/local latent temperatures of $0.1, 0.1$ and evaluated $0.1, 0.4$, $0.4, 0.1$, and $0.4, 0.4$.

Overall, the campaign required more than 140,000 GPU hours and produced a curated set of approximately 1.4 million binders. Because Proteina-Complexa was run across more than 1,200 simultaneous tasks, we standardized beam search hyperparameters to generate roughly 27,000 binders per task. Under this setup, 80% of tasks finished within 20–32 GPU hours, with higher costs for longer target crops. To match compute across methods, each baseline received approximately 32 GPU hours per top-hotspot target pair. The baseline allocation within the 1.4 million-binder budget was distributed evenly across methods. Because downstream experimental processing introduced modest differences in final tested counts, we report tested sequence counts alongside hit counts and hit rates. For the initial designed set on the 75 solvable baseline targets (Fig. 2), Proteina-Complexa used 26 GPU hours on average, with 80% of targets under 32 GPU hours and a maximum runtime of approximately 40 GPU hours. We further restricted analysis to the shared set of 127 design and screen targets to focus on on-design specificity. Additional baseline implementation details are provided in Appendix H.

B.2. Experimental Methods: Manifold Bio Large-Scale All-by-All Binder Experiment

B.2.0.1 Protein sourcing. Recombinant proteins were obtained from commercial vendors (Tab. 1) and accompanied by vendor quality-control documentation (e.g., SDS-PAGE purity and activity/binding checks). Proteins were reconstituted and stored according to manufacturer specifications before use.

Table 1 | List of commercially sourced proteins used in the Manifold Bio large-scale all-by-all binder screen.

Target_UniProtID	Target	Catalog No.	Vendor
ACE2_Q9BYF1	ACE2	10108-H08H	Sino Biological
ADAM10_O14672	ADAM10	936-AD	R&D Systems
ALCAM_Q13740	ALCAM	10045-H08H	Sino Biological
ANTXR1_Q9H6X2	ANTXR1	13367-H08H	Sino Biological
AOC3_Q16853	AOC3	3957-AO	R&D Systems
BACE2_Q9Y5Z0	BACE2	10783-H08H	Sino Biological
BCAM_P50895	BCAM	10238-H08H	Sino Biological
Bet v1_Q42499	Bet v1	NA-BV1-1	In Bio
BGN_P21810	BGN	10447-H08H	Sino Biological
CA12_O43570	CA12	10617-H08H	Sino Biological
CCR5_P51681	CCR5	13022-H91H-NA	Sino Biological
CD200_P41217	CD200	10886-H08H	Sino Biological
CD46_P15529	CD46	12239-H08H	Sino Biological
CD47_Q08722	CD47	12283-H08H	Sino Biological
CD55_P08174	CD55	10101-H08H	Sino Biological
CD59_P13987	CD59	12474-H08H	Sino Biological
CD81_P60033	CD81	14244-H07H	Sino Biological
CEACAM1_P13688	CEACAM1	10822-H08H	Sino Biological
CEACAM5_P06731	CEACAM5	11077-H08H	Sino Biological
CEACAM6_P40199	CEACAM6	10823-H08H	Sino Biological

Continued on next page

Table 1 – *continued from previous page*

Target_UniProtID	**Target**	**Catalog No.**	**Vendor**
CEMIP2_Q9UHN6	CEMIP2	CSB-EP023791HU	Cusabio
CLMP_Q9H6B4	CLMP	5019-AM	R&D Systems
CLU_P10909	CLU	11297-H08H1	Sino Biological
CNTFR_P26992	CNTFR	11012-H08B	Sino Biological
CR1_P17927	CR1	5748-CD	R&D Systems
CSPG4_Q6UVK1	CSPG4	2585-PG	R&D Systems
CXADR_P78310	CXADR	10799-H08H	Sino Biological
CXCR3_P49682	CXCR3	12979-H95H-NA	Sino Biological
CXCR4_P61073	CXCR4	11325-H95H-NA	Sino Biological
DCN_P07585	DCN	10189-H08H	Sino Biological
DPP4_P27487	DPP4	10688-H08H	Sino Biological
EFNB1_P98172	EFNB1	10894-H08H	Sino Biological
ENPEP_Q07075	ENPEP	10554-H07B	Sino Biological
ENTPD1_P49961	ENTPD1	16020-H08B	Sino Biological
FcRn_P55899	FcRn	CT009-H08H	Sino Biological
FLRT2_O43155	FLRT2	11296-H08H	Sino Biological
FMOD_Q06828	FMOD	11514-H02H	Sino Biological
FN1_P02751	FN1	10314-H08H	Sino Biological
HER2_P04626	HER2	10004-H08H	Sino Biological
HLA-B_P01889	HLA-B	ab225646-100ug	Abcam
HLA-E_P13747	HLA-E	HL0-H82E7-25ug	Acro
HYAL2_Q12891	HYAL2	11012-GH	R&D Systems
ICAM1_P05362	ICAM1	10346-H08H	Sino Biological
IFNAR2_P48551	IFNAR2	10359-H08H	Sino Biological
IGSF8_Q969P0	IGSF8	13435-H08H	Sino Biological
IL-13_P35225	IL-13	10369-H08H	Sino Biological
IL4R_P24394	IL4R	10402-H08H	Sino Biological
ITGA6_P23229	ITGA6	CT069-H2508H	Sino Biological
JAM2_P57087	JAM2	1074-VJ	R&D Systems
KIRREL1_Q96J84	KIRREL1	15752-H08H1	Sino Biological
LTBP4_Q8N2S1	LTBP4	CSB-EP850788HU	Cusabio
MCAM_P43121	MCAM	10115-H08H	Sino Biological
MRC2_Q9UBG0	MRC2	10483-EN	R&D Systems
NECTIN2_Q92692	NECTIN2	10005-H08H	Sino Biological
NEO1_Q92859	NEO1	8607-NE-050	R&D Systems
NRG3_P56975	NRG3	16071-H08H	Sino Biological
NTM_Q9P121	NTM	15169-H08H	Sino Biological
NTRK2_Q16620	NTRK2	10047-H08H	Sino Biological
PD-1_Q15116	PD-1	10377-H08H	Sino Biological
PD-L1_Q9NZQ7	PD-L1	10084-H08H	Sino Biological
PDGFRA_P16234	PDGFRA	10556-H08H	Sino Biological
POSTN_Q15063	POSTN	10299-H08H	Sino Biological
PSG1_P11464	PSG1	15378-H08H	Sino Biological
RECK_O95980	RECK	10309-RE	R&D Systems
SEMA4D_Q92854	SEMA4D	11825-H08H	Sino Biological
SEMA7A_O75326	SEMA7A	2068-S7	R&D Systems
SERPINA3_P01011	SERPINA3	10307-H08H	Sino Biological
SIRPA_P78324	SIRPA	30014-H08H	Sino Biological
SLIT2_O94813	SLIT2	11967-H08H	Sino Biological
SLIT3_O75094	SLIT3	9067-SL	R&D Systems
THBD_P07204	THBD	11864-H08H	Sino Biological
THSD1_Q9NS62	THSD1	IT-010-017p	Immune Technology
TSPAN13_O95857	TSPAN13	11551-TS	R&D Systems
TSPAN14_Q8NG11	TSPAN14	11552-TS	R&D Systems
TSPAN31_Q12999	TSPAN31	11553-TS	R&D Systems
TSPAN8_P19075	TSPAN8	15683-H07H	Sino Biological
VSTM4_Q8IW00	VSTM4	2086-VT	R&D Systems

Continued on next page

Table 1 – *continued from previous page*

Target_UniProtID	**Target**	**Catalog No.**	**Vendor**
Additional targets (short-form identifiers)			
AHSP	AHSP	14391-HNAE	Sino Biological
AMBP	AMBP	13141-H05H1	Sino Biological
CALR	CALR	13539-H02H	Sino Biological
CD226	CD226	10565-H02H	Sino Biological
CEAM6	CEAM6	10823-H08H-B	Sino Biological
CRTAM	CRTAM	11975-H08H-B	Sino Biological
CSF1	CSF1	11792-H02H	Sino Biological
DAF	DAF	10101-H02H	Sino Biological
EFNA1	EFNA1	10882-H02H	Sino Biological
EFNA5	EFNA5	10192-H02H	Sino Biological
EFNB2	EFNB2	10881-HCCH	Sino Biological
EPCR	EPCR	13320-H02H	Sino Biological
FCG3B	FCG3B	11046-H27H-B	Sino Biological
FGFR1	FGFR1	10616-H02H	Sino Biological
FST	FST	10685-H02H	Sino Biological
HDAC8	HDAC8	10864-H09B	Sino Biological
IL1R1	IL1R1	10126-H02H	Sino Biological
IL20	IL20	13060-HNAE	Sino Biological
IL3	IL3	11858-HNAE	Sino Biological
LEP	LEP	10221-HNAE	Sino Biological
MK08	MK08	10795-H09B	Sino Biological
MMP2	MMP2	10082-HNAH	Sino Biological
NECT1	NECT1	11611-H02H	Sino Biological
PARVA	PARVA	13919-H09E	Sino Biological
PD1L2	PD1L2	10292-H02H	Sino Biological
PIN1	PIN1	10282-HNCE	Sino Biological
PTN2	PTN2	10570-HNCB	Sino Biological
RNF43	RNF43	16108-H02H	Sino Biological
RSPO1	RSPO1	11083-HNAS	Sino Biological
S10A4	S10A4	10185-H01H	Sino Biological
SAE1	SAE1	13921-HNCB	Sino Biological
SORCN	SORCN	14547-HNCB	Sino Biological
STC2	STC2	13653-H02H	Sino Biological
SYUA	SYUA	12093-HNAE	Sino Biological
TACT	TACT	11202-H08H-B	Sino Biological
TF65	TF65	12054-H09E	Sino Biological
TNFL4	TNFL4	13127-H04H	Sino Biological
TNFL9	TNFL9	15693-H01H	Sino Biological
UBC9	UBC9	13205-HNCE	Sino Biological
UBE2B	UBE2B	51243-MNCE	Sino Biological
UBP7	UBP7	11681-HNCB	Sino Biological
VATF	VATF	15605-H09E	Sino Biological
XIAP	XIAP	10606-H17E	Sino Biological
Benchmark targets			
BHRF1_2wh6	BHRF1	LS-G64115-20	LS Bio
SC2RBD_6m0j	SC2RBD	40592-V08H-100	Sino Biological
IL-7RA_3di3	IL-7RA	10975-H02H-100	Sino Biological
PD-L1_5o45	PD-L1	10084-H02H-100	Sino Biological
TrkA_1www	TrkA	11073-H08H-100	Sino Biological
Insulin_4zxb	Insulin	11038-HNAY-100	Sino Biological
H1N1_5vli	H1N1	11684-V08H-100	Sino Biological
VEGF-A_1bj1	VEGF_A	11066-HNAH-20	Sino Biological
IL-17A_4hsa	IL-17A	12047-HNAS-100	Sino Biological
TNFa_1tnf	TNFa	10602-H08H-50	Sino Biological

B.2.0.2 Library cloning, phage production, and display. All binder polypeptide sequences were back-translated using Manifold's internal codon optimization tool and ordered as oligopools (Twist Biosciences). Libraries were cloned into a pADL-100 derivative phagemid (Antibody Design Laboratories) using Gibson Assembly. Insert/vector mixtures were assembled under standard conditions, electroporated into bacterial cells, and assessed for library coverage. Libraries with adequate coverage were prepared and stored at $-80\,^{\circ}\mathrm{C}$. Phage display protocols followed Swanson et al. [2025]. Sequencing libraries were indexed and run on an Illumina NovaSeq X (Broad Clinical Labs, MA). Paired-end reads (2×150 bp) were merged with BBMerge [Bushnell et al., 2017]. Mean sequence quality was Q40 (error rate $\approx 10^{-4}$ per base). Only exact matches to designed binder variants were counted.

B.2.0.3 Computational binder selection and hit calling. Candidate hits were classified by the number of distinct screen targets called positive, relative to the intended design target. Single-target hits were labeled *on-design specific* (match to design target) or *off-design specific* (single non-design target). Hits spanning two to four targets were labeled *on-design cross-reactive* (design target plus one to three additional targets) or *off-design cross-reactive* (two to four non-design targets). These labels were assigned independently for each hit-calling method. Final hits were then defined as the union across methods, followed by reclassification on the merged target set.

First, we applied the quantile-ratio method from Swanson et al. [2025]. For binder i and target group g, we computed the minimum replicate signal

$$v_{i,g} = \min_r x_{i,g,r}. \tag{2}$$

Let $Q_{i,g}^{(q+1)}$ be the $(q+1)$-th largest value across all N samples in the screen. A hit was called when

$$\frac{v_{i,g}}{Q_{i,g}^{(q+1)}} > 1. \tag{3}$$

We used $q = 20$, corresponding to the 21st-largest sample among approximately $N \approx 640$ measurements ($\sim 3.3\%$ upper tail). This setting balances outlier sensitivity and conservativeness in large multiplexed screens (129 targets, ~ 5 replicates each). Specificity was assessed from the number of distinct targets exceeding threshold.

Second, we applied a complementary enrichment-based method to account for library abundance effects and recover low-input true binders. For binder i and target group g, enrichment was

$$R_{i,g} = \frac{c_{o,i,g}}{c_{i,i}}, \tag{4}$$

where c_o is post-selection output frequency and c_i is pre-selection input frequency. A robust null was estimated per sub-library:

$$R_{\text{null}} = \text{median}(R), \qquad \sigma_{\text{null}} = 1.4826 \times \text{MAD}(R), \tag{5}$$

with stratification by input concentration to reduce low-abundance false positives. We then used a one-sided z-test:

$$z_{i,g} = \frac{R_{i,g} - R_{\text{null}}}{\sqrt{e_{R_{i,g}}^2 + \sigma_{\text{null}}^2}}. \tag{6}$$

Within each target group, p-values were adjusted by Benjamini–Hochberg FDR control to produce q-values. Hits were called at $q < 0.75$ and $R > 1$. We used this relaxed q-threshold because downstream specificity classification across 129 simultaneous targets provides an orthogonal filter against false positives in noisy phage-display readouts.

B.3. Per-target hotspot analysis

To characterize hotspot-conditioning effects on design success, we analyzed hit distributions across hotspot combinations for targets with multiple tested hotspots (Fig. 11). For each target, cells report the fraction of designs yielding on-design specific hits for a given hotspot, with in silico hotspot rank annotated. Targets are

Table 2 | Temperature ablation for Proteina-Complexa (self) on baseline solvable targets ($n = 75$). On-design hit counts, specificity (specific/all on-design), and target coverage across backbone and local latent temperature combinations. New and lost columns indicate targets gained or lost relative to the default (0.1, 0.1) setting; unique denotes targets solved exclusively by that condition and no other.

Temperature	**On-Design**			**Targets**				**# Seqs.**
(backbone, local latent)	**All**	**Specific**	**Specificity**	**Solved**	**New**	**Lost**	**Unique**	
(0.1, 0.1)	**691**	**630**	**91.2%**	33	–	–	2	25,707
(0.1, 0.4)	522	476	91.2%	36	+11	−8	4	23,075
(0.4, 0.1)	502	448	89.2%	34	+9	−8	2	23,546
(0.4, 0.4)	300	268	89.3%	29	+8	−12	4	26,791

ordered by hotspot entropy, which quantifies how evenly success is distributed across hotspot choices. Targets at the top are robust to hotspot choice, whereas targets at the bottom are hotspot-sensitive. Notably, the top-ranked computational hotspot is not always experimentally optimal, supporting multi-hotspot evaluation per target.

B.4. Model temperature ablation

To assess sensitivity to sampling temperature, we ran an ablation across four backbone/local latent temperature pairs on the 75 solvable baseline targets, designing against each target's top-ranked in silico hotspot. The default setting $(0.1, 0.1)$ yielded the highest on-design hit count (691) and specificity (91.2%). Increasing local latent temperature to 0.4 while keeping backbone temperature at 0.1 solved the most targets (36), despite fewer total hits (522), indicating a diversity–coverage tradeoff. Each temperature setting uniquely solved 2–4 targets not solved by the others. Specificity remained stable (89.2–91.2%), suggesting elevated temperatures did not materially reduce selectivity. These results support a mixed-temperature strategy to expand solvable-target coverage while preserving hit quality.

B.5. Compositional dropout from the phage input pool

Approximately 1.4 million minibinder sequences were designed and ordered across four length-stratified pools. Deep sequencing of the phage input pool before selection recovered 73.2% of designs at one or more reads. Two sequence features strongly predicted dropout: sequence length and fractional lysine+glutamate (K+E) content.

Across amino acids, lysine and glutamate showed the strongest negative Spearman correlation with detection (K: $\rho = -0.268$, E: $\rho = -0.217$; Fig. 12B). Joint binning by length and K+E fraction (Fig. 12A) showed detection rates from 97.9% at 50 aa with K+E < 0.08 to 0.19% at 88 aa with K+E > 0.50.

We tested three possible explanations. (i) Net charge alone does not explain the trend: arginine (+1, $\rho = +0.090$) and aspartate (−1, $\rho = +0.063$) were positively correlated with detection. (ii) Helix propensity alone does not explain the trend: alanine, methionine, and leucine have comparable/high P_α values without similarly negative correlations ($\rho = +0.071, -0.038, -0.050$). (iii) Adenine-rich codons alone do not explain the trend: asparagine and glutamine share related codon motifs yet show no corresponding negative behavior ($\rho = +0.011, +0.145$). No single property recapitulated the K+E dropout pattern.

In a PaCMAP embedding of all 1,405,522 designs in 20-dimensional amino-acid composition space (Fig. 12C,D), undetected designs concentrated in the high-K+E region. Among the 1,028,458 detected designs, input abundance spanned 0.011× to 12.3× the median (2nd–98th percentile; 1092-fold), with lower abundance in the same high-K+E region.

An enrichment of K+E among non-recovered de novo designs was independently observed in a CAR-T proliferation assay [Kosonocky et al., 2026], where combined helical-region K+E content was the strongest predictor of non-recovery (ROC-AUC = 0.91). This suggests a shared compositional liability across both bacterial phage display and mammalian surface-expression contexts.

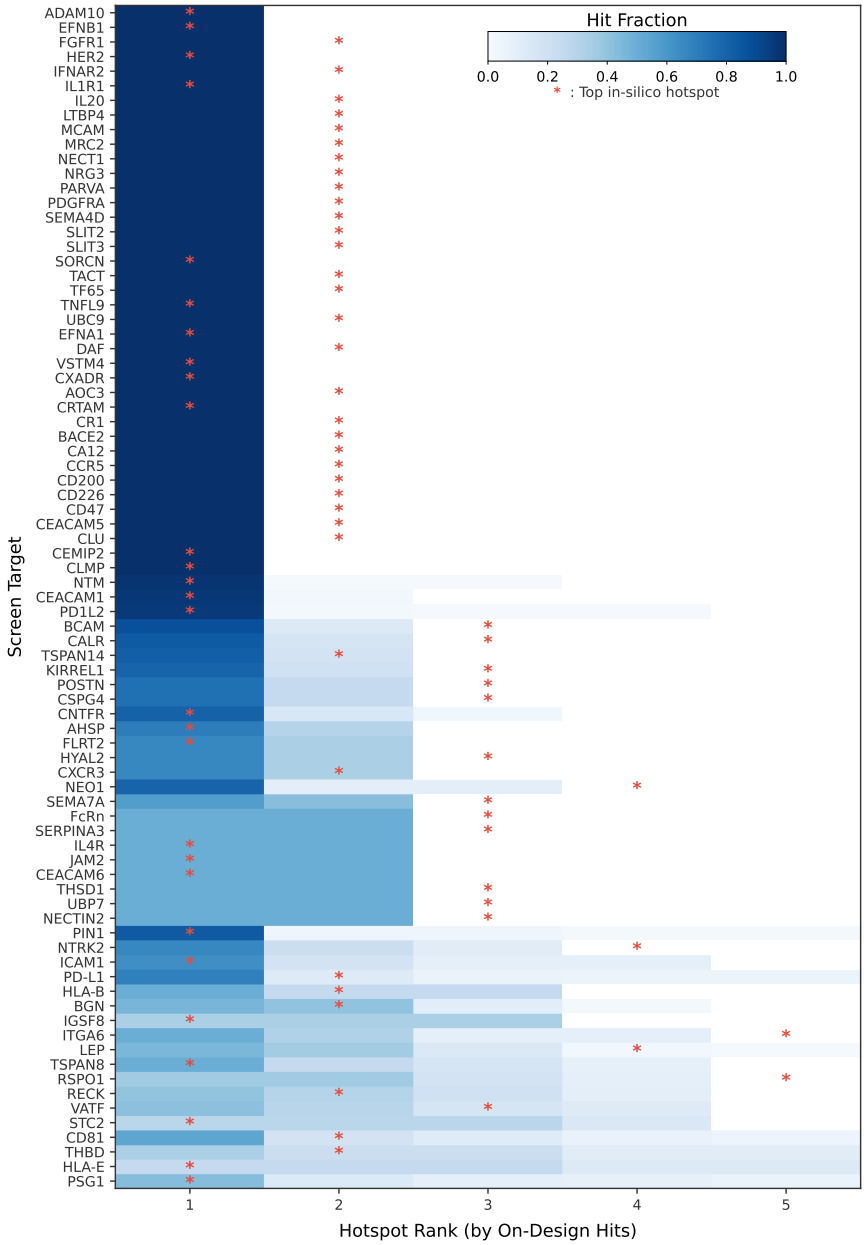


Figure 11 | Per-target hotspot hit concentration analysis. Heatmap of on-design hit fraction across the top five hotspots for targets with at least one on-design hit. Targets are ordered by hotspot entropy. Red asterisks denote the top-ranked in silico hotspot per target. Hit fractions are normalized over the top 10 hotspots per target.

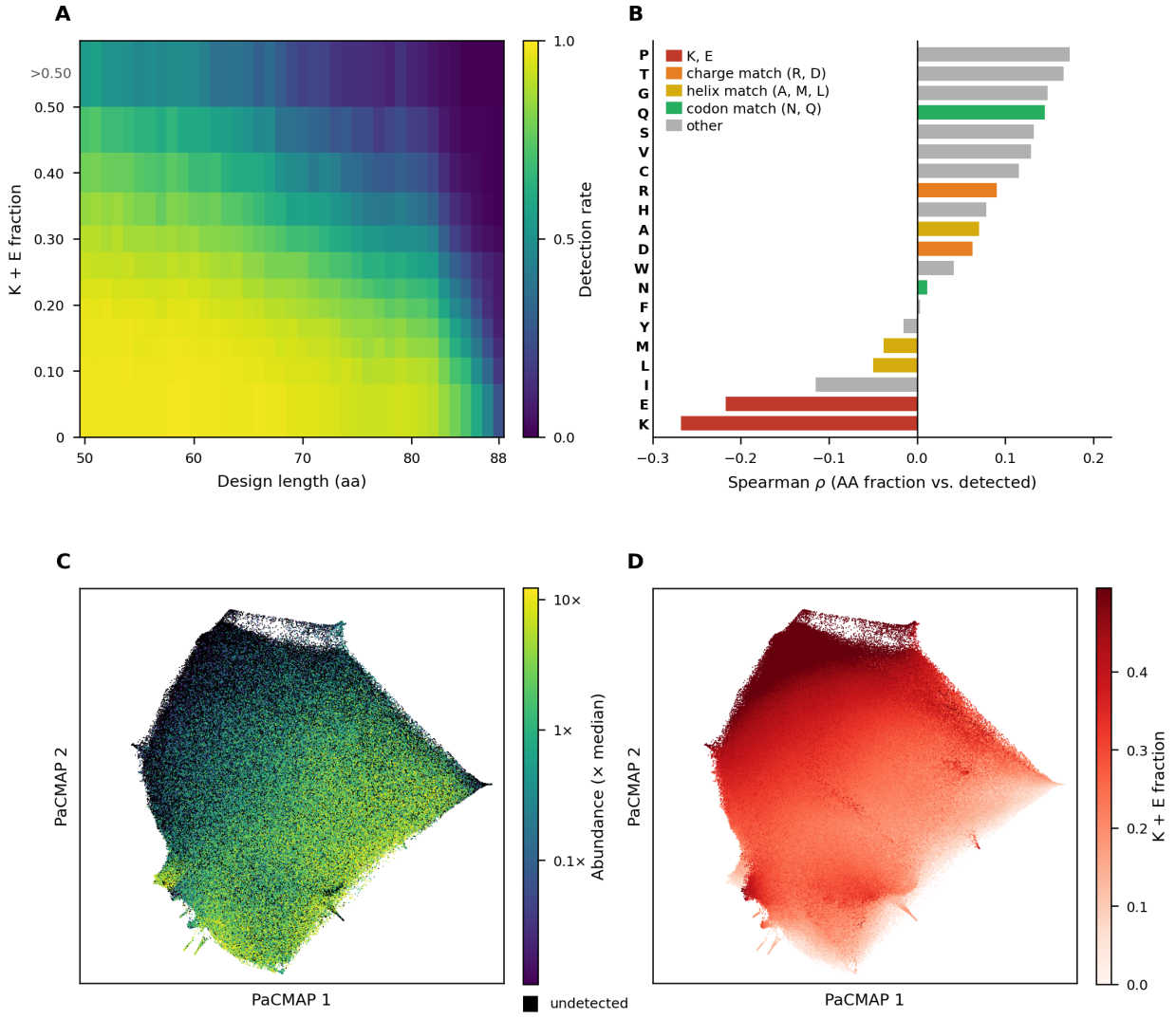


Figure 12 | Lysine and glutamate content correlates with loss from the phage input pool. (A) Detection rate as a function of design length and combined K+E fraction across 1,405,522 de novo designs in four length-stratified libraries. (B) Spearman correlation between per-residue fractional composition and detection. (C) PaCMAP embedding of all designs, highlighting undetected sequences and relative abundance among detected sequences. (D) Same embedding colored by K+E fraction.

C. Biophysical Characterization of Designed Monomers: Experimental Methods

Design filtering. Structure predictions for La-Proteína co-designed and ProteinMPNN-designed sequences for the unconditional design benchmark set were generated in single-sequence mode with 3 recycles using the *alphafold2_ptm_model_1* of AF2 in its ColabFold implementation. Designs were considered successful if the predicted structures had a mean pLDDT > 90 and an RMSD < 2 Å from the design model. All successful designs were subsequently manually screened; a subset of 49 designs were selected for experimental analysis. La-Proteína co-generated sequences were preferred whenever metrics matched those of the respective ProteinMPNN sequence. Sequence identity and structural similarity were further analysed using HHblits and the Foldseek web server, respectively.

For the high-temperature generation set, designs were predicted both in single-sequence mode and with MSA. Designs were selected for further analysis if they were confidently predicted with MSA (mean pLDDT > 80 and RMSD < 2 Å from the design model), while being poorly predicted in single-sequence mode (mean pLDDT < 80 and RMSD > 4 Å). Successful designs were manually screened and filtered based on sequence identity (not higher than 40%), yielding a final set of 13 designs.

Cloning, expression, purification, and yield quantification. Genes coding for the designed proteins were codon-optimised for *Escherichia coli* K-12 and synthesised by GenScript. Gene fragments were cloned into the LM627 bacterial expression vector using BsaI-mediated Golden Gate assembly, replacing the vector's *ccdB* selection marker. The assembly products were transformed into *E. coli* NEB5α cells (New England Biolabs) through heat-shock transformation and selected on LB agar (Carl Roth) supplemented with kanamycin (VWR). Plasmids were purified from overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen) and verified through Sanger sequencing (Eurofins Genomics). The sequence-verified plasmids were subsequently transformed into *E. coli* BL21(DE3) cells (New England Biolabs) for protein expression.

Proteins were expressed by auto-induction in TBII medium (MP Biomedicals), supplemented with kanamycin (VWR), a 50×5052 autoinduction mixture, 20 mM $MgSO_4$, and a trace metal mix. The cultures were initially grown at 37 °C for 8 hours, followed by incubation at 18 °C for 24 hours with shaking. Cells were harvested by centrifugation at 5,000 *g*, resuspended in lysis buffer (50 mM Tris, pH 8.0, 250 mM NaCl, 20 mM imidazole) containing lysozyme and DNase I (Sigma-Aldrich), and lysed by sonication. Clarified lysates obtained by centrifugation were purified using immobilised metal affinity chromatography with HisPur Ni-NTA resin (Thermo Fisher Scientific). Samples were washed with wash buffer (50 mM Tris, pH 8.0, 250 mM NaCl, 20 mM imidazole) and eluted using the same buffer containing 500 mM imidazole. Eluted proteins underwent further purification by size-exclusion chromatography (SEC) on a Superdex 75 Increase 10/300 GL column (Cytiva), using an ÄKTA Pure system (Cytiva) equilibrated in 25 mM Tris, pH 8.0, and 150 mM NaCl.

Protein yields were estimated from the SEC chromatograms recorded at 280 nm by integrating the absorbance peak corresponding to the expected elution of the target species. Peak areas were converted to protein mass using sequence-derived extinction coefficients, corrected for the ÄKTA flow-cell path length, and scaled from the injected sample to the total Ni-NTA eluate. To compare soluble expression yields between different sequence design strategies, a subset of constructs sharing the same backbone but differing in sequence (co-generated by La-Proteína or fixed-backbone design with ProteinMPNN) was expressed and purified in three independent replicates.

Size-exclusion chromatography coupled to multi-angle light scattering (SEC–MALS). Where needed, SEC-purified proteins were concentrated prior to further analysis to ensure sufficient scattering signal. Molecular weight was determined by SEC–MALS using an ÄKTA Pure system equipped with a Superdex^{TM} 75 10/300 GL column and a Wyatt miniDAWN^{TM} detector. Extinction coefficients were calculated from the protein sequence and used by the ASTRA software to determine molecular weights.

Circular dichroism (CD). Proteins were diluted and/or buffer-exchanged into PBS (100 mM KCl, 10 mM phosphate) buffer to concentrations between 0.1–0.3 mg/mL prior to analysis and measured in a 2 mm Hellma QS cuvette. A JASCO J-1500 spectrometer equipped with a Xe lamp was used in combination with the Spectra Manager^{TM} CFR software. Spectra between 200 and 250 nm were initially obtained for each sample at 25 °C to confirm protein folding. Subsequently, temperature-ramp measurements at three distinct wavelengths (208, 222, and 230 nm) were recorded in steps of 10 °C between 25 °C and 95 °C. Once overall stability was confirmed, full spectra in the range of 200–250 nm were obtained at the start and end points

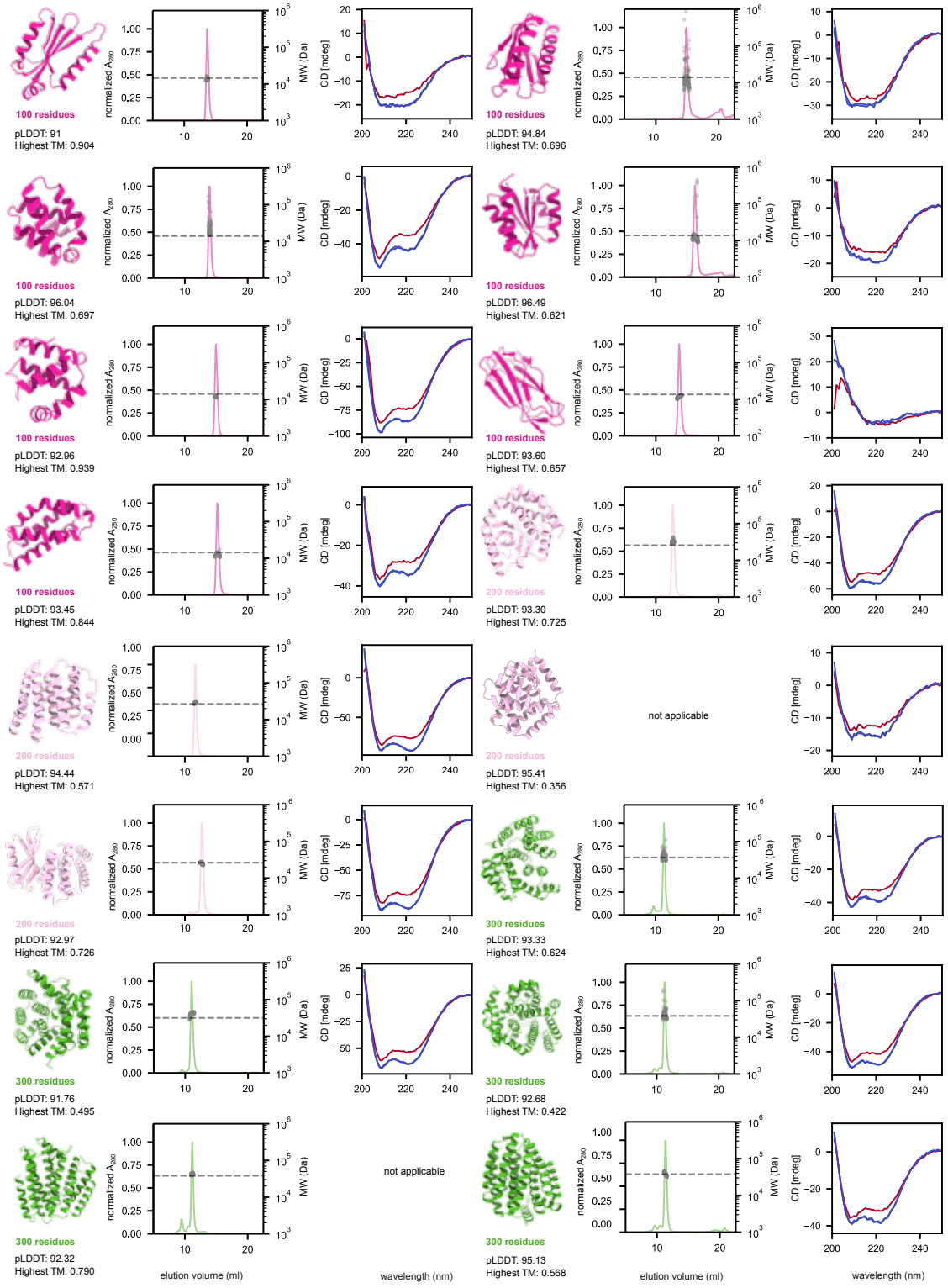


Figure 13 | Biophysical characterization of designed monomeric folds: 100–300 residue designs. For each design, the predicted structure model (left), SEC profile overlaid with MALS-determined molecular weight (middle), and CD spectra at 20 °C (blue) and 95 °C (red) (right) are shown. Protein models are colored according to chain length following the scale in Fig. 4. The average pLDDT from AF2 single-sequence predictions and the Foldseek TM-score to the closest known fold are reported for each design.

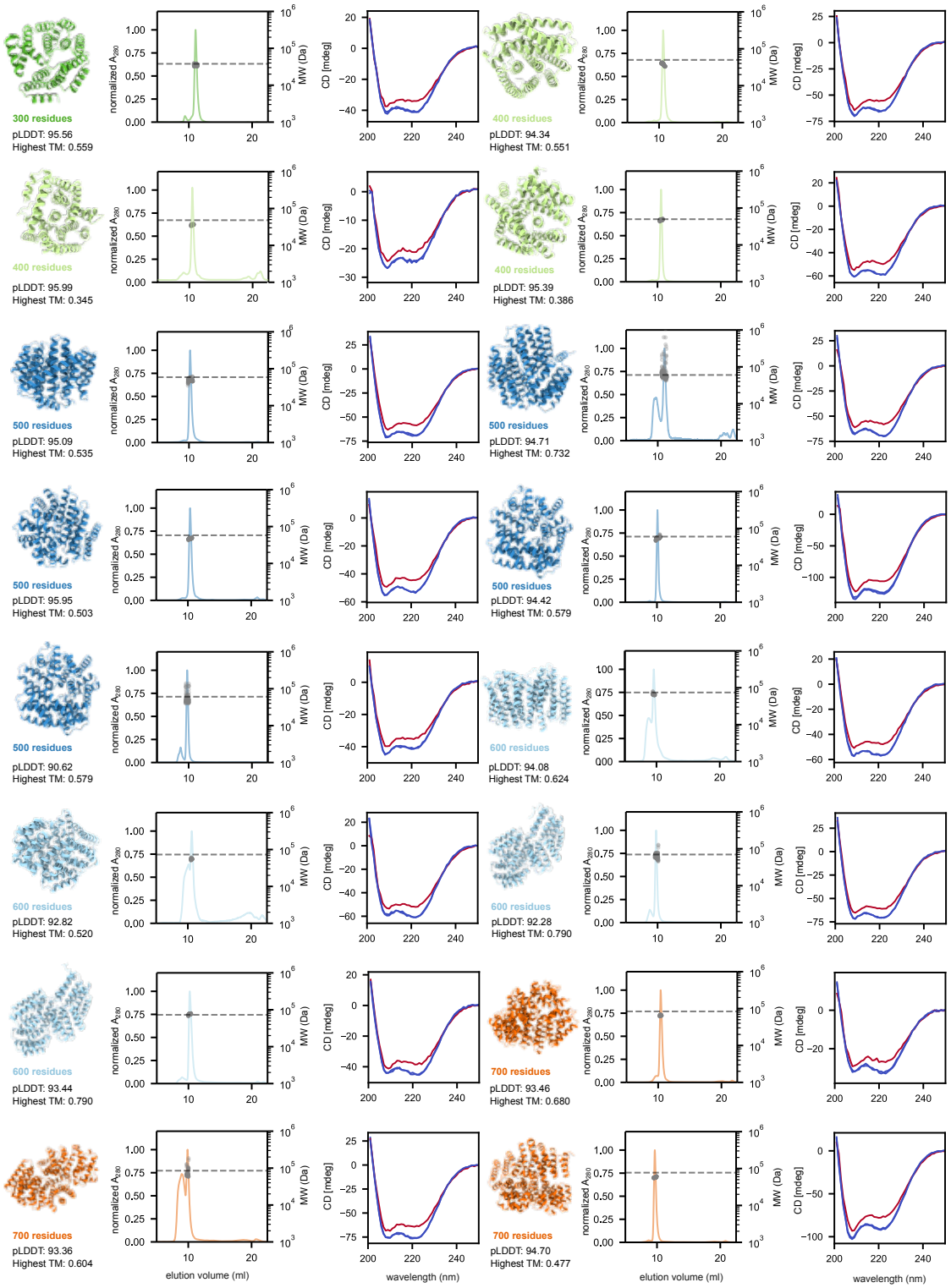


Figure 14 | Biophysical characterization of designed monomeric folds: 300–700 residue designs. Layout as in Fig. 13.

(25 °C, 95 °C, and return to 25 °C). All data were baseline-corrected with PBS.

Data processing. All data were processed and plotted using Python.

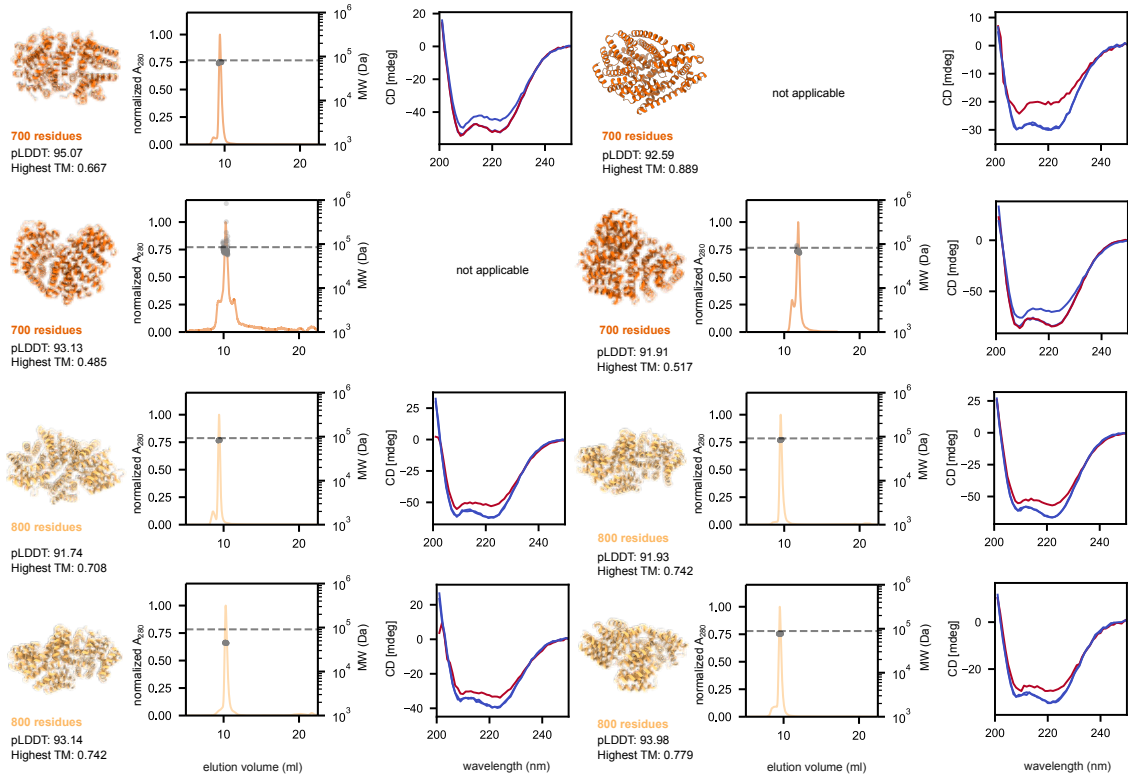


Figure 15 | Biophysical characterization of designed monomeric folds: 700–800 residue designs. Layout as in Fig. 13.

D. PDGFR and PD-L1 Binder Design: Filtering and Hotspot Analysis

D.1. Design filtering pipeline

As described in Sec. 5, a two-stage filtering pipeline was applied to Proteina-Complexa-generated candidates for PDGFR and PD-L1. In the initial screening stage, designs were required to meet the following quantitative thresholds:

Isoelectric Point ≤ 6.0, selected to ensure favorable solubility and charge profiles under physiological conditions.

ESM Pseudolikelihood ≤ 10.0, utilized as a metric for sequence fitness and evolutionary plausibility.

Minimum IPAE ≤ 1.9, ensuring high-confidence relative positioning and orientation of the binder at the target interface.

Complex iPTM ≥ 0.80, prioritizing designs with a high probability of correct interface assembly.

Complex pLDDT ≥ 0.88, ensuring high local structural confidence across the designed binder.

Cysteine Count exactly 0, to prevent the formation of off-target disulfide bonds and simplify experimental expression and purification.

To ensure broad exploration of the structural landscape, the remaining candidates were clustered at a TM-score threshold of 0.8, with a single representative from each cluster retained for final selection.

D.2. Hotspot conditioning analysis

Multiple hotspot combinations were evaluated for both targets to assess the influence of interface residue selection on design success (Tab. 3). For PDGFR, hotspot sets that included residues A77, A82, and

Table 3 | Evaluation of hotspot output and results.

Input ID	Hotspots	Percent Binders	Best -log(Kd)
NN-PDGFR-1	A139, A77, A82, A11, A116, A80, A142, A8, A134	73.5%	9.61
NN-PDGFR-2	A77, A82, A11, A80, A8	77.5%	9.75
NN-PDGFR-3	A139, A116, A142, A134	31.3%	7.42
NN-PDGFR-5	A139, A77	58.5%	9.24
NN-PDGFR-6	A139, A77, A82	71.1%	10.03
NN-PDL1-1	A39, A98, A51, A108, A49, A106, A59, A37, A32	5.5%	6.97
NN-PDL1-2	A39	5.3%	6.10
NN-PDL1-3	A39, A98	9.1%	6.33
NN-PDL1-4	A39, A98, A51	17.9%	6.75
NN-PDL1-5	A98, A108, A106	8.3%	6.03
NN-PDL1-6	A39, A51, A49, A37	11.1%	5.02
NN-PDL1-7	A39, A98, A108, A106	13.9%	6.91

A11—which contribute hydrophobic contacts at the core of the interface—consistently yielded the highest binder rates and affinities. The best-performing combination (NN-PDGFR-6: A139, A77, A82) achieved a 71.1% hit rate with the tightest binder reaching $-\log(K_D) = \textit{10.03}$, corresponding to sub-100 pM affinity. Conversely, the set restricted to peripheral residues (NN-PDGFR-3: A139, A116, A142, A134) yielded a substantially lower hit rate (31.3%) and weaker affinities, confirming that hotspot selection critically shapes design outcomes.

For PD-L1, overall hit rates were lower (5–18%), consistent with the more challenging, flatter interface. Nevertheless, hotspot conditioning had a marked effect: the combination NN-PDL1-4 (A39, A98, A51) achieved the highest hit rate of 17.9%, while the single-residue set NN-PDL1-2 (A39 alone) yielded only 5.3%. This indicates that even on difficult targets, multi-residue hotspot conditioning concentrating on key interface positions substantially improves design success.

E. ActRIIA Binder Design: Experimental Methods

Expression and purification protocol. Plasmids encoding the top 200 designs were transformed into *E. coli* BL21(DE3) cells. A single colony was inoculated into 100 mL LB medium supplemented with 100 μg/mL ampicillin and cultured at 37 °C until the optical density at 600 nm (OD_{600}) reached approximately 0.6 (3–4 hours). The temperature was then reduced to 18 °C, and protein expression was induced with 0.1 mM IPTG. Cultures were incubated at 18 °C for an additional 16 hours. The overall timeline from gene synthesis to purified protein was approximately three weeks.

To obtain high-quality protein samples for downstream analysis, affinity tags were removed for selected constructs. Cell pellets were resuspended in lysis buffer (1× PBS, pH 7.4), lysed by sonication, and centrifuged to remove cell debris. The clarified supernatant was loaded onto a nickel-affinity column for purification. Eluted fractions were treated with TEV protease to remove the affinity tag and subsequently passed through a second nickel column to separate cleaved protein from uncleaved protein and the His-tagged TEV protease. The target protein was then buffer-exchanged into PBS, and protein concentrations were determined prior to surface plasmon resonance (SPR) analysis. A summary of expression yields for both tagged and tagless constructs is provided in Tab. 4; of the 200 designs, 192 expressed solubly and 110 achieved yields exceeding 50 mg/L.

SPR protocol. The surface plasmon resonance (SPR) experiments were conducted on a Biacore 8K+ instrument at 25 °C (both sample and analysis temperature) to characterize the binding kinetics of 80 analytes and a Bimagrumab control to the ligand ActRIIA(20–135)-Avi. The ligand was immobilized on all eight channels of a Streptavidin (SA) sensor chip at a concentration of 0.5 μg/mL with a flow rate of 5 μL/min for 52 seconds. Assays were performed using a running buffer of 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05% Tween-20, and 2% DMSO. The experimental workflow included 15 start-up cycles to stabilize the surface prior to analysis, and data were collected at a flow rate of 30 μL/min using multi-cycle kinetics (MCK) with association/dissociation times of 90/180 seconds. Two independent SPR runs were performed; binding measurements were largely consistent across both experiments. The only design showing

Table 4 | Expression and purification yields for the 200 ActRIIA binder designs. Soluble expression yields are reported after the first nickel-affinity purification step for both His-tagged (tagged) and TEV-cleaved (tagless) constructs. Of the 200 designs, 192 expressed solubly and 110 achieved yields exceeding 50 mg/L in the tagged format. Tag removal by TEV protease was unsuccessful for 22 constructs.

Yield (after 1^{st} nickel)	**Tagged Constructs**	**Tagless**
No soluble expression	9	9
Tag can't be cleaved	–	22
<10 mg/L	2	59
10–20 mg/L	9	52
20–30 mg/L	12	36
30–50 mg/L	58	20
>50 mg/L	110	2
Total	**200**	**200**

noticeable variability was #51, which exhibited fluctuations in both K_D and R_{max} values, possibly arising from nonspecific surface interactions or partial self-aggregation during the assay. Representative sensorgrams from both runs are shown in Fig. 17, and the full Activin A functional dose-response panel across all eight lead designs is shown in Fig. 16.

Smad2/3 luciferase reporter assay protocol. A Smad2/3-responsive luciferase reporter assay was performed using a stable HEK293 TGF-β/Activin/Myostatin-responsive reporter cell line in a 384-well format to evaluate the inhibitory activity of candidate binders. Reporter cells were seeded at 10,000 cells per well in 20 μL of DMEM-based culture medium in white, clear-bottom 384-well plates and incubated at 37 °C with 5% CO_2 for approximately 18 h to allow cell attachment. Test proteins, including engineered mini-binder candidates and the reference antibody bimagrumab, were added in 5 μL volumes and pre-incubated with the cells for 4 h. Subsequently, 5 μL of ligand solution was added to stimulate the pathway, using either human Activin A (20 ng/mL final concentration) or GDF8/myostatin (200 ng/mL final concentration). Test molecules were evaluated in three-fold serial dilutions across the plate. Control wells included ligand-stimulated cells (assay medium + ligand) and unstimulated cells (assay medium only) to define maximal activation and basal signaling levels, respectively. After overnight incubation (∼18 h) to allow transcriptional activation of the Smad2/3-responsive reporter, 30 μL of luciferase detection reagent was added to each well, and luminescence was measured using a plate luminometer. Inhibition of pathway signaling by test proteins was quantified as the reduction in luminescence relative to ligand-stimulated control wells.

F. PAK1 and CK1δ Kinase Binder Design: Experimental Methods

F.1. De novo binder design

Candidate binders targeting the PAK1 catalytic domain (PAK1c; Uniprot Q13153, amino acids 273–498) or the CK1δ catalytic domain (CK1δc; Uniprot P48730, amino acids 9–277) were designed using Proteina-Complexa. For the split-NeoR complementation assay, a library of 50 PAK1c binders (49–74 amino acids) was generated. Binder open reading frames were flanked by constant sequences for downstream PCR amplification and cloned into the split-NeoR selection vector. For CK1δc, 18 short peptide binders (<31 amino acids) were designed and synthesized as biotin-conjugated peptides.

F.2. Split-NeoR bacterial selection

The split-neomycin resistance (split-NeoR) protein-fragment complementation assay was used to select PAK1 binders based on reconstitution of aminoglycoside phosphotransferase (APH(3′)-II) activity [He et al., 2025, Paschon et al., 2005]. Binder candidates were fused to the C-terminal fragment of NeoR, and PAK1c was fused to the N-terminal fragment; both fusion proteins were co-expressed in *E. coli* from a single plasmid. The pooled binder library was transformed into NEB Stable Competent *E. coli* (New England Biolabs, C3040) by chemical transformation. Transformants were plated on LB agar without kanamycin (input control) or LB agar supplemented with 25 μg/mL kanamycin (selection) and incubated overnight. Colonies from both conditions were pooled and plasmid DNA was isolated by miniprep for downstream NGS analysis.

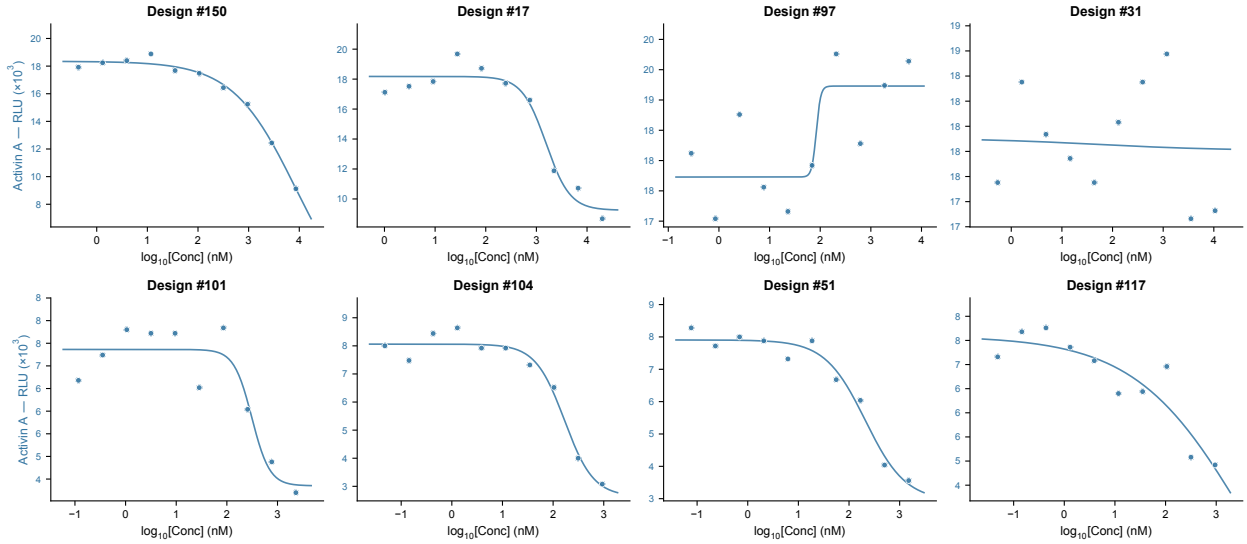


Figure 16 | Activin A functional assay across all eight ActRIIA binder designs. Dose-response curves from the Smad2/3 luciferase reporter assay are shown for all eight designs tested in this campaign (blue curves). Luminescence values are plotted against $\log_{10}$ concentration (nM). This panel complements the focused main-text figure by showing the full set of experimentally characterized designs under Activin A stimulation conditions.

F.3. Next-generation sequencing and enrichment analysis

Binder-encoding amplicons were prepared by PCR using KAPA HiFi 2× ReadyMix (Roche, KK2601) with library-specific primers flanking the binder open reading frames. PCR conditions: 98 °C for 3 min; 15 cycles of 98 °C for 20 s, 65 °C for 15 s, 72 °C for 30 s; 72 °C for 1 min. Template DNA was normalized to 3.0 ng/μL (30 ng input per reaction). Amplicon libraries were submitted for paired-end sequencing (Azenta AmpliconEZ). Reads were mapped to the reference library using a 30 bp ORF-prefix matching strategy anchored to the 5′ constant sequence (5′-`CACTCAGGGTCCGGT`-3′), achieving mapping rates of 91.7–94.7% across samples. Raw read counts were normalized to counts per million (CPM). Enrichment was calculated as $\log_2$ fold change:

$$\log_2\big((\text{selection CPM} + 1) \,/\, (\text{input CPM} + 1)\big),$$

where a pseudocount of 1 CPM was added to avoid division by zero. To classify binders as enriched or depleted, a two-component Gaussian mixture model (GMM) was fitted to the $\log_2$ fold change values using expectation-maximization. Two components were chosen based on the observed bimodal distribution. Each binder was assigned to the component with the higher posterior probability, and the decision boundary was defined as the $\log_2$FC value at which the posterior probabilities of the two components were equal.

F.4. Structure prediction

Predicted structures of binder–PAK1 complexes were generated using AlphaFold2 Multimer (v2.3.2) with five models and five seeds per model (25 predictions total per complex). The combined interface predicted TM-score and predicted TM-score (ipTM+pTM) from the best-ranked model was used as the metric of predicted binding confidence.

F.5. GST-tagged protein purification

GST-tagged CK1δc was expressed (pGEX-GST-CK1δ) as an N-terminal fusion in BL21(DE3) *E. coli*, induced with 0.1 mM IPTG, and grown at 20 °C for two nights. Cells were lysed by sonication in lysis buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM EGTA, 2% IGEPAL CA-630, 3 mM DTT, protease inhibitors) and clarified by centrifugation and 0.45 μm filtration. Proteins were purified using glutathione-Sepharose resin (Glutathione Sepharose$^{\text{TM}}$ 4B, Cytiva) at 4 °C and eluted in 50 mM Tris pH 7.5, 150 mM NaCl, 3 mM

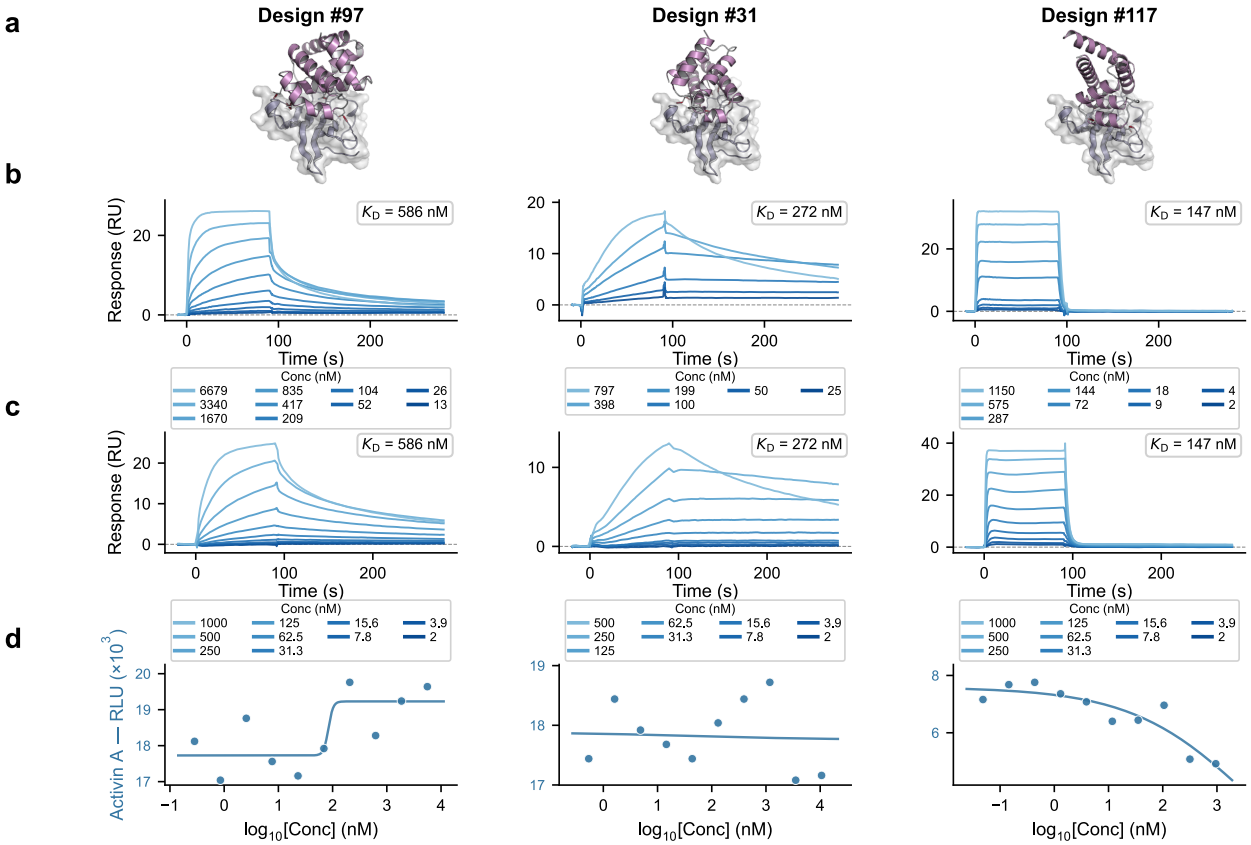


Figure 17 | Experimental characterization of the three ActRIIA binder designs (#97, #31, #117) that achieved sub-μM binding affinity but did not exhibit significant functional inhibition. **a)** Predicted complex structures of each binder (purple) engaged with the ActRIIA receptor (gray surface). **b)** SPR sensorgrams from the first independent measurement (raw traces only), with equilibrium K_D values indicated. **c)** SPR sensorgrams from the second independent measurement (raw traces only). **d)** Dose-response curves from the Smad2/3-responsive luciferase reporter assay under Activin A stimulation (20 ng/mL). Despite nanomolar-to-sub-μM affinity for ActRIIA, these designs did not produce measurable functional inhibition at the tested concentrations, suggesting that binding orientation and/or epitope overlap was insufficient to block native ligand signaling.

DTT, and 25 mM reduced glutathione, as previously described [Carlson et al., 2011]. Protein-containing fractions were identified by Bradford assay, pooled, concentrated by ultrafiltration with PBS washes, and stored at −80 °C.

F.6. Biotinylated peptide synthesis and streptavidin bead pull-down

Biotinylated peptides were synthesized on a MultiPep 1 automated peptide synthesizer (CEM corporation) using standard Fmoc solid-phase chemistry. Biotin was incorporated at the N-terminus via a biotinylated lysine followed by two glycine spacers. For pull-down assays, all steps were carried out at 4 °C unless otherwise noted. Streptavidin magnetic beads (Pierce^{TM} Streptavidin Magnetic Beads) were washed three times with 500 μL Bead Wash Buffer (BWB; 25 mM HEPES pH 7.4, 150 mM NaCl, 0.01% Tween-20, 1 mM DTT). For peptide capture, 17 μL of bead slurry was incubated with 1 mL of 3 μM biotinylated peptide with end-over-end mixing for 30–45 min at 4 °C, then washed three times for 3 min with 500 μL BWB. Purified GST-tagged kinase domain (0.75 μg) was added to peptide-loaded beads in 1 mL BWB+ (BWB supplemented with protease inhibitor) and incubated for 30–40 min at 4 °C with end-over-end rotation. Beads were washed four times with 500 μL BWB+ (10 min each). Bound proteins were eluted in 20 μL of 2× Laemmli Sample Buffer (Bio-Rad) containing β-mercaptoethanol at 95 °C for 5 min.

F.7. Co-immunoprecipitation

HEK293T cells were seeded in 12-well plates and transfected 24 hours later using PEI MAX 40000 (Polyethylenimine HCl MAX, Linear, Mw 40,000; Kyfora Bio, 24765-100). Each well received 500 ng of binder plasmid (pDS261: pCAG-GFP-3×Flag-P2A-3×HA-binder) and 500 ng of bait plasmid (pDS217: pCAG-PAK1-Myc), with PEI MAX at 20 μL per condition. DNA–PEI MAX complexes were formed in Opti-MEM (15 min, room temperature) and added to cells. Medium was replaced 6 hours post-transfection; cells were harvested 48 hours post-transfection. Cells were washed once with cold PBS and lysed in 300 μL of lysis buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.5% NP-40, protease inhibitor cocktail) on ice for 5–10 min. Lysates were clarified by centrifugation (15,000×g, 10 min, 4 °C) and 5–10% of clarified lysate was saved as input. HA-tagged proteins were immunoprecipitated using 15 μL ChromoTek HA-Trap Magnetic Agarose (Proteintech, cat# atma) with rotation for 90 min at 4 °C. Beads were washed three times with 500 μL wash buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.1% NP-40). Bound proteins were eluted in 30 μL 2× Laemmli sample buffer at 95 °C for 5 min. An empty pDS261 vector served as the negative control. Three independent biological replicates were performed.

F.8. Western blotting

For PAK1 co-IP experiments, input and eluate samples were resolved by SDS-PAGE on Mini-PROTEAN precast gels (4–20%; Bio-Rad, 4561096) and transferred to membranes. Membranes were probed with anti-Myc chicken polyclonal antibody (Invitrogen, A-21281) to detect Myc-PAK1, followed by IRDye 800CW donkey anti-chicken IgY secondary antibody. Anti-Flag antibody was used to detect GFP-Flag binder expression (IRDye 680RD secondary, 700 nm channel). Blots were imaged using a LI-COR Odyssey system with simultaneous dual-color detection (700 nm and 800 nm channels). For CK1δ pull-down experiments, proteins were transferred to membrane and probed with anti-GST antibody (Bethyl Laboratories GST Tag Polyclonal Antibody, HRP, 1:1000). Standard western blot procedures were followed for blocking, primary antibody incubation, and chemiluminescent detection (SuperSignal$^{\text{TM}}$ West Femto Maximum Sensitivity Substrate).

F.9. Image quantification and statistical analysis

Band intensities were quantified using Fiji/ImageJ. For PAK1 co-IP experiments, co-IP efficiency was calculated as the ratio of anti-Myc signal in the eluate to the anti-Myc signal in the input (elution/input ratio), then normalized to the empty vector control within each replicate to obtain fold enrichment values. Statistical significance was assessed by one-sample t-test (two-tailed) against a null hypothesis of fold enrichment $= 1.0$. Data are presented as mean ± SD from three independent biological replicates. For CK1δ pull-down experiments, samples were normalized to the scrambled control for their respective run. Statistical significance was assessed by one-sample t-test (one-tailed) against a null hypothesis of fold enrichment $= 1.0$. Data are presented as mean ± SD from three to four independent biological replicates.

Table 5 | Properties of selected PAK1 mini-protein binder candidates. NGS $\log_2$FC represents $\log_2$ fold-change in read frequency relative to the input library after kanamycin selection. AF2 ipTM+pTM is the combined score from AlphaFold2 Multimer (best of 25 models).

Binder	Length (aa)	NGS $\log_2$FC	AF2 ipTM+pTM	Amino acid sequence
Pk3	57	1.83	0.848	IVAQVIYHRLSPELRQEFEEKYKGNKSNAKLFAFARQKDP SLTQESVARVLFRQIVA
Pk4	51	1.05	0.765	IFKALARKLLPPEVYEKIKDSGPVTILNAISQYTGKSKEA LARPLFRQVVA
Pk5	52	1.73	0.626	FIEIIVKTKFRKEYEAELEKGKSPFQAANTVLAKKGLSRE SVARPLFRQLVA
Pk6	55	−0.20	0.538	FRFHIIKRKAQKNPQYSHLSDAELLALARKLSLLSSHPLA PTEKEYRTLFRQIVA

Table 6 | Predicted CK1δ peptide binders. All peptides were synthesized with an N-terminal biotinylated lysine followed by two glycine spacers.

Name	Length (aa)	Sequence
CK1	25	YTEAMDLVRKGIYRREEVLAQYGEK
CK2	27	EEQYKVSLQYAKEKKLEEYIKSINQLI
CK3	26	LSLAEYLVKHLKLPKNLAEKIASSMN
CK4	26	SLSEATIQLIAKQLGVDASQLEQLLS
CK5	25	DAVYLKSILQSGDLLRLKEYLLSLN
CK6	26	NFLEDGLSPSTALSFAIEKLCIQYGV
CK7	25	KQATEDRIRQEAEQKAAQERQQLAA
CK8	26	NLQEIKQQYREKEQQLQQEQQQRQQQ
CK9	28	FPCSYPGCGKTFTSSSSLRNHMKTHSRS
CK10	30	EEYKFELERLKKGELSPTEITLLEENLKRA
CK11	26	ELLKENPDLSVDFIRHQLSLLRSVGV
CK12	26	SLAEELFLRKLVDVHTANLISEKLTE
CK13	25	QLTSYQLEFMKKHGLTTEEELTAFM
CK14	25	YTSVYSSIYGPCPSCDEEERRRFGL
CK15	29	DLSDLSLGSLELIARLTGTNLSDLLAQRK
CK16	28	SDIKKLREFFTEKGYNPSDNLLELMVKL
CK17	30	GGASDEFLARLLGVSVEEIKRRRSLAASKK
CK18	30	EFAKNRQLLEEDLKRRGFSPEEISYRLSFV

G. Carbohydrate Binder Design: Experimental Methods

G.1. Design generation and filtering

Candidate binders targeting the blood group B trisaccharide (α-D-Gal-(1→3)-[α-L-Fuc-(1→2)]-D-Gal) were generated using Proteina-Complexa conditioned on the sugar ligand structure. To promote precise hydrogen-bonding networks and deep pocket burial—both critical for overcoming the large desolvation penalty of carbohydrate binding—inference-time rewards included pocket burial scores and hydrogen bond counts at the binder–ligand interface. After filtering, 24 designs (64–88 amino acids) were selected for experimental characterization.

G.2. Protein cloning and expression

All 24 designs (designated NV1–NV24) were gene-optimized for *E. coli* expression and ordered as gene strings with 20 bp overhangs compatible with the acceptor vector T7-pRSF. The T7-pRSF vector was linearized by PCR, and residual circular template was removed by DpnI (NEB) digestion. Linear product was purified by SPRI beads (Beckman) and eluted at a concentration of 50 ng/μL. Genes were ligated into the vector using Gibson assembly and transformed into BL21(DE3) cells (NEB) under kanamycin selection. Colonies were verified by whole-plasmid sequencing.

Cells were grown at 37 °C in 5 mL LB-kanamycin medium to an OD_{600} of 0.4–0.8, at which point expression was induced with 1 mM IPTG. Cultures were transferred to 20 °C with agitation at 1,000 rpm on a flat shaker-incubator, and expression was carried out for 14 hours. Cells were pelleted by centrifugation (5 min, 4,000×g) and lysed in 2 mL lysis buffer (10× BugBuster (Millipore), diluted in PBS supplemented with $CaCl_2$ and $MgCl_2$ (Thermo)). Lysates were agitated at 1,000 rpm for 30 min at 20 °C and clarified by centrifugation (10 min, 4,000×g).

The supernatant was collected and incubated with Ni-NTA agarose beads (200 μL slurry per protein) with agitation at 1,000 rpm for 30 min at 20 °C. After centrifugation (3 min, 300×g), the supernatant was discarded and the bead pellet washed with 1 mL PBS. Proteins were eluted in 200 μL elution buffer (PBS, 250 mM imidazole) with agitation for 10 min. The eluate was subjected to two rounds of buffer exchange into PBS using 3 kDa size-exclusion filters (Amicon), with 10-fold dilution and re-concentration per round. Final protein concentrations, quantified by A_{280} (NanoDrop), ranged from 0.8 to 3.5 mg/mL. Proteins were stored at 4 °C and assayed within three days of purification.

G.3. Red blood cell agglutination assay

Reagent red blood cells (type B) were obtained from Bio-Rad. The neat stock was diluted 1:100 in PBS and washed once by centrifugation (3 min, 300×g) without further dilution. A Ni-NTA agarose bead suspension was prepared at a ratio of 2:48 (neat resin:PBS, v/v). For each sample and replicate, 50 μL of bead suspension was loaded with 2 μL of purified binder solution, and 50 μL of the B-RBC suspension was added. Negative controls included RBCs + beads (no protein), RBCs + protein (no beads), and beads + protein (no RBCs). Samples were incubated in U-bottom 96-deep-well plates (Greiner), with a separate plate per replicate. After 10 min, samples were inspected macroscopically and microscopically for visible agglutination. After 2 hours of incubation, samples were imaged using an EVOS-FL microscope at the region of highest bead density per well, then agitated and OD_{600} measured using a plate reader (SpectraMax). Data are plotted as mean ± standard deviation of three independent replicates. A construct containing the B-binding CBM51 domain served as the positive control, and RBCs + beads without protein served as the negative control baseline.

G.4. Glycan–DNA conjugation for BLI

To produce a glycan conjugate of sufficient molecular weight to elicit a detectable biolayer interferometry (BLI) response, the blood group B pentasaccharide (type II; Biosynth) was conjugated to a modified DNA oligonucleotide. An oligonucleotide bearing three amine-modified positions ([AmC6]GA[AmC6dT]GA[AmC6dT]CTCGACGCTCTCCCTTATGCGACTCC; Integrated DNA Technologies) was mixed with the pentasaccharide at a 1:5 molar ratio in 50% DMSO containing 100 mM 2-picoline borane (Sigma). The reductive amination reaction was incubated for 3 hours at 37 °C. The product was purified and buffer-exchanged into PBS using PD-10 columns (Sephadex G-25, 3 mL elution). Recovered DNA yield was estimated by A_{260} (NanoDrop).

G.5. Biolayer interferometry

BLI experiments were performed on an Octet K2 (Sartorius) two-channel system using HIS1K biosensors (Sartorius, 18-5120). Interactions were measured in flat black 96-well plates (Nunc; 200 μL/well) with shaking at 1,000 rpm. The assay sequence consisted of: baseline (PBS, 120 s), loading (protein in PBS at 0.1 mg/mL, 120 s), baseline (PBS, 120 s), association (four sequential steps of 120 s each at glycan–DNA conjugate concentrations of 2, 4, 6, and 8 μM in PBS), and dissociation (PBS, 120 s). A reference sensor was treated identically but without protein in the loading step. Data were acquired at 5 Hz and reference-subtracted.

G.6. Circular dichroism

CD spectroscopy was performed on an Aviv 410 CD spectrometer using a quartz cuvette (1 mm path length). Design NV15 (S15) was loaded at a final concentration of 0.25 mg/mL in PBS. A wavelength scan (200–250 nm) was carried out at 25 °C with a bandwidth of 1.00 nm to confirm the presence of secondary structure elements. For thermal stability analysis, CD signal was monitored at 222 nm with a bandwidth of 10 nm. Dynode voltage was recorded concurrently to check for aggregation, which was not observed. An initial scan from 40 to 95 °C identified the inflection region, followed by a higher-resolution scan from 70 to 98 °C. Secondary structure elements remained stable beyond 100 °C, precluding accurate determination of a melting temperature within the accessible range.

H. Details on Use of Prior Methods for Benchmarking

BindCraft We used the code and checkpoints provided in the public BindCraft repository. We followed the default `default_4stage_multimer_hardtarget` configuration for setting and filters. For each target, we used the top-ranked hotspot for the hotspot conditioning of BindCraft. To ensure fair comparison under the constraint of 32 GPU-hr computation budget while maintaining a minimum number of sample for meaningful statistical analysis, we adopted the following procedure of sample selection. First, we generated M binders with 32 GPU-hrs budget. Second, we re-designed and evaluated the generated binder sequence 8 times using soluble weight of ProteinMPNN [Dauparas et al., 2022] and interface residues fixed. We re-designed 8 times to be consistent with the re-designing protocol of all models benchmarked in this paper. Then we randomly trim the generated binder set size to N ($N \leq M$) such that the total generation+redesign+evaluation pipeline could fit in the computation budget. After that, we selected the top-1 re-designed sequence from each

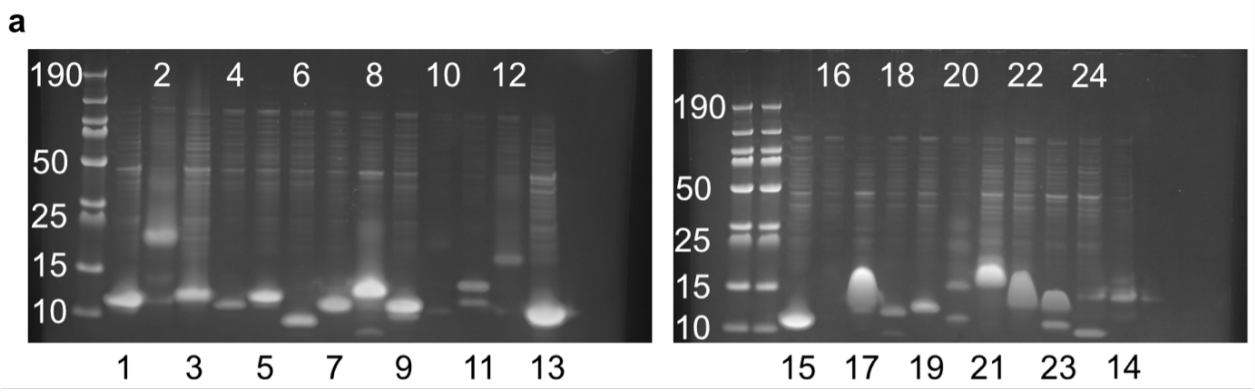


Figure 18 | SDS-PAGE analysis of all 24 carbohydrate binder designs expressed in *E. coli*. Lanes S1–S24 correspond to the 24 designs; molecular weight markers (kDa) are shown. Most designs express as soluble protein at the expected size (7–10 kDa).

generated binder based on the best *minimum ipSAE* [Dunbrack Jr, 2025] metric. If $N < 400$ for large target which required more compute, we would backfill to 400 samples using the best top-2 (or even top-3 if need) re-designed sequences. In this way, we would have at least 400 samples per target in the final BindCraft sample set for meaningful statistical analysis.

RFDiffusion We utilized the open-source implementation of RFdiffusion in the public RFdiffusion repository, and followed the protein-protein interaction (PPI) design settings provided in the official documentation. For each target, we first generated backbone structures and subsequently employed the soluble weights of ProteinMPNN to generate 8 binder sequences per backbone. We performed 8 redesign cycles to maintain consistency with the benchmarking protocol used for all models in this study. A total computational budget of 32 GPU-hours was allocated for the full pipeline, including backbone generation, sequence redesign, and AlphaFold2-based evaluation. From the resulting pool, we selected the top-1 redesigned sequence for each generated binder based on the best *minimum ipSAE* metric, yielding a final set of 400 binders per target.

RFDiffusion3 We adopted the protein binder design protocol following the production settings in the RosettaCommons Foundry repository for RFdiffusion3. The procedure involved an initial generation of binder structures, followed by the application of soluble ProteinMPNN to produce 8 redesigned sequences per structure. We evaluated both the self-generated sequences and the 8 redesigned variants. Under the 32 GPU-hour budget, we selected the top 400 self-generated sequences and the top 400 redesigned sequences based on the *minimum ipSAE* metric. This procedure resulted in a total of 800 binders per target.

BoltzGen We employed the BoltzGen `protein-anything` protocol to design binders using the code from the public BoltzGen repository. This unified all-atom diffusion approach integrates structure generation (BoltzGen core) and inverse folding (BoltzIF). For each design trajectory, we generated one sequence via Boltz inverse folding and an additional 8 sequences using the soluble weights of ProteinMPNN. We evaluated the self-generated sequences, the Boltz inverse-designed sequences, and the 8 ProteinMPNN-redesigned sequences. The 32 GPU-hour budget encompassed the entire pipeline: self-generation, Boltz inverse design, ProteinMPNN redesign, and the subsequent evaluation of all resulting sequences. Following the standard selection rules based on *minimum ipSAE*, we identified the top 400 sequences from each of the three categories (self-generated, Boltz inverse-designed, and soluble inverse-designed), resulting in a total of 1,200 binders per target.